The investigation of internal ribosome entry in the c-*myc* an

genes.

by

Joanne R.Evans BSc (Southampton)

Thesis submitted for the degree of Doctor of Philosophy at th

University of Leicester

January 2003

UMI Number: U169875

UMI U169875

The investigation of internal ribosome entry in the c-*myc* and c-*myb* genes.

Joanne R.Evans

The c-*myc* gene contains an internal ribosome entry site (IRES) within its 5' untranslated region. The IRES was shown to have different activities between cell lines suggesting a requirement for protein *trans*-acting factors that are present in these cell lines in varying amounts. In addition a number of proteins have been shown to interact with the IRES by north-western and UV cross-linking analysis.

Investigation of the protein factors involved in c-*myc* IRES translation identified PCBP1 (Poly (rC) binding protein 1), PCBP2, HnRNPK (heterogeneous nuclear ribonucleoprotein K), UNR (upstream of N-ras) and UNRIP (unr interacting protein) as having a role in c-*myc* IRES translation. PCBP1, PCBP2, HnRNPK and UNR were found to directly interact with the IRES RNA by UV cross-linking and electrophoretic mobility shift assays (EMSAs). Investigation of the proteins effect on c-*myc* IRES activity showed stimulation of IRES activity in HeLa cells by PCBP1 and PCBP2. The factor HnRNPK was found to have a slight stimulatory effect *in vivo*. In addition PCBP1 and PCBP2 were found to stimulate IRES activity *in vitro* in combination with UNR and UNRIP.

Using the yeast three-hybrid system a number of additional proteins were found to interact with the c-*myc* IRES RNA. A novel Fibrillarin–like protein was identified and shown to strongly interact with the IRES by EMSA. Studies to determine a direct role of this factor in c-*myc* IRES translation were inconclusive.

The study of translation of the c-*myb* gene identified an IRES within its 5'UTR. Investigation of the role of *trans*-acting factors in its translation showed a possible role of the factors PCBP2, HnRNPK and ITAF45 (IRES *trans*-acting factor 45).

Acknowledgements

Table of contents

List of figures

Chapter 5. Internal ribosome entry in the c-*myb* 5'UTR.

Abbreviations

3-AT	3-Aminotriazole
Apaf-1	apoptotic protease-activating factor 1
ATP	adenosine 5'-triphosphate
BAG-1	Bcl-2-associated athanogene-1
bp	base pairs
BSA	bovine serum albumin
CIAP	calf intestinal alkaline phosphatase
cpm	counts per minute
CrPV	cricket paralysis virus
CSFV	classical swine fever virus
CTP	cytidine 5'-triphosphate
dATP	deoxyadenosine 5'-triphosphate
DAP5	Death associated protein 5
dCTP	deoxycytidine 5'-triphosphate
ddNTP	dideoxynucleotide
dGTP	deoxyguanosine 5'-triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxythymidine 5'-triphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetate
eIF	eukaryotic initiation factor
EMCV	encephalomyocarditis virus
EMSA	electrophoretic mobility shift assay
5-FOA	5-Flouroorotic acid
FGF2	fibroblast growth factor 2
FMDV	foot and mouth disease virus
GAPDH	glyceraldehyde 3'phosphate dehydrogenase
GDP	guanosine diphosphate

GTP	guanosine 5'-triphosphate
hr	hours
HAV	hepatitis A virus
HCV	hepatitis C virus
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid
hnRNP	heterogeneous nuclear ribonucleoprotein
HRV	human rhinovirus
IRE	Iron Response Element
IRES	internal ribosome entry segment
IRP	IRE-binding Protein
ITAF	IRES *trans*-acting factor
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobases
LB	Luria-Bertani broth
met	methionine
met-tRNA$_i$	initiator methionyl-tRNA
min	minutes
Mnk	MAP-kinase-interacting kinase
MOPS	3-[N-morpholino]-2-hydroxypropanesulfonic acid
mRNA	messenger ribonucleic acid
nt	nucleotides
ODC	ornithine decarboxylase
ORF	open reading frame
PABP	poly(A) binding protein
Paip	poly(A) binding protein interacting protein
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCBP1/2	poly(rC) binding protein 1/2
PCR	polymerase chain reaction
PEG	polyethylene glycol
PIPES	piperazine-N,N'-bis[2-ethanesulphonic acid]

Pfu	*Pyrococcus furiosus*
PTB	polypyrimidine tract binding protein
PV	poliovirus
RNA	ribonucleic acid
RNase	ribonuclease
RNAsin	ribonucleic acid hydrolase inhibitor
rRNA	ribosomal RNA
SD	synthetic dropout media
SDS	sodium dodecyl sulphate
TAE	tris acetate EDTA
TBE	tris borate EDTA
TE	tris EDTA
4-thioUTP	4-thiouridine 5'-triphosphate
tRNA	transfer ribonucleic acid
UNR	upstream of N-*ras*
UNRIP	UNR interacting protein
uORF	upstream open reading frame
UTP	uridine 5'-triphosphate
UTR	untranslated region
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YPAD	yeast-extract peptone adenine dextrose media

Chapter 1

Introduction

1.1 Eukaryotic protein synthesis

Eukaryotic protein synthesis can be divided into three stages of initiation, elongation, and termination. The initiation stage involves the binding of the 40S ribosomal complex at the mRNA cap, the scanning of the 40S complex to the initiation codon, and the joining of the 60S ribosomal subunit followed by the subsequent release of the initiation factors. In the elongation stage the ribosome moves along the mRNA in a 5' to 3' direction with the aid of elongation factors to synthesize the polypeptide chain. Once a termination codon is reached release factors bring about the dissociation of the ribosomal subunits, and the polypeptide chain release. The process of translation is highly regulated in eukaryotic cells and it is at the initiation stage where the majority of this regulation occurs (Pain, 1996).

1.2 Cap-dependent initiation

1.2.1 Introduction

The main process of translation initiation is known as cap-dependent initiation. All cellular mRNAs contain a nuclear modification at their 5' end that occurs at the time of transcription (Shatkin, 1976). This methyl guanosine (m7GpppN) cap acts in many aspects of transcript metabolism including protection against 5'-3' exonucleases, splicing, nuclear export, and translation (McKendrick et al., 2001). In translation initiation, the binding of initiation factors to the cap brings the 40S ribosomal subunit to the 5' end of the mRNA. Initiation in eukaryotes is a highly controlled process and involves at least 13 initiation factors. These initiation factors carry out a wide range of functions from initiation codon recognition to unwinding of RNA structure, and are targets for the control of initiation.

1.2.2 Formation of the 43S pre-initiation complex

The first stage of initiation involves eukaryotic initiation factor 2 (eIF2) joining with GTP and methionine-charged initiator tRNA (Met-tRNA$_i$) to form the ternary complex (Figure 1.1A). The formation of this complex is considered a major rate-limiting step for translation initiation. eIF2 is a large protein made up of three polypeptide chains, α, β, and γ, and is subject to phosphorylation on its α subunit. This phosphorylation
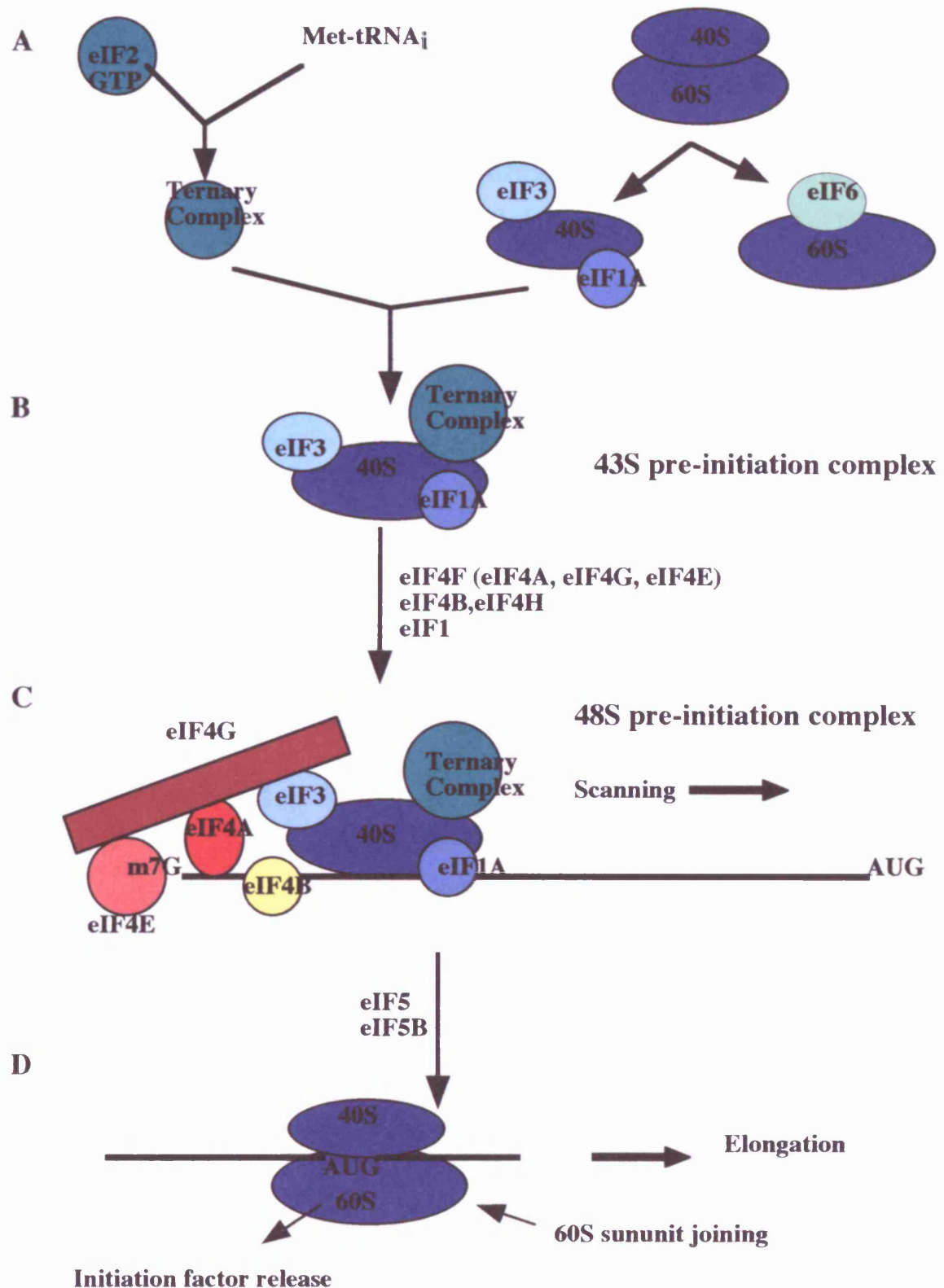
Figure 1.1. Cap-dependent translation. A) eIF2.GTP joins with Met-tRNA$_i$ to form the ternary complex. B) The ternary complex binds to eIF3/eIF1A bound 40S ribosomal subunits to form the 43S pre-initiation complex. C) The 43S complex is brought to the mRNA through interactions with eIF4F (eIF4A, eIF4G, eIF4E). D) Scanning occurs to the initiation codon, where eIF5 and eIF5B cause factor dissociation and 60S ribosomal subunit joining. Elongation then occurs.

prevents the recycling of eIF2.GDP by the GDP/GTP exchange factor eIF2B, and therefore decreases the availability of eIF2.GTP for ternary complex formation (Kimball, 1999). Both the β and γ subunits of eIF2 have been implicated in Met-tRNA$_i$ binding, and the β subunit of the protein contains the binding sites for eIF2B and also eIF5 (Kimball, 1999). The main role of eIF2 is to transfer Met-tRNA$_i$ to the 40S ribosomal subunit, however studies on eIF2 show it can also bind mRNA and plays a role in identification of the initiation codon (Bommer et al., 1991).

Free 40S ribosomal subunits are recruited for the next round of translation by the binding of the factors eIF3 and eIF1A. These factors are believed to bind to the 40S ribosomal subunits at the end of translation following the dissociation of the ribosomal subunits. This was considered to delay the formation of 80S ribosomes and maintain free 40S subunits for initiation (Pain, 1996). However, recent work shows that eIF3 and eIF1A only prevent the formation of 80S ribosomes in the presence of the ternary complex (Chaudhuri et al., 1999). The 60S ribosomal subunits are stabilised by the binding of eIF6 (Si et al., 1997). The eIF3/eIF1A bound 40S ribosomal subunits join with the ternary complex to form the '43S pre-initiation complex' (Figure 1.1B). eIF3 and eIF1A are critical for formation of this complex- eIF1A acts to transfer the ternary complex to the 40S ribosomal subunit, and eIF3 stabilizes the 43S pre-initiation complex (Chaudhuri et al., 1999). eIF1A and eIF3 also have a role in the binding of the 43S pre-initiation complex to the mRNA (Pestova et al., 2001; Bommer et al., 1991).

1.2.3 mRNA binding by the 43S pre-initiation complex

The 43S pre-initiation complex is recruited to the mRNA by the eIF4F group of initiation factors, made up of eIF4E, eIF4A and eIF4G. These factors perform a central role in initiation of translation. eIF4E is the cap binding protein that brings the translation components to the 5'end of the mRNA. Structural studies show it binds around the 5'cap structure like a cupped hand, with the cap structure binding in a hydrophobic pocket of the protein through interactions with several tryptophan residues (McKendrick et al., 1999). eIF4E binding to the cap structure is an important control point for cap-dependent translation. The factor can be controlled both by its own phosphorylation, and its association with several binding proteins (4E-BPs).

eIF4A is a 46kDa RNA helicase and RNA-dependent ATPase. It belongs to the DEAD box family of helicases, which have a role in a variety of cellular processes

including splicing, ribosome biogenesis, RNA degradation, and translation (Lorsch and Herschlag, 1998). eIF4A acts to unwind secondary structure in the 5' untranslated region (UTR) using energy from ATP hydrolysis, thereby allowing the initiation complex contact with the mRNA and enabling the scanning of the complex to the start codon. There are three mammalian isoforms of the protein named eIF4AI, eIF4AII and eIF4AIII. eIF4AI and eIF4AII share 89% identity, are cytoplasmic and are both known to bind eIF4G. These two forms of the protein are functionally indistinguishable. eIF4AIII shares only 65% identity to eIF4AI and cannot substitute for its activity (Li et al., 1999). eIF4AIII is mainly nuclear, and its role in translation is unknown. The factor eIF4A works together with the factors eIF4B, and eIF4H. eIF4B acts to stimulate eIF4A helicase activity and binding of the 40S ribosomal subunit to mRNA (Gingras et al., 1999). An eIF4B homolog, called eIF4H, has also been discovered which has 39% identity to eIF4B. This protein was found to stimulate the activity of eIF4A, eIF4B and eIF4F (Richter-Cook et al., 1998).

The factor eIF4G has an important role in initiation by acting as a molecular scaffold protein to bring several other factors together at the cap. eIF4G has interactions with the factors eIF4A, and eIF4E as part of the eIF4F complex . It also has interactions with eIF3, PABP (poly A binding protein), and the eIF4E kinase Mnk1 (MAP-kinase-interacting kinase-1) and contains an RRM (RNA-recognition motif) like RNA binding domain (Figure 1.2A). It is eIF4G's interaction with both eIF3 in the 43S pre-initiation complex and the cap binding factor eIF4E that brings the 43S pre-initiation complex to the 5'end of the mRNA. It also acts to bring in the helicase eIF4A and circularises the mRNA through interactions with PABP (Figure 1.2B).

There are 2 isoforms of eIF4G in mammalian cells: eIF4GI and eIF4GII, which have 46% identity (Gradi et al., 1998a). Both proteins have been found to share similar activities. There are also proposed to be multiple isoforms of eIF4GI from the use of alternative start codons, although the function of these is unknown (Byrd et al., 2002). Two more distant relatives of the protein are also known: Paip-1 (Poly A binding protein interacting protein-1) and DAP5 (Death associated protein 5). Paip-1 shares 25% identity with the central domain of eIF4G that contains an eIF4A binding site, and has been shown to bind eIF4A and the factor PABP (Craig et al., 1998). Over expression of the protein stimulated translation *in vivo*, and it is thought it may act to circularise the mRNA. DAP5, also called p97 or NAT-1, shares 28% identity with the C-terminal region of eIF4G. It contains binding sites for eIF3, eIF4A and Mnk1, but not eIF4E or PABP. It is thought

A

eIF4G

| PABP | | eIF4E | | eIF4A/ RNA | eIF3 | | eIF4A | Mnk |

B
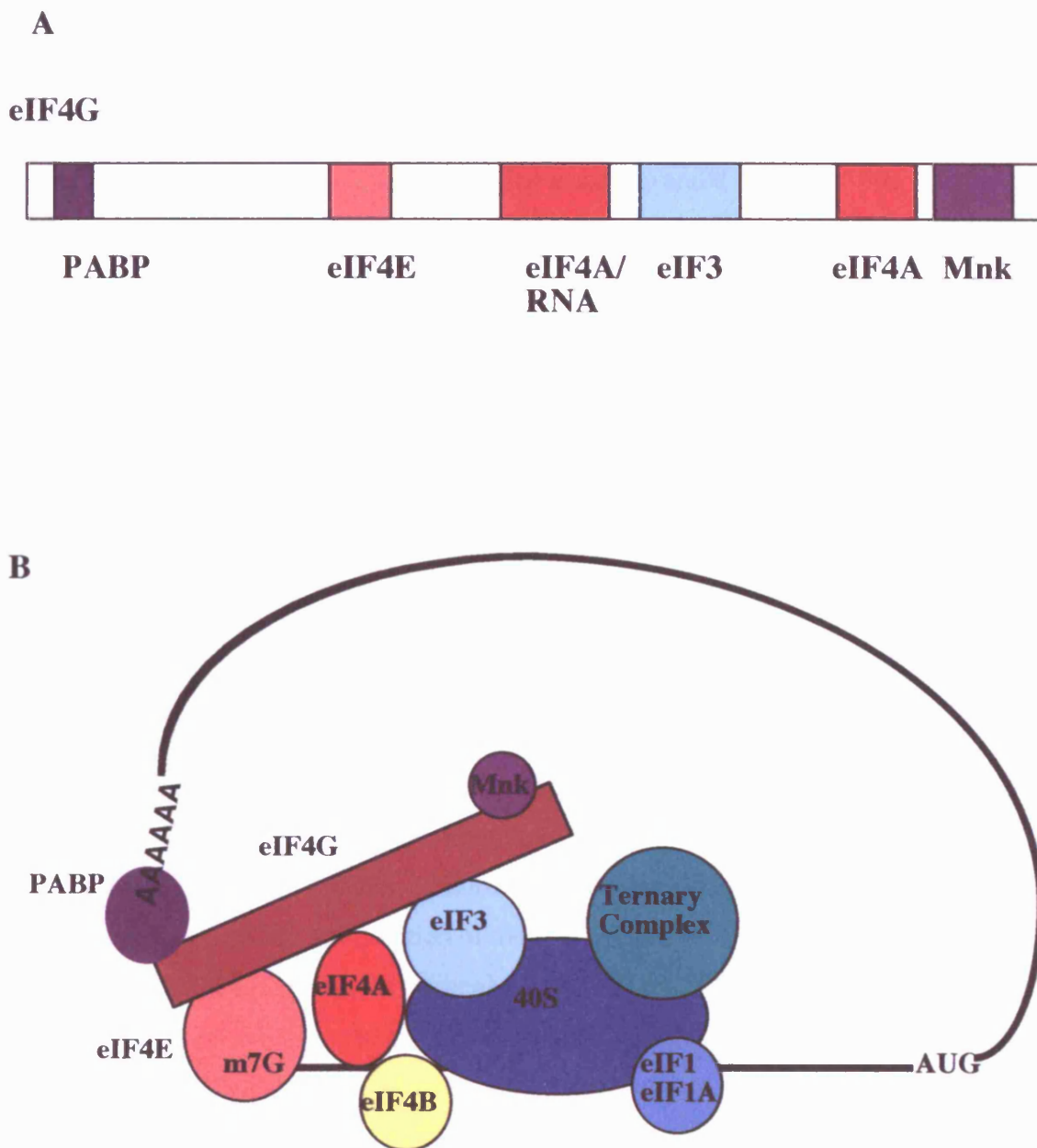
Figure 1.2. The structure of eIF4G and the closed loop model of translation initiation. A) Structure of eIF4G indicating domains that interact with PABP, eIF4E, eIF4A, eIF3 and Mnk. B) The closed loop model of translation initiation. The 43S pre-initiation complex is recruited to the mRNA by the eIF4F group of initiation factors. This includes eIF4G, eIF4A, and eIF4E. PABP is also recruited to the complex and acts to bring the poly(A) tail to the cap, and circularize the mRNA.

DAP5 may act as an inhibitor of translation by sequestering eIF3, eIF4A and Mnk1 (Gingras et al., 1999; Pyronnet et al., 1999). However, the protein has also been shown to have a role in cap-independent translation (see section 1.5.2).

The exact sequence of events that occur in the process of mRNA binding is not known. Whether the eIF4F complex assembles at the cap and then recruits the 43S pre-initiation complex, or if eIF4F associates with the 43S pre-initiation complex before binding to the cap structure is debated. Recent work in yeast suggests that a multifactor complex consisting of eIF1, eIF3, eIF5 and the eIF2.GTP.Met-tRNA$_i$ ternary complex exists independently of the 40S ribosomal subunit (Asano et al., 2000). Therefore these factors may bind 40S ribosomal subunits as a complex rather individually. The assembly of all these factors at the cap is known as the 48S pre-initiation complex (Figure 1.1C).

1.2.4 Initiation codon selection

The recruitment of the 40S ribosomal complex to the 5'end of the mRNA does not position it at the translational start codon, which is usually 50-100 nucleotides downstream. The initiation complex must move in a 5' to 3' direction until the start codon is located in a process known as scanning. Scanning is ATP dependent, but it is not known if this ATP hydrolysis is required for ribosomal movement or to unwind secondary structure in the mRNA (Pestova et al., 2001). The initiation codon is normally the first AUG encountered, but the composition of the surrounding bases is also important (Kozak, 1996). Investigation found that the optimal start codon context was GCCACC(AUG)G, with the bases at +4, -3 being the most conserved. Occasionally, non AUG codons, such as CUG, are used but these are rare and again context dependent.

Several factors are involved in scanning and start codon selection. The factor eIF1A acts very closely with another factor called eIF1. Both act synergistically to enable scanning of the 43S pre-initiation complex to the initiation codon (Pestova et al., 1998a). Any complexes formed on the mRNA that are not competent for scanning are dissociated by these factors, and then rebind the mRNA in a competent state (Pestova et al., 1998a). Once the initiation codon is reached the factor eIF5 causes the release of initiation factors from the complex. eIF5 is a 49kDa factor and binds specifically to eIF2 and eIF3. It promotes the hydrolysis of eIF2-bound GTP, which is thought to be induced by base pairing between the Met-tRNA$_i$ and the initiation codon (Pestova et al., 2001). GTP

hydrolysis leads to the dissociation of eIF2.GDP (Pestova et al., 2001) and eIF3 (Chakrabarti and Maitra, 1991), and leaves the initiator tRNA in the ribosome P site.

The joining of the 60S subunit is dependent on eIF5 and eIF5B (Pestova et al., 2000). eIF5B is a homologue of the prokaryotic initiation factor IF2, and like IF2 it has a ribosomal dependent GTPase activity that is essential for its function. Binding and hydrolysis of GTP could be required for it to adopt an active structure, or for the release of it or other factors from assembled 80S ribosomes. GTP hydrolysis of eIF2-bound GTP and of eIF5B-bound GTP are a prerequisite for 60S subunit joining (Chakrabarti and Maitra, 1991; Pestova et al., 2000). There is also evidence that initiation factors such as eIF2, eIF1, and eIF1A play a role in start codon recognition. The 80S ribosome then begins the elongation stage of translation (Figure 1.1D).

1.2.5 The role of the mRNA poly(A) tail

In addition to the 5' cap modification, cellular mRNAs are also modified in the nucleus at their 3'end by the addition of a poly (A) tail of up to 200 adenosine residues (Day and Tuite, 1998). This modification protects against degradation by 3'-5' exonucleases, and has a role in translational control. The poly (A) tail can be bound by multiple copies of PABP (Poly A Binding Protein), which also interacts with the N-terminal portion of eIF4G (Imataka et al., 1998). As described, eIF4G interacts with the cap binding protein eIF4E, therefore bringing the poly (A) tail to the 5' cap structure and circularising the mRNA transcript. This is described as the closed loop model (Figure 1.2B). The poly (A) tail and 5' cap act synergistically to stimulate cap dependent translation, and this synergy is dependent on the eIF4G-PABP interaction (Michel et al., 2000). It is thought that via the closed loop model ribosomes are delivered back to the 5'end of the mRNA. Alternatively, it may provide a way in which control elements in the 3'UTR are able to effect initiation (Gingras et al., 1999).

PABP also interacts with Paip-1 and Paip-2. As described earlier Paip-1 can bind eIF4A and PABP, and stimulates translation *in vivo* (Craig et al., 1998). In contrast Paip-2 is a much smaller protein that binds to PABP and inhibits its function. Paip-2 decreases the binding of PABP to the poly (A) tail, and competes with Paip-1 for PABP binding (Khaleghpour et al., 2001). Paip-2 inhibits translation both *in vitro* and *in vivo*.

1.3 Mechanisms of translational control

1.3.1 Introduction

The process of translation is highly controlled, and most of this control is at the stage of translation initiation. One way in which this is achieved is through the regulation of initiation factor activity. Several initiation factors can be controlled via phosphorylation, and through their interaction with binding proteins. In addition several factors have proteins that compete with them for binding, and therefore inhibit their activity. Two central rate-limiting steps are the formation of the ternary complex and the binding of the 43S pre-initiation complex to the mRNA (Pain, 1996). Further regulation occurs through *cis*-acting elements in the 5' and 3' UTR of the mRNA.

1.3.2 Regulation of initiation factors

The regulation of initiation factor activity effects total protein synthesis in the cell. This allows rapid changes in protein synthesis to occur in response to cellular conditions. Translation is promoted in situations of cell growth, and inhibited in conditions of cell stress such as heat shock, serum deprivation, mitosis, and viral infection (Feigenblum and Schneider, 1996).

1.3.2.1 Control of ternary complex formation

An important control point for translation initiation is in the formation of the ternary complex. This is controlled through changes in the activity and binding properties of eIF2 and eIF2B (Figure 1.3). As discussed, eIF2 can be phosphorylated on its α subunit, which occurs at Ser51. This phosphorylation increases the affinity of eIF2 for eIF2B, and has been found to prevent eIF2B carrying out its GDP/GTP exchange function (Pain, 1996). eIF2 has been shown to be in excess of eIF2B in the cell, and therefore the binding of eIF2(αP) acts to diminish free eIF2B. This leads to a reduction in GDP.GTP exchange, and formation of the ternary complex, and therefore inhibits translation. eIF2 phosphorylation is carried out in response to cellular stress by four kinases - HCR (Heme-controlled repressor), PKR (Protein Kinase activated by double-stranded RNA), GCN2 (general control non-derepressible) and PERK (PKR-like ER kinase) (Krishnamoorthy et al., 2001). The kinases increase eIF2α phosphorylation under conditions of heme deficiency, in viral infection, in amino acid and serum starvation, and following heat shock
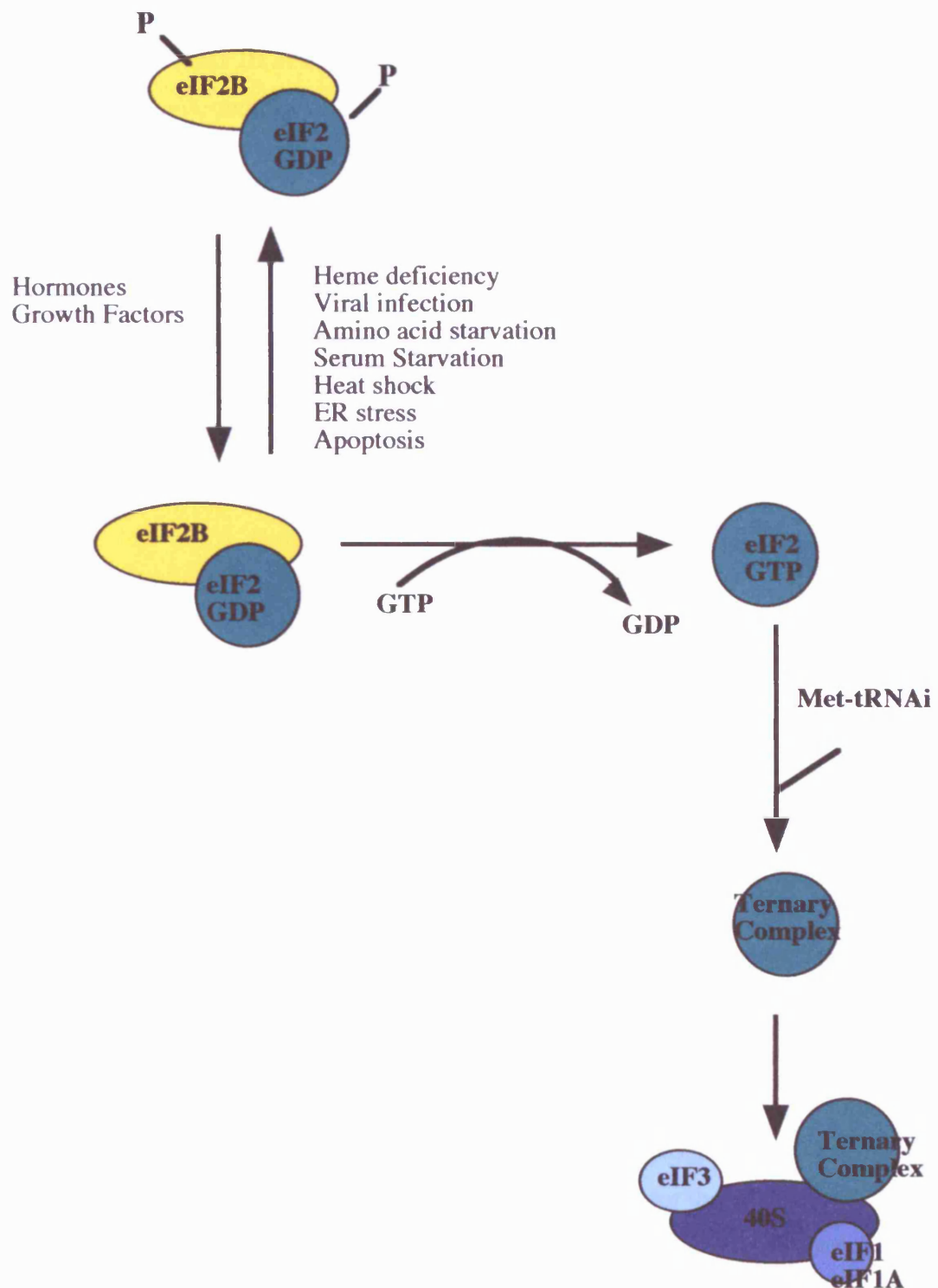
Figure 1.3. Control of ternary complex formation through phosphorylation changes in eIF2 and eIF2B. Phosphorylation of eIF2 causes increased affinity for eIF2B and prevents recycling by eIF2B. Phosphorylation of eIF2B decreases its GDP.GTP recycling activity. These changes act to decrease eIF2.GTP for ternary complex formation.

and ER stress leading to a general decrease in cap-dependent translation (Pain, 1996). eIF2α phosphorylation also occurs in apoptosis.

A second control mechanism at this stage is the phosphorylation of eIF2B. As described previously, eIF2B acts as an eIF2 GDP/GTP exchange factor. It is a large protein made up of 5 subunits, and can be phosphorylated at 5 sites on its catalytic subunit (ε) by Glycogen synthase kinase-3 (GSK3), and casein kinases I and II (CKI and II) (Wang et al., 2001). GSK3 phosphorylates eIF2Bε at Ser535, which inhibits the proteins function. Stimuli such as insulin and growth factor treatment inactivate GSK3, causing a dephosphorylation of eIF2B. This increases its activity and therefore acts to increase translation initiation. The effect of phosphorylation by Casein Kinase I and II is not thought to be important in control of eIF2B activity (Wang et al., 2001). The phosphorylation of both eIF2 and eIF2B inhibits GDP/GTP exchange and the formation of the ternary complex, and hence translation initiation.

1.3.2.2 Control of eIF4E activity

Another central control mechanism is exerted though changes in the activity of the cap-binding protein eIF4E. eIF4E is controlled both by its own phosphorylation, and its association with several binding proteins (4E-BPs) (Figure 1.4). eIF4E itself is phosphorylated at Ser209, which has been found to correlate with enhanced translation (Kaspar et al., 1990). This phosphorylation is carried out by Mnk1 (MAP kinase-interacting protein kinase-1) and Mnk2 in response to stimuli that promote cell growth (Scheper et al., 2001; Waskiewicz et al., 1999). Both Mnk1 and Mnk2 interact with the C-terminal region eIF4G, bringing them in the vicinity of their substrate eIF4E. The importance of this interaction is verified by experiments showing mutations in the eIF4E binding site of eIF4G decreases eIF4E phosphorylation (Pyronnet et al., 1999). Originally the phosphorylation of eIF4E was reported to increase its affinity for the cap (Minich et al., 1994), however recently contradictory results have been shown (Scheper et al., 2002). The phosphorylation of eIF4E has been found to reduce its affinity for the cap. The mechanism by which this enhances translation is not understood. It is proposed that the dissociation from the cap may increase movement of eIF4F and the ribosome complex towards the initiation codon (Scheper et al., 2002). Due to the role of eIF4G in eIF4E phosphorylation, eIF4E may only be phosphorylated and therefore dissociate from the cap once eIF4F is assembled. Situations such as growth arrest, heat shock, mitosis, viral infection and amino
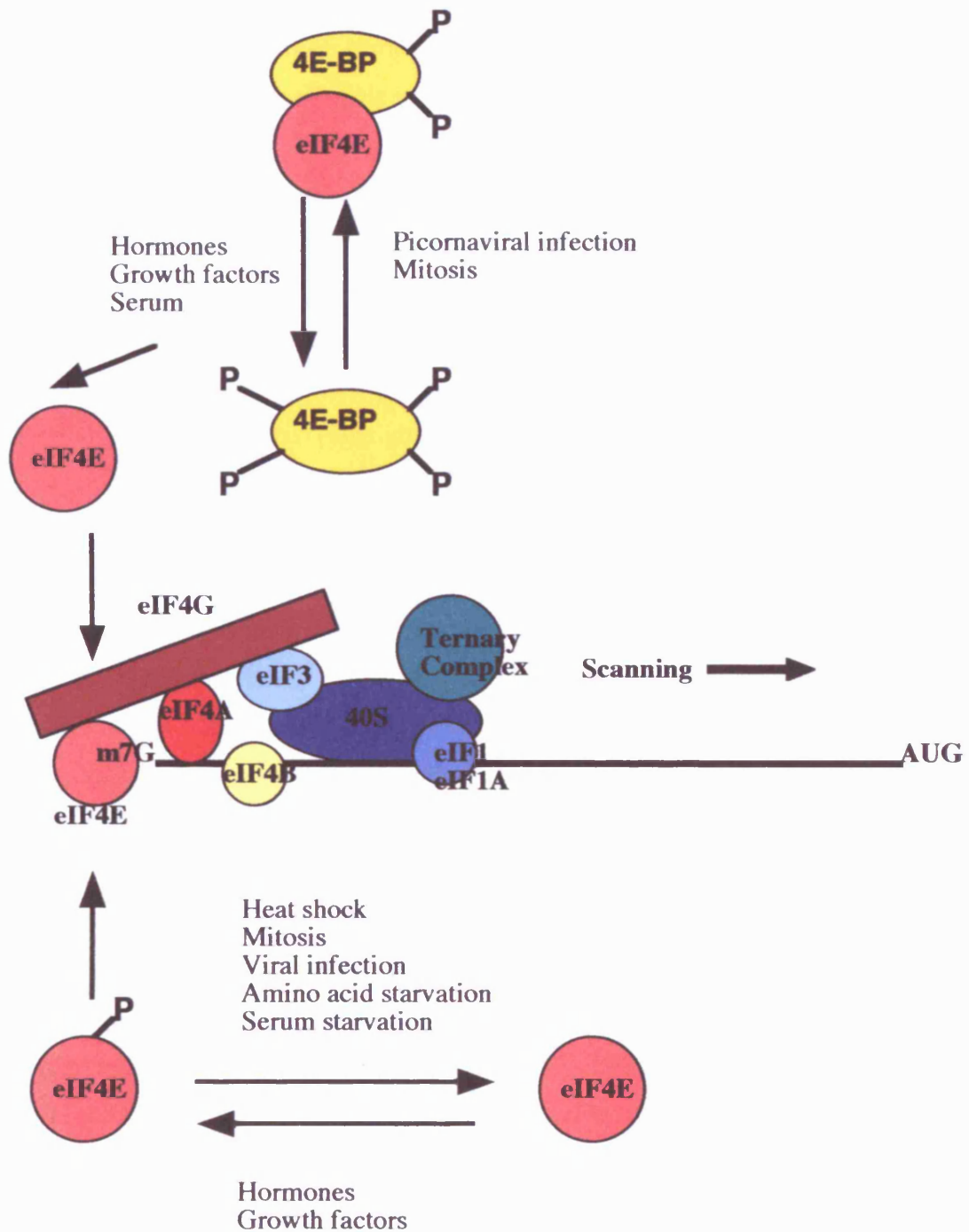
Figure 1.4. Control of translation through changes in the phosphorylation state of eIF4E and 4E-BPs. The hyper-phosphorylation of 4E-BP decreases its affinity for eIF4E, causing more eIF4E to become available to bind eIF4G. The phosphorylation of eIF4E acts to stimulate translation.

acid/serum deprivation all act to reduce protein synthesis through the dephosphorylation of eIF4E, which acts to reduce eIF4F complexes. The phosphorylation of eIF4E has also been found to alter throughout the cell cycle in relation to translational activity (Gingras et al., 1999).

There is further control of eIF4E activity by eIF4E binding proteins (4E-BPs). Three such proteins have been discovered: 4E-BP1, 4E-BP2, and 4E-BP3 (Pause et al., 1994; Poulin et al., 1998). The 4E-BPs are small proteins (10-12kDa) that have been found to have similar functions, and share identity (4E-BP1 and 4E-BP2 share 56% identity). These factors have no effect on the cap binding affinity of eIF4E, but have been found to compete with eIF4G for eIF4E (Haghighat et al., 1995). 4E-BP1 is phosphorylated at 6 sites, 2 of which are serum sensitive and thought to have a role in the control of eIF4E binding (Gingras et al., 2001). The kinases involved in the phosphorylation of 4E-BP1 have not been identified, however the phosphoinositide 3-kinase (PI3K) and FRAP/mTOR signalling pathways are known to be involved. The hyperphosphorylation of 4E-BP1 decreases its affinity for eIF4E, and hence more eIF4E is available to form eIF4F, and increase translation. The 4E-BPs are controlled by hyperphosphorylation in response to cell growth, serum and hormones, and by dephosphorylation in picornaviral infection, nutrient deprivation and mitosis (Gingras et al., 2001; Pyronnet et al., 2001).

1.3.2.3 Control of other initiation factors

The factor eIF4G is controlled both through its phosphorylation and by its cleavage. Increased phosphorylation of the factor has been observed *in vivo* in response to stimuli that promote cell growth (Gingras et al., 1999). eIF4G can be phosphorylated *in vitro* by protein kinase A, protein kinase C and protease-activated kinase II, however the effect of these phosphorylations on eIF4G function is unknown (Gingras et al., 1999). eIF4G is also reported to be phosphorylated by S6 kinase, which is said to stimulate *in vitro* protein synthesis and binding of mRNA to the 43S complex (Rhoads, 1999).

Cleavage of eIF4G occurs during apoptosis and following picornaviral infection. In apoptosis there is rapid inhibition of protein synthesis, caused by the activation of specific proteases called caspases. It was found that caspase 3 acted to degrade eIF4G (Bushell et al., 1999). The factor is cleaved into 3 fragments called N-FAG, M-FAG and C-FAG (N-terminal, Middle, or C-terminal-Fragment of Apoptotic cleavage of eIF4G) (Figure 1.5). The M-FAG fragment contains the eIF4E, eIF3, and one eIF4A site. It is thought this

8
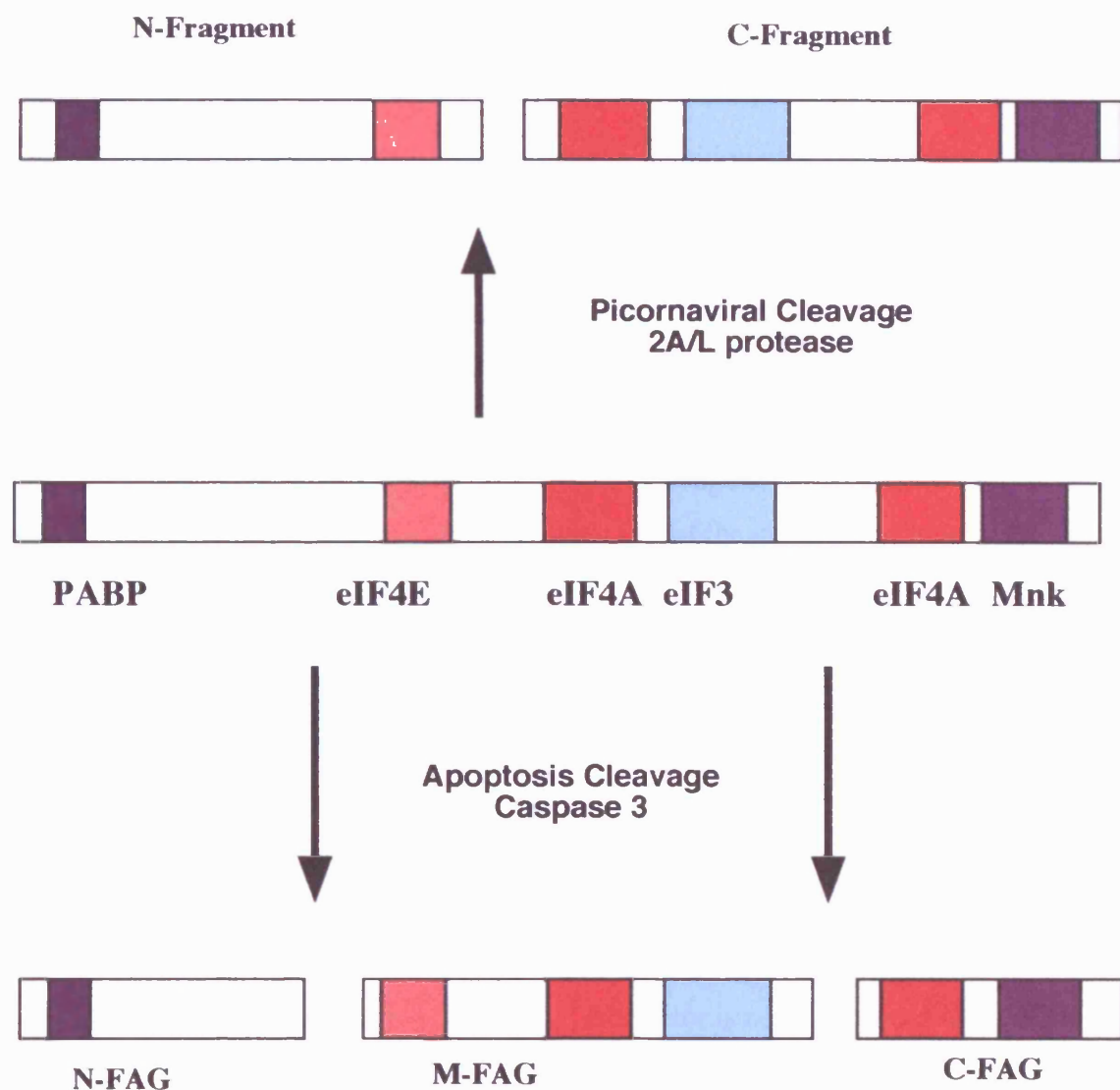
Figure 1.5. eIF4G cleavage during picornaviral infection and apoptosis. In picornaviral infection cleavage by proteases produces 2 fragments, removing the eIF4E and PABP sites. In apoptosis cleavage by caspase 3 produces 3 fragments.

fragment may still be able to support cap-dependent translation but at a lower efficiency due to the loss of one eIF4A site, and the PABP and Mnk sites. In addition the ability of eIF4E phosphorylation may be reduced due to the loss of the Mnk site (Bushell et al., 2000b). This cleavage has been found to occur for both eIF4GI and eIF4GII (Bushell et al., 2000a).

Picornaviral infection also causes a general inhibition of cap-dependent translation. This occurs through the cleavage of initiation factors by picornaviral proteases. eIF4G cleavage is carried out in rhinovirus and enterovirus infection by protease 2A, and in aphthoviruses by the leader (L) protease. The cleavage sites for these proteases lie very close (7aa apart) producing almost identical products (Kirchweger et al., 1994). eIF4G is cleaved into two fragments, removing the eIF4E and PABP binding sites from the eIF3, eIF4A, and Mnk-1 binding sites (Figure 1.5). This cleavage event is considered to inhibit cellular cap-dependent translation while allowing that of the virus to continue. Viral translation can continue with the C-terminal cleavage product, which can support picornaviral IRES translation, whereas cellular cap-dependent translation is inhibited as there is loss of the eIF4E and PABP sites. The factor eIF4GII has been found to be more resistant to cleavage by viral proteases than eIF4GI, and its cleavage coincides with the shut off of host cell protein synthesis (Gradi et al., 1998b). eIF4G cleavage by picornaviral proteases produces different products than apoptotic cleavage.

eIF4G activity may also be controlled by its homolog DAP-5. As described earlier, DAP5 contains binding sites for eIF3, eIF4A and Mnk1, but not eIF4E or PABP (Imataka et al., 1997; Pyronnet et al., 1999). The role of this factor is not fully understood. It has been reported to inhibit both cap-dependent translation, and cap-independent translation of the EMCV IRES (Imataka et al., 1997). This led to the suggestion that it may sequester factors such as eIF3 to decrease 43S complex formation or Mnk to block eIF4E phosphorylation (Pyronnet et al., 1999). However, it has also been reported to stimulate cap-independent translation (Henis-Korenblit, 2000) (see section 1.5.2). Finally, eIF4G is also found to be inhibited in heat shock by the binding of Hsp27, and the immobilisation of the protein in heat shock granules (Cuesta et al., 2000).

The role of other initiation factors in translational control is not fully understood. eIF3 has been found to be regulated by the protein P56. P56 is induced by interferon and dsRNA, and the binding of P56 to eIF3 was found to inhibit translation *in vitro* and *in vivo* (Guo et al., 2000). A possible regulation of eIF4B and eIF3 by phosphorylation has been

reported (Gingras et al., 1999). There is cleavage of eIF4B and eIF3(p35), and some eIF2α, and 4E-BP1 in apoptosis to produce discrete cleavage products (Bushell et al., 2000a). There are also reports of translational control through phosphorylation of ribosomal protein S6 (Rhoads, 1999).

1.3.3 Control of translation by mRNA elements

In addition to translational control via initiation factors there is translational control by *cis*-acting elements present in mRNA. These act to control the translation of specific messages, rather than the control of total cell protein synthesis. Elements exist in both the 5' and 3' UTR of mRNAs

1.3.3.1 Control by elements in the 5'UTR

Most eukaryotic mRNAs contain a 5'UTR sequence of 20-100 nucleotides (Kozak, 1987). Many features of the 5'UTR can affect translation. There is control of translation through leaky scanning, upstream AUGs, upstream open reading frames (uORFs) and secondary structure elements. In addition the 5'UTR length itself can affect the efficiency of translation initiation. It has been shown that longer 5'UTRs increase translation efficiency, possibly due to increased loading of ribosomes onto the RNA. Many mRNAs encoding genes involved in cell growth and differentiation such as growth factors, transcription factors and proto-oncogenes have been found to have 5'UTRs exhibiting translational control (Kozak, 1991).

Leaky scanning is the process when translation occurs both from the main initiator codon and downstream AUG codons (Figure 1.6A). It can occur if the AUG start codon is in a poor context or present close to the cap structure. The optimal consensus start site is GCCACC(AUG)G (Kozak, 1996). If a poor context is present it may be missed by some ribosomes, which will initiate at an AUG further downstream. Alternative start codons such as CUG are also poorly recognised and may result in leaky scanning. In addition a start codon present less than 40 nucleotides from the cap may not be recognised (Geballe and Morris, 1994). Leaky scanning can produce proteins with varying N-terminal extensions, or even the translation of two entirely different proteins from the translation of alternative reading frames.

5'UTR regions can contain upstream AUG codons that act to regulate translation. Upstream AUGs are present in less than 10% of all eukaryotic genes and are particularly found in genes involved in cell growth and differentiation (Morris and Geballe, 2000).

A. Leaky scanning

Main AUG
in poor
context

Some ribosomes
continue scanning

Downstream
AUG used

m7G —— AUG —— AUG —— AAAA

B. uORFs

Ribosomes must reinitiate

m7G —— uORF / AUG —— ORF —— AAAA

C. Secondary structure

40S binding/scanning inhibited

m7G —— AUG —— AAAA

D. Shunting

m7G —— uORF / AUG —— ORF —— AAAA
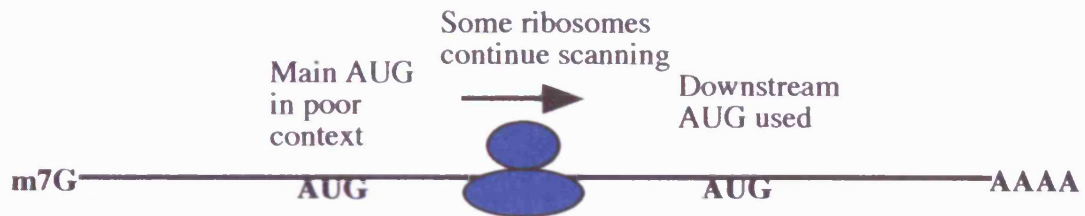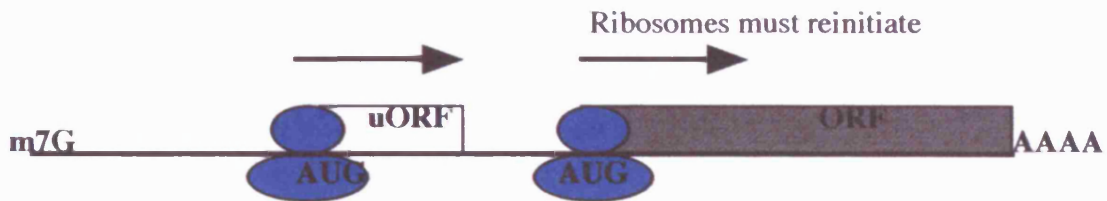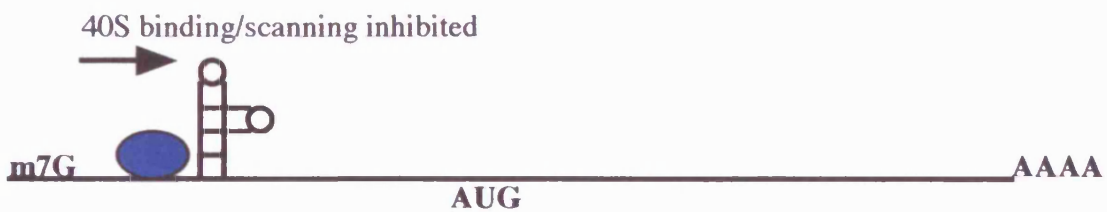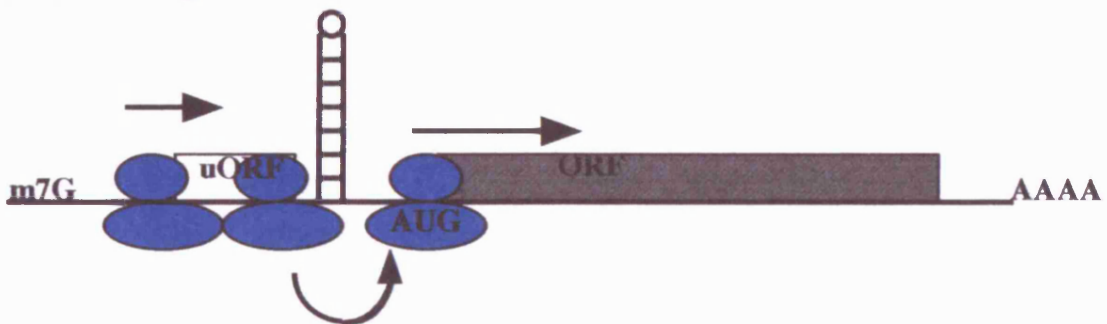
**Figure 1.6. Translational control through *cis*-acting elements in the mRNA 5'
untranslated region.** A)Leaky scanning acts to produce proteins with varying N-
terminal extensions. B) uORFs control the rate of translation from the main ORF. C)
Secondary structure in the 5'UTR acts to inhibit translation. D) Shunting allows the
ribosome to bypass uORFs.

Some upstream AUGs are closely followed by in frame stop codons and represent upstream open reading frames (uORFs) (Figure 1.6B). Most uORFs are short which favours the reinitiation at the main initiator AUG codon. When uORF lengths are increased there is a decrease in reinitiation, which may be due to a gradual loss of initiation factors from the ribosomal complex (Kozak, 2001). Not all of the ribosomes reinitiate at the main initiator codon thereby decreasing translation of the downstream cistron. It has been shown that increasing the length between the uORF and main ORF acts to increase reinitiation. This is thought to allow the ribosome time to join with the necessary initiation factors required for reinitiation (Geballe and Morris, 1994). The mechanism of translational control of the yeast gene GCN4 has been well documented. This gene is controlled by 4 uORFs in response to amino acid availability (Geballe and Morris, 1994). After translation of uORF1 ribosomes can reinitiate down stream with a high efficiency. However, uORFs 3 and 4 are inhibitory to reinitiation. It has been shown that uORF1 is translated and ribosomes resume scanning. Where translation recommences on the mRNA depends upon the ribosomes ability to become competent for translation again. If ribosomes quickly regain competency then initiation occurs at other uORFs and translation of GCN4 is inhibited. However, if ribosomes take longer to regain competency then translation begins at the GCN4 start codon. The ability of ribosomes to regain function depends on the availability of the ternary complex. Therefore in amino acid starvation the ternary complex is limited, ribosomes take longer to reacquire translation ability after translation of uORF1, and there is increased translation of GCN4 (Morris and Geballe, 2000).

Secondary structure elements in the 5'UTR effect translation efficiency (Figure 1.6C). The inhibitory effect seen is determined by the position and stability of the structure (Day and Tuite, 1998). Many genes encoding growth factors, transcription factors, signal transduction components and receptor proteins have GC rich 5'UTRs with the ability to form strong secondary structure (Kozak, 1991). If a strong secondary structure is formed which cannot easily be unwound by the helicase eIF4A, it will inhibit scanning and therefore initiation. Secondary structure can also inhibit 43S pre-initiation complex binding (Gray and Hentze, 1994).

Shunting can occur when the ribosome bypasses a secondary structure element and continues scanning downstream (Figure 1.6D). The CaMV (cauliflower mosaic virus) 35S RNA is translated via this mechanism (Futterer et al., 1993). The 5'UTR contains a hairpin structure and several uORFs, which would be expected to inhibit translation.

However, it has been shown that the ribosome translates uORF A and bypasses the hairpin structure to continue scanning downstream (Hemmings-Meiszczak and Hohn, 1999). Several other virus RNAs have also been found to be translationally controlled in this manner including adenovirus, and plant-infecting pararetroviruses. It is proposed that ribosome shunting acts to allow the ribosome to bypass uORFs, and allows translation in situations where cap-dependent translation is compromised when there is reduced secondary structure unwinding ability (Yueh and Schneider, 2000). Ribosome shunting has also been shown to occur in eukaryotic genes. For example, the heat shock protein hsp70 can be translated by this mechanism in heat shock (Yueh and Schneider, 2000).

The 5'UTR can contain specific sequences that act to control translation. An example is the Iron Response Element (IRE), which is present near the cap in a number of mRNAs. This element forms a stem loop structure that binds IRE-binding proteins (IRPs) to inhibit translation under certain conditions. The mRNA for the iron storage protein Ferratin contains an IRE sequence, which acts to inhibit translation when iron levels are low and the protein is not required (Hentze et al., 1987). The binding of IRP blocks the binding of the 43S pre-initiation complex to the mRNA (Pain, 1996). Two IRP proteins have been identified in eukaryotes. IRP1 and 2 are both regulated by iron availability and nitric oxide, whereas IRP1 is also regulated by oxidative stress (Gray and Wickens, 1998).

Finally, 5'UTRs can contain internal ribosome entry segments (IRESs). These elements allow translation to occur in a cap-independent manner, by recruiting ribosomes directly to an internal site within the mRNA. This method allows translation to occur in situations where cap-dependent translation is compromised, and occurs in both viral RNAs and cellular mRNAs. Internal ribosome entry is discussed in detail in section 1.4.

1.3.3.2 Control by elements in the 3'UTR

In addition to 5'UTR mechanisms, control elements also exist in the 3'UTR. This form of translational control is used during development, but has also been found to occur in other situations. The main mechanism of 3'UTR control is through effects on poly adenylation. Some mRNAs are repressed early in development by shortening of the poly (A) tail. The translation of these mRNAs is then activated by the action of 3'UTR elements. These elements are the conserved poly adenylation motif (AAUAAA) and U rich cytoplasmic polyadenylation elements (CPEs) (Pain, 1996). A protein called CPE-binding protein (CPEB) can bind to the CPE and promote poly adenylation.

The 15-lipoxygenase (LOX) mRNA is translationally controlled through a 3'UTR element. This mRNA contains a CU rich sequence known as a differentiation control element (DICE) that binds HnRNPK and PCBP1. The binding of these proteins acts to translationally silence the message until erythrocyte maturation when the enzyme is required (Ostareck et al., 1997).

1.4. Internal ribosome entry

1.4.1 Introduction

Translational control can occur by a structural element in the 5'UTR known as an internal ribosome entry site (IRES). The element directs ribosome entry at an internal site close to or upstream of the main initiator codon (Figure 1.7A). Ribosome entry has been found to be independent of the 43S pre-initiation complex binding the cap structure, and therefore it is also referred to as cap-independent translation. This mechanism of translation was discovered in viral RNAs, but is also used by a number of cellular mRNAs to allow translation in situations where cap-dependent translation is compromised.

1.4.2 Picornaviral IRES translation

Internal ribosome entry was first discovered in the picornaviral family. These viruses possess an RNA genome, containing a long (610-1200 nucleotide), GC rich 5'UTR with the capacity to form extensive secondary structure (Belsham and Sonenberg, 1996). In addition picornaviral RNAs do not possess a m7GpppG cap structure at their 5' end and their 5'UTRs contain a high number of upstream AUG codons. These qualities all act to decrease initiation factor association and inhibit ribosome scanning to the initiation codon. This prompted investigation into the translation mechanism involved.

Studies showed that an internal site within the 5'UTR was responsible for ribosome binding. This was determined using a dicistronic mRNA construct (Figure 1.7B). The fragment of interest of the poliovirus 5'UTR was placed into the intercistronic region, and its effect on translation monitored. Under conditions of cap-dependent translation, there is high expression of the upstream cistron and low expression of the downstream cistron as few ribosomes reinitiate at the downstream start codon. When the region of the poliovirus 5'UTR was inserted into the intercistronic region there was still high expression of the upstream cistron, but also enhanced translation of the second cistron (Pelletier and Sonenberg, 1988). This is due to the recruitment of ribosomes through the IRES element. Further experiments verified that internal ribosome entry was occurring and the RNA

13
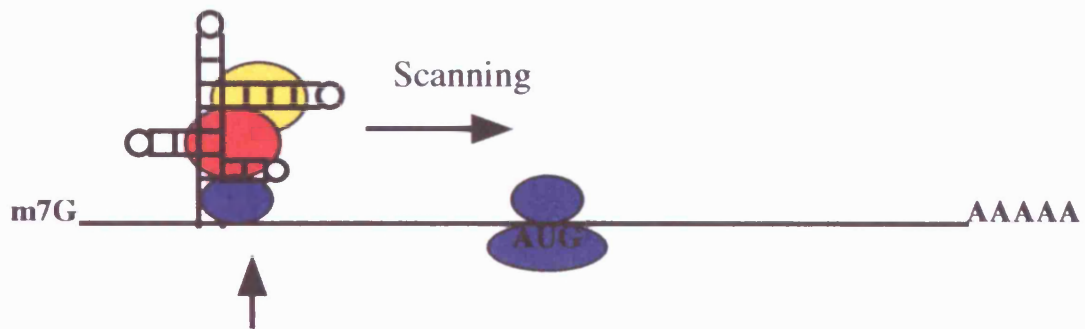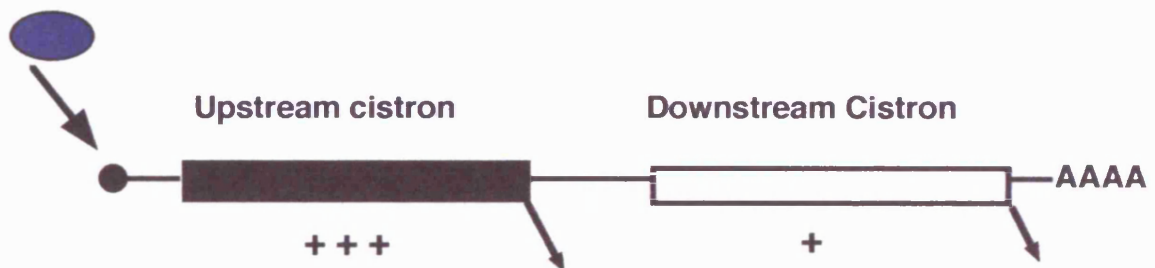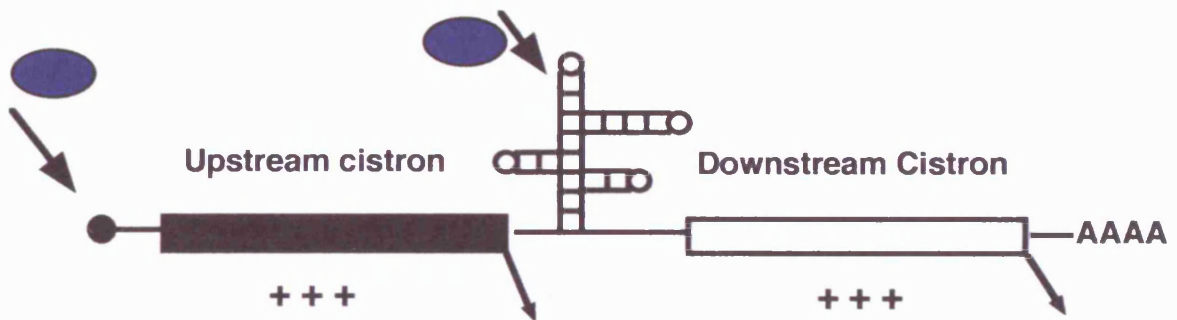
Figure 1.7. Internal ribosome entry and the dicistronic vector system. A) Internal ribosome entry. The 43S ribosomal subunits enter directly at the IRES structure, which can be at the AUG or up to 150 nucleotides upstream. The binding of the ribosomal subunit involves other initiation factors and *trans*-acting factors. B) The dicistronic vector system used to identify IRES elements. i) Under conditions of cap-dependent translation there is high translation of the upstream cistron and low translation of the downstream cistron. ii) When an IRES element is present in the intercistronic region there is enhanced translation of the downstream cistron as ribosomes are directly recruited.

region responsible was named an internal ribosome entry segment (IRES). The mechanism was then found to occur in other picornaviruses such as EMCV (encephalomyocarditis virus) (Jang et al., 1988), and it is now known that all picornaviruses are translated via this mechanism. Internal ribosome entry is independent of a free mRNA 5' end, with the 43S pre-initiation complex binding at the IRES site. This has been verified by experiments showing that the EMCV IRES can direct translation in a circular RNA (Chen and Sarnow, 1995).

The picornaviral IRESs are placed into two main groups due to their sequence and structural similarities, and their protein factor requirements. Type 1 IRES elements include enteroviruses (poliovirus) and rhinoviruses (Human rhinovirus), and type 2 the cardioviruses (encephalomyocarditis virus) and aphthoviruses (Foot and Mouth disease virus). There is also a third minor group of hepatitis A virus (Hellen and Sarnow, 2001).

1.4.3 Internal ribosome entry segments

Investigation of picornaviral IRESs showed that there was very little sequence homology within the virus groups, but a strong conservation of secondary structure (Jackson et al., 1990). The IRES region was found to consist of a pattern of stem loops organised into structural domains. The type I enterovirus and rhinovirus IRESs have been shown to consist of a series of 5 stem loops structures (Figure 1.8), and type II IRESs approximately 9 stem loop structures (Stewart and Semler, 1997). Within this structure several conserved secondary structure motifs were identified as being important for IRES function. These conserved motifs may act to maintain the correct tertiary structure for recognition of the IRES by the translational machinery (Ramos and Martinez-Salas, 1999). This maintenance of tertiary structure could occur through RNA-RNA interactions or with the involvement of protein factors.

There was also identification of a conserved polypyrimidine tract at the 3'end of picornaviral IRESs, followed by a spacer region of average length 15-25nt, and an AUG site. For type 2 IRESs this AUG is the initiation codon, whereas for type 1 IRESs the initiation codon is located 30-150 nucleotides downstream requiring scanning or shunting to occur.

Other common features have also been identified. GNRA motifs have been identified in stem loop IV of type 1 IRESs and stem loop I of type 2 IRESs, and A/C rich sequences in stem loops IV and V of type 1 IRESs and stem loop 1 of type 2 IRESs
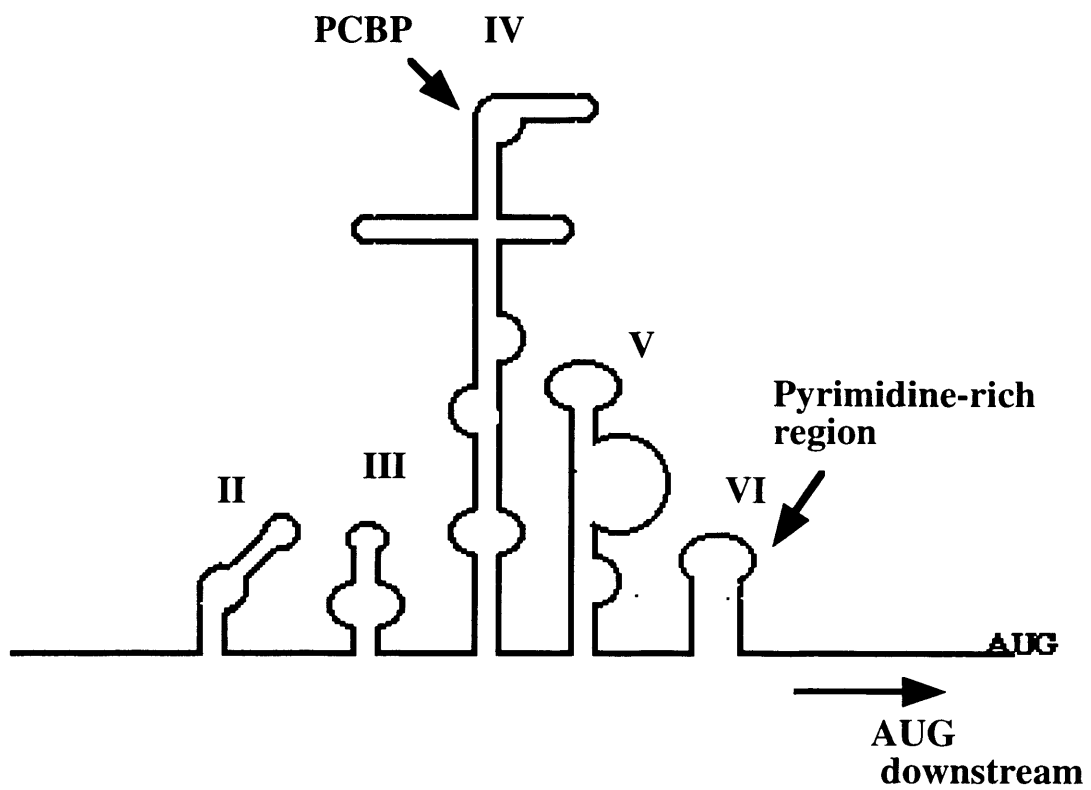
Figure 1.8. The poliovirus IRES structure . The region contains 5 stem loop structures, with a conserved polypyrimidine tract at the 3'end. The initiaton codon is located 150 nucleotides downstream requiring scanning or shunting to occur.

(Stewart and Semler, 1997). Mutations in the GNRA motifs in EMCV and PV IRESs have been shown to inhibit IRES activity (Robertson et al., 1999).

1.4.4 The involvement of canonical initiation factors

Considerable work has been centred on the protein factors required for picornaviral IRES translation. From this work it was found that almost the same set of factors that are important for cap-dependent translation are also important for cap-independent translation (Hunt et al., 1999). The EMCV IRES was found to require the factors eIF2, eIF3, and eIF4F for 48S complex formation *in vitro*. This complex formation was stimulated by eIF4B (Pestova et al., 1996a). As well as initiation factors, EMCV IRES translation also required Met-tRNA$_i$, GTP and ATP. Further work showed that the eIF4F could be replaced by eIF4A and the central third of eIF4G (457-932), and the interaction between these two factors was required to form an active complex for IRES translation (Lomakin et al., 2000; Pestova et al., 1996b). This indicates that the helicase function of eIF4A is important, which is verified by the requirement for ATP. eIF4A may act to unwind mRNA structure or have some function in structural IRES rearrangements for ribosome binding (Hellen and Sarnow, 2001). It was found that the central domain of eIF4G bound directly to the J-K domain of the EMCV IRES. eIF4G is thought to promote the binding of the 43S complex to the IRES (Pestova et al., 1996a). The FMDV IRES has been found to have the same requirements for eIF2, eIF3, eIF4A, eIF4B and eIF4F (Pilipenko et al., 2000). Again an interaction of eIF4G and eIF4A was required for IRES recognition, and eIF4F was found to bind the J-K domain of the IRES.

As described earlier, picornaviral infection causes a general inhibition of cap-dependent translation. One of the major ways this is achieved is through cleavage of eIF4G by picornaviral proteases into two fragments (see figure 1.5). The C-terminal cleavage product produced includes the central fragment of eIF4G that can support EMCV translation. Therefore, this cleavage event is considered to inhibit cellular cap-dependent translation while allowing that of the virus to continue. EMCV and PV infection also causes a dephosphorylation of 4E-BPs, which acts to decrease free eIF4E (Gingras et al., 1996). This also inhibits cellular cap-dependent translation, and allows picornaviral cap-independent translation to continue.

The only IRES among the picornaviral family that does not agree with this factor requirement is the Hepatitis A virus IRES. This IRES is strongly inhibited when eIF4G is

cleaved, suggesting a different mechanism operates (Borman and Kean, 1997). The HAV IRES requires components of the eIF4F complex that interact with the N-terminal fragment of eIF4G, such as eIF4E (Ali et al., 2001; Borman et al., 2001).

1.4.5 The role of IRES *trans*-acting factors

The two classes of picornaviral IRES were discovered to have different requirements for efficient translation. Type 2 picornaviral IRESs were found to be very efficient in *in vitro* translation systems such as rabbit reticulocyte lysate, whereas type 1 IRESs (such as poliovirus) needed HeLa cell lysate for efficient translation. This suggested that type 1 IRESs had a requirement for additional protein factors over and above those found in rabbit reticulocyte lysate and prompted investigation into these additional protein factors (Jackson et al., 1995). One of the first IRES *trans*-acting factors (ITAFs) identified was a 57kDa protein that UV cross-linked to stem loop H of the EMCV IRES, and was required for IRES mediated translation (Jang and Wimmer, 1990). Further work showed this protein to be a nuclear RNA binding protein called PTB (Polypyrimidine tract binding protein) (Borman et al., 1993; Hellen et al., 1993). PTB was required for efficient PV and EMCV IRES translation. Using HeLa cell-free lysates that had been immunodepleted of PTB the translation from these IRESs was found to be inhibited (Hellen et al., 1993). PTB was also identified separately as an activity that enhanced HRV and poliovirus IRES translation (Hunt and Jackson, 1999). PTB has other roles in the cell as a negative regulator of alternative splicing, in 3' end processing and RNA localization. The protein contains 4 RNA recognition motifs and has been found to interact with short pyrimidine motifs such as UCUUC/CUCUCU (Yuan et al., 2002). PTB is also known as heterogeneous nuclear ribonucleoprotein I (hnRNP I).

A number of other cellular *trans*-acting factors have since been discovered. The requirement for these factors differs considerably between IRESs. They are La, UNR (upstream of N-ras), PCBP 1 and 2 (Poly (rC) binding protein 1 and 2), ITAF 45 (IRES *trans*-acting factor 45), and Nucleolin. UNR was discovered through investigation of an activity in HeLa cells required for PV and HRV IRES activity (Hunt et al., 1999). This activity was found to be from proteins of 38kDa and 97kDa. The 97kDa protein was found to be UNR, a known RNA binding protein with 5 cold shock domains. UNR was identified in this study along with PTB. Using *in vitro* translation assays the addition of UNR, PTB and another factor PCBP2, was found to stimulate HRV IRES activity to the level

observed with the addition of HeLa cell extract (Hunt et al., 1999). However, it was found that UNR had only a small effect on PV IRES translation, indicating PV and HRV may have different requirements for *trans*-acting factors. The 38kDa protein was found to have no RNA binding activity and named UNRIP (UNR interacting protein) due to its UNR binding ability. This is a novel member of the GH-WD repeat family that contains 6 WD-40 repeats. Proteins containing these repeats are often associated with other proteins and are thought to act as bridging proteins for protein complexes. However, no effect on IRES translation was observed for recombinant UNRIP and its role in IRES translation has yet to be determined.

The La autoantigen was discovered from its binding and stimulation of the poliovirus IRES (Meerovitch et al., 1993). This is a known 52kDa RNA binding protein, which has roles in transcription termination and mRNA maturation. Interestingly it was found that La protein was nuclear, but was redistributed to the cytoplasm after poliovirus infection and in apoptosis. La was also found to be required for efficient translation of the EMCV IRES (Kim, 1999).

PCBP 1 and 2 are 38kDa proteins that were found to bind to stem loop IV of the poliovirus IRES (Gamarnik and Andino, 1997). These proteins were also shown to effect poliovirus IRES efficiency, which was reduced in PCBP depleted HeLa extracts. It was found that PCBP2 had an important role in internal initiation directed by PV and HRV IRESs (type 1), but was dispensable for type 2 IRESs of EMCV and FMDV (Walker, 1999). PCBP1/2 are also known as hnRNP E1/2, or αCP1/2 (α-globin mRNA poly(C) rich segment binding protein). They belong to the KH domain family of proteins, which have 3 KH (hnRNP K homologous) RNA binding domains and a binding preference for poly C (Leffers et al., 1995). The PCBPs are involved in both the stabilization and translational regulation of several cellular mRNAs, and are known to dimerize *in vivo*. PCBP2 has several known splice variants.

ITAF 45 is a 45KDa protein found to be required by the FMDV IRES (Pilipenko et al., 2000). It is the known murine proliferation-associated protein (MPP1), whose translation is up regulated in response to proliferation signals. This protein has homology with type II methionine aminopeptidases, but does not have any classical RNA binding motifs. ITAF45 was found to be required, along with PTB, for 48S complex assembly on the FMDV IRES. The protein bound the IRES at multiple sites around the base of domain I, and along with PTB enhanced eIF4G and eIF4A binding.

Nucleolin is a 110kDa protein, also called C23. It was found to stimulate rhinovirus IRES mediated translation *in vivo* and poliovirus IRES translation *in vitro* (Izumi et al., 2001). Nucleolin is a nuclear RNA binding protein that has been shown to translocate to the cytoplasm following poliovirus infection (Waggoner and Sarnow, 1998).

Another factor GAPDH (glyceraldehyde 3'phosphate dehydrogenase) has been found to have a role in IRES translation, but does not have the stimulatory role observed for other *trans*-acting factors. GAPDH binds to multiple sites of the Hepatitis A virus IRES (Schultz et al., 1996). Interestingly this factor competed with PTB for binding, and was found to destabilize stem loop IIIa of the IRES. Further work showed that GAPDH suppressed HRV IRES activity *in vivo* (Yi et al., 2000). GAPDH is an abundant cytoplasmic protein, and has been shown to bind AU rich RNA sequences (Nagy and Rigby, 1995).

ITAFs are considered to aid IRES elements to adopt and maintain the correct structure for initiation factor and 40S ribosomal subunit binding. It is thought that there are such a wide variety of ITAFs due to the range of IRES structures (Pestova et al., 2001). Many of these factors have multiple RNA binding domains, and may therefore contact the RNA at multiple points in the IRES to stabilize the secondary and tertiary structure (Hellen and Sarnow, 2001). Several IRES elements have been found to require more than one *trans*-acting factor for maximum activity. It has also been found that many of these factors are nuclear, suggesting that they may bind the IRES element in the nucleus or relocate to the cytoplasm (Vagner et al., 2001).

1.4.6 Mechanism of translation of other viral IRESs

IRES elements have also been identified in other viral RNAs. These non-picornaviral IRESs have been found to have different mechanisms and initiation factor requirements. One group includes Hepatitis C virus (HCV), GB virus B (GBV-B), classical swine fever virus (CSFV), and Bovine viral diarrhoea virus (BVDV). These viruses are all members of the *Flaviviridae* family and their IRESs share common features. HCV and CSFV IRESs have both been shown to contain pseudoknot structures upstream of the initiation codon that are important for IRES function (Wang et al., 1995; Rijnbrand et al., 1997). In addition HRV IRES function is reported to be effected by the downstream coding sequence (Reynolds et al., 1995). Recent reports suggest this requirement is not dependent on exact sequence, but a requirement for an unstructured region to allow IRES interaction with the 40S ribosomal subunit at the initiation codon (Rijnbrand et al., 2001).

HCV and CSFV IRESs were found to bind 40S ribosomal subunits without the presence of eukaryotic initiation factors (Pestova et al., 1998b). The 40S ribosomal subunit bound with the initiator AUG positioned in the ribosomal P site. eIF3 was found not to be required for formation of the 48S complex, but due to its 40S binding is likely to exist in the complex *in vivo*, and has been shown to independently bind to the IRES (Pestova et al., 1998b). The assembly of functional 48S complexes required eIF2, Met-tRNA$_i$ and GTP, but not eIF4A, eIF4B or eIF4F. eIF3 was required for formation of 80S ribosomes. The assembly of a functional complex involved multiple IRES interactions, showing the IRES structure itself is important. Further work shows that the HCV IRES binding to the 40S ribosome induces a conformational change in the ribosomal subunit. This could play some role in the activation of translation without initiation factors (Spahn et al., 2001).

HCV IRES translation has been found to involve *trans*-acting factors. La was found to be required for efficient translation of the HCV IRES (Ali and Siddiqui, 1997). The binding of La to HCV was found to involve the initiation codon, and it was proposed that La could act as a molecular chaperone to position the 48S complex at the initiation codon (Ali et al., 2000). In addition a protein called HnRNP L (heterogeneous nuclear ribonucleoprotein L) was reported to bind to the 3' border of the HCV IRES and its binding was related to the IRES efficiency (Hahm et al., 1998). HnRNP L is a 68Kda protein closely related to PTB (hnRNP I), and contains the same 4 RNA recognition motifs. hnRNP L is reported to be mainly localised in the nucleus, though it has been found in the cytoplasm during poliovirus infection. There are also reports of PCBP1 and PCBP2 interacting with the HCV IRES (Spangberg and Schwartz, 1999). *Trans*-acting factors may act to maintain the HCV IRES in the correct structure for the binding of the ribosomal subunit.

Another example of a different IRES mechanism has been discovered in the insect picorna-like virus family. This family includes Cricket paralysis virus (CrPV), and Plautia stali intestine virus (PSIV), and Himetobi P virus (HiPV). These viral RNAs contain two separate open reading frames coding for the non-structural and structural proteins. Work on PSIV found that the structural protein translation from ORF2 was under the control of an IRES, and translation began from an unusual initiation codon of CUU (Sasaki and Nakashima, 1999). Further work showed that initiation did not begin with methionine, but began with glutamine encoded by the second codon (CAA). The IRES region was predicted to form a pseudoknot structure immediately upstream of ORF2, and involving

the CUU codon. This structure was found to be essential for IRES translation. It was proposed that the ribosome could bind the IRES structure in the P site, allowing the translation to begin with glutamine(CAA) at the A site(Sasaki and Nakashima, 2000).

Further work showed this unusual mechanism also occurred in CrPV. It was found that both CrPV ORFs were preceded by IRES elements, with the second intergenic region IRES (IGR-IRES) showing higher activity to produce higher amounts of structural proteins in infected cells (Wilson et al., 2000b). This IGR-IRES was found promote translation of a protein whose first amino acid was alanine encoded by a GCU codon. Again the preceding codon (CCU) was found to be important, and that the interaction of this CCU with the IRES was important for IRES function. These results raised the possibility that there was no requirement for ternary complex, which introduces Met-tRNA$_i$. It was found that 40S subunits could bind stably to CrPV IGR IRES in the absence of any initiation factors including eIF2 and Met-tRNA$_i$ (Wilson et al., 2000a). Further work showed that 80S ribosomes could form without eIF3, eIF5, eIF5B or GTP hydrolysis. It was found that ribosomes bound the IRES at the pseudoknot structure, with the CCU codon positioned in the P site and the GCU in the A site.

1.5 Cellular IRESs

1.5.1 Introduction

Further work brought about the discovery of cellular mRNAs containing IRESs. The first cellular IRES discovered was in the immunoglobulin heavy chain binding protein (BiP) mRNA (Macejak and Sarnow, 1991). This was discovered due to the continued translation of BiP in picornaviral infected cells when cap-dependent translation was inhibited (Sarnow, 1989). The 5'UTR of BiP was found to be 220nt, demonstrating that shorter regions of RNA could promote internal ribosome entry. This work also established that uninfected mammalian cells had the protein factors necessary for IRES translation, and no viral proteins were required. Further work on the BiP IRES showed it to be present in the 3'end of the 5'UTR at residues 129-220 (Yang and Sarnow, 1997).

After the discovery of the BiP IRES there was identification of other cellular IRESs. In the last few years a high number have been identified (Table 1.1), suggesting the mechanism may be an important control point for cellular translation. Many of these IRES containing mRNAs encode proteins involved in cell regulation and growth, such as transcription factors and growth factors. The IRES elements appear to maintain translation

of the proteins in situations where cap-dependent translation is inhibited, and the protein is required. Some mRNAs, such as c-*myc*, have been shown to be translated by both cap-dependent and cap-independent translation, with the IRES maintaining expression of the protein in some situations of decreased cap-dependent translation (Stoneley et al., 2000a; Stoneley et al., 2000b).

Interestingly genes such as eIF4G, DAP5 and La that are involved in translation have been found to contain IRES elements (Johannes and Sarnow, 1998; Henis-Korenblit, 2000; Carter and Sarnow, 2000). This suggests the levels of these proteins may be maintained in some situations when there is decreased cap-dependent translation, and they may act to maintain translation of other mRNAs.

Gene Type	Name
Transcription factors	Antennapedia, Ultrabithorax, c-myc, N-myc, MYT2, AMLI/RUNXI, Gtx, c-Jun, Mnt, Nkx6.1, NRF, YAP1, Smad5, HIP-1alpha
Translation factors	La, eIF4GI, TIF4631, DAP5
Apoptotic factors	XIAP, APC, Apaf-1, BAG-1
Growth Factors and receptors	FGF2, PDGF2, VEGF-A, IGF-II, Oestrogen receptor alpha, IGF-I receptor, Notch2
Cytoskeleton proteins	ARC, MAP2
Kinases and related	Pim1, p58/PITSLRE, alpha CaM kinase II, CDK inhibitor p27, PKCδ
Channels/Transporters	KV.14, BetaF1-ATPase, Cat-1
Other	BiP, Connexin 43 and 32, Cyr61, ODC, Dendrin, Neurogranin/RC3, NBS1, FMR1, Rbm3, NDST

Table 1.1. Cellular genes reported to be controlled by internal ribosome entry.

Cellular IRESs have been found to be shorter than picornaviral IRESs, and do not appear to share any secondary structure motifs. A conserved Y-type stem loop motif has been identified in the BiP, Antennapedia and FGF2 (Fibroblast growth factor-2) IRESs (Le and Maizel Jr, 1997). However this is based solely on computer secondary structure predictions rather than any experimental data. The only published cellular IRES structure

from experimental data is for the c-*myc* IRES, which was found to consist of 2 domains: a highly folded region of 210 nucleotides containing a pseudoknot structure, and a separate stem loop structure further 3' (Le Quesne et al., 2001) (Figure 1.9). This suggests a more complex structure is required for internal ribosome entry. The IRES was found to have a modular structure as no single mutations or deletions were found to completely inactivate the IRES.

There are other reports of cellular IRESs having a modular structure. Work on the Gtx IRES found that several non-overlapping regions of the 5'UTR displayed IRES activity. Further work showed that a region of 9nt that is 100% complementary to 18S rRNA could promote internal ribosome entry (Chappell et al., 2000a). One copy of this module was found to promote IRES activity to only 2.5-fold. However, when 10 copies of this module were used in a dicistronic reporter construct there was a 570-fold increase in IRES activity. This 9nt module was also shown to cross-link to the 40S ribosomal subunit. The action of sequences complementary to rRNA in translation has similarities with the prokaryotic translation initiation mechanism, where ribosomes are recruited due to Shine-Dalgarno sequence complementary to 16S RNA. It is suggested that cellular IRESs may be made up of several modules that act to recruit the ribosome to the mRNA by direct base pairing between the rRNA and mRNA. Sequences complementary to 18S RNA have also been discovered in other IRESs although the effect of them has not been investigated.

1.5.2 Protein factor requirements

Investigation of canonical initiation factor requirements has concentrated on the picornaviral IRESs, and very little is known for the cellular IRESs. It has been shown that overexpression of eIF4G slightly increased ODC (ornithine decarboxylase) IRES activity (Hayashi et al., 2000). However, in some situations where cellular IRES translation is maintained there is cleavage of this factor, suggesting the cleaved form of eIF4G that supports picornaviral translation may support cellular IRES translation. The eIF4G homolog DAP5 may have a role in cap-independent translation of cellular messages. It was found that the addition of recombinant DAP5 to rabbit reticulocyte lysate assays stimulated translation from its own IRES (Henis-Korenblit, 2000). During apoptosis the translation of DAP5 is maintained by IRES translation, and also it is cleaved in its C-terminus to an 86kDa protein that can still complex with eIF4A and eIF4E. DAP5 IRES translation was found to be stimulated *in vitro* to a greater degree by the addition of the apoptotic caspase
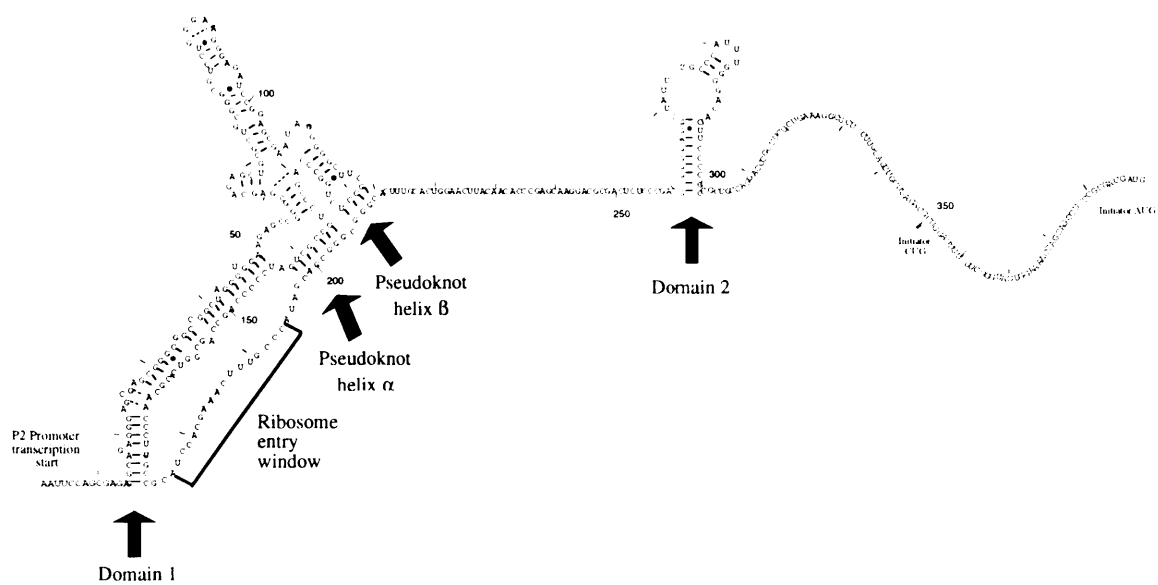
Figure 1.9. Structure of the c-*myc* IRES. Domain 1 contains a highly folded region including a pseudoknot structure. Domain 2 contains a stem loop structure. The ribosome entry window is located between nucleotides 177-194. Diagram adapted from (Le Quesne et al., 2001).

cleaved form of DAP5 (p86) than by the addition of the full length form. More recently there has been report of the DAP5 (p86) effecting translation from other IRESs that are active in apoptosis, such as XIAP (X-linked inhibitor of apoptosis), c-*myc*, and Apaf-1 (apoptotic protease activating factor 1) (Henis-Korenblit et al., 2002). The expression of DAP5 (p86) stimulated translation from these IRES elements *in vivo*. In addition the apoptotic caspase cleaved form of eIF4G (M-FAG) was not found to stimulate IRES translation in this system. However, it is not known if DAP5 (p86) is acting directly on the IRES (by replacing eIF4G) to stimulate translation. Another report demonstrates that IRES translation of c-*myc* and DAP5 can continue in DAP5 -/- cells showing DAP5 is not essential for IRES translation of these mRNAs (Yamanaka et al., 2000).

The majority of work on the factor requirements of cellular IRESs has concentrated on the *trans*-acting factors required. Cellular IRESs have been found to have different activities between cell lines, suggesting a requirement for factors present in these cell lines in varying quantities. In addition very few cellular IRES have been found to be active *in vitro* in rabbit reticulocyte lysate assays. This is similar to results seen for type 1 IRESs, suggesting a requirement for additional *trans*-acting factors. Several cellular IRESs have been reported to require a nuclear event. The c-*myc* IRES was found not to be active when mRNA was introduced directly into the cytoplasm (Stoneley et al., 2000b). Therefore the binding of a nuclear factor may be required for efficient translation.

The Apaf-1 IRES has been shown to require PTB and UNR for activity (Mitchell et al., 2001). Both factors were found to stimulate IRES activity *in vitro,* and were found to have an additive effect when added together. UNR and PTB were also found to stimulate Apaf-1 IRES activity *in vivo*. UV cross-linking experiments showed that PTB only bound the IRES in the presence of UNR, suggesting that the binding of UNR allowed the IRES to adopt the correct structure for PTB binding.

The XIAP IRES has been found to require the *trans*-acting factor La. The protein was found to bind to the IRES *in vivo* and *in vitro*, and a decrease in La availability was found to cause a decrease in XIAP IRES activity (Holcik and Korneluk, 2000). La is also required by the BiP IRES. The protein bound to the IRES element, and acted to enhance translation *in vivo* and *in vitro* (Kim et al., 2001).

The protein hnRNP C was found to bind the c-*sis* (B chain of Platelet derived growth factor) IRES (Sella et al., 1999). The cytoplasmic binding of the factor was found to be differentiation induced, suggesting a role for this protein in c-*sis* IRES translation,

which has been found to increase in differentiation. However, no effect on the activity of the IRES has been reported. The 43kDa factor was found to bind to a U-rich region of the IRES. This factor is a novel factor that has not previously been shown to interact with a picornaviral IRES.

As is found with picornaviral IRESs, cellular IRESs appear to differ in their requirements for *trans*-acting factors, and the identification of hnRNP C suggests that novel *trans*-acting factors may be involved whose upregulation or relocation may occur in some cellular situations to promote IRES translation.

1.5.3 Function of cellular IRESs

Many cellular IRESs have been found to be active in situations where cap-dependent translation is inhibited, such as heat shock, apoptosis, cellular stress, mitosis, and following DNA damage. Cellular IRESs appear to maintain translation of proteins that are required in these situations. Cap-dependent translation is inhibited by the control of initiation factors, especially through control of eIF2 and eIF4E activity (section 1.3.2). eIF2α phosphorylation acts to decrease protein synthesis by reducing ternary complex formation. This phosphorylation is carried out by 4 kinases which increase eIF2α phosphorylation under conditions of low heme, in viral infection, in amino acid and serum starvation, and following heat shock (Pain, 1996). The activity of eIF4E also decreases protein synthesis by effecting formation of the 48S pre-initiation complex. There is decreased eIF4E activity in situations such as growth arrest, heat shock, mitosis, viral infection, amino acid starvation, and serum deprivation.

Several IRESs have been found to be active under these conditions. The VEGF (vascular endothelial growth factor) IRES has been found to remain functional in hypoxia, which inhibits cap-dependent translation (Stein et al., 1998). This is significant as the VEGF protein is required in this situation to promote angiogenesis. The Rbm3 IRES is found to have enhanced activity in mild hypothermia (Chappell et al., 2001), the BAG-1 (Bcl-2-associated athanogene-1) IRES is active following heat shock (Coldwell et al., 2001) and the c-*myc* IRES following DNA damage (Subkhankulova et al., 2001).

PDGF2/c-sis (B chain of Platelet derived growth factor) IRES activity is increased in differentiating cells (Bernstein et al., 1997). A two-fold enhancement of IRES activity was seen in TPA-induced cell differentiation. It was also found that a 43KDa factor, hnRNP C, bound to the IRES, and that its binding activity relocated to the cytoplasm in

differentiated cells (Sella et al., 1999). This suggested hnRNPC may be a differentiation induced *trans*-acting factor for c-*sis* IRES translation.

Several cellular IRESs have been found to have activity in G2/M of the cell cycle when cap-dependent translation is inhibited. The ODC IRES maintains expression of the gene in the G2/M stage of the cell cycle (Pyronnet et al., 2000). The continued expression of ODC would promote elevated levels of polyamines, which are implicated in mitotic spindle formation and chromatin condensation. The p58^{PITSLRE} protein has been found to be under the control of an IRES, which acts to maintain expression of it in G2/M of the cell cycle (Cornelis et al., 2000). The PITSLRE mRNA produces two different isoforms from the use of alternative AUG codons, and it is expression of the smaller isoform that is under control of an IRES. This isoform has a role in growth inhibition.

The cat-1 (cationic amino acid transporter-1) IRES has been found to be active in amino acid starvation when cap-dependent translation is compromised (Fernandez et al., 2001). Cat-1 encodes a transporter of lysine and arginine, whose up-regulation would be required in amino acid starvation. It was found that after 12 hours of amino acid starvation cap-dependent translation decreased by 70% whereas cat-1 IRES activity increased by 3.5 fold. Amino acid starvation caused a dephosphorylation of eIF4E, which correlated with the decrease in cap-dependent translation. In addition there was increased phosphorylation of eIF2α, which reduces formation of ternary complex. This increase in eIF2α phosphorylation occurred during the first hour of amino acid starvation and then decreased gradually after this. Therefore it did not correlate with IRES activity and could not be directly regulating the cat-1 IRES. Further work found that increased phosphorylation of eIF2α by GCN2 kinase was required for the increase of IRES activity in amino acid starvation (Fernandez et al., 2002b). It was also found that glucose starvation activated the cat-1 IRES (Fernandez et al., 2002a). This activation was dependent on the phosphorylation of eIF2α by PERK. In addition dsRNA stimulated the IRES. The dsRNA replicates viral infection, and this activation required phosphorylation of eIF2α by PKR. Therefore multiple pathways were responsible for eIF2α phosphorylation, and IRES activation. It is not known how the activation of eIF2α phosphorylation, and the decrease in ternary complex would act to stimulate cellular IRES translation. The cat-1 IRES region also contains an uORF, the translation of which is proposed to be required for IRES activity.

25

Picornaviral infection leads to a decrease in cap-dependent translation through the cleavage of initiation factors such as eIF4G and the dephosphorylation of 4E-BP1. The natural mRNAs of c-*myc*, BiP, and eIF4G were found to be polysome associated in picornaviral infected cells, suggesting their translation is maintained in picornaviral infection (Johannes and Sarnow, 1998). Several cellular IRESs have been shown to be active under conditions of picornaviral infection. The XIAP IRES is maintained under conditions of increased viral protease 2A, which cleaves eIF4G (Holcik et al., 2000).

During apoptosis there is an inhibition of protein synthesis. This occurs through the activation of caspases that cleave initiation factors such as eIF4G, and through phosphorylation changes. The c-*myc*, and DAP5 IRESs are known to produce continued expression of the proteins during apoptosis (Stoneley et al., 2000a; Henis-Korenblit, 2000). The XIAP IRES acts to increase expression of the gene after low dose irradiation, and maintain expression following serum starvation (Holcik et al., 1999). XIAP is an anti-apoptotic gene, and the IRES mediated translation of it would protect cells against apoptosis induced by serum starvation and radiation.

Several cellular IRESs have been found to be involved in developmental regulation. The activity of the Ultrabithorax and Antennapedia IRESs were investigated in transgenic Drosophila and found to be highly developmentally regulated (Ye et al., 1997). Both genes encode developmentally important genes, suggesting a role for IRES elements in translational control during development. FGF2 and c-*myc* IRESs were also found to be involved in developmental control. The FGF2 IRES was found to be active during embryogenesis in transgenic mice, showing strong tissue specificity (Creancier et al., 2000). The FGF2 IRES also showed high activity in adult brain. These findings were said to fit with the proposed role of FGF2 in embryogenesis and the adult CNS. The c-*myc* IRES was found to be active in E11 and E16 transgenic mouse embryos, but inactive in adult tissues (Creancier et al., 2001). c-*myc* is widely expressed in embryogenesis for the promotion of cell proliferation, suggesting its regulation in embryogensis may be controlled by its IRES element.

IRES elements have also been suggested to be involved in local control of translation in the cell. The translation of dendritically localised neuronal mRNAs has been found to be controlled by IRES elements (Pinkstaff et al., 2001). The mRNAs of ARC, alpha CaM Kinase II, MAP2, dendrin, RC3 are all localised in dendrites, and have been shown to contain IRESs. The RC3 IRES was more efficient in dendrites than in the cell

body. This could act to rapidly and locally synthesize proteins required to strengthen active synapses.

1.6. c-*myc*

1.6.1 Introduction

c-*myc* is part of the Myc family of genes, which encode transcription factors. The family also includes N, L, S and B-*myc*. c-*myc* was originally discovered as the cellular homolog of an oncogene in avian myelocytomatosis virus MC29 (Vennstrom et al., 1982). It is an immediate early gene and acts as a heterodimer with the protein Max in the Myc/Max/Mad network of transcription factors (Grandori et al., 2000). This network plays a role in cell proliferation, differentiation and death.

1.6.2 c-*myc* structure

The c-*myc* gene consists of 3 exons, the first of which is mainly non-coding (Figure 1.10A). It encodes 3 protein isoforms with varying N-terminal extensions from the use of alternative start sites. The main start sites used are CUG and AUG, which lie 15 codons apart on the DNA, and produce Myc-1 (67kDa) and Myc-2 (64kDa) proteins. Both isoforms contain the active transcriptional transactivator domain at the N-terminus, as well as the Leucine Zipper, Helix-Loop-Helix, and region of basic amino acids at the C-terminus (Dang, 1999) (Figure 1.10B). The basic region is involved in DNA binding through the DNA major groove, whereas the Helix-Loop-Helix and Leucine Zipper motifs are involved in dimerization of the protein with its partner Max. These proteins have both Myc Box I and Myc Box II in the transactivation domain, which are conserved regions and found to be important for function. Myc box 2 has been shown to be essential for c-*myc* function and Myc Box I is a major phosphorylation site of the protein (Sakamuro and Prendergast, 1999). The other 45kDa isoform of the protein is known as Myc-S (Spotts et al., 1997). This protein is missing most of the transactivation domain except for Myc Box 2. The role of the different isoforms in the cell is not fully understood, but they have been shown to have different activities. Overexpression of Myc-1 in the absence of Myc-2 inhibits cell growth, whereas the normally predominant Myc-2 is responsible for growth and proliferation (Ryan and Birnie, 1996). The ratio of Myc-1: Myc-2 in the cell is therefore important. Myc-S has been shown to have weak repression activity and some transactivation activity (Hirst and Grandori, 2000).

A
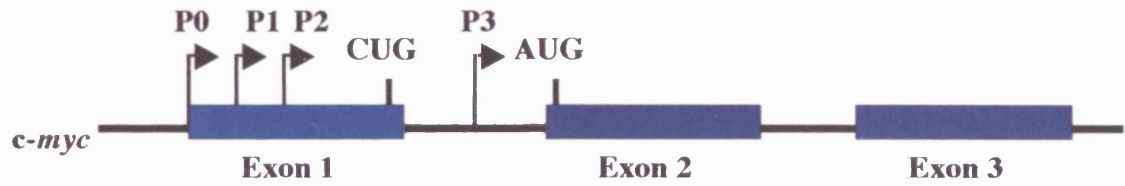
B
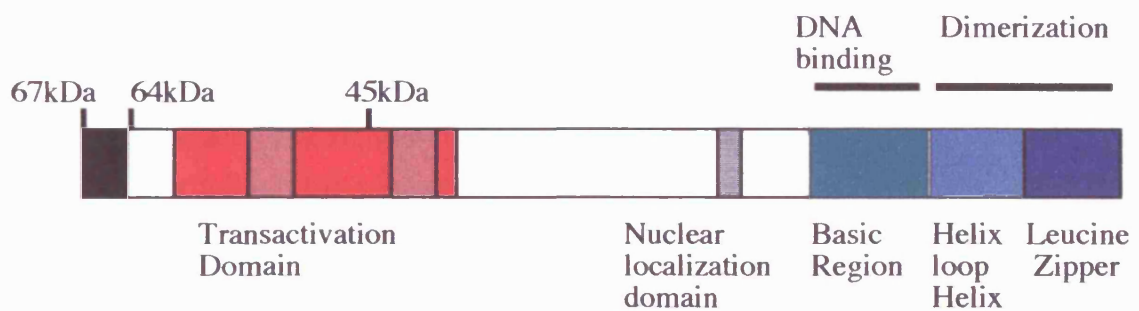
Figure 1.10. The *c-myc* gene and c-Myc protein structure. A) The *c-myc* gene structure consists of three exons. Exon 1 is almost entirely non-coding and contains the CUG start codon. B) c-Myc protein structure showing functional domains. Three isoforms of the protein are known of 67, 64 and 45kDa.

c-Myc acts as a heterodimer with its partner Max and binds to E boxes (CACGTG) to activate transcription of its target genes (Amati et al., 1992; Kato et al., 1992). However, the transactivation observed is weak compared to the activities for other transcription factors, and it is thought its activity may be modulated by other factors (Dang et al., 1999). c-Myc also recognizes non-consensus sites, and it has been shown that the nucleotides flanking the E box, as well as methylation within it, can affect the binding of Myc.Max heterodimers (Grandori et al., 2000). Max can also form heterodimers with Mad, Mnt and Mga. Max itself doesn't contain a transactivation domain and its transcriptional activity is exerted through its binding partners (Kato et al., 1992; Grandori et al., 2000). The different heterodimers have varying cellular roles, for example the Mad family of proteins act in cellular differentiation and growth arrest (Dang, 1999). As well as its role as a transcription activator c-Myc has also been shown to represses the transcription of certain mRNAs. This is reported to be due Myc.Max interacting with and effecting the activity of other transcription factors (Eisenman, 2001).

How c-Myc exerts is role in the cell is not fully understood. As well as interacting with Max, it can also interact with a number of other proteins through its C-terminal region such as Nmi, YY-1 and AP-2. In addition the N-terminal domain is implicated in the interaction with a large range of proteins including p107, TRRAP and α-tubulin (Sakamuro and Prendergast, 1999). Therefore as well as a transcription factor role, c-Myc appears to have other roles beyond this via interactions with regulatory and effector proteins (Sakamuro and Prendergast, 1999).

1.6.3 c-*myc* function

c-Myc effects a wide range of target genes and effects multiple pathways leading to cell proliferation, differentiation and apoptosis. It is an immediate early gene whose expression is induced by growth factors, cytokines and mitogens. The level of c-*myc* expression in the cell directly correlates with cell proliferation. c-*myc* mRNA was found to be low in quiescent fibroblast cells, and increased following stimulation of cell growth by serum (Dean et al., 1986). A role in cell growth was also demonstrated using c-*myc* null fibroblast cell lines, which were found to have significantly impaired growth (Mateyak et al., 1997). Fitting with its role in cell proliferation, c-*myc* is broadly expressed in embryogenesis. It has been shown to be critical for development, as the homozygous knockout mouse results in embryonic lethality (Davis et al., 1993). The cell growth effects

of c-Myc are considered to be exerted through c-Myc target genes such as genes involved in rRNA processing and transcription (BN51, nucleolin, fibrillarin), ribosomal proteins (Rps11) and translation initiation factors (eIF5A, eIF4E, eIF2α) (Schmidt, 1999; Nasi et al., 2001).

c-Myc was also found to effect the cell cycle. In c-myc null fibroblast cell lines it was found that the G1 and G2 phases of the cell cycle were lengthened, and entry of the cells into S phase was delayed (Mateyak et al., 1997). Other work showed that c-myc was activated in G_0 to G_1 transition of the cell cycle (Rabbitts et al., 1985). Therefore, its regulation of the cell cycle occurs in the early G1 phase or G1/S transition. The cell cycle is regulated by cyclins and cyclin-dependent kinases (CDKs), and c-myc has been reported to effect the cell cycle through these proteins (Obaya et al., 1999). cdc25A and cyclin E have been reported as c-Myc target genes (Galaktionov et al., 1996; Perez-Roger et al., 1997). c-Myc also has a negative effect on cell differentiation. When c-myc was expressed in MEL cells it was found to inhibit DMSO induced differentiation (Dmitrovsky et al., 1986).

The overexpression of c-myc has been shown to induce apoptosis when there is a block in cell proliferation. Fibroblasts overexpressing c-myc underwent apoptosis in conditions that block proliferation, such as serum or isoleucine deprivation, interferon arrest, and cycloheximide treatment (Evan et al., 1992). c-Myc was also found to be a required factor for apoptosis . The use of antisense c-myc oligonucleotides inhibited TNFα induced cell death of HeLas (Janike et al., 1994). c-Myc is also required for apoptotis in response to other stimuli including transcription and translation inhibitors, hypoxia, glucose deprival, heat shock, cheomotoxins, DNA damage and cancer therapeutics (Prendergast, 1999). How c-myc effects apoptosis is not fully understood. Activation of c-myc triggers the release of cytochrome c from the mitochondria, an effect that is necessary for c-myc induced apoptosis (Juin et al., 1999). This cytochrome c release does not require CD95/Fas and p53. However it was found that apoptosis required CD95/Fas signalling. p53 has also been shown to be required for c-myc induced apoptosis as p53 -/- fibroblast cells could not undergo apoptosis (Hermeking and Eick, 1994). Therefore both p53 and Cd95 appear to be required for c-myc induced apoptosis, and it is suggested that the cytochrome c release sensitizes the cell to these apoptotic triggers. c-myc induced apoptosis has also been shown to be dependent on the downstream apoptotic regulator Bax,

as fibroblasts lacking bax are not susceptible to cytochrome c release and c-*myc* induced apoptosis (Juin et al., 2002).

1.6.4 Deregulated c-*myc* expression

The expression of c-*myc* is tightly controlled, and increases in this expression have been implicated in many cancers including Burkitts lymphoma, lung carcinoma, breast carcinoma, and colon carcinoma (Dang, 1999). Increased expression is caused by gene amplification, and by deregulated expression due to genetic rearrangements including viral insertions and chromosomal translocations (Grandori et al., 2000). Deregulation is considered to increase c-*myc* expression rather than alter the activity of the c-Myc protein through mutations. For example Burkitts Lymphoma involves a translocations of c-*myc* with regulatory elements of the immunoglobulin heavy chain μ or light chains λ and κ (Saito et al., 1983). In the Burkitts Lymphoma Manca, the exons 2 and 3 are translocated downstream of the immunoglobulin enhancer/promoter element. This was predicted to enhance translation of c-*myc* by removing the 5'UTR containing translational control elements (Saito et al., 1983). However, the role of the immunoglobulin regulatory domains in the development of cancer was shown by the use of transgenic mice. Mice expressing c-*myc* under the control of immunoglobulin μ or κ enhancer elements developed early pre-B cell and B cell lymphoid tumours , whereas mice expressing the natural c-*myc* mRNA or coding region only showed no incidence of tumours (Adams et al., 1985). In addition some point mutations found in burkitts lymphomas are thought to have a role in its development. These mutations cluster in transactivation domain and cause increased transforming activity (Nesbit et al., 1999).

1.6.5 Control of c-*myc* expression

Control of c-*myc* occurs by several mechanisms. At the level of transcription c-*myc* is controlled by 4 promoters (Figure 1.10A). The majority of transcripts in the cell are from P1 and P2, and use of promoters has been shown to vary between cell lines (Ryan and Birnie, 1996). Control of transcriptional elongation also occurs, as RNA polymerase II initiation complexes are paused at the end of exon 1(Bentley and Groudine, 1986). c-*myc* expression is controlled through mRNA and protein stability, as both have very short half lives of 15 and 30 minutes respectively (Ryan and Birnie, 1996). There is also

autoregulation with Myc.Max heterodimers binding to the P2 promoter and suppress c-*myc* transcription.

As well as these mechanisms there is translational control. Work on c-*myc* identified an IRES within its 5'UTR (Stoneley et al., 1998; Nanbru et al., 1997). The IRES site was found to be within a section of the 5'UTR 390 nucleotides upstream of the AUG start site and therefore present in the majority of transcripts. The IRES segment was found to be very efficient in initiating internal ribosome entry. When placed in the centre of a dicistronic construct it increased the expression of the downstream cistron by up to 50 fold (Stoneley et al., 1998). Further work on the c-*myc* IRES discovered that the IRES remains functional in situations where cap-dependent translation is compromised. It was shown that there is continuous expression of the c-myc protein in apoptosis, and following DNA damage due to IRES translation (Stoneley et al., 2000a; Subkhankulova et al., 2001). Recently work on the c-*myc* IRES found that it controlled gene expression throughout development, and was down regulated in adult tissues where most tissues are in a non-proliferating state (Creancier et al., 2001). The c-*myc* IRES has been shown to contain a mutation in Multiple Myeloma cell lines, which causes enhanced binding of protein factors (Paulin et al., 1998). This mutant IRES has also been shown to display higher IRES activity than the wild type IRES (Chappell et al., 2000b).

1.7. c-*myb*

1.7.1 Introduction

c- *myb* was discovered as the cellular homolog of an oncogene in avian myeloblastosis virus (AMV) (Klempnauer et al., 1982). The gene codes for a transcription factor, and belongs to a family that also contains A-Myb and B-Myb. The Myb family of genes share strong homology but differ in their tissue expression patterns. c-*myb* is involved in proliferation, cell cycle regulation, and differentiation of haematopoietic cells.

1.7.2 c-*myb* structure

c-Myb has two isoforms- of 75kDa and 89kDa (Figure 1.11). The 89kDa protein results from alternative splicing, and the addition of exon 9A in the C-terminus of the protein. c-Myb contains three 50 amino acid repeats in its N-terminal region termed R1-3, which consist of a series of conserved residues including 3 tryptophans. It was found that repeats R2 and R3 formed two helix-turn-helix motifs and were essential for DNA-

75kDa
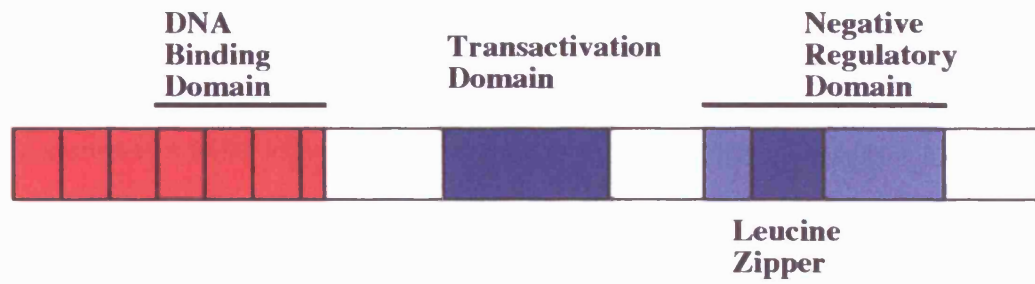
89kDa
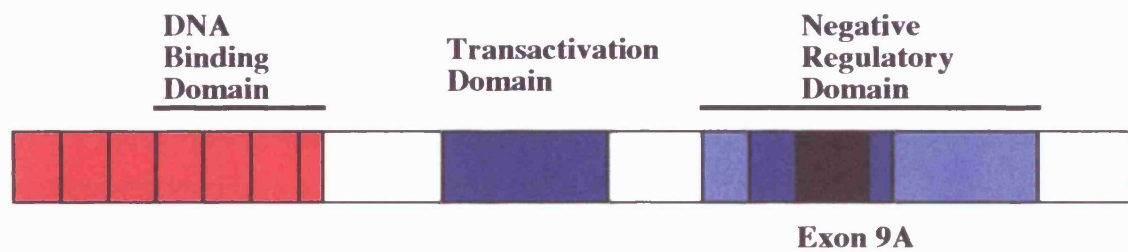
Figure 1.11. The c-Myb protein structure. Two protein isoforms exist for the protein. The 89kDa isoform has an additional exon 9A that disrupts the Leucine zipper motif.

binding. The R1 repeat is thought to be involved in the stabilization of DNA binding (Tanikawa et al., 1993). c-Myb binds to the sequence PyAACG/TG *in vitro* as a monomer (Biedenkapp et al., 1988). The central region of the protein contains the transcriptional activation domain, which includes clusters of acidic residues. The C-terminal domain was found to have a negative regulatory function as deletions in this region were found to increase c-*myb* transactivation activity (Sakura et al., 1989). In this region there is a leucine zipper motif, which is thought to be involved in the proteins negative regulation. In the p89 form of the protein the insertion of exon 9A abolishes the leucine zipper and this isoform has been shown to exhibit higher transactivation activities (Woo et al., 1998). It is also proposed that this domain may interact with other proteins involved in its negative regulation. The C-terminal region of the protein also contains an EVES domain, which interacts with the N-terminus of the protein and is thought to have an inhibitory role (Dash et al., 1996).

c-Myb acts in combination with a number of other proteins, which both increase and inhibit its function. In some situations it interacts with other transcription factors such as c/EBP and Ets2 to exert its function. For example both c-Myb and either c/EBP or Ets2 are required for mim-1 activation (Dudek et al., 1992; Burk et al., 1993). It also interacts with transcriptional coactivators such as CBP/p300 and p100 to activate transcription. Other proteins have been found to inhibit its function such as the transcription factors RARα and c-Maf, and proteins such as p160/p67 (Ness, 1999).

A number of c-Myb target genes have been identified. It is known to transactivate mim-1, c-*myc*, several T-cell surface markers such as CD4 and CD34 and is implicated in activating the cell cycle regulated gene cdc2 (Oh and Reddy, 1999). It has also been shown to repress transcription of some genes such as PMCA1 (plasma membrane Ca^{2+}-ATPase) (Afroze and Husain, 1999).

1.7.3 c-*myb* function

The expression of c-*myb* is mainly limited to immature hematopoietic tissues (Oh and Reddy, 1999). However, it is also expressed in other cell lines such as smooth muscle and colonic cells. It is involved in the control of cell growth and differentiation. c-*myb* was found to be critical for hematopoietic cell proliferation, as the use of antisense oligonucleotides on human bone marrow cells resulted in a decrease in cell proliferation (Gewirtz and Calabretta, 1988). Antisense oligonucleotides also blocked T-cell

proliferation of PBMC cells (peripheral blood mononuclear cells) (Gewirtz et al., 1989). c-Myb is also involved in vascular smooth muscle cell (VSMC) proliferation, which is a major pathway in atherogenesis. The use of antisense oligonucloetides suppressed VSMC growth *in vivo* (Simons et al., 1992). c-*myb* has a role in cell proliferation in development. It was found that homozygous c-*myb* mutant mice showed a failure in adult-type erythropoiesis in the fetal liver resulting in embryonic death (Mucenski et al., 1991).

In more differentiated cells the expression of c-*myb* is cell cycle related. Levels peak in mid to late G1 or S phase. The use of antisense oligonucleotides blocked T-cell cell cycle progression in late G1 or early S phase (Gewirtz et al., 1989). This cell cycle regulation was found to be related to the intracellular ionised calcium levels. It was found that in VSMC there was a 2-fold rise in Ca^{2+} levels in the G1/S phase of the cell cycle that was preceded by an increase in c-*myb* mRNA levels. The use of antisense c-*myb* oligonucleotides suppressed this calcium increase (Simons et al., 1993). It was found that c-Myb decreased the levels of PMCA1 (plasma membrane Ca^{2+}-ATPase). Suppression of c-*myb* led to an increase in PMCA1 levels and a decrease in intracellular Ca^{2+}, S phase entry and cell proliferation of VSMC. c-Myb was found to bind to the PMCA1 promoter to repress its transcription (Afroze and Husain, 1999).

c-*myb* is also considered to have a role in cell differentiation. The induced differentiation of murine erythroleukemia cells was found to be associated with a down regulation of c-*myb* expression (Ramsay et al., 1986). Also the expression of c-*myb* was found to block induced differentiation of MEL cells (McClinton et al., 1990). It is suggested that the downregulation of c-*myb* is required for cell differentiation.

1.7.4 Deregulated c-myb expression

The deregulation of c-*myb* is closely linked to cancer, though in contrast to c-*myc* mainly haematopoietic cells are transformed. This deregulation is due to retroviral insertions, and gene amplifications, which act to increase expression and also produce mutant forms of the protein. In a group of murine tumours called ABPL (Aberson plasmacytoid lymphosarcomas) there is rearrangement of the c-*myb* locus due to viral integration into the 3rd exon resulting in expression of a protein with a truncated N-terminus (Lavu and Reddy, 1986). There are also reports of rearrangements in other myeloid tumor lines such as the leukemia cell line NFS-60 that has a c-*myb* rearrangement due to retroviral integration (Weinstein et al., 1986). In addition there are abnormalities in

the c-*myb* locus in acute myelogenous leukemias, T-cell leukemias, colon carcinomas, and melanomas due to gene amplification and enhanced expression of the gene (Oh and Reddy, 1999).

1.7.5 Control mechanisms

c-*myb* has multiple levels of control. As mentioned above, the protein contains a negative regulatory domain that affects its function, probably through the interaction with other proteins. The protein is phosphorylated at several sites *in vivo*, which appears to have functional importance. Casein kinase II has been found to phosphorylate 2 of these sites at Ser 11, Ser 12 *in vitro* which reduced DNA binding and transactivation by the protein (Oelgeschlager et al., 1995). c-*myb* also has autoregulation by acting on its own promoter, which contains 3 myb binding sites. c-Myb protein has a very short half life and there is also regulation of the stability of mRNA. Finally there is control at transcription. c-*myb* is reported to be transcribed from multiple transcription sites, some of which appear to be tissue specific. There is no clear picture regarding when certain transcription start sites are used. The main transcription start site appears to be 201bp upstream of the AUG codon in human transcripts. There is also control via transcriptional elongation.

1.8 Project aims

The project aim was to investigate the protein factor requirements of the c-*myc* IRES. This was achieved by analysing the effects of known viral IRES interacting factors and by performing a yeast three-hybrid screen to identify novel factors. The work on known factors investigated both canonical factors and *trans*-acting factors involved in internal initiation. Results describe the role of a number of known *trans*-acting factors in c-*myc* IRES translation, as well as the discovery of novel factors that may be involved. In addition the identification of internal ribosome entry in the cellular c-*myb* gene and the role of protein factors in its translation was investigated.

Chapter 2
Materials and Methods

2.1 General reagents

Unless otherwise stated all chemical reagents were of analytical grade and were obtained from BDH laboratory supplies (Lutterworth, Leicestershire), Fisher Scientific (Loughborough, Leicestershire), ICN Flow Ltd (Thame, Oxfordshire), Sigma Chemical Company Ltd (Poole, Dorset), Oxoid (Unipath, Basingstoke, Hampshire), Pierce (c/o Perbio Science UK, Tattenhall, Cheshire) or BIO101(c/o Anachem, Luton, Bedfordshire). Products for molecular biological techniques were routinely purchased from Calbiochem (c/o CN Biosciences UK, Beeston, Nottingham), Gibco-BRL (Paisley, Scotland), MBI Fermentas (c/o Helena Biosciences, Sunderland, Tyne and Wear), New England Biolabs (NEB) (c/o CP Labs, Bishops Stortford, Hertfordshire), Pharmacia Biotech (Milton Keynes, Buckinghamshire), Promega (Southampton), QIAGEN (Crawley, West Sussex), Roche UK Ltd (Lewes, East Sussex) and Stratagene Ltd (Cambridge). Radiolabelled chemicals were purchased from Amersham International (Little Chalfont, Buckinghamshire) and NEN Dupont (Hounslow).

2.2 Tissue culture techniques

2.2.1 Media and supplements

DMEM medium: Dulbecco's modified eagle medium, without sodium pyruvate (Gibco-BRL).

RPMI 1640 medium: Rose Park Memorial Institute 1640 medium, with L-glutamine (Gibco-BRL).

MEM medium: Minimal essential medium (Sigma)

Media were supplemented with foetal calf serum (FCS) (Helena Biosciences), L-glutamine (Sigma) and non-essential amino acids (NEAA) (Sigma) as indicated in table 2.1.

Cell line	Origin	Media
Balb/c-3T3	Murine embryonic fibroblast	DMEM + 10% FCS
BAE	Bovine aortic endothelial cells	DMEM + Glutamax + 10% FCS

Cell line	Origin	Media
Cos7	Monkey epithelial cells immortalised with SV40 DNA	DMEM + 10% FCS
GM03201	Human EBV immortalised lymphoblastoid cell line	RPMI +10% FCS
GM01953	Human lymphoblast	RPMI +10% FCS
H9c2	Rat Heart Myoblast cells	DMEM + Glutamax + 10% FCS
HEK293	Human embryonic kidney	DMEM + Glutamax + 10% FCS
HeLa	Human cervical carcinoma	DMEM + 10% FCS
HepG2	Human hepatocyte carcinoma	DMEM + 10% FCS
Jurkats	Human T lymphocyte	RPMI + 10% FCS
MCF7	Human breast carcinoma	DMEM + Glutamax + 10% FCS
MRC5	Human Lung	MEM + 10% FCS+ 1% NEAA
PAE	Porcine aortic endothelial cells	DMEM + Glutamax + 10% FCS
Raji	Human burkitts lymphoma	RPMI + 10% FCS

Table 2.1 Name, tissue origin, and media requirements of cell lines used.

2.2.2 Maintenance of cell lines

The cell lines in table 2.1 were grown at 37°C in a humidified atmosphere containing 5% CO_2. Cells were cultured using the growth medium indicated, in sterile plasticware (TPP, c/o Helena Biosciences). Adherent cell lines were grown until confluent in 75cm^2 flasks and treated with 0.05% Trypsin/0.5mM EDTA (Gibco BRL). Approximately 6.5×10^5 cells were diluted into fresh medium and re-plated into a new flask. Cells grown in suspension were maintained at concentrations of 5×10^5-1×10^6 cells/ml.

2.2.3 Calcium Phosphate-mediated DNA Transfection

Calcium phosphate-mediated DNA transfection of adherent cells was performed using a modified protocol of (Jordan et al., 1996). Cells were seeded 24 hours prior to

transfection at 1×10^5 cells per well using 2ml of medium in a 6-well plate. A solution of 10μl of 2.5M $CaCl_2$ and 1-1.5μg of plasmid DNA was diluted with sterile deionised water to a final volume of 100μl. An equal volume of 2x HEPES buffered saline (50mM HEPES, pH7.05, 1.5mM Na_2HPO_4, 140mM NaCl) was added to this solution whilst bubbling air through the mixture. The calcium phosphate-DNA co-precipitate was allowed to form for 5 min and was then added dropwise to the medium covering the cells. After exposing the cells to the precipitate for 15-20 hours at 37°C, the medium was removed. The cells were washed twice with phosphate buffered saline (PBS) (4.3mM Na_2HPO_4, 1.5mM KH_2PO_4, 137mM NaCl, 2.7mM KCl, pH 7.4) before the addition of fresh media. Cells were left for a further 24 hours before harvesting.

2.2.4 FuGENE 6-mediated DNA transfection

Cells were seeded as described for calcium phosphate-mediated transfection. For each 1μg Plasmid DNA, 3μl of FuGENE 6 transfection reagent (Roche) was used. Each 3μl of FuGENE 6 reagent was added directly to 100μl of serum-free DMEM and incubated at room temperature for 5 min. This mixture was added dropwise to DNA solution and incubated at room temperature for a further 15 min. The DNA solution was then added dropwise to the medium covering the cells and cells were grown for a further 48 hours before harvesting.

2.2.5 DEAE-Dextran-mediated DNA transfection

COS 7 cells were seeded at a density of 1×10^6 per 10cm plate in 10ml medium 24 hours prior to transfection. The medium was replaced with serum-free medium and the cells were returned to 37°C. 2μg of plasmid DNA was added to a mixture of 3.2ml of serum-free DMEM, 0.8ml of 0.25M Tris-HCl, pH7.5 and 0.5ml of 10mg/ml Dextran. The medium was removed and replaced with the mixture then the cells incubated at 37°C for 20 min. Chloroquinine was added to the mixture covering the cells to a final concentration of 40μg/ml, and the cells were incubated for a further 3-4 hours. The medium was removed and the cells were washed twice with PBS. Fresh complete medium was added and the cells grown for a further 48 hours before harvesting.

2.2.6 DNA transfection by electroporation

Suspension cell lines were transfected by electroporation. 8×10^6 cells and 4-6µg plasmid DNA were used per transfection. Cells were washed in serum free RPMI media and then resuspended in 250µl serum free RPMI plus DNA. Electroporation was carried out using 0.5cm cuvettes (Bio-Rad) at 0.3KV, 975µF, Max 1000Ω with a time constant of 60msec. Complete RPMI medium was added and the cells grown for 48 hours before harvesting.

2.3 Bacterial techniques

2.3.1 Media and supplements

LB Medium : 10g Bacto-tryptone, 5g bacto-yeast extract, 10g NaCl dissolved in 1l of deionised water.

LB Agar: LB Medium with the addition of 15g Agar.

5 x M9 Salts: 33.9g $NaHPO_4$, 15g KH_2PO_4, 2.5g NaCl, 5g NH_4Cl dissolved in 1l of deionised water pH7.2.

M9 Plates: 100ml M9 Salts, 1ml of 1M $MgSO_4$, 50µl of 1M $CaCl_2$, 5ml of 40% Glucose, 10g Agar and deionised water to 500ml. After autoclaving Thiamine was added to a final concentration of 2µg/ml and Proline to 40µg/ml.

SOC Medium: 2g bacto-tryptone, 0.5g bacto-yeast extract, 0.05g NaCl, 1ml of 250mM KCl, 0.5ml of 2M $MgCl_2$, 2ml of 1M Glucose in 100ml deionised water.

Ampicillin: Stock solution of 100 mg/ml of ampicillin in sterile deionised water. This was used at a final concentration of 100µg/ml.

Bacterial strains:

E.coli strains DH5α, HB101 and BL21 used.

DH5α: *F', φ80dlacZΔM15, Δ(lacZYA-argF), U169, deoR, recA1, endA1, hsdR17, phoA, supE44, λ–, thi-1, gyrA96, relA1*

HB101: *F', mcrB, mrr, hsdS20, recA13, supE44, ara-14, galK2, lacY1, proA2, rpsL20, xyl-5, λ–, leuB6, mtl-1, thi-1*

BL21(DE3): *F, ompT, hsdS_B, (r_B¯, m_B¯), dcm, gal, λ(DE3)*

2.3.2 Preparation of competent cells for Heat Shock Transformation

A single colony from an LB agar plate was inoculated into 2.5ml of LB medium and incubated overnight at 37°C with shaking. The entire overnight culture was inoculated into

250ml of LB medium supplemented with 20mM MgSO$_4$ and incubated at 37°C until the A$_{595}$ reached 0.4-0.6. Cells were pelleted by centrifugation at 4,500x g for 5 min at 4°C. The pellet was gently resuspended in 100ml of ice-cold filter sterile TFB1 (30mM KAc, 10mM CaCl$_2$, 50mM MnCl$_2$, 100mM RbCl, 15% glycerol, pH 5.8). After incubating on ice for 5 min, the cells were centrifuged at 4,500x g for 5 min at 4°C. The pellet was resuspended in 10 ml of ice-cold filter sterile TFB2 (10mM MOPS pH 6.5, 75mM CaCl$_2$, 10mM RbCl, 15% glycerol) and incubated on ice for 1 hour. Finally, the cells were rapidly frozen in an isopropanol/dry ice bath in 200μl aliquots and stored at –80°C.

2.3.3 Transformation of competent cells by Heat Shock

Ligation products or plasmid DNA (5-10ng) were added to 50μl of competent cells and incubated on ice for 25 min. After heating the mixture at 42°C for 1.5 min the cells were returned to ice. 150μl of LB medium was added and the cells were incubated with shaking at 37°C for 30 min. The cells were then spread onto a pre-warmed LB agar plate containing appropriate antibiotic, inverted and incubated at 37°C for 16-20 hours.

2.3.4 Transformation of cells by Electroporation

A single colony of HB101 cells from an LB agar plate was inoculated into 10ml of LB medium and incubated overnight at 37°C with shaking. The entire overnight culture was inoculated into 200ml of LB medium and incubated at 37°C until the A$_{595}$ reached 0.5. Cells were pelleted by centrifugation at 4000rpm for 5 min at 4°C. The pellet was washed with 50ml ice-cold deionised water and re-pelleted. This was repeated with 25ml and 10ml of deionised water, followed by 5ml and 1ml of 10% Glycerol. Cells were then resuspended in 500μl 10% Glycerol. 3μl of plasmid DNA from yeast was added to 50μl cells on ice. Electroporation was carried out in chilled 0.2cm cuvettes (Bio-Rad) at 2.0KV, 25μF, 200Ω, (10 KV/cm) with a time constant of 5msec. Cells were rescued in 1ml SOC and incubated with shaking at 37°C for 1hour. Cells were then pelleted and resuspended in 100μl M9 medium before plating onto pre-warmed M9 plates containing ampicillin.

2.3.5 Amplification of cDNA library

The cDNA library used for the yeast-three hybrid screen was amplified prior to screening. The library titre was determined and the cell stock diluted and plated accordingly. One hundred 15cm plates were plated to obtain twice the number of

independent clones of the library. Plates were incubated at 37°C for 16hrs, then cells were harvested by scraping into 5ml LB/25% glycerol per plate. All colonies were pooled and a third of the resulting culture used for plasmid preparation.

2.4 Yeast techniques

2.4.1 Media and supplements

YPAD (Yeast-Extract Peptone Adenine Dextrose): 10g bacto-yeast extract, 20g bacto-peptone, 0.1g Adenine, 100ml of 20% Dextrose dissolved in 1l deionised water.

YPAD Plates: YPAD with the addition of 20g Agar.

SD (Synthetic Dropout) Media: 2.68g Yeast Nitrogen Base without amino acids, 8g Dextrose and 0.65-0.77g complete supplement mixture with appropriate amino acid dropout (- Histidine, Leucine and/or Uracil) (BIO101) dissolved in 400ml deionised water.

SD Plates: SD with the addition of 8.8g Agar.

Yeast Strain: L40-Coat : *MATa, ura3-52, leu2-3, 112, his3Δ200, trp1Δ1, ade2, LYS2::(LexA op)-HIS3, ura3:(LexA-op)-LacZ, LexA-MS2 coat(TRP1)*. (a gift of Dr M.Wickens, University of Wisconsin).

cDNA Library: Human Testis MATCHMAKER cDNA Library (Clontech). Cloned into pACTII in XhoI/EcoRI.

3AT (3-Aminotriazole): Used at final concentration of 2-10mM

5-FOA (5-Flouroorotic acid): Used at a final concentration of 0.1%.

2.4.2 Small Scale Yeast Transformation

Yeast cells were inoculated into 10ml of YPAD medium and incubated overnight at 30°C with shaking. The entire overnight culture was inoculated into 50ml of YPAD medium and incubated at 30°C until the A_{595} reached 0.4-0.6. Cells were pelleted by centrifugation at 2500rpm, washed with 40ml TE, repelleted and resuspended in 2ml LiAc/TE (100mMLiAc, 5mM Tris-HCl pH7.5, 0.5mM EDTA). Cells were then incubated for 10 min at room temperature. For each transformation 1μg plasmid DNA and 100μg denatured sheared Salmon Sperm DNA were added to 100μl yeast cell s. 700μl of LiAc/PEG/TE (100mM LiAc pH 7.5, 40% PEG 3350, 10mM Tris-HCl pH7.5, 1mM EDTA) was added and after thorough mixing the reaction incubated at 30°C for 30 minutes. 88μl of DMSO was added and the reaction heat shocked at 42°C for 7 minutes.

Cells were then pelleted, washed in 1ml TE and resuspended in 100µl TE. Cells were then plated onto appropriate selective plates and incubated at 30°C for 2-5 days.

2.4.3 Yeast Protein Extraction

Yeast cells were inoculated into 20ml of SD medium and incubated overnight at 30°C with shaking. Cells were pelleted by centrifugation at 2500rpm and resuspended in 200µl breaking buffer (50mM Tris-Hcl pH8.0, 400mM KCl, with the addition of protease inhibitors). Glass beads (425-600µm) were added and the samples vigorously shaken using a vortex for 1min before returning to ice for 1min. This was repeated two times. Cell debris was pelleted and the supernatant collected and stored at -70°C.

2.4.4 Yeast RNA Extraction

Yeast cells were inoculated into 20ml of SD medium and incubated overnight at 30°C with shaking. Cells were pelleted by centrifugation at 2500rpm, washed with deionised water, and resuspended in 600µl RNA Extraction Buffer (0.1M NaCl, 10mM EDTA, 5% SDS, 50mM Tris-HCl pH7.5). An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was immediately added and the samples left at room temperature for 5 min. Glass beads were added and the samples were vigorously shaken using a vortex for 2 min . The phases were then separated by centrifugation at 13,000rpm for 3min. The aqueous phase was further phenol:chloroform extracted twice and chloroform extracted. RNA was ethanol precipitated, washed and resuspended in 30µl deionised water.

2.4.5 Large Scale Yeast Transformation

This technique was used for transformation of the cDNA library into yeast cells already transformed with the RNA hybrid pIIIAMS2-2'M251. An overnight yeast culture containing the plasmid was inoculated into 500mls of SD media and incubated at 30°C until the A_{595} reached 0.4-0.6. Cells were pelleted and washed with 100ml TE followed by 50ml TE/LiAc. Cells were then resuspended in 3ml LiAc/TE and left for 10 min at room temperature. For the transformation 100µg library DNA and 4mg denatured sheared Salmon Sperm DNA were added to yeast cells. 15ml of LiAc/PEG/TE was added and the reaction incubated at 30°C for 30 min. 1.5ml of DMSO was added prior to heat shocking cells at 42°C for 15 minutes. Cells were washed in TE and resuspended in 10ml TE. Cells

were then plated onto 30x appropriate 15cm selective plates and incubated at 30°C for 2-5 days.

2.4.6 Preparation of plasmid DNA

Yeast cells were inoculated into 10ml of YPAD medium and incubated overnight at 30°C with shaking. Cells were pelleted by centrifugation at 2500rpm for 5min and washed in 1ml of deionised water. Cells were then repelleted and resuspended in 250μl lysis buffer (2% Triton X-100, 625mM EDTA, 50mM Tris-HCl pH8.0). Glass Beads were added and the samples vigorously shaken using a vortex for 2 min. 300μl phenol was added and samples were then vortexed for a further 1min before centrifugation at 13,000rpm for 10-20min. The aqueous phase was removed, and nucleic acid ethanol precipitated. The final sample as resuspended in 100μl deionised water, and 3μl used for electroporation.

2.4.7 β-Galactosidase Overlay Assay

Yeast transformants were grown on appropriate selective plates for 1-3 days. A solution of 0.5% of low-melting agarose in 0.5M potassium phosphate buffer pH7.0, 6% DMF, 0.1%SDS was prepared. This was cooled to 60°C before β-Mercaptoethanol and X-Gal were added to 50mM and 0.5mg/ml respectively. The agar was poured onto plates and left to solidify, before plates were inverted and incubated at 30°C. Plates were then monitored for colour changes.

2.4.8 Liquid β-Galactosidase Assays

Yeast cells were inoculated into 15ml of medium and incubated overnight at 30°C with shaking. Cells were pelleted by centrifugation at 2500rpm and washed with 1ml LacZ Buffer (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$, 2.7ml/l β-Mercaptoethanol). Cells were repelleted and resuspended in 200μl LacZ Buffer. Glass beads were added, and the reaction incubated on ice for 5 min. Samples were vigorously shaken using a vortex for 2 min, and cell debris pelleted at 13,000rpm for 10 min. 150μl of supernatant was transferred to fresh tubes. The activity of β-galactosidase in the samples was determined using a Galacto-Light Plus assay system (Tropix). For each assay 5μl of sample was added to 25μl of Galacton Plus reaction buffer (Galacton-Plus substrate diluted 1:100 with Reaction Buffer diluent) and incubated at room temperature for 1 hour. 37.5μl of Light Emission Accelerator II was then added and enzyme activity was determined by

immediately measuring the light emission over 10 s. This was carried out using an Optocomp I luminometer (MGM Instruments). β-Galactosidase light readings were normalised to protein concentrations of the samples. Protein concentrations were calculated by the Bradford Assay technique using Coomassie Protein Assay Reagent (Pierce). Standards were prepared from stock BSA solution (2mg/ml), and absorbance measured at 595nm. Protein concentration was determined using a standard curve.

2.5 Molecular biology techniques

2.5.1 Buffers and solutions

TE: 10mM Tris-HCl pH7.5, 1mM EDTA

1xTAE: 40mM Tris, 40mM acetic acid, 1mM EDTA, pH 8.0

1x TBE: 89mM Tris, 89mM boric acid, 2mM EDTA, pH 8.0

5x TBE loading buffer: 50% v/v glycerol, 200mM Tris, 200mM acetic acid, 5mM EDTA, 0.1% Bromophenol blue, 0.1% Xylene cyanol FF

DNA sequencing formamide loading dyes: 100% deionised formamide, 0.1%(w/v) Xylene cyanol FF, 0.1%(w/v) Bromophenol blue, 1mM EDTA

RNA formamide loading buffer: 80% deionised formamide, 10 mM EDTA, 0.1% SDS, 0.1% Xylene cyanol FF, 0.1% Bromophenol blue

RNA formaldehyde loading buffer: 50% glycerol, 1mM EDTA, 0.4% Bromophenol blue, 0.4% Xylene cyanol FF

2.5.2 Plasmids

pGL3', pGML, phpL, pRF, pRMF, phpRMF , pRemcvF (all described in (Stoneley et al., 2000b))

pSKL, pSKML (Stoneley et al., 1998)

pSKGAP:E/H (Paulin et al., 1998)

pMint1L (Le Quesne et al., 2001)

pBluescript II SK (+) (Stratagene)

pcDNA3 (Invitrogen)

pJ7*lac*Z (a gift of Dr D.Heery, University of Leicester)

pIIIA/MS2-2, pACTII, pIIIA/IRE-MS2, pAD-IRP (a gift of Dr M.Wickens, University of Wisconsin)

pET21a (Novagen)

pcDNA-UNR, pcDNA-PTB, pETUNR , pETUNRIP, pETPTB, pJHRV(a gift of

Dr.R.Jackson, University of Cambridge)

pSG5.K, pET16b.K (a gift of Dr.M.Hentze, EMBL, Germany)

2.5.3 Ethanol precipitation of nucleic acids

Nucleic acids were precipitated by the addition of 0.1 volume of 3M sodium acetate, pH5.2 and 2.5 volumes of absolute ethanol. The sample was incubated at –20°C for 15-30 min and the nucleic acid was pelleted by centrifugation at 12,000x g for 10 min. Excess salt was removed by washing with 70% ethanol before the pellet was dried and resuspended in sterile deionised water.

2.5.4 Phenol/chloroform extraction

Solutions of nucleic acid were separated from contaminating proteins by the addition of an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). After vigorous mixing, the phases were separated by centrifugation at 12,000x g for 5 min. The upper aqueous phase was transferred to a separate tube and an equal volume of chloroform:isoamyl alcohol (24:1) was added. Following extraction and separation of the phases, the aqueous layer was transferred to a new tube and the nucleic acid was precipitated.

2.5.5 Agarose gel electrophoresis

Fragments of DNA or RNA were fractionated according to their molecular weight by electrophoresis through agarose gels. Agarose was melted in 1x TBE or TAE buffer and cooled, after which 2μl of 10mg/ml ethidium bromide was added and the gel was cast. The set gel was submerged in 1x TBE or TAE buffer in a horizontal electrophoresis tank. Samples were mixed with 0.2 volume of 5x TBE loading buffer and separated in the gel at up to 8V/cm. After electrophoresis, the nucleic acid was visualised on a UV transilluminator. DNA fragments that required gel elution were separated as described and agarose containing the required fragment was excised from the gel. DNA was purified from the excised gel using Qiaquick columns.

2.5.6 Purification of DNA using Qiaquick columns

The Qiaquick gel extraction kit (Qiagen) was used to isolate DNA from agarose gel fragments or to purify DNA fragments. The manufacturer's protocols were followed and the DNA was eluted from the column in 30μl of sterile deionised water.

2.5.7 Oligonucleotides

Oligonucleotides were purchased from the Protein and Nucleic Acid Chemistry Laboratory (University of Leicester) or Gibco BRL. After ethanol precipitation where necessary, oligonucleotides were resuspended in sterile deionised water. Sequences of the oligonucleotides used are given in Table 2.2.

Oligonucleotide Name	Sequence
LAFOR	CGATGAATTCAGATAGCCGTAATGGCTG
LAREV	CGATCTCGAGAACTACTGGTCTCCAGCAC
PETLAF	CGATGAATTCAAGATAGCCGTAATGGCTG
PETLAR	CGATCTCGAGACTGGTCTCCAGCACAT
PCBP1F	CGATCCATGGATGCCGGTGTGAC
PCBP1R	CGATCTCGAGTCTAGCTGCACCCCATGC
PETP1F	CGATGAATTCATGGATGCCGGTGTGACTG
2PETP1R	CGATCTCGAGGCTGCACCCCATGCCCTTC
PCBP2F	CGATCCATGGACACCGGTGTGATTG
PCBP2R	CGATCTCGAGTCTAGCTGCTCCCCATGC
PETP2F	CGATGAATTCATGGACACCGGTGTGATTG
2PETP2R	CGATCTCGAGCTGGCTGCTCCCCATGCCAC
ITAFF	CGATGAATTCGGAAGATGTCGGGCGAG
ITAFR	CGATCTCGAGAGATGGGACGCACCTCAGTC
PETITAFF	CGATGAATTCAAGATGTCGGGCGAGGAC
PETITAFR	CGATCTCGAGCACCACAGTCCCCAGCTTC
UNRNCO1	CGATCCATGGAGATGAGCTTTGATCCAAAC
UNRXHO1	CGATCTCGAGGGTTAGTCAATGACACCAGC
PTBFOR	CAGTGAATTCCCATGGACGGCATTGTCCAG
PTBREV	TGCACTCGAGCCCTAGATGGTGGACTTGGAG

UNRIPF	CGATGGATCCCCATGGCAATGAGACAG
UNRIPR	CGATCTCGAGGCTCAGGCCTTAACATCAG
4GFOR	CGATCCATGGTGGATCCCACTAGACTAC
4GREV	CGATCTCGAGGGTACCGCTTGTTGAAG
PET4GF	CGATGAATTCATGGATCCCACTAGACTAC
PET4GR	CGATCTCGAGGTACCGCTTGTTGAAGGGC
NDAP5F	AGCTGAATTCAAGGCCGCCAAAATGGAGAGTGC
NDAP5R	GCATCTCGAGGCAGTTACATCATAGCAACAG
DAP86R	GCATCTCGAGATGAATCTGTTTCATCGC
PETDAPF2	CGATGAATTCAAAATGGAGAGTGCGATTGC
PETDAPR	CGATCTCGAGTGGTTCTTTCGTCAGCTTC
NDAP86R	GCATCTCGAGTGAATCTGTTTCATCGCTG
MYC211	TGCCCGCCCGCTGCTATG
MYC257	GGGAGAGTCGCGTCCTTG
T7FOR	TAAAACGACGGCCAGTGAGC
XHOMF1	CTATCTCGAGTAATTCCAGCGAGAGGCAGA
CLAMIR2	TCATATCGATGTCGCGTCCTTGCTCGGGTG
XHOMIF3	CTATCTCGAGGGAACTTACAACACCCGAGC
NHEMR3	CTATGCTAGCTCCAGCGTCTAAGCAGCTGC
MINTFOR	CTATCTCGAGGGCAGAGGGAGCGAGCG
MINTREV	CTATGCTAGCTGCCCGCCCGCTGCTATG
HRVFOR	CTATCTCGAGATCCAGGTTGTTCCCACCTG
HRVREV	CTATGCTAGCACGGACACCCAAAGTAGTTG
EMCVFX	CTATCTCGAGAGGGCCCGGAAACCTGGC
EMCVRN	CTATGCTAGCAGCTTCCAGAGGAACTGC
GAPDHFOR	CGATCTCGAGATGGTGAAAGTCGGAGTC
GAPDHREV	CGCTGCTAGCGTGTAGAATCATATTTGA
VR1	AGATGGTGCACGATGCACAG
VF1	TATGGCTTACCCATACGATG
FLFOR	CGAAGCTTCCGCGTCGATGGGCGGCGC
FLREV	CGATCTCGAGGCTGTTTGCTGCTGCTCTTG
PETFLREV	CGATCTCGAGTTTGCTGCTGCTCTTGGGAAG

46

33F	CGATGAATTCATGGGGTGCGGTTCAGTCGGTCG
33R	CGATCTCGAGACAAATCGGTCCGTGCACTG
36F	CGATAAGCTTATGCACAATAAAAGGAAACGGC
36R	CGATGGATCCAGAAGAAATCCATAGCCCTG
19F	CGATGAATTCATGCCTTACGGGAAAGTCCAG
19R	CGATCTCGAGATTTGCCCTCTGTGACTCTC
YH100F	CGATGAATTCATGAAGCCAGGATTCAGTC
YH100R	CGATCTCGAGACTTCGGTTCTTCACCTTG
MYBNCO	CGATCCATGGCGCGGCGGGCGGC
MYBECO	CGATGAATTCATATCAACCTGTTTCCTC
BRNASEF	GTGGATCCGCAAGAAGATGCAGATG
HLUC3'	ATAAGCTTGCGTATCTCTTCATAGCCTT

Table 2.2. Oligonucleotide sequences.

2.5.8 Polymerase chain reaction (PCR)

Standard PCR reactions were performed in a final volume of 50μl containing 1x *Pfu* reaction buffer (Stratagene), 50ng of template DNA, 200ng of primers, 0.2mM of each dNTP and 2.5 units of *PfuTurbo* DNA polymerase (Stratagene). Reactions were performed in a Techne Genius Thermal Cycler. DNA was initially denatured by heating at 95°C for 2 min. The reactions then underwent cycles of denaturation, annealing and extension that were variable in temperature and time depending on template amplified. They were approximately 95°C for 30s (denature), 55-68°C for 30s (anneal) and 72°C for 60s per kb (extend), for 25-35 cycles. This was followed by a final extension at 72°C for 7 min. PCR reactions for GC rich templates were supplemented with 0.5-1M Betaine and 5% DMSO.

2.5.9 cDNA synthesis

cDNA was synthesised by reverse transcription from poly (A)+ RNA extracted from cultured cells. Reactions were performed in a final volume of 20μl containing 1x Superscript II Buffer (Gibco-BRL), 1μl poly (A)+ RNA, 0.5mM dNTPs, 2.5μM random nonamers (Sigma), 10mM DTT, 200 units Superscript II. Reactions containing random nonamers, RNA and dNTPs were prepared, and heated to 70°C for 10 min. The remaining

reaction components were added and the reaction heated to 25°C for 15 min, followed by 42°C for 50 min. The enzyme was heat inactivated at 70°C for 10 min.

2.5.10 Restriction enzyme digestion

DNA was digested with restriction enzymes in a total volume of 10-50µl under the conditions recommended by the suppliers. Reactions were typically incubated at the 37°C for 1-2 hours and the enzyme heat activated by incubation at 65°C for 10 min.

2.5.11 Alkaline phosphatase treatment of DNA

Linearised plasmids were treated with calf intestinal alkaline phosphatase (CIAP) to remove the 5' phosphate groups and prevent self-ligation. Following restriction digestion, the restriction enzyme was inactivated by heating at 65°C for 10 min. Dephosphorylation was performed in a final volume of 50µl in 1x restriction enzyme buffer, using 1 unit of CIAP. The reaction was incubated for 30 min at 37°C for DNA fragments with overhanging 5' ends, and incubated at 37°C for 15 min followed by 56°C for 15 min for DNA fragments with blunt ends. The DNA was then phenol/chloroform purified before ligation.

2.5.12 Ligations

Ligations were performed in a total volume of 10µl. Vector DNA (50ng) was mixed in a 1:3 molar ratio with insert DNA in a reaction containing 1x T4 DNA ligase buffer (MBI Fermentas) and 2.5 units of T4 DNA ligase. The reaction was incubated at 16°C for 2-16 hours, the enzyme heat inactivated at 65°C for 10 min, after which 5µl of the ligation reaction was used to transform competent *E.coli*.

2.5.13 Small scale preparation of plasmid DNA

A single colony was inoculated into 5ml of LB media containing ampicillin and incubated at 37°C for 12-16 hours with shaking. 2ml of the culture was transferred into a fresh tube and the bacteria were pelleted by centrifugation at 12,000x *g* for 1 min. The pellet was resuspended in 100µl of ice-cold solution I (25mM Tris-HCl, 10mM EDTA, 50mM Glucose, pH 8.0) and incubated for 5 min on ice. 200µl of solution II (1% SDS, 0.2M NaOH) was added, solutions mixed gently and the sample was incubated on ice for 5 min. 150µl of solution III (7.5M ammonium acetate pH7.6) was added, solutions mixed

and the sample was incubated on ice for a further 5 min. The precipitated matter was pelleted by centrifugation at 12,000x g for 5 min and the supernatant was removed to a fresh tube. Plasmid DNA was precipitated from this solution by the addition of 1 volume of isopropanol, followed by centrifugation and ethanol washing as described in section 2.5.3. The pellet was dried and resuspended in 30μl of 1x TE buffer. Diagnostic restriction digests were performed using 5μl of this solution. Alternatively, DNA was prepared using the Wizard Plus SV miniprep kit (Promega) following the manufacturers protocol. Plasmid DNA was eluted from columns in 60μl of deionised water.

2.5.14 Large scale preparation of plasmid DNA

To prepare larger quantities of plasmid DNA, an overnight culture of *E. coli* containing the plasmid was inoculated into 250mls of LB media supplemented with ampicillin. The culture was grown for 12-16 hours at 37°C. Cells were harvested by centrifugation at 6,000x g for 10 min at 4°C. Plasmid DNA was prepared using Plasmid Midi Kit (Qiagen) following the manufacturers protocol. The resulting pellet was resuspended in a volume of 200μl of deionised water.

2.5.15 DNA sequencing

2-3μg of plasmid DNA was denatured in a 10μl volume by incubating with 0.1 volumes of 2mM NaOH, 2mM EDTA pH8.0 at 37°C for 15 min. The solution was then neutralised with 0.1 volumes of 7.5M NH$_4$Ac pH7.5 and 1 volume of isopropanol was added. Following incubation at -80°C for at least 30 min, the single stranded DNA was pelleted by centrifugation at 12,000g for 10 min and the pellet was dried. The pellet was resuspended in 10μl of a 5ng/μl solution of sequencing primer and 2μl of annealing buffer (280mM Tris-HCl, pH7.5, 100mM MgCl$_2$, 350mM NaCl). The reaction was heated at 65°C for 2 min, followed by incubation at room temperature for 10 min and on ice for 5 min to achieve primer annealing. Samples were then labelled at 20°C for 5 min, in a reaction containing 0.3μl [α-^{35}S] dATP (12.5mCi/ml), 3μl of labelling mix A (2μM dGTP, 2μM dCTP, 2μM dTTP), and 1 unit of T7 DNA polymerase. Labelling was terminated by the addition of 2.5μl of each termination mix (150μM dNTPs, 10mM MgCl$_2$, 40mM Tris-HCl, pH 7.5, 50mM NaCl, 15μM ddNTP G, A, T, or C) to 4μl of reaction. The reactions were then incubated at 42°C for 5 min before the reaction was stopped by the addition of 4μl of DNA formamide loading dye. The labelled DNA fragments were fractionated on a

6% polyacrylamide/7M urea gel, following which the gel was dried for 1 hour at 80°C and exposed to x-ray film (Fuji) for 16-48 hours. ABI sequencing was carried out by Protein and Nucleic Acid Chemistry Laboratory (University of Leicester).

2.5.16 Radiolabelled DNA markers

1µg of pBR322 DNA was digested with 5 units of HpaII for 1hr in a volume of 10 µl and the reaction was stopped by heating at 90°C for 2 min. The DNA fragments were radiolabelled using the Klenow fragment of $E.$ $coli$ DNA polymerase I in a reaction volume of 15µl containing 1x restriction buffer, 100 µg/ml of BSA, 1mM dCTP, 10µCi of [α-^{32}P] dCTP (800 Ci/mmol) and 5 units of Klenow DNA polymerase. The reaction was incubated at 30°C for 15 min and stopped by the addition of RNA formamide loading buffer.

2.5.17 Isolation of total cellular RNA

Total cellular RNA was isolated using TRI Reagent (Sigma). 5-10x10^6 suspension cells were pelleted by centrifugation at 1000x g for 5 min and washed with PBS. Cells were lysed by resuspension in 1ml of TRI Reagent and incubation at room temperature for 5 min. Adherent cells were washed with PBS and treated with 1ml of TRI Reagent in $situ$ and lysates were removed with a cell scraper and transferred to a fresh tube. 200µl of chloroform was added to the lysate and the mixture was vigorously mixed using a vortex for 15s then left for 2-15min at room temperature. Samples were then centrifuged at 12,000x g for 15 min. The upper aqueous phase was transferred to a fresh tube and 0.5 volumes of isopropanol was added. The sample was incubated at room temperature for 10 min and the precipitated RNA was pelleted by centrifugation at 12,000x g for 10 min and washed with 75% ethanol. After briefly drying the pellet, it was resuspended in 100µl of deionised water and stored at -80°C.

2.5.18 Purification of poly(A)$^+$ mRNA from total cellular RNA

Poly(A)$^+$ mRNA was purified from total cellular RNA using oligo[dT]$_{25}$ magnetic DynaBeads (Dynal). The manufacturer's protocols were followed, except that 50µl of oligo[dT]$_{25}$ beads were used to isolate poly(A)$^+$ RNA from 50µg of total RNA. Isolated RNA was stored at -80°C and beads were reconditioned according to the manufacturer's instructions and stored at 4°C.

2.5.19 *In vitro* run-off transcription

10μg of vector DNA was linearised by restriction digestion using a site downstream of the sequence of interest. Subsequently, the protein was removed by phenol/chloroform extraction and, following ethanol precipitation, the DNA was resuspended in 10μl of filter sterile deionised water. To synthesise uncapped transcripts, a reaction was assembled containing 1x Transcription buffer (MBI Fermentas), 13 units of RNAguard, 0.5mM of each NTP, 1μg of DNA template, and 40 units of T7, T3, or SP6 RNA polymerase in a final volume of 50μl. After incubation at 37°C for 1 hour, a further 20 units of RNA polymerase were added and the sample was incubated at 37°C for 30 min. The DNA template was then digested with 10 units of RNase-free DNase I for 15 min at 37°C. Immediately after digestion unincorporated nucleotides were removed by passing the solution through a Sephadex G-50 column.

Capped transcripts were synthesised in an identical reaction to which 1mM m^{7}G(5')ppp(5')G was added and only 0.25mM GTP used. The RNA was synthesised and isolated as described above.

Radiolabelled transcripts for use in UV crosslinking assays were synthesised in a 20μl reaction volume containing 1x transcription buffer, 1mM ATP and GTP, 0.75mM UTP, 0.25mM 4-thioUTP, 0.1mM CTP, 50μCi of [α-^{32}P] CTP (400Ci/mmol), 13 units of RNAguard, 1μg of template DNA and 40 units of T7 or T3 RNA polymerase. The RNA was synthesised at 37°C for 2 hours, a further 10 units of RNA polymerase were added and the sample was incubated at 37°C for 1hr. RNA was isolated in the same manner as unlabelled RNA.

Radiolabelled transcripts for use in electrophoretic mobility shift assays were synthesised in a 50μl reaction contained 1x transcription buffer, 0.5mM ATP, GTP and UTP, 0.1mM CTP, 50μCi of [α-^{32}P] CTP (400Ci/mmol), 13 units of RNAguard, 1μg of template RNA and 40 units of T7 or T3 RNA polymerase. The RNA was synthesised at 37°C for 2 hour, a further 10 units of RNA polymerase were added and the sample was incubated at 37°C for 1hr. RNA was isolated in the same manner as unlabelled RNA.

Radiolabelled transcripts for use in RNase protection assays were synthesised in a 10μl reaction volume containing 1x transcription buffer, 4mM KOH, 20 units of Rnasin, 50μCi of [α-^{32}P] CTP (400Ci/mmol), 0.67mM ATP, UTP and GTP, 5μM unlabelled CTP, 1μg of template DNA and 20 units of T7 RNA polymerase. The RNA was synthesised at 37°C for 1 hour, a further 10 units of RNA polymerase were added and the sample was incubated at

37°C for 1hr. The template DNA was digested and the RNA purified as described previously. 0.2 volumes of 5 x TBE loading Dye was added and the RNA fractionated on a 1.5% agarose gel. RNA was detected by exposure to x-ray film for 30 s and a gel slice containing only full length transcripts was excised from the gel. RNA was extracted by phenol/chloroform extraction. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and the sample put at -80°C for 16hr. This was then defrosted and centrifuged at 12,000g for 5 min to separate the phases. The procedure was then continued as standard and the RNA was finally resuspended in 200µl sterile filtered water. Radiolabelled transcript concentrations were determined by Cerenkov scintillation counting.

2.5.20 RNase protection

RNA samples were combined with 4×10^5 cpm of radiolabelled riboprobe and precipitated with 0.1 volume of 3M NaAc, pH4.6 and 2.5 volumes of ethanol for 30min at -20°C. The RNA was pelleted by centrifugation, washed with 75% ethanol, and briefly dried, after which it was resuspended in 30µl of hybridisation buffer (80% deionised formamide, 40mM PIPES, pH6.4, 0.4M NaCl, 1mM EDTA). The samples were heated at 85°C for 10 min and then incubated at 50°C for 16 hours to allow annealing to occur. 300µl of RNase digestion buffer (10mM Tris-HCl, pH 7.5, 5mM EDTA, 200mM sodium acetate) was then added and single stranded RNA was digested for 120 min at 37°C with RNase ONE™ (Promega) at a concentration of 1 unit/µg of RNA. The reaction was terminated by the addition of 10µl of 20% SDS and proteins were digested with 2.5µl of 20mg/ml proteinase K at 37°C for 15 min. The sample was extracted once with phenol/chloroform and the aqueous phase was removed to a separate tube containing 10µg of yeast carrier tRNA. The sample was ethanol precipitated, washed and resuspended in 8µl of RNA formamide loading buffer. RNA was denatured by heating at 85°C for 5 min and fractionated by electrophoresis through a 4% polyacrylamide/7M urea gel. Finally, radiolabelled RNA fragments were detected by analysis of the dried gel using a Molecular Dynamics phosphorimager.

2.5.21 Denaturing RNA agarose gel electrophoresis and Northern blotting

Samples of RNA were denatured by incubation in 20µl of 1x MOPS buffer (20mM MOPS, pH7.0, 5mM NaAc, and 1mM EDTA), 6.5% formaldehyde, and 50% deionised

formamide at 55°C for 15 min. A 1% agarose gel was prepared in 1x MOPS buffer and 6% formaldehyde. After setting the gel was submerged in 1x MOPS buffer. 4μl of RNA formaldehyde loading buffer was added to each sample prior to fractionation by electrophoresis at 120V for 2-3 hours. After electrophoresis was complete, the portion of gel containing the markers was removed, stained with ethidium bromide (5μg/ml) for 10 min and destained for 10 min using 1x MOPS buffer. The markers were visualised using a UV transilluminator and photographed for later reference. The remainder of the gel was washed in deionised water before being soaked in 0.05 M NaOH for 20 min. After rinsing with deionised water, the gel was incubated in 20x SSC (3M NaCl, 0.3M tri-sodium citrate) for 30 min. The RNA was transferred from the gel to Zetaprobe membrane (Biorad) using capillary blotting for 16 hours and was fixed to the membrane by baking at 80°C for 2 hours before hybridisation.

The DNA fragment required was isolated by gel extraction and a random-primed radiolabelled DNA probe prepared. 30ng of the DNA fragment was heated at 100°C for 5 min in 11μl of sterile deionised water. A 20μl reaction containing 11μl sample, 1x labelling buffer (Promega), 0.25mM BSA, 0.4mM dATP, dGTP and dTTP, 20μCi of [α-^{32}P] dCTP (3000 Ci/mmol) and 5 units of Klenow DNA polymerase was prepared and incubated at 37°C for 1.5 hours. Unincorporated nucleotides were removed by passing the probe through a Sephadex G-50 column.

The RNA bound to the filter was pre-hybridised with 10ml of Church-Gilbert buffer (180mM Na_2HPO_4, 70mM NaH_2PO_4, 7% SDS) supplemented with 0.2mg/ml denatured Salmon sperm DNA and 50μg/ml yeast tRNA (Sigma) for 1 hour at 65°C. The random-primed radiolabelled DNA probe was denatured by heating at 95°C for 5 min and added directly to the pre-hybridisation buffer, followed by hybridisation at 65°C for 16-24 hours. After this the filter was washed once for 20 min at 65°C with Church-Gilbert buffer followed by 2-4 washes for 20 min with Church Wash buffer 1 (14.4mM Na_2HPO_4, 5.6mM NaH_2PO_4, 1mM EDTA, 5% SDS). Further washes using Church Wash buffer 2 (14.4mM Na_2HPO_4, 5.6mM NaH_2PO_4, 1mM EDTA, 1% SDS) were performed if the background counts on the filter remained high after the initial washes. Excess moisture was removed from the filter and radiolabelled probe was detected by Molecular Dynamics phosphorimager.

2.6 Biochemical techniques

2.6.1 Buffers and Solutions

1x SDS sample buffer: 50mM Tris pH 6.8, 10% glycerol, 4% SDS, 0.1% bromophenol blue, 10% β-mercaptoethanol, 1mM EDTA

SDS-PAGE resolving buffer: 1.5M Tris, 0.24% TEMED, 1% SDS pH 8.8

SDS-PAGE stacking buffer: 0.25M Tris, 0.12% TEMED, 0.2% SDS pH 6.8

1x SDS running buffer: 25mM Tris, 192mM glycine, 0.1% SDS pH 8.3

TBST (Tris buffered saline, Tween): 10mM Tris pH 8.0, 0.9% NaCl, 0.1% Tween

Protease inhibitors: 19μg/ml Aprotinin, 1μg/ml Leupeptin, 1μg/ml TLCK, 20μg/ml PMSF, pepstatin

2.6.2 *In vitro* translation reactions

In vitro translation reactions were performed using the Flexi® rabbit reticulocyte lysate system (Promega) with minor modifications to the manufacturer's recommendations. Reactions contained 8.25μl of reticulocyte lysates, 0.6mM MgOAc, 60mM KCl, 0.16mM complete amino acid mixture and 100ng capped RNA template in a final volume of 12.5μl. Protein factors were added at 125-250ng. Reactions were incubated at 37°C for 1.5 hours before analysis of samples by Luciferase assay.

2.6.3 Preparation of total cell extract

1 x 10^7 cells were pelleted by centrifugation at 1000 rpm for 5 min, washed in PBS and resuspended in 200μl of polysome buffer (300mM KCl, 5mM MgCl$_2$, 10mM HEPES pH7.4, 0.5% NP-40 supplemented with protease inhibitors). Cells were incubated on ice for 1 hour and nuclei were sheared by drawing the extract through a 21G needle several times. Cell debris was pelleted by centrifugation at 13000 rpm for 10 min at 4°C. The supernatant was removed to fresh tubes and stored in aliquots at -80°C.

2.6.4 Preparation of cell lysates from transfected cells

After transfection, the medium was aspirated and the adherent cells were washed twice with PBS. Cells were lysed by the addition of 200μl of 1x Passive lysis buffer (Promega) per well of a 6 well plate and lysates were removed from wells with a cell scraper. Lysates were transferred to a tube and the insoluble matter was pelleted by centrifugation.

2.6.5 Luciferase assays

The activity of firefly luciferase in lysates prepared from cells transfected with monocistronic reporter vectors was measured using a luciferase reporter assay system (Promega). Lysates were prepared using 1x Passive lysis buffer and 5µl of lysate was added to 25µl of luciferase assay reagent I. Light emission was measured over 10 s using an Optocomp I luminometer (MGM Instruments).

The activity of both firefly and *Renilla* luciferase in transfections and *in vitro* translation reactions using dicistronic luciferase plasmids was measured using a Dual-luciferase reporter assay system (Promega). Lysates were prepared using 1x Passive lysis buffer and 5µl of lysate was added to 25µl of luciferase assay reagent II (LARII). Light emission was measured as described previously. 25µl of Stop &Glo® was then added and a further light emission reading determined for Renilla luciferase.

2.6.6 β-Galactosidase assays

The activity of β-galactosidase in lysates prepared from cells transfected with pJ7*lacZ* was measured using a Galacto-Light Plus assay system (Tropix). 5µl of cell lysate was added to 25µl of Galacton Plus reaction buffer (Galacton-Plus substrate diluted 1:100 with Reaction Buffer diluent) and incubated at room temperature for 1 hour. 37.5µl of Light Emission Accelerator II was then added and enzyme activity was determined by immediately measuring the light emission from the reaction in a luminometer as previously described.

2.6.7 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein extracts were denatured by the addition of an equal volume of 2x SDS sample buffer containing protease inhibitors and heated at 95°C for 5 min prior to loading. SDS-polyacylamide gels were prepared in a total volume of 5ml containing 1.25ml resolving buffer, 50µl 25% ammonium persulphate (APS), and the appropriate volume of 30%:0.8% acrylamide:bisacrylamide solution. Resolving gels of 5-17.5% acrlyamide were used. Stacking gels were prepared with 0.33ml 30%:0.8% acrylamide:bisacrylamide solution, 1.25ml stacking buffer, 0.9ml water and 50µl 25% APS. Electrophoresis was carried out at 150V in a Bio-Rad Protean II system using SDS running buffer.

2.6.8 Coomassie staining of SDS-polyacrylamide gels

Gels were stained in a Coomassie staining solution (0.1% Coomassie brilliant blue R-250 in 5:1:5 methanol:acetic acid:water) for 30min at room temperature and destained in destaining solution (5:1:5 methanol:acetic acid:water) for 3-5 hours. Gels were then washed in deionised water for 30min before drying.

2.6.9 Transfer of proteins on to nitrocellulose membranes

Cell extracts separated by SDS-PAGE were transferred on to nitrocellulose (Schleicher and Schuell) by semi-dry blotting. This was carried out using transfer buffer (50mM Tris, 192mM glycine, 20% methanol) for between 30 and 90 min at 10V. Protein transfer was visualised temporarily by staining with Ponceau-S solution (0.5% w/v in 5% w/v TCA).

2.6.10 Western blotting

Nitrocellulose membranes from semi-dry blotting were incubated in a 5% dried milk/TBST solution for 1 hour at room temperature to block non-specific binding. Membranes were then incubated in 10ml 5% milk/TBST containing the appropriate concentration of primary antibody for 12-16 hours at 4°C. After three 10 min washes in TBST, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies to mouse IgG (Dako A/S, Denmark) or rabbit IgG (Sigma), diluted 1:2000 and 1:10,000 respectively in 5% milk TBST, for 1 hour at room temperature. Three 10 min washes in TBST solution were carried out and protein-antibody complexes were detected using an enhanced chemiluminescence (ECL) technique. For this, 1ml Luminol solution (50mg Luminol (5-amino-2,3-dihydro-1,4-phthalazinedione) in 0.1M Tris-HCl pH8.6), 10μl Enhancer (11 mg para-coumaric acid in 10 ml DMSO) and 3.1μl 3% hydrogen peroxide were mixed and incubated on the membrane for 60 s. Chemiluminescence was visualised by exposing the membrane to X-ray film for periods of between 10 s and 30 min.

Nitrocellulose membranes were stripped of existing protein-antibody interactions by incubation in a solution of 100mM β-mercaptoethanol, 2% SDS and 62.5mM Tris-HCl pH 6.7 for 10 min at 50°C. Membranes were then washed in TBST and re-probed with a different primary antibody as described above.

2.6.11 Electrophoretic mobility shift assays (EMSAs)

2.5×10^4 cpm of radiolabelled RNA was combined with protein in a total volume of 20μl, containing 0.5 x transcription buffer, 1mM ATP and 13 units of RNAguard. The reaction was incubated at room temperature for 10min after which 0.2 volumes of 5x TBE loading buffer was added. Reactions were run on 5% polyacrylamide/0.5% TBE/ 5% Glycerol gels at 150V in 0.5% TBE using the Bio-Rad Protean II system. After electrophoresis gels were dried and results visualised by phosphorimager.

2.6.12 UV cross-linking

Reactions were performed in a final volume of 30μl containing 4×10^5 cpm of radiolabelled RNA, 15μl 2x UV cross-linking buffer (20mM HEPES pH7.4, 6mM MgCl2, 2.6mM ATP, 10mM creatine phosphate, 2mM DTT, 200mM KCl, 12% glycerol, 0.2 μg/μl yeast tRNA), and appropriate amount of protein and cold competitor RNA. Reactions were incubated at room temperature for 15 min before being irradiated at 312nm using a UV source (UVP) at a distance of 5cm, for 30 min on ice. After irradiation, unbound RNA was digested with 0.2mg/ml pancreatic RNase A (Sigma) and 600units RNase T1 (Ambion) for 30 min at 37°C. An equal volume of 2x SDS sample buffer was added to each sample and cross-linked RNA-protein complexes separated by SDS-PAGE. Gels were fixed, dried and visualised by a phosphorimager.

2.6.13 His-Tagged protein expression

DNAs of proteins of interest were cloned into pET21a in frame with His Tag extension and plasmids transformed into BL21 bacterial cells. An overnight culture of transformed cells was inoculated into 500mls of LB media supplemented with ampicillin and incubated at 37°C with shaking until the A_{595} reached 0.6. Protein expression was induced by the addition of 1mM IPTG (isopropyl β-D-thiogalactopyranoside) for 3hrs at 30/37°C. Cells were then pelleted by centrifugation at 4000rpm for 10min, washed with PBS and resuspended in 5ml sonication buffer (50mM sodium phosphate buffer pH7.8, 300mMNaCl, 20mM Imidazole). Sonication was carried out, with the addition of proteinase inhibitors, and samples incubated for 30 min on ice. Cell debris was then pelleted by centrifugation at 12,000g for 5min at 4°C and lysates transferred to a fresh tube, before Imidazole concentrations of lysates were adjusted to a 30mM. Ni-NTA (nickel-nitrilotriacetic acid) agarose columns were prepared and washed with 20ml

sonication buffer prior to loading samples. The sample eluates were collected and reloaded onto the column. Columns were then washed with 10ml sonication buffer containing 30mM Imidazole, and 5ml sonication buffer containing 50mM Imidazole, before proteins were eluted with the addition of 3ml sonication buffer containing 250mM Imidazole. Eluted proteins were dialysed overnight against 2l H100 buffer (20mM HEPES pH7.5, 100mM KCl, 2mM DTT). Proteins were analysed by SDS-PAGE, and concentrations determined by Bradford assay, before aliquoting and freezing at -80°C.

Chapter 3

Investigation of protein factors that interact with the c-*myc* IRES.

3.1 Introduction

The majority of information regarding the protein factor requirements for internal initiation comes from the picornaviral family. It was found that internal initiation of the EMCV IRES had almost the same requirement for canonical initiation factors as cap-dependent initiation. The EMCV IRES was found to require eIF2, eIF3, and eIF4F for translation (Pestova et al., 1996a). Further work showed that the eIF4F required could be substituted by the central third of eIF4G and eIF4A. Type 1 picornaviral IRESs have also been shown to have a requirement for additional factors for their activity called IRES *trans*-acting factors (ITAFs). A large number of these have been discovered and the requirements for them differ between IRESs. Several type 1 IRESs require more than one of these *trans*-acting factors for activity.

Investigation of other IRES elements has identified different protein factor requirements. HCV, which is a member of the *Flaviviridae* family, has been found to bind 43S complexes directly at the initiation site and has no requirement for the eIF4F, eIF4A or eIF4B group of initiation factors (Pestova et al., 1998b). However, HCV initiation has also been found to involve *trans*-acting factors. More recently CrPV, an insect picorna-like virus, has been shown to have no canonical initiation factor requirement, even for eIF2 and the ternary complex (Wilson et al., 2000a).

Significantly less is known about the requirements of cellular IRESs. It has been found that they appear to require both canonical and *trans*-acting factors. The evidence of canonical factor requirement is sparse. There are reports of eIF4G stimulating ODC IRES translation (Hayashi et al., 2000), and DAP5, an eIF4G homolog, acting on translation from its own IRES (Henis-Korenblit, 2000). More work has centred on *trans*-acting requirements for cellular IRESs, for example the Apaf-1 IRES has been shown to require both UNR and PTB for function (Mitchell et al., 2001).

Little is known about the protein factors requirement of the c-*myc* IRES. This IRES has been found to have very different *in vivo* activities between cell lines (Stoneley et al., 2000b). The IRES activity is high in HeLa cells, followed by decreasing activity in MRC5, HepG2, GM637, HEK293, and Cos-7. The activity is very low in MCF7 and Balb/c-3T3. This suggests there may a requirement for protein factors that are present in these cell lines

in differing amounts. The IRES has also been found to be less efficient when expressed from the CMV promoter/enhancer (Stoneley et al., 2000b). This could be due to limiting amounts of required *trans*-acting factors under these conditions.

North-western blotting and UV cross-linking identified the binding of a large number of proteins to the IRES (Paulin et al., 1998). Bands of the size 160, 138, 98, 82, 61, 57, 46 and 38kDa were shown to bind by North-western blotting, and bands of 105, 98, 86, 61, 57 and 38kDa were detected by UV cross-linking. Further work has also shown that the 98kDa *trans*-acting factor UNR cross-links to the c-*myc* IRES (Mitchell et al., 2001). To investigate proteins that interact with the c-*myc* IRES we studied known canonical and *trans*-acting factors involved in internal initiation.

3.2 UV cross-linking of the c-*myc* IRES RNA to HeLa cell extract.

In order to identify proteins interacting with the c-*myc* IRES in HeLa cell extract UV cross-linking was performed. HeLa cell extract was chosen due to the high activity of the IRES in this cell line, which could signify a high abundance of protein factors necessary for efficient IRES translation. The radiolabelled c-*myc* IRES transcript was prepared by *in vitro* run-off transcription from pSKML linearised with the restriction enzyme *Nco*I (Figure 3.1A). Radiolabelled RNA was incubated in the appropriate buffer with 10μl of HeLa total cell extract for 15 minutes at room temperature. UV cross-linking was then performed for 30 minutes on ice, before unbound RNA was digested, and RNA-protein complexes separated by SDS-PAGE. The c-*myc* IRES RNA was found to bind a number of proteins in HeLa cell extract (Figure 3.1B-Lane 0) ranging in size from 25-100kDa. Bands of approximately 28, 35, 38, 42, 50, 70 and 100kDa were observed. Poly (C) was used as a competitor in the UV cross-linking reaction at 1, 5, and 10 fold molar excess to c-*myc* IRES RNA. It was found that several of the proteins of the sizes 38-70kDa were competed off with the addition of Poly (C) (Figure 3.1B). These corresponded to the sizes of members of the poly (rC) binding protein family, which includes PCBP1(42KDa), PCBP2(42KDa), PCBP3(36kDa), PCBP4(42kDa), and hnRNP K(68KDa). This family of proteins is known to be involved in picornaviral IRES translation, for example PCBP2 is a known *trans*-acting factor involved in polioviral translation (Silvera et al., 1999). Therefore these proteins were of interest as potential c-*myc* IRES *trans*-acting factors.
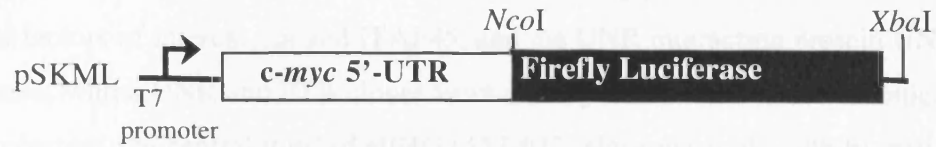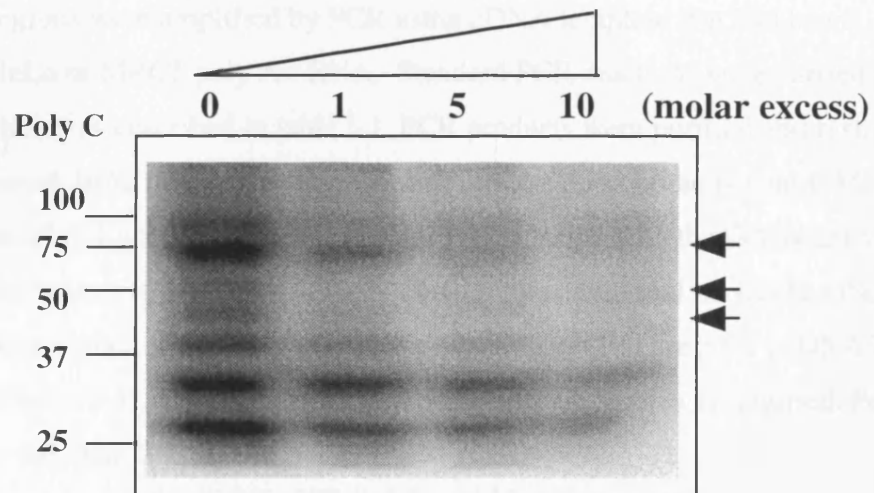
Figure 3.1 UV cross-linking of c-*myc* IRES RNA and HeLa cell extract. A) pSKML was linearised with *Nco*I and transcripts produced by *in vitro* run-off transcription from the T7 promoter. B) UV crosslinking of c-*myc* IRES RNA and HeLa cell extract. Poly (C) was added at 1, 5 and 10 fold molar excess to IRES RNA. Bands of the sizes 38-70 kDa were competed off with the addition of poly (C) as indicated by arrows.

3.3 Cloning of known canonical and *trans*-acting factors.

In order to further investigate the poly (C) binding protein family several factors from the family were obtained either by cloning or from other sources. The factors PCBP1, and PCBP2 were chosen as known picornaviral *trans*-acting factors (Gamarnik and Andino, 1997), and the protein hnRNP K, which is recognised to have a role in c-*myc* transcription (Michelotti et al., 1996). Further *trans*-acting proteins were chosen, and two canonical factors of interest. La and ITAF45, and the UNR interacting protein UNRIP were cloned, whilst UNR and PTB clones were already available. For the canonical initiation factors, the central third of eIF4G (457-932) (Pestova et al., 1996b), and both the p97 and p86 forms of DAP5 were cloned (Henis-Korenblit, 2000). The full list of proteins obtained is shown in table 3.1.

Coding regions were amplified by PCR using cDNA template that had been produced from HeLa or MRC5 poly A+ RNA. Standard PCR reactions were carried out using specific primers as described in table 3.1. PCR products were purified and restriction digestions performed, before fragments were cloned into the appropriate plasmid. Most products were initially cloned into pACTII (Figure 3.2) in frame with the GAL4 activation domain and fully sequenced, before subsequent cloning into additional plasmids. DNA coding regions were cloned into pET21a for protein expression (Figure 3.3), pcDNA3 for *in vivo* studies (Figure 3.4), and pACTII for yeast three-hybrid studies as required. For primer sequences see table 2.2.

Protein	Reason obtained	Obtained	Subsequent cloning.
La	IRES *trans*-acting factor	PCR from MRC5 template (LAFOR + LAREV). Cloned into pACTII.	PCR from pACTII (PETLAF + PETLAR) and cloned into pET21a. Restriction digest from pACTII (*Eco*RI + *Xho*I) and cloned into pcDNA3.
PCBP1	IRES *trans*-acting factor	PCR from HeLa template (PCBP1F + PCBP1R) and cloned	PCR from pACTII (PETP1F + 2PETP1R) and

		into pACTII.	cloned into pET21a. Restriction digest from pACTII (*Eco*RI + *Xho*I) and cloned into pcDNA3.
PCBP2 (All splice variants)	IRES *trans*-acting factor	PCR from HeLa template (PCBP2F + PCBP2R) and cloned into pACTII.	Full length PCR from pACTII (PETP2F + 2PETP2R) and cloned into pET21a. All splice variants restriction digest from pACTII (*Eco*RI + *Xho*I) and cloned into pcDNA3.
ITAF45	IRES *trans*-acting factor	PCR from HeLa template (ITAFF + ITAFR) and cloned into pcDNA3 and pACTII.	PCR from pACTII (PETITAFF + PETITAFR) and cloned into pET21a.
UNR	IRES *trans*-acting factor	pET28a and pcDNA3 clones available.	PCR from pcDNA3 (UNRNCO1 + UNRXHO1) and cloned into pACTII.
PTB	IRES *trans*-acting factor	pET21a and pcDNA3 clones available	PCR from pcDNA3 (PTBFOR + PTBREV) and cloned into pACTII.
hnRNPK	Poly (rC) binding protein	pET and pSG5 clones available	
UNRIP	UNR interacting protein	PCR from HeLa template (UNRIPF + UNRIPR) and cloned into pcDNA3. pET28a clone available.	

62

eIF4G (457-932)	Canonical translation factor	PCR from HeLa template (4GFOR+ 4GREV) and cloned into pACTII.	PCR from pACTII (PET4GF + PET4GR) and cloned into pET21a. Restriction digest from pACTII (*Eco*RI + *Xho*I) and cloned into pcDNA3.
DAP5 (p97)	Canonical translation factor	PCR from MRC5 template (NDAP5F + NDAP5R) and cloned into pACTII.	PCR from pACTII (PETDAPF2 + PETDAPR) and cloned into pET21a. Restriction digest from pACTII (*Eco*RI + *Xho*I) and cloned into pcDNA3.
DAP5 (p86)	Canonical translation factor	PCR from p97 template (NDAP5F + DAP86R) and cloned into pACTII and pcDNA3.	PCR from pACTII (PETDAPF2 + NDAP86R) and cloned into pET21a.

Table 3.1. DNAs obtained encoding proteins of interest. Plasmids, primers and templates used are shown.

3.4 UV cross-linking of known protein factors to the c-*myc* IRES

Recombinant proteins were obtained by His-Tag purification using Nickel affinity columns. DNAs of interest were cloned into the pET21a plasmid in frame with the C-terminal His-Tag extension. Plasmids were transformed into BL21 bacterial cells and overnight cultures obtained, before protein expression was induced by the addition of 1mM IPTG for 3hr at 37°C. The cell lysates were added to Ni-NTA agarose columns and proteins were eluted with the addition of 250mM Imidazole. HnRNPK was obtained in the plasmid pET16b with an N-terminal His-Tag extension, and protein expression induced at
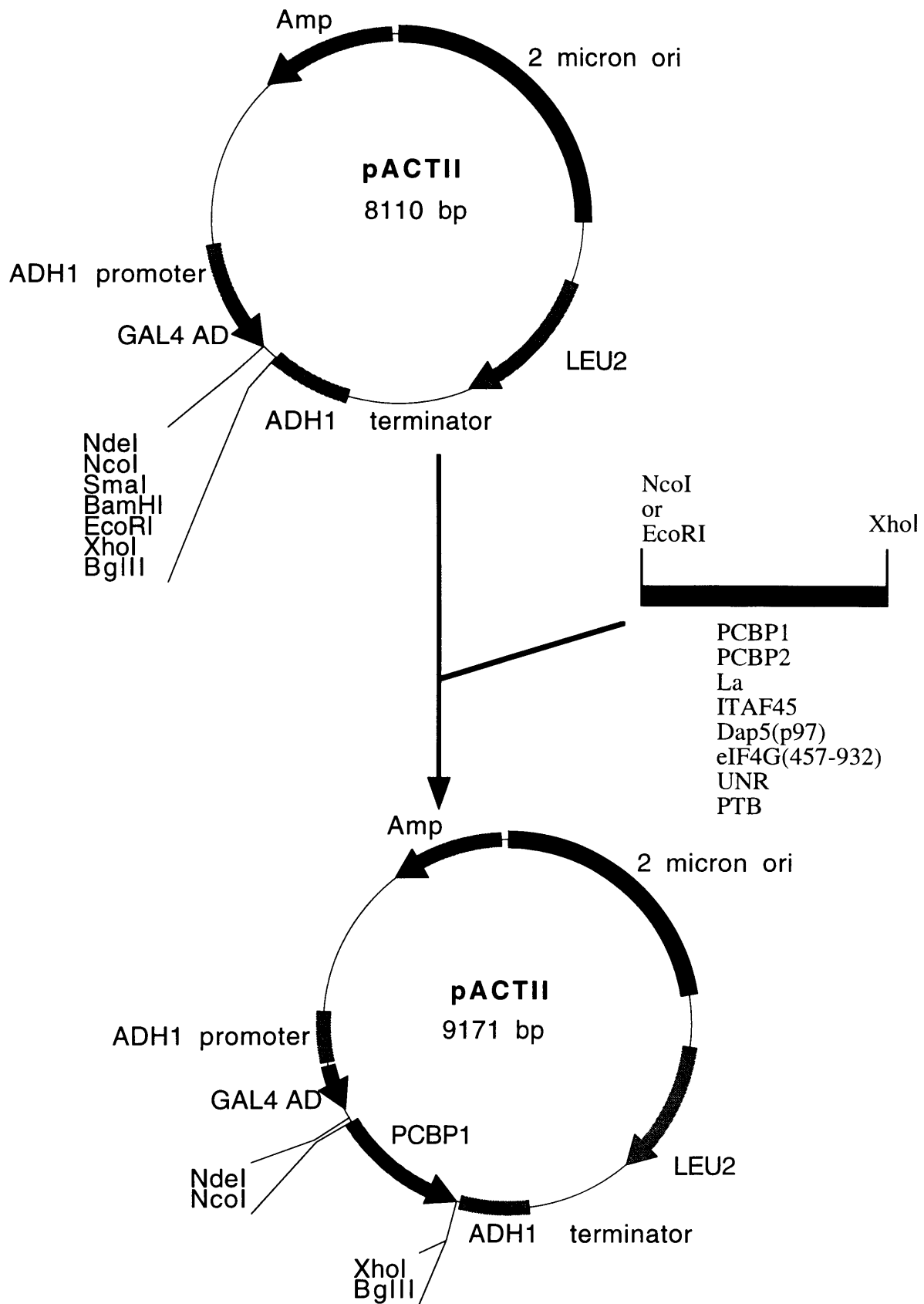
Figure 3.2 Cloning of pACTII constructs. Products were PCR amplified
with gene specific primers containing required restriction sites and cloned
into the polylinker site of pACTII in frame with the GAL4 activation
domain.
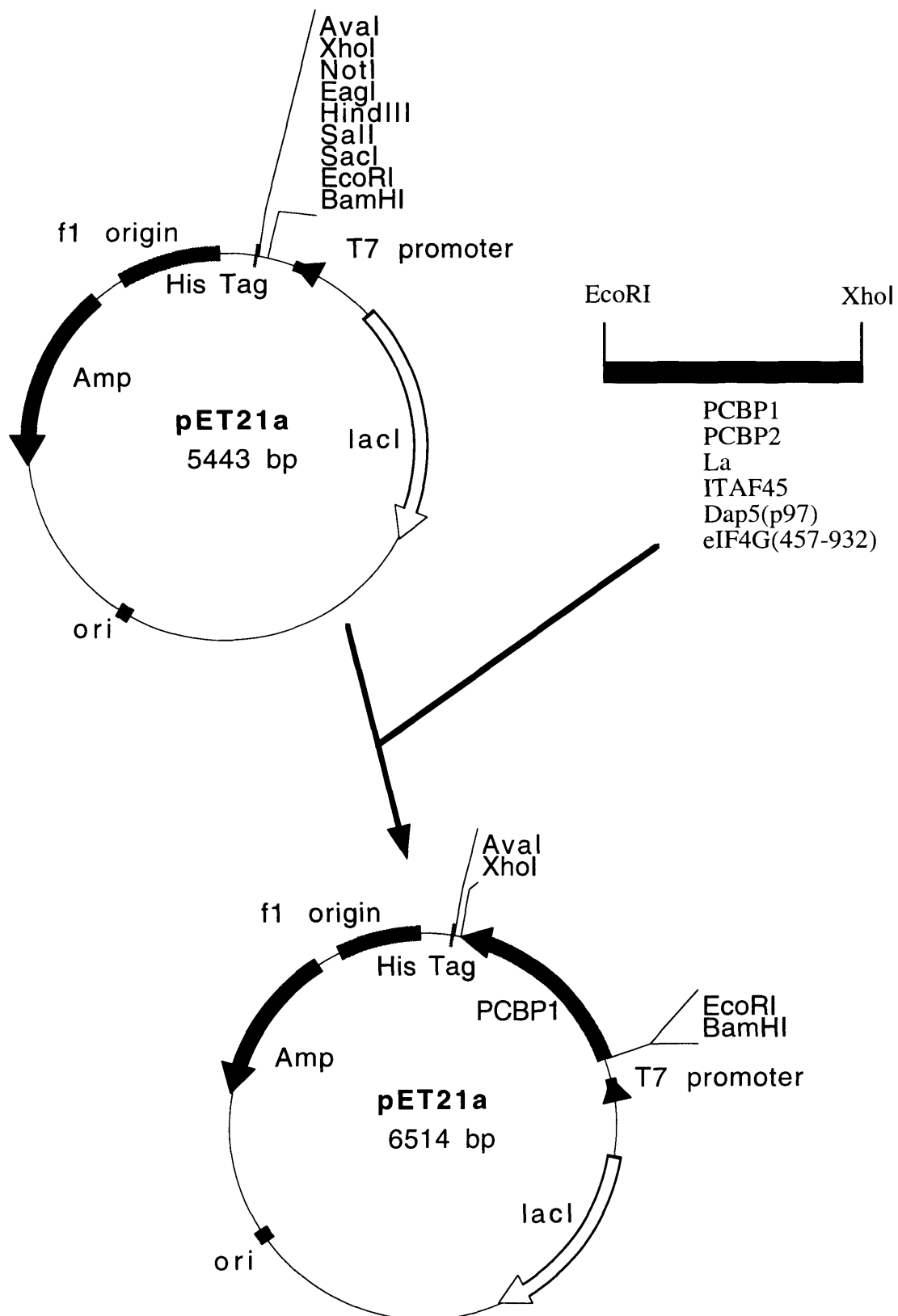
Figure 3.3 Cloning of pET21a constructs. Products were PCR amplified with gene specific primers containing required restriction sites and cloned into the polylinker site of pET21a in frame with the His Tag sequence.
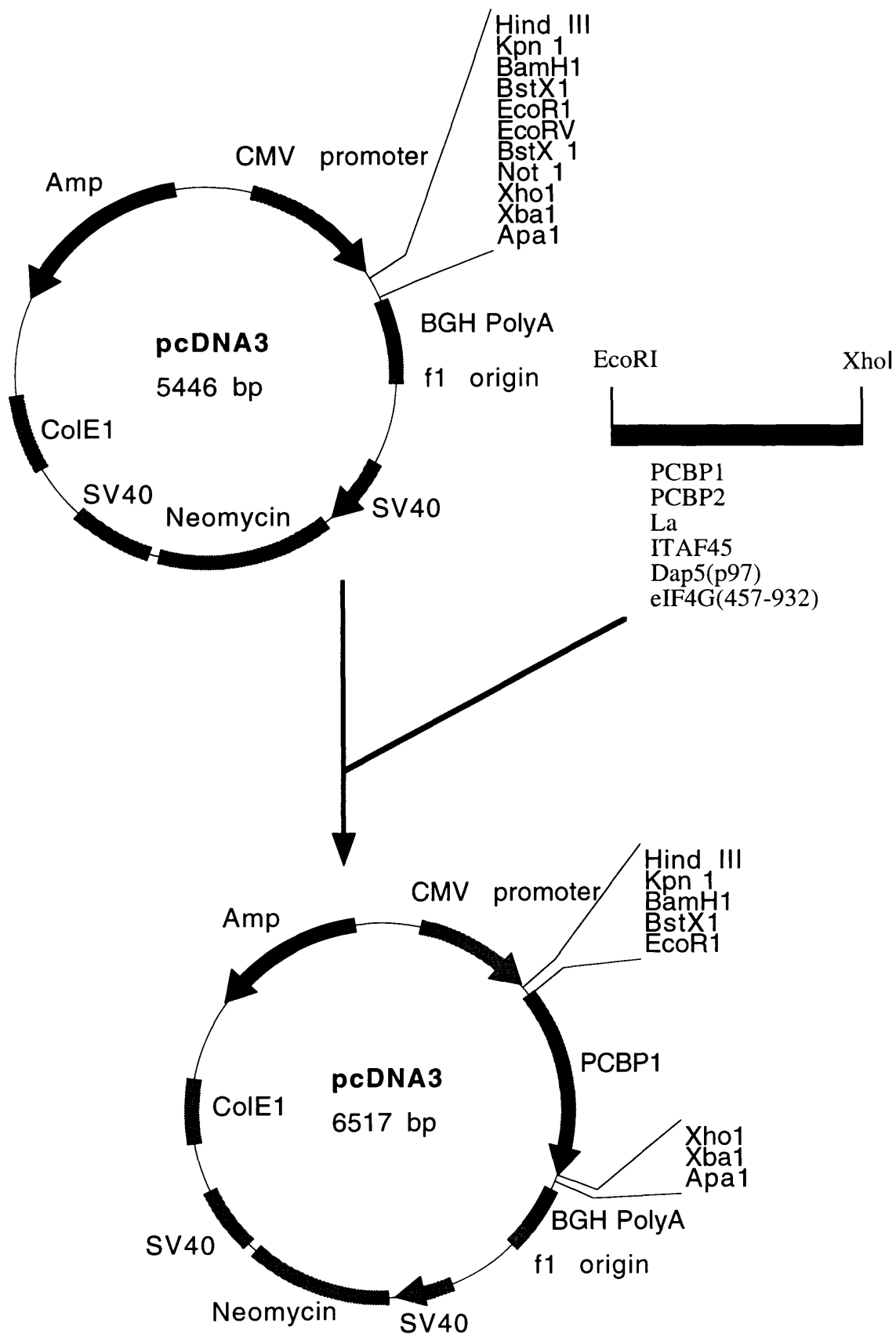
Figure 3.4 Cloning of pcDNA3 constructs. Products were excised from pACTII with appropriate restriction enzymes or PCR amplified with gene specific primers containing required restriction sites and cloned into the polylinker site of pcDNA3.

30°C. Recombinant UNR and its interacting protein UNRIP were already available (Mitchell et al., 2001). Eluted proteins were analysed by SDS-PAGE (Figure 3.5).

UV cross-linking reactions were performed using radiolabelled c-*myc* IRES RNA in the appropriate buffer with 0.5-1μg protein. UV cross-linking was observed between the c-*myc* IRES and HnRNPK, PCBP1, PCBP2 and UNR (Figure 3.6). No UV cross-linking was observed for La, ITAF45, eIF4G (457-932), DAP5 (p97) or UNRIP. Competition reactions were then performed using unlabelled c-*myc* IRES and control GAPDH RNA. Cold competitor RNAs were produced by *in vitro* run-off transcription from pSKML linearised with *Nco*I for c-*myc* IRES RNA, and pSKGAP:E/H linearised with *Hind*III for GAPDH RNA respectively. Reactions were supplemented with 1, 5, or 10 fold molar excess of unlabelled competitor RNA. For PCBP1, PCBP2, and HnRNPK the interaction of the c-*myc* IRES with the protein could be competed off with increasing amounts of unlabelled c-*myc* IRES RNA (Figure 3.6A) but not with unlabelled GAPDH RNA (Figure 3.6B). This showed that the proteins were interacting with increased specificity for the c-*myc* IRES RNA than for the GAPDH RNA. The interaction between the c-*myc* IRES and UNR was competed off to the same degree with c-*myc* IRES and GAPDH RNA, showing there was no difference in its specificity for the two RNAs.

UV cross-linking of the c-*myc* IRES RNA to PCBP1 produced both a band of the expected 38kDa, and a faint band of higher molecular weight (Figure 3.6). The higher band is of approximately 100kDa. This higher band could be due to 2 or more PCBP1 molecules binding close on the RNA, or PCBP1 binding the RNA as a dimer. Members of the poly (C) binding protein family, including PCBP1 are known to be able to form homodimers (Makeyev and Liebhaber, 2002; Gamarnik and Andino, 1997).

UV cross-linking was then carried out using combinations of proteins. Combinations of HnRNPK and PCBP1, or HnRNPK and PCBP2 were used, and competitions performed as previously described. In each case both proteins could be UV cross-linked to the c-*myc* IRES RNA together (Figure 3.7A). However, in competition experiments with unlabelled c-*myc* IRES RNA, PCBP1 and PCPB2 proteins were competed off before HnRNPK (Figure 3.7 Ai). This shows that there is a difference in the strength of binding between the PCBP proteins and HnRNPK with the c-*myc* IRES RNA. The interactions could not be competed off with unlabelled GAPDH RNA (Figure 3.7 Aii). It was also found that both UNR and PCBP1, or UNR and PCBP2 could be UV cross-linked to the c-*myc* IRES in combination (Figure 3.7 B). Experiments using PCBP1 and
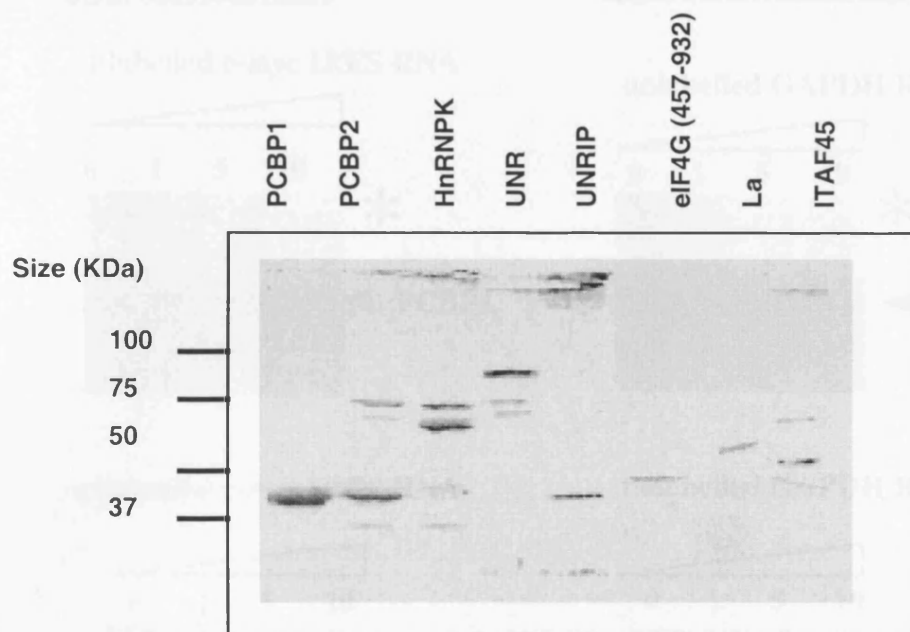
Figure 3.5. SDS-PAGE analysis of recombinant proteins.
Proteins were run on SDS-PAGE gels and visualised by
coomassie staining.
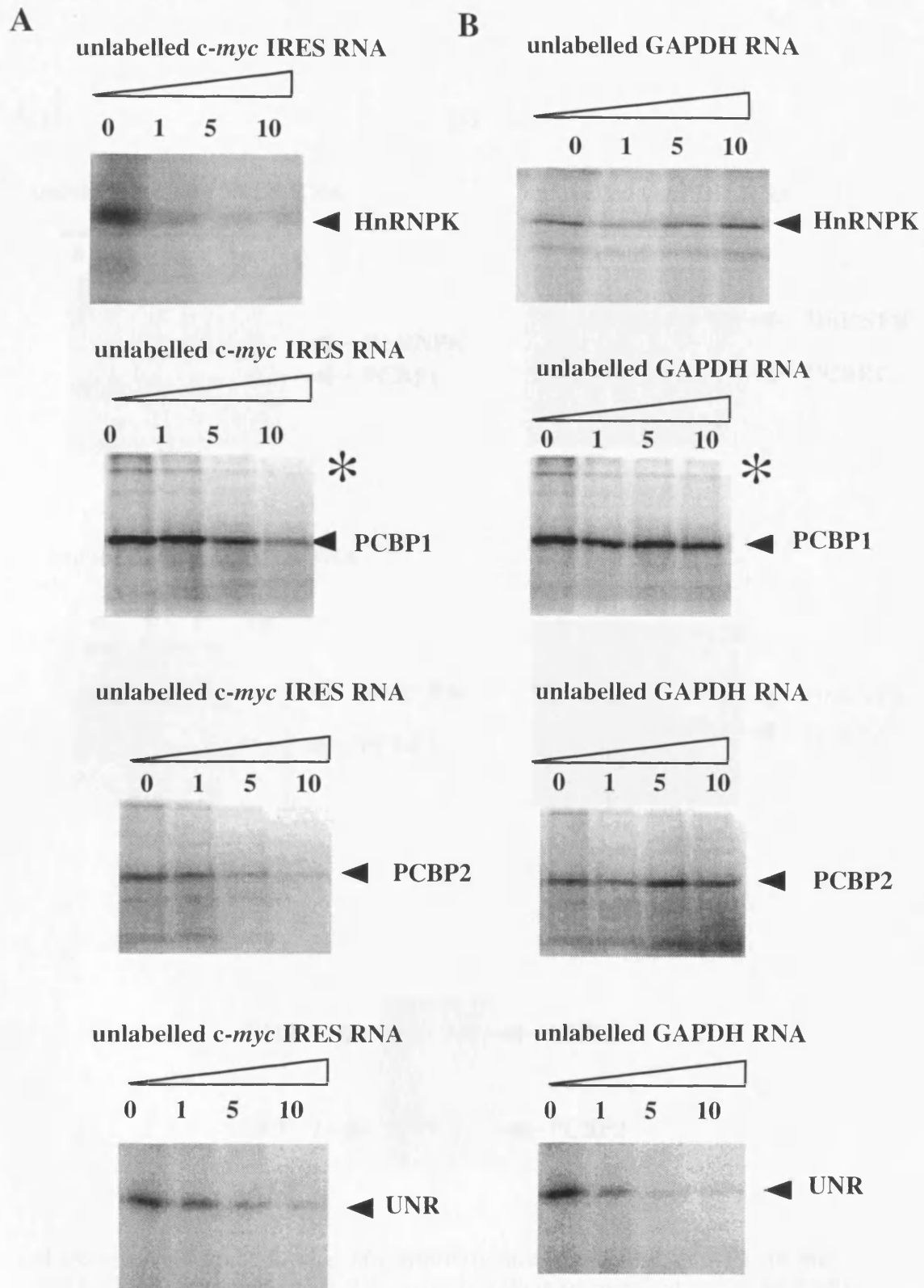
Figure 3.6. UV cross-linking of HnRNPK, PCBP1, PCBP2 and UNR to the c-*myc* IRES RNA. 0.5μg of HnRNPK, PCBP1 and PCBP2 protein was used, and 1μg UNR. Competitions were performed with 1, 5, and 10 fold molar excess of unlabelled c-*myc* IRES RNA (A) or unlabelled GAPDH RNA (B). * A faint band of a higher molecular weight was detected when the c-*myc* IRES RNA was crosslinked to PCBP1.

A i)   **ii)**

unlabelled c-*myc* IRES RNA   unlabelled GAPDH RNA

0   1   5   10   0   1   5   10
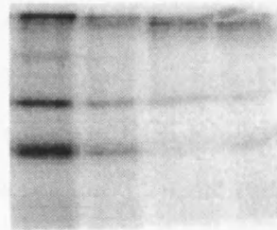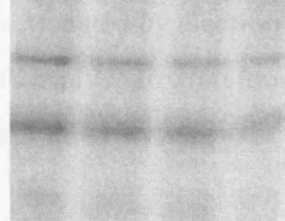
← HnRNPK   ← HnRNPK

← PCBP1   ← PCBP1

unlabelled c-*myc* IRES RNA   unlabelled GAPDH RNA

0   1   5   10   0   1   5   10

← HnRNPK   ← HnRNPK

← PCBP2   ← PCBP2

B

UNR →   ← UNR
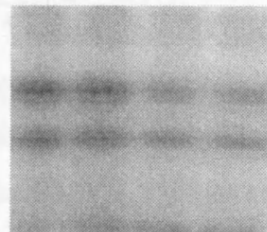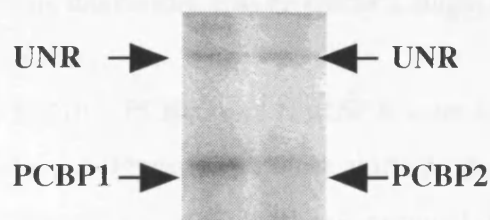
PCBP1 →   ←PCBP2

Figure 3.7 UV cross-linking of combinations of HnRNPK, PCBP1, PCBP2 and UNR to the c-*myc* IRES. A) UV cross-linking of combinations of HnRNPK and PCBP1 or PCBP2 to the c-*myc* IRES. Competitions were performed with 1, 5, or 10 fold molar excess of unlabelled c-*myc* IRES RNA (i) or GAPDH RNA (ii). B) UV cross-linking of combinations of UNR and PCBP1 or PCBP2 to the c-*myc* IRES.

PCBP2 together could not be performed as the proteins have comparable sizes and could not be distinguished by SDS-PAGE.

3.5 EMSAs of the c-*myc* IRES with known protein factors

EMSAs were performed using radiolabelled c-*myc* IRES RNA and recombinant protein factors. Radiolabelled IRES RNA was incubated in the appropriate buffer with 0.5-1μg protein at room temperature. Reactions were fractionated by electrophoresis on 5% polyacrylamide/TBE gels. HnRNPK, PCBP1, PCBP2 and UNR were found to interact with the c-*myc* IRES RNA (Figure 3.8). Differences were observed in the band shifts patterns obtained for the four proteins.

HnRNPK showed two shifts on the gel. Approximate molar ratios of RNA to protein were calculated for the reactions. With a 1:1 molar ratio of protein to RNA a single shift was observed, however with a 2:1 ratio of protein to RNA a higher shift on the gel was seen. This suggests that HnRNPK may be able to bind the RNA as a monomer or dimer. A similar effect is seen for PCBP1. With a 1:1 ratio a smear is observed and no clear shift is seen. This suggests that a stable structure is not being produced, and PCBP1 is only binding weakly. When a 2:1 ratio of protein to RNA is used there is a clear shift higher in the gel, indicating that PCBP1 is binding the RNA as a dimer and strong binding is only obtained under these conditions. This agrees with the UV cross-linking results obtained for PCBP1 where a faint higher band is observed suggesting 2 or more PCBP1 molecules are binding (Section 3.4). The PCBP2 EMSA only produces a lower single shift in the gel. Therefore PCBP2 is likely to be binding as a monomer. UNR interacts with the c-*myc* IRES by EMSA. This interaction also produces a single shift in the gel of the expected size to be a monomer.

Combinations of PCBP1, PCBP2 and HnRNPK were found to produce only higher shifts of c-*myc* RNA on the gel (Figure 3.9A). The shifts produced could not be distinguished in size and therefore it could not be determined if they were due to homodimers or heterodimers of the proteins binding.

Other proteins UNRIP, ITAF45, DAP5, eIF4G (457-932) and La did not interact with the c-*myc* IRES RNA by EMSA (Figure 3.9B).

Protein (molar excess)	O	1X	2X
c-*myc* IRES RNA	+	+	+

HnRNPK

PCBP1

PCBP2

UNR
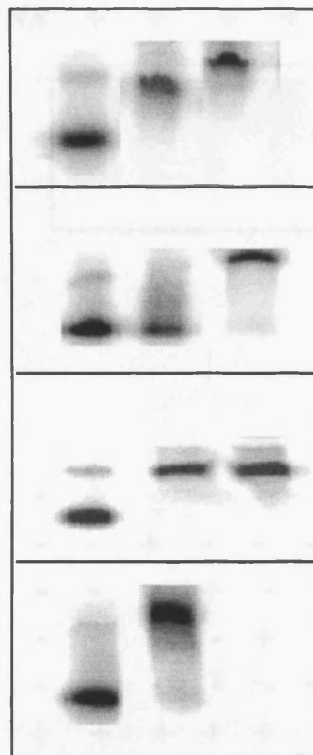
Figure 3.8. Interaction of the c-*myc* IRES with HnRNPK, PCBP1, PCBP2, and UNR by EMSA. Protein was added at approximately 1 fold, and 2 fold molar excess to c-*myc* IRES RNA. UNR crosslinking was carried out with 1 fold molar excess of protein only.

A

PCBP1	-	+	+	-
PCBP2	-	+	-	+
HnRNPK	-	-	+	+
c-*myc* IRES RNA	+	+	+	+

B

UNRIP	-	+	-	-	-	-
ITAF45	-	-	+	-	-	-
Dap5 (p97)	-	-	-	+	-	-
eIF4G (457-932)	-	-	-	-	+	-
La	-	-	-	-	-	+
c-*myc* IRES RNA	+	+	+	+	+	+
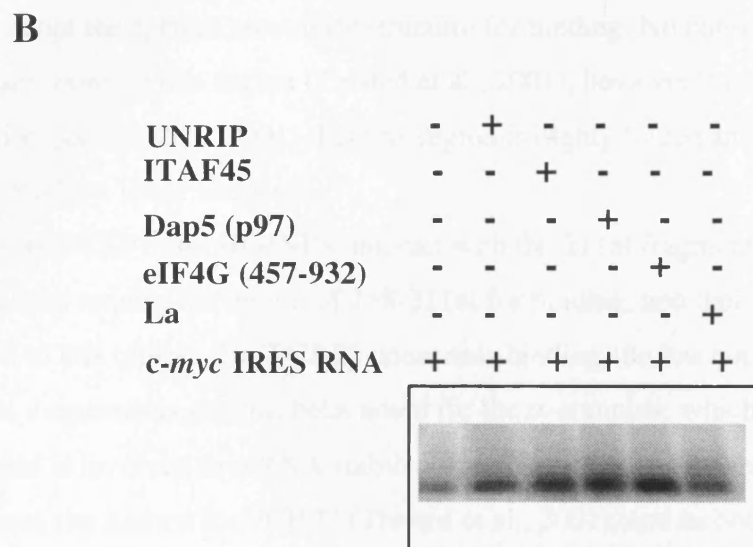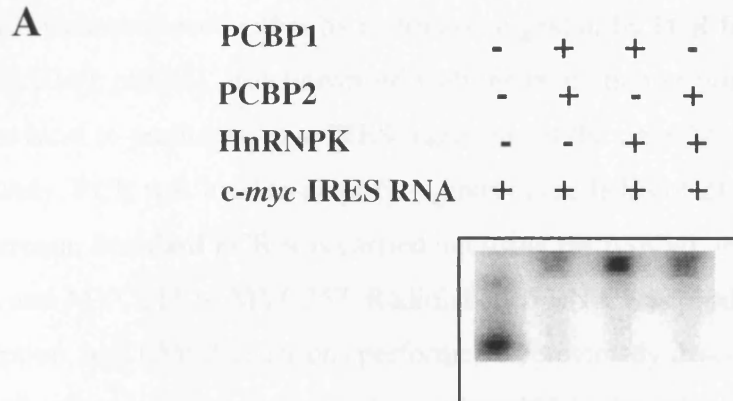
Figure 3.9. Interaction of the c-*myc* IRES with combinations of proteins, and UNRIP, ITAF45, Dap5 (p97), eIF4G (457-932) and La by EMSA. A)EMSAs of combinations of HnRNPK, PCBP1, and PCBP2 with the c-*myc* IRES RNA . Reactions were performed with 0.75µg of protein. B) EMSAs of the c-*myc* IRES with UNRIP, ITAF45, Dap5 (p97), eIF4G (457-932) and La. Reactions were performed with 1µg protein.

3.6 Identifying the binding regions for PCBP1, PCBP2 and HnRNP K in the c-*myc* IRES

Further EMSAs were undertaken in order to identify the regions of the c-*myc* IRES that were interacting with the proteins PCBP1, PCBP2 and HnRNPK. Regions of the c-*myc* IRES were produced either by restriction digest or by PCR from the plasmid pSKML (Figure 3.10A). pSKML was linearised with the restriction enzymes *Eag* I, *Kpn* 2I, *Rsr*II, *Pvu*II, or *Nco*I to produce c-*myc* IRES fragments of the sizes 34, 98, 158, 340, 396nt respectively. PCR was used to amplify regions of the IRES of 211 and 257nt with the T7 site upstream. Standard PCR was carried out using the pSKML template and primers T7FOR and MYC211 or MYC257. Radiolabelled RNA was produced by *in vitro* run-off transcription, and EMSA reactions performed as previously described.

The 34nt fragment was not observed to shift in the gel with the addition of any proteins (Figure 3.10B). PCBP2 interacted with the 98nt fragment and therefore requires the region 34-98nt for binding. It may bind in this region or require this region for the IRES to adopt the correct secondary structure for binding. No consensus binding sites for PCBP2 are found in this region (Thisted et al., 2001), however it is known from structural studies (Le Quesne et al., 2001) that this region is highly folded and therefore the binding site may not be a linear sequence.

Both PCBP1 and HnRNPK interact with the 211nt fragment of the IRES. Therefore these proteins require the region of 158-211nt for binding, and their binding sites may be contained in this region. The PCBP1 consensus binding site has not been determined. However, a consensus site has been noted for the α-complex, which includes PCBP1 and PCBP2 and is involved in mRNA stabilization (Holcik and Liebhaber, 1997). There is also a consensus site known for PCBP2 (Thisted et al., 2001), and as both PCBP1 and PCBP2 are able to independently bind the same site in α-globin mRNA it is known PCBP1 and PCBP2 share similar binding specificities (Chkheidze et al., 1999). The α-complex consensus is (C/U)CCA N_x CCC(U/A) Py_x UC(C/U)CC, and like the known PCBP2 consensus is CU rich and consists of a series of three C repeat regions. A C/U region fitting this general consensus is present in the c-*myc* IRES at 128-150nt (Figure 3.11). This is contained within the 158nt fragment that did not shift with PCBP1, but has been noted as a potential PCBP1 binding site as the protein may need additional sequence after this site to bind. It may also be that extra sequence is required to adopt the correct secondary structure for protein binding.
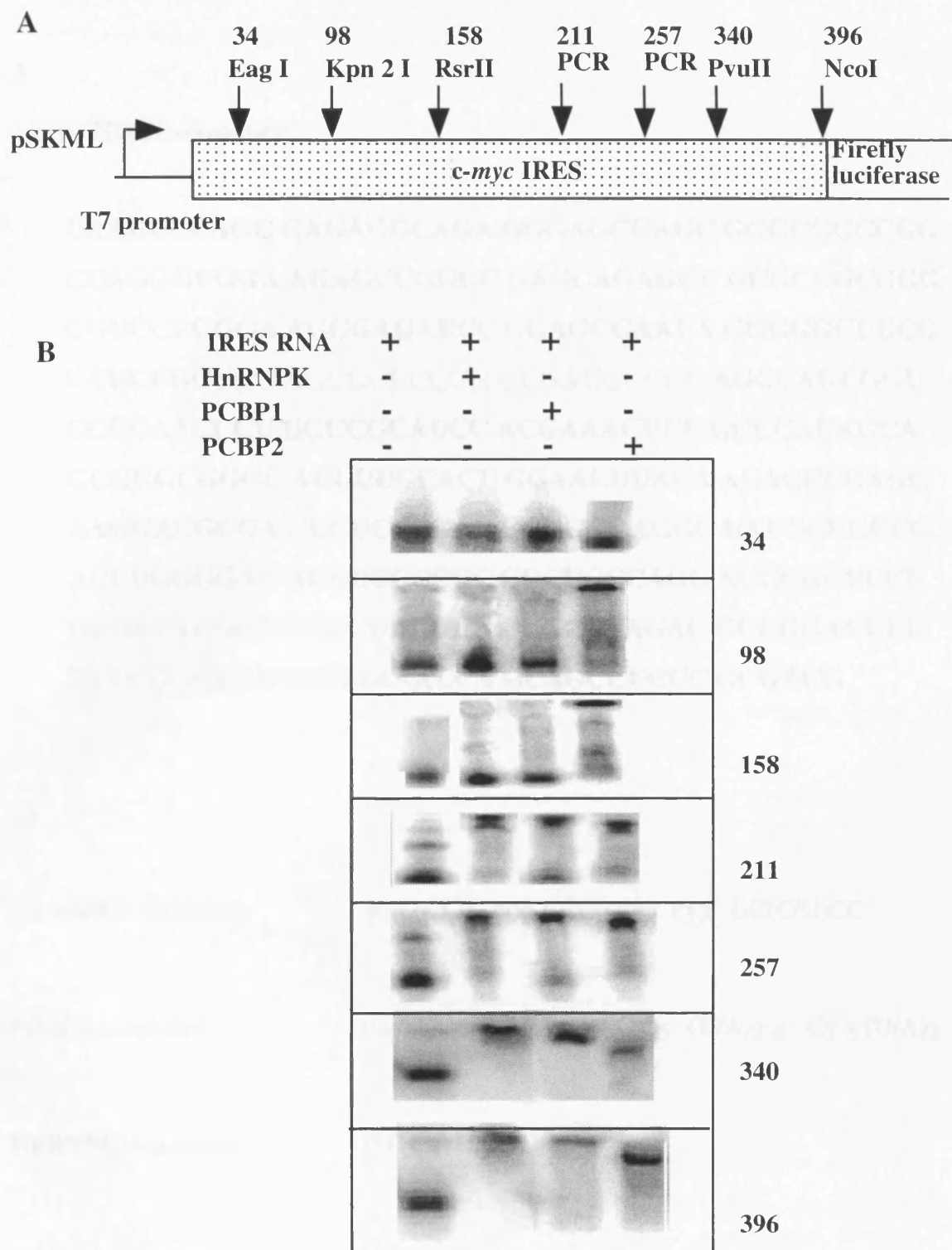
Figure 3.10. Production of c-*myc* IRES RNA fragments and interactions with HnRNPK, PCBP1 and PCBP2 by EMSA. A) c-*myc* IRES fragments were produced by *in vitro* run-off transcription reactions on pSKML that had been restriction digested or used in PCR as shown. B)EMSAs of resulting c-*myc* IRES fragments with HnRNPK, PCBP1 and PCBP2. Reactions were performed with 1μg of protein.

A

c-*myc* IRES sequence

UAAUUCCAGC GAGAGGCAGA GGGAGCGAGC GGGCGGCCGG

CUAGGGUGGA AGAGCCGGGC GAGCAGAGCU GCGCUGCGGG

CGUCCUGGGA AGGGAGAUCC GGAGCGAAUA GGGGGGCUUCG

CCUCUGGCCC AGCCCUCCCG CUGAUCCCCC AGCCAGCGGU

CCGCAACCCU UGCCGCAUCC ACGAAACUUU GCCCAUAGCA

GCGGGCGGGC ACUUUGCACU GGAACUUACA ACACCCGAGC

AAGGACGCGA CUCUCCCGAC GCGGGGAGGC UAUUCUGCCC

AUUUGGGGAC ACUUCCCCGC CGCUGCCAGG ACCCGCUUCU

CUGAAAGGCU CUCCUUGCAG CUGCUUAGAC GCUGGAUUUU

UUUCGGGUAG UGGAAAACCA GCAGCCUCCC GCGACG

B

α-complex consensus	(C/U)CCA N_X CCC(U/A) Py_X UC(C/U)CC
PCBP2 consensus	$(U/A)_2C_{3-5}$ $(U/A)_{2-6}$ C_{3-5} $(U/A)_{2-6}$ $C_{3-5}(U/A)_2$
HnRNPK consensus	$U\ C_{3-4}\ (UA/AU)$

Figure 3.11. Potential PCBP1 and HnRNPK binding sites in the c-*myc* IRES. A) Possible PCBP1 site as determined from α-complex and PCBP2 consensus binding sequence at 128-150nt shown in red. Possible HnRNPK consensus sites in region of interest at 166-171nt and 191-196nt shown in blue. B) Consensus sites for proteins as reported in the literature.

There are two potential HnRNPK sites in the region 158-211nt (Figure 3.11). The HnRNPK consensus site is $UC_{3-4}(UA/AU)$ (Thisted et al., 2001), and there is a sequence ACCCUU at 166-171nt and GCCCAU at 191-196nt. It is interesting that all proteins appear to bind to the c-*myc* IRES RNA up to 211nt as it is this region of the IRES that has been found to be highly structured and to contain the ribosome landing site (Le Quesne et al., 2001).

3.7 The effect of protein factors on c-*myc* IRES activity *in vivo*

To study the proteins effect on IRES activity *in vivo* the dicistronic construct pRMF was used. pRMF contains the c-*myc* IRES between the *Renilla* luciferase and Firefly luciferase genes (Figure 3.12A). The *Renilla* luciferase expression shows cap-dependent translation, whereas the firefly luciferase is under the control of the c-*myc* IRES. Protein factors of interest were introduced into mammalian cells by co-transfection of pcDNA3 protein constructs with pRMF, and their effect on translation from pRMF measured. Translation levels of Firefly and *Renilla* luciferase were determined using a dual luciferase assay. All transfections were carried out with the addition of pJLacZ, and results shown relative to β-galactosidase activity to adjust for changes in transfection efficiency. The plasmid pRF was used as a control.

HeLa cells were co-transfected with pRMF and protein factor contructs (Figure 3.12B). As described c-*myc* IRES activity is high in HeLa cells compared to other cell lines. Addition of UNR, ITAF45, La, eIF4G (457-932) and DAP5 (p97) had no stimulatory effect on IRES activity. HnRNPK has a slight stimulatory effect of 1.2 fold when compared to the no protein control. The factors PCBP1 and PCBP2 were found to stimulate c-*myc* IRES activity at approximately 1.7 fold and 1.4 fold respectively. (For separate firefly and renilla data see appendix.)

Co-transfections with PCBP1 and PCBP2 were repeated and levels of protein construct transfected were titrated from 0.1-1μg (Figure 3.13). PCBP1 was found to have a maximal stimulation on IRES activity of 1.9 fold, and then decline with the transfection of higher amounts. PCBP2 was found to have a maximal stimulation of 1.4 fold. These are significant increases in c-*myc* IRES activity as the IRES is already highly active in HeLa cells. It is more active in this cell line than any others tested and enhances translation from the downstream cistron by approximately 70 fold (Stoneley et al., 2000b). Therefore to
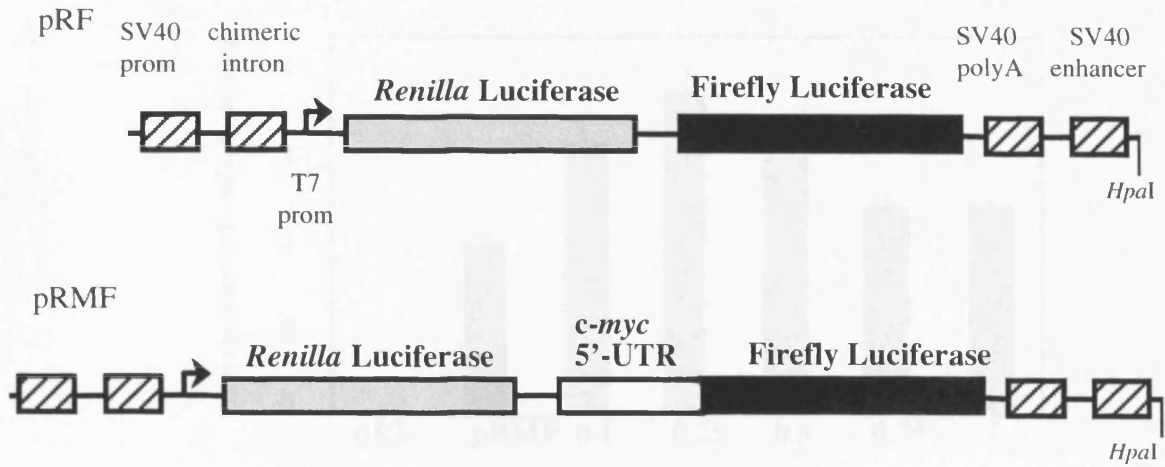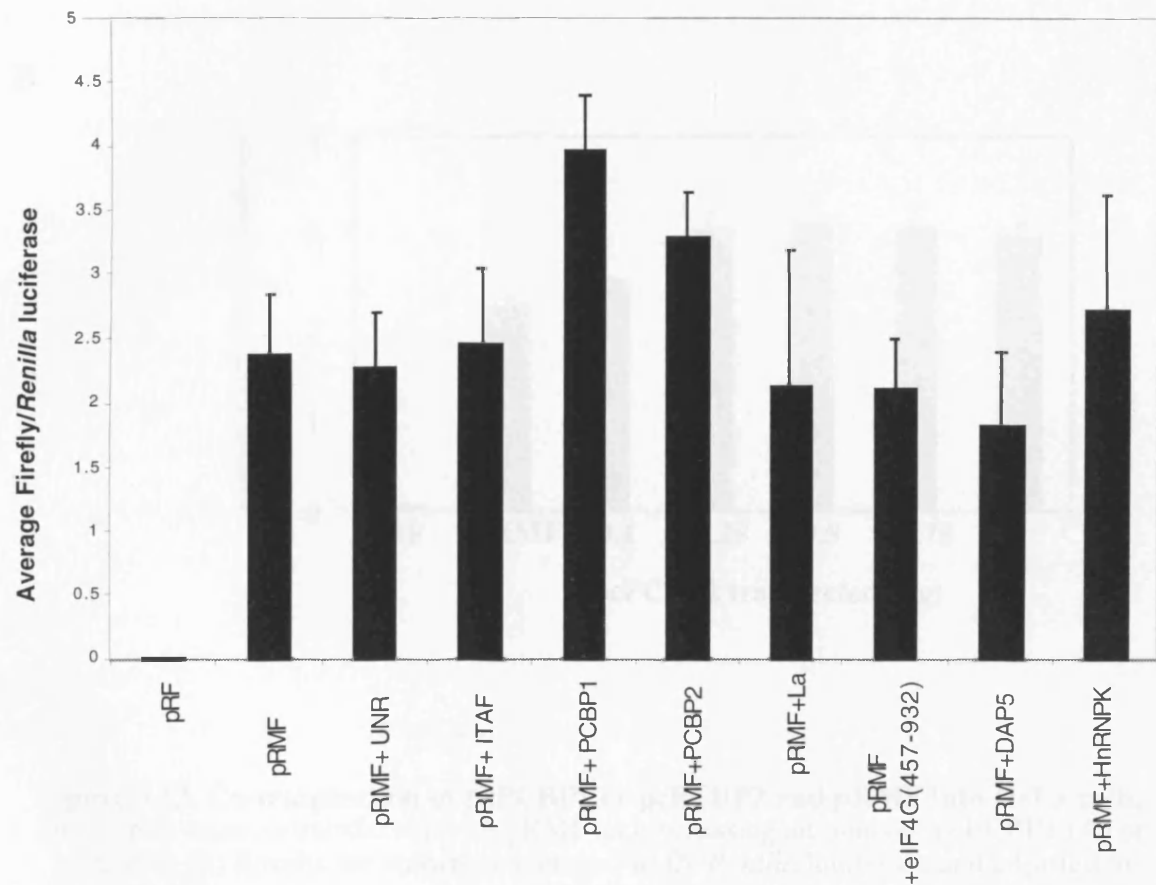
Figure 3.12. Co-transfection of protein factor constructs and pRMF into HeLa cells.
A) pRMF dicistronic construct containing the c-*myc* IRES between the *Renilla* and
Firefly luciferase genes. B) Co-transfection of pcDNA3 protein factor constructs and
pRMF into HeLa cells. Results are shown as average Firefly/*Renilla* luciferase and
adjusted to relative β-galactosidase activity. Error bars shown are for standard deviation.
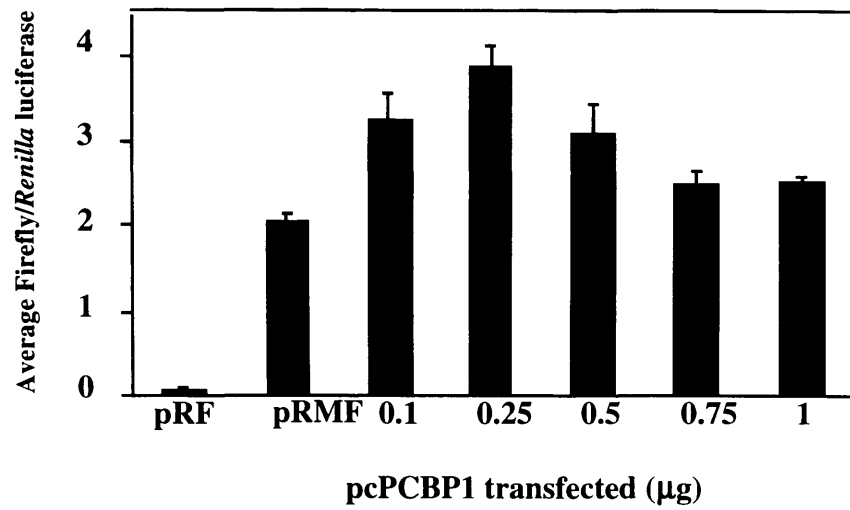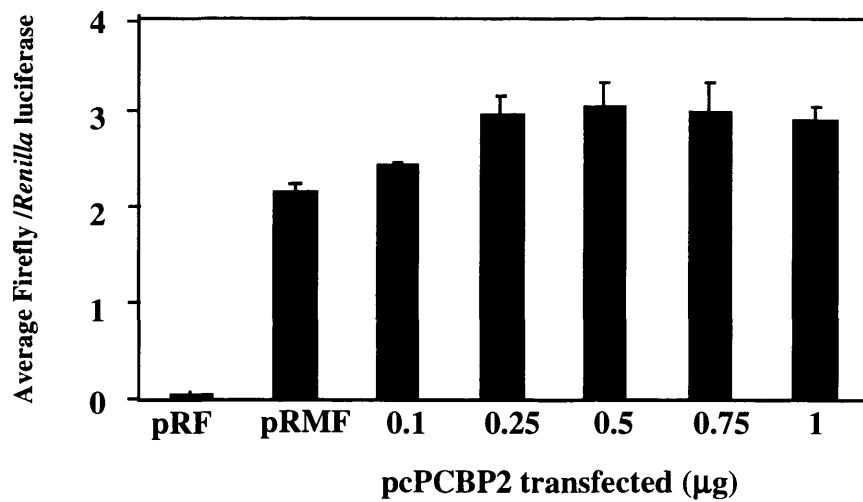
Figure 3.13. Co-transfection of pcPCBP1 or pcPCBP2 and pRMF into HeLa cells.
HeLa cells were co-transfected with pRMF and increasing amounts of pcPCBP1 (A) or pcPCBP2 (B) Results are shown as average Firefly/*Renilla* luciferase and adjusted to relative β-galactosidase activity. A representative set of data is shown. Error bars shown are for standard deviation.

observe an almost 2 fold increase in translation in this cell line demonstrates a very high level of IRES activity.

Expression levels from pcPCBP1 were verified by western blotting (Figure 3.14). HeLa cell extracts from pcPCBP1 transfected and non-transfected cells were blotted with a PCBP1 specific antibody. Results showed a detectable increase in pcPCBP1 protein in the transfected cell line. No antibodies were available for PCBP2 or HnRNPK.

Combinations of protein factors known to interact with the c-*myc* IRES were co-transfected (Figure 3.15A). These were combinations of UNR, PCBP1, and PCBP2. No additive effect of adding combinations of proteins was observed, and maximal stimulation did not exceed that of the individual proteins. The maximum activity observed when adding both proteins may be due to other factors required for efficient IRES translation, which are limiting in these cells.

Co-transfections were then performed in Raji cells with proteins found to have a stimulatory effect in HeLa cells. Raji cells are known to have a lower level of c-*myc* IRES activity and may therefore show different results. Again PCBP1 and PCBP2 were found to have a stimulatory effect (Figure 3.15B). However, this time the level of stimulation seen for PCBP2 was higher than that seen for PCBP1. PCBP2 showed an almost 3 fold stimulation, and PCBP1 again showed approximately 2 fold. HnRNPK was seen to stimulate to approximately 1.3 fold. Again combinations of these proteins had no effect above that seen for the individual proteins.

3.8 PCBP2 splice variants show similar effects on c-*myc* IRES activity

A number of splice variants have been reported for PCBP2, which vary in the region between KH domain 2 and KH domain 3 (Makeyev and Liebhaber, 2002). There are a large number of possible variations for the gene due to 6 alternatively spliced regions. At least 6 variants have been reported in humans (Makeyev et al., 1999).

The protein cloned for these studies was a splice variant missing a region of 36nt. This is not the most abundant variant reported and may show different properties. An additional 4 splice variants of PCBP2 were cloned (Figure 3.16A). These included the full-length protein. The most abundant variant reported (αCP-2KL) which has only the 93nt region missing was not identified, however other variants missing this region in addition to other variations were cloned.

pcPCBP1 transfection   **-**   **+**

PCBP1 ➤

Actin ➤
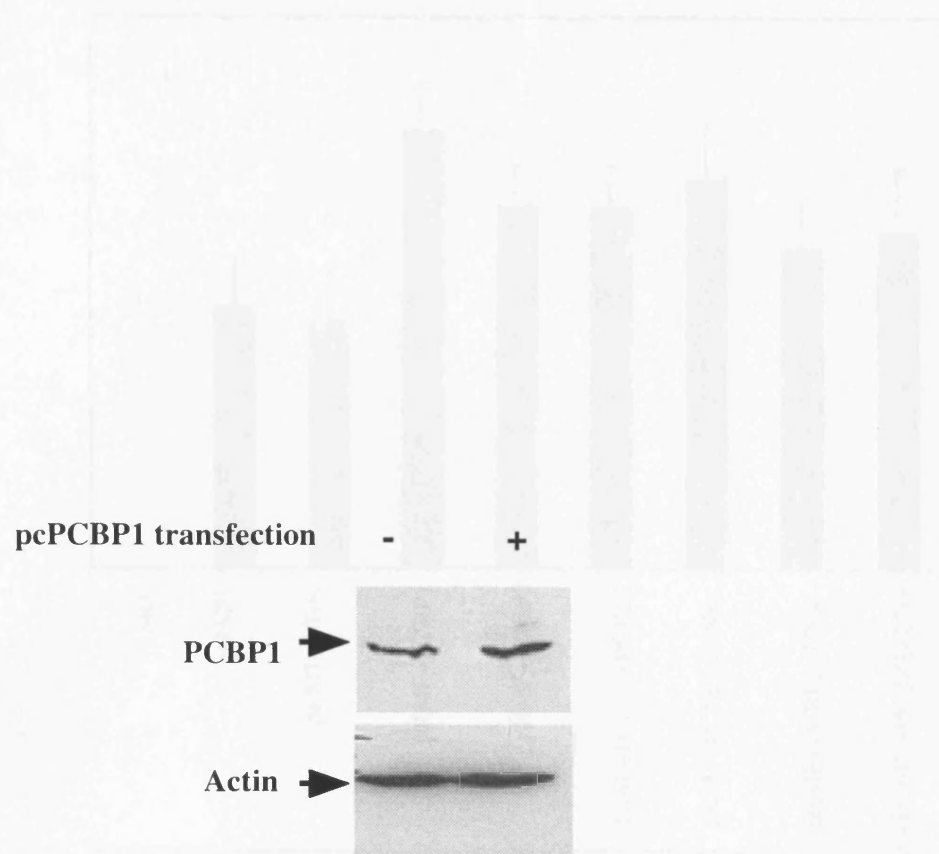
Figure 3.14. PCBP1 levels in HeLa cells. Western blotting was carried out from HeLa cell extracts. Levels of PCBP1 were compared between pcPCBP1 transfected and non-transfected HeLa cells.
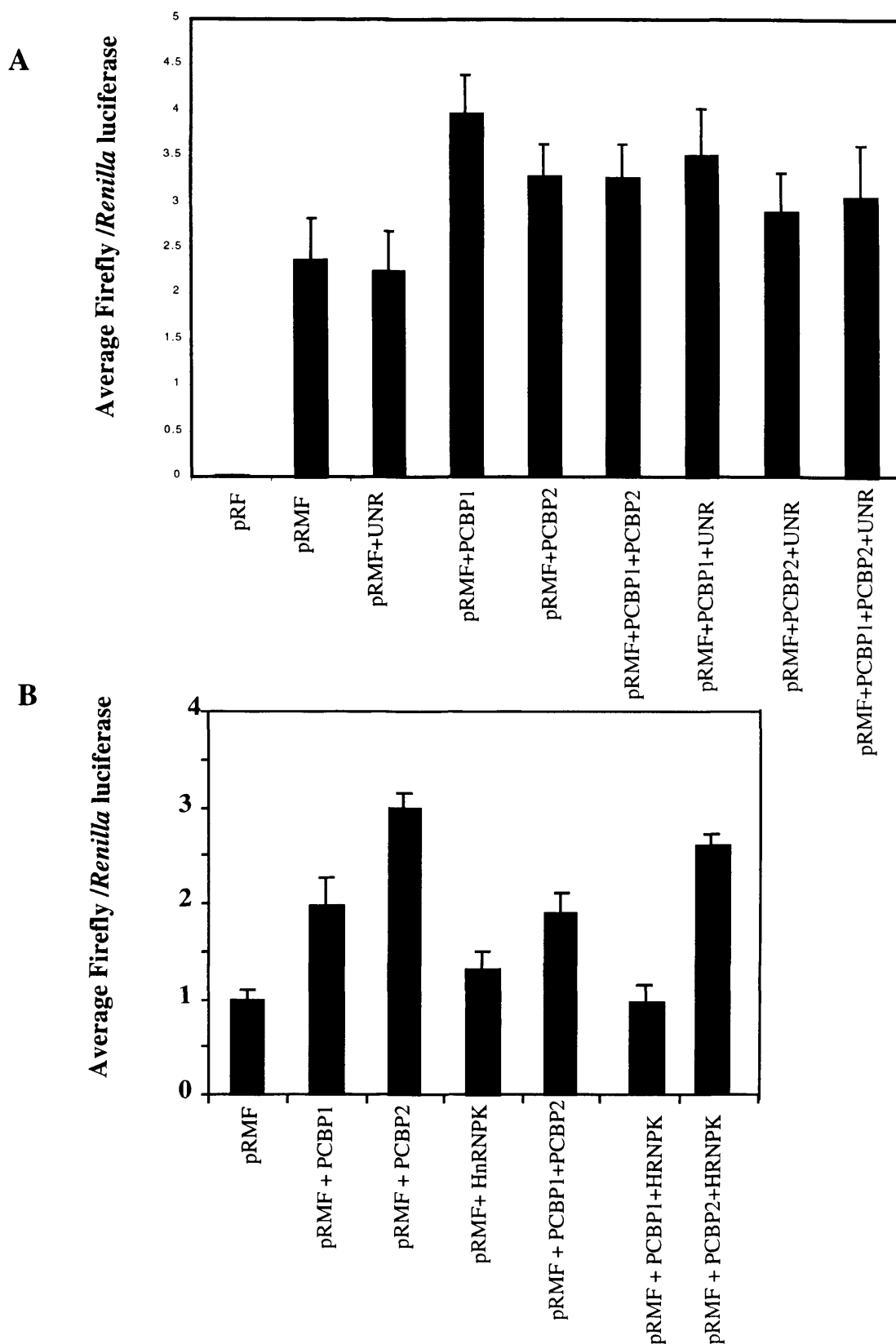
Figure 3.15. Co-transfection of pRMF and combinations of protein factors into HeLa and Raji cells. A) Combinations of PCBP1, PCBP2 and UNR were co-transfected into HeLa cells with pRMF. B) Co-transfection of pRMF and protein factors into Raji cells. All results are shown as average Firefly/*Renilla* luciferase values and are adjusted to relative β-galactosidase activity. Error bars shown are for standard deviation.
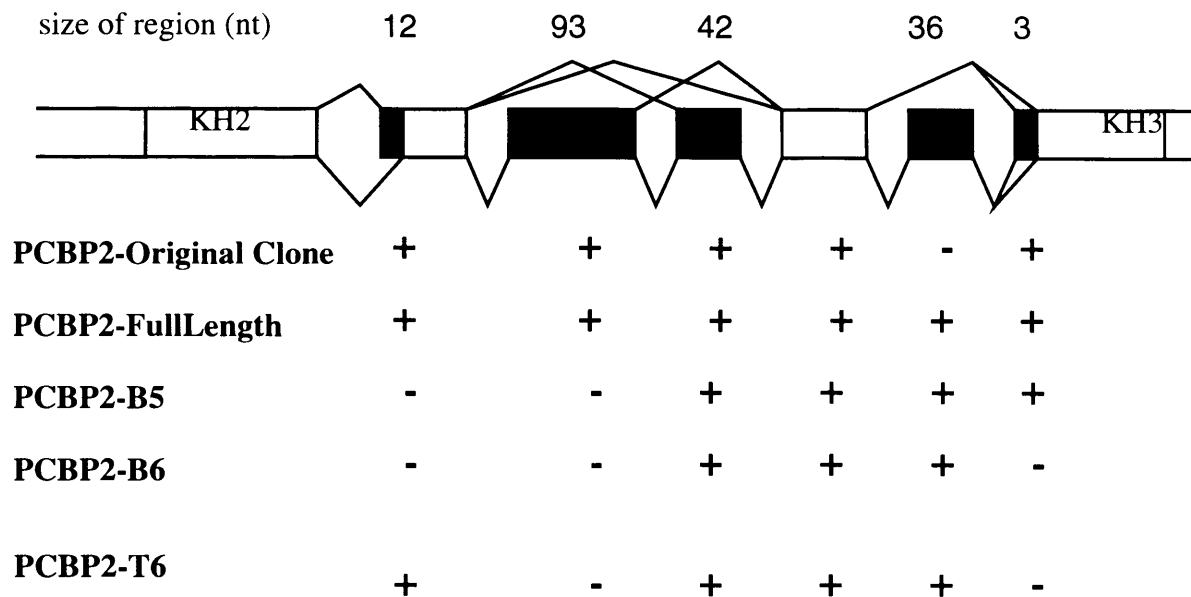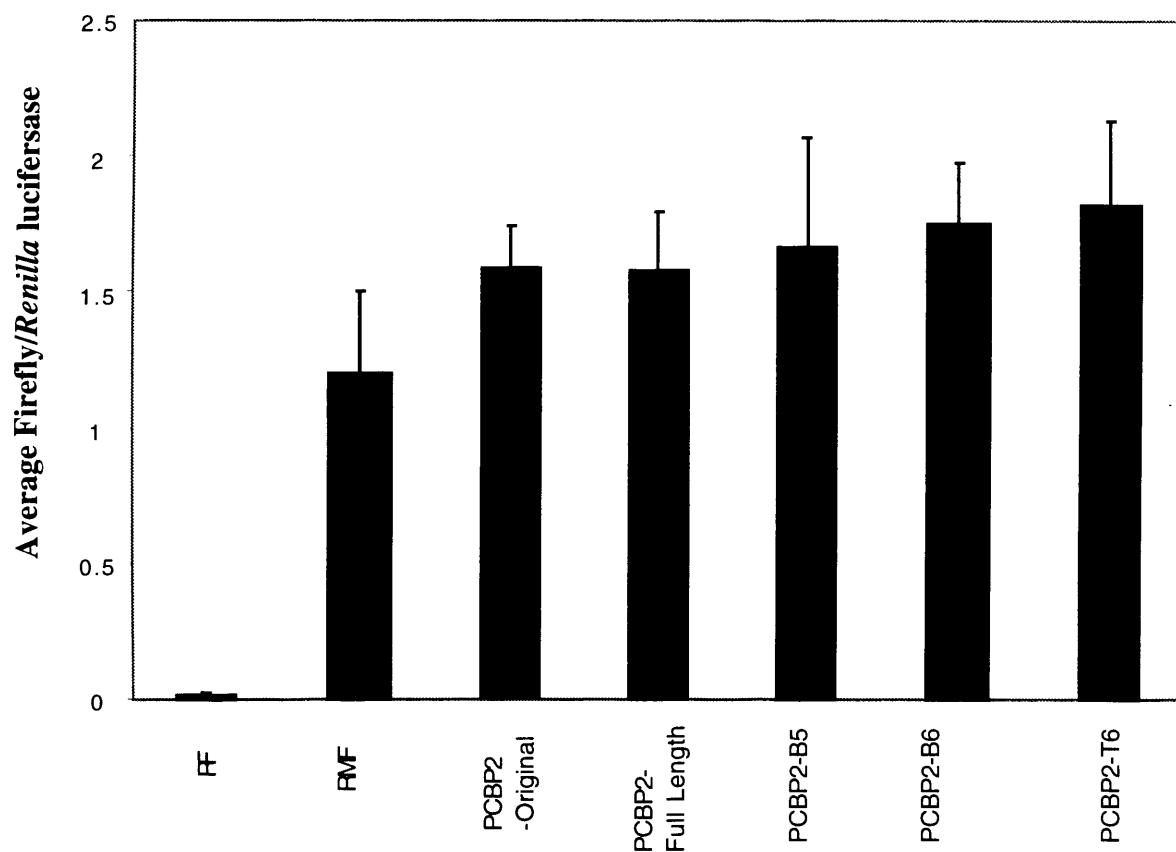
Figure 3.16. Co-transfection of PCBP2 splice variants and pRMF into HeLa cells.
A) Splice variants cloned for PCBP2. Variants differ in the region between KH domains
2 and 3 which contains 6 exons. The black exon regions represent alternatively spliced
sequences of the sizes shown. B) Co-transfection of PCBP2 splice variants into HeLa
cells with pRMF. Results are shown as an average of Firefly/*Renilla* Luciferase and
adjusted to relative β-galactosidase activity. Error bars shown are for standard deviation.

Co-transfections were performed to identify any differences in activation of c-*myc* IRES activity by the PCBP2 splice variants. All PCBP2 variants were cloned into pcDNA3 and co-transfected with pRMF as previously described. No significant variation was seen in the activation of the c-*myc* IRES *in vivo* (Figure 3.16B). All splice variants showed a similar 1.4 fold activation in HeLa cells.

3.9 The effect of protein factors on c-*myc* IRES activity *in vitro*

Cellular IRESs function very inefficiently in *in vitro* systems. This is true for the c-*myc* IRES which is known to have a low efficiency in the rabbit reticulocyte lysate system when compared to picornaviral IRESs (Stoneley et al., 2000b). This is thought to be due to limiting of protein factors required for efficient translation.

In vitro translations were carried out using the rabbit reticulocyte lysate system and the effect of the addition of protein factors analysed. RNA transcripts were produced by *in vitro* run-off transcription from the dicistronic plasmid pRMF linearised with *Hpa*I or pRF linearised with *Bam*HI (Figure 3.17A). Transcripts were added to rabbit reticulocyte lysate with protein factors of interest and incubated at 30°C for 1.5 hours. Approximately 150ng UNR and 100ng of all other proteins were added. Reactions were then assayed for luciferase activity.

Little effect on IRES activity was observed for the protein factors when added individually (Figure 3.17B). There were slight effects seen for UNR, and UNR with its interacting protein UNRIP. Significant activations were seen with combinations of proteins including PCBP1, PCBP2, UNR and UNRIP. This increased to a 3 fold activation when PCBP1, PCBP2, UNR and UNRIP were added together. There appears to be a cumulative effect of adding combinations of PCBP1, PCBP2, UNR and UNRIP. The additive effect suggests that the c-*myc* IRES needs numerous factors that may be required to produce the correct IRES structure for maximal activity. These factors must all be limiting in rabbit reticulocyte lysate for this effect to occur.

3.10 Protein expression levels in cell lines do not correlate to IRES activity

Cell lines with known c-*myc* IRES activity were blotted for proteins of interest. Cell extracts were blotted for PCBP1, UNR and UNRIP. No antibodies were available for PCBP2 or HnRNPK. PCBP1 levels did not vary greatly between cell lines (Figure 3.18A) and did not correlate to activities found in the different cell lines (Stoneley et al., 2000b).
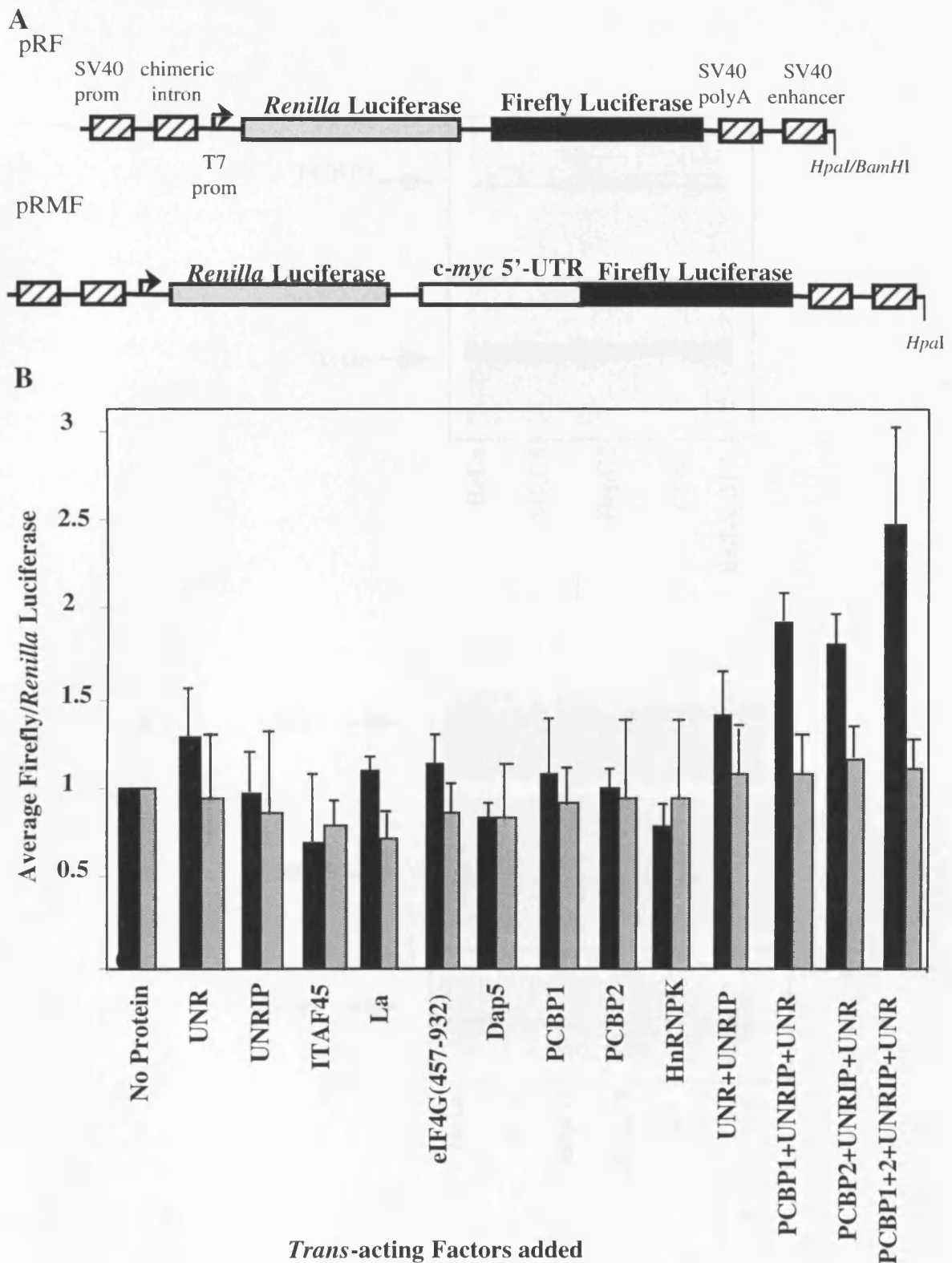
Figure 3.17. Addition of recombinant protein factors to *in vitro* translation reactions.
A) Transcripts were produced by *in vitro* run-off transcription from pRF or pRMF
linearised plasmids. B) 100-150ng recombinant protein factors were added to *in vitro*
translation reactions with RNA transcripts. Results are shown as normalised
Firefly/*Renilla* luciferase readings. ■ pRMF transcript □ pRF transcript.
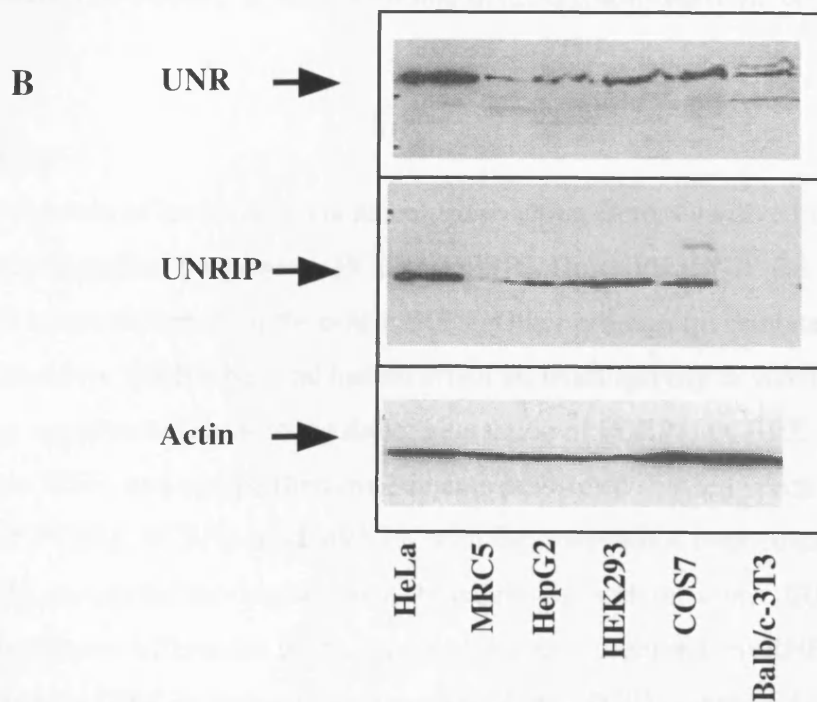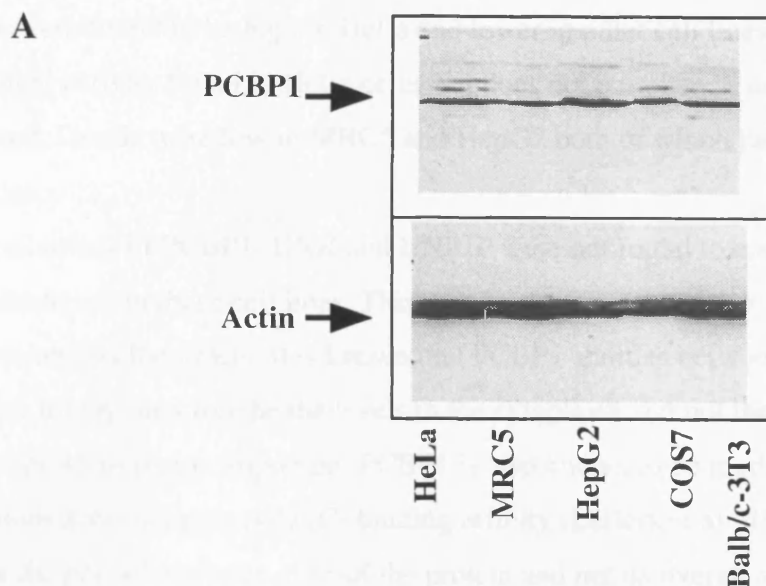
Figure 3.18. Western blotting for PCBP1, UNR and UNRIP in a variety of cell lines. Cell extracts from HeLa, MRC5, HepG2, HEK293, Cos7 and Balb/c-3T3 cells were used for western blotting to determine the levels of PCBP1(A), and UNR and UNRIP (B) present.

c-*myc* IRES activity was reported to be high in HeLa cells, followed by decreasing activity in MRC5, HepG2, GM637, HEK293, and Cos-7 and very low activity in MCF7 and Balb/c-3T3. UNR and UNRIP levels were found to vary more between cell lines (Figure 3.18B). Both were found to be high in HeLa and lower in other cell lines tested. This agrees with high activity found in HeLa cells, but does not correlate to activities in other cell lines tested. Levels were low in MRC5 and HepG2 both of which have high c-*myc* IRES activities.

Overall levels of PCBP1, UNR and UNRIP were not found to correlate to c-*myc* IRES activates found in these cell lines. This may be due to the fact that the cell extracts prepared were total cell extracts. It is known that PCBP1 shuttles between the nucleus and cytoplasm and it may therefore be the levels in the cytoplasm and not the total cellular level of these proteins that is important. PCBP1 is also known to be modified in the cell by phosphorylation decreasing its poly (C) binding affinity (Leffers et al., 1995), and this suggests that the phosphorylation state of the protein and not its overall level may be important. The PCBP family are known to have a wide variety of roles in the cell and therefore it would be unlikely if the level found in the cell correlated directly to IRES activity.

3.11 Discussion

Investigation of known canonical and *trans*-acting factors involved in IRES translation has identified the proteins PCBP1, PCBP2, HnRNPK, UNR and UNRIP as having a role in translation from the c-*myc* IRES. Other proteins investigated showed no binding to the c-*myc* IRES RNA and had no effect on IRES activity *in vivo* or *in vitro*. UV cross-linking experiments showed the direct interaction of PCBP1, PCBP2, HnRNPK and UNR with the IRES, and competition experiments performed showed specificity in the interaction of PCBP1, PCBP2 and HnRNPK with the c-*myc* RNA over control GAPDH RNA. EMSAs confirmed the interaction of these proteins with the c-*myc* IRES RNA. However, there were differences in the band shift patterns observed. HnRNPK appeared to be able to bind the IRES as both a monomer and a dimer. PCBP1 appeared only to bind the RNA as a dimer, and PCBP2 only as a monomer. The presence of dimers binding to the IRES was not unexpected, as it is known that all three proteins are able to form both homodimers and heterodimers (Gamarnik and Andino, 1997; Makeyev and Liebhaber, 2002).

EMSAs of fragments of the c-*myc* IRES also showed differences in where the proteins were binding to the RNA. PCBP2 appeared to bind in the first 98nt, and PCBP1 and HnPNPK in the first 211nt. PCBP1 and PCBP2 share 90% protein similarity and are known to have similar binding properties. Therefore it was interesting to note differences in their binding to the RNA suggesting both proteins may be involved in c-*myc* IRES translation. A region fitting the general α-complex and PCBP2 consensus binding site was found in the c-*myc* IRES between 128-150nt. This would agree with the EMSA data for PCBP1 and was therefore noted as a possible PCBP1 binding site. Possible HnRNPK sites were also found in this region. Interestingly the region of the IRES up to 211nt that was found to bind all three proteins is highly structured and has been found to contain the ribosome landing site (Le Quesne et al., 2001). The PCBP1 site in particular covers the IRES region where a pseudoknot structure has been identified. It is possible that proteins binding to this region could disrupt the structure and open up the IRES allowing the ribosome direct access to the ribosome landing site. In agreement to this model mutations of this c-*myc* pseudoknot helix α region have been shown to enhance translation from the IRES showing it is inhibitory to IRES function (Le Quesne et al., 2001). A similar model has recently been reported for the role of UNR and PTB in Apaf-1 IRES translation, which are reported to maintain the IRES structure in the correct conformation for ribosome recruitment [S.Mitchell- Personal communication].

PCBP1, PCBP2 and HnRNPK were found to stimulate c-*myc* IRES activity *in vivo*. PCBP1, PCBP2 and HnRNPK were all found to enhance IRES activity in both HeLa and Raji cells. The stimulatory effect of PCBP1 and PCBP2 in HeLa cells was significant due to the initial high c-*myc* IRES activity in this cell line. UNR and UNRIP were however not found to have an effect *in vivo*. This may be due to levels already present in the cells, as both UNR and UNRIP were found to have very high levels in HeLa cells. PCBP1 and PCBP2 were also found to have an effect on IRES activity *in vitro*. This was in combination with UNR and UNRIP, and stimulation of up to 3 fold was observed. This again is significant as cellular IRESs function very inefficiently *in vitro* and this is the first time any enhancement of c-*myc* IRES activity has been observed in this system.

The different *in vivo* enhancements of c-*myc* IRES activity by PCBP1 and PCBP2 indicates that the proteins cannot functionally replace each other. This is also confirmed by the different binding patterns of the proteins with the c-*myc* IRES RNA. Therefore PCBP1 and PCBP2 are likely to have distinct roles in c-*myc* IRES translation. The enhancements

seen *in vitro* for UNR and its binding partner UNRIP also show a direct role in c-*myc* IRES translation. The direct role of HnRNPK has yet to be determined as only slight enhancements of IRES activity are observed *in vivo*.

Chapter 4

Identifying novel protein factors that interact with the c-*myc* IRES using the yeast three-hybrid system.

4.1 Introduction

To investigate the protein factor requirements of the c-*myc* IRES initial experiments concentrated on known canonical and *trans*-acting proteins. This identified the interaction of PCBP1, PCBP2, HnRNPK, UNR and UNRIP with the c-*myc* IRES. However, further proteins may also be involved in c-*myc* IRES translation. Using the rabbit reticulocyte lysate system it was shown that the c-*myc* IRES was inefficient *in vitro* when compared to the high activity observed *in vivo*. This suggested that additional *trans*-acting factors could still be required for efficient translation in addition to those described. Therefore a more direct approach using the yeast three-hybrid system was undertaken to identify novel factors interacting with the c-*myc* IRES. This system is an adaptation of the yeast two-hybrid system, which examines RNA-protein interactions, and has been used to identify a large number of interactions (Bernstein et al., 2002).

4.2 Cloning of yeast three-hybrid components

The yeast three-hybrid system is a variation of the yeast two-hybrid system that allows the interaction of RNA and protein to be studied (Figure 4.1). Three components are required: a DNA binding domain hybrid, an RNA hybrid, and a transcriptional activation domain hybrid. The DNA binding domain hybrid is a LexA-MS2 fusion and is present in the genome of the yeast strain L40-coat. The remaining two components need to be introduced into the yeast by transformation. The RNA of interest is expressed as a hybrid with the MS2 RNA binding sites, and is brought to the promoter through the interaction of these sites with the MS2 coat protein. When the RNA of interest comes together with its binding protein, the Gal4 activation domain is brought close to the promoter and activates transcription of His3 and LacZ reporters. The expression of these reporters is detected by growth of yeast on Histidine deficient media or by use of a β–Galactosidase assay.

Fragments of the c-*myc* IRES RNA were cloned into the plasmid pIIIA/MS2-2' (Figure 4.2). For the yeast three-hybrid library screen a region of the c-*myc* IRES of 251nt was cloned into this construct. This was chosen due to the constraints of the three-hybrid
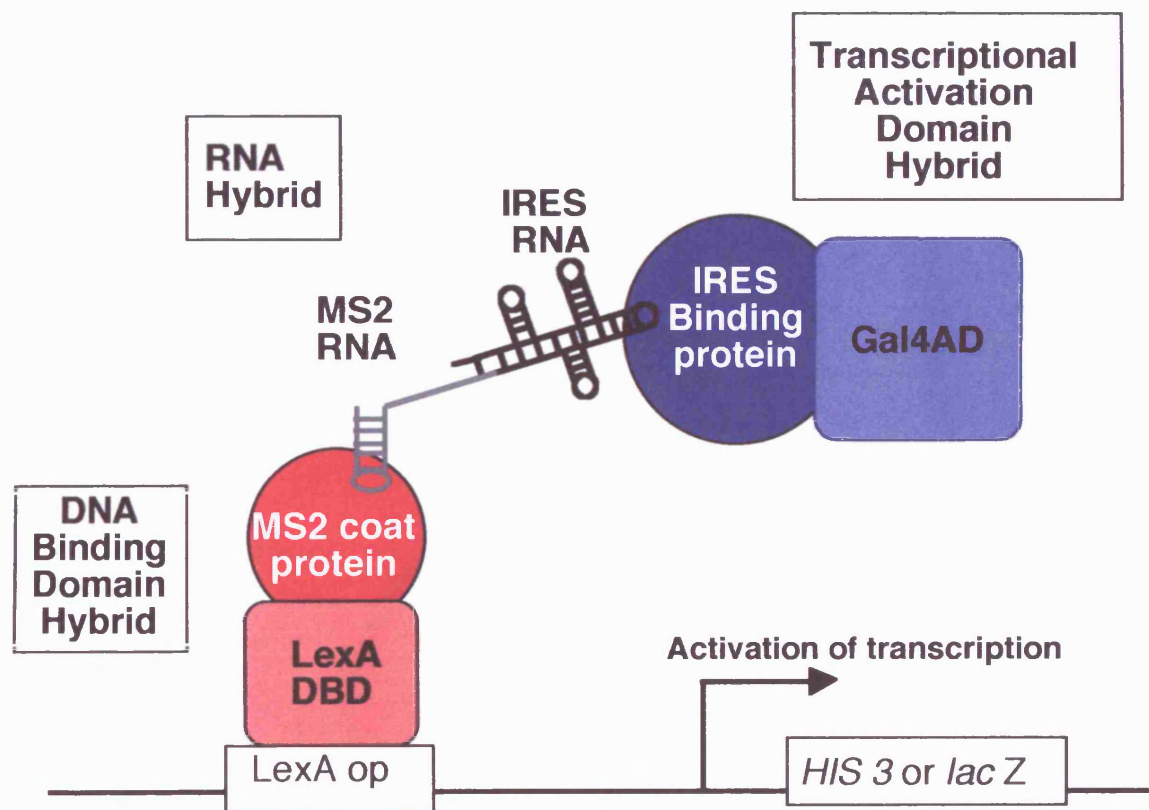
Figure 4.1. The yeast three-hybrid system. Three components are required: the DNA binding domain hybrid, the RNA hybrid and the transcriptional activation domain hybrid. The DNA binding domain hybrid is a LexA-MS2 fusion present in the genome of the yeast strain L40-coat. The other two components are introduced into the system by transformation.
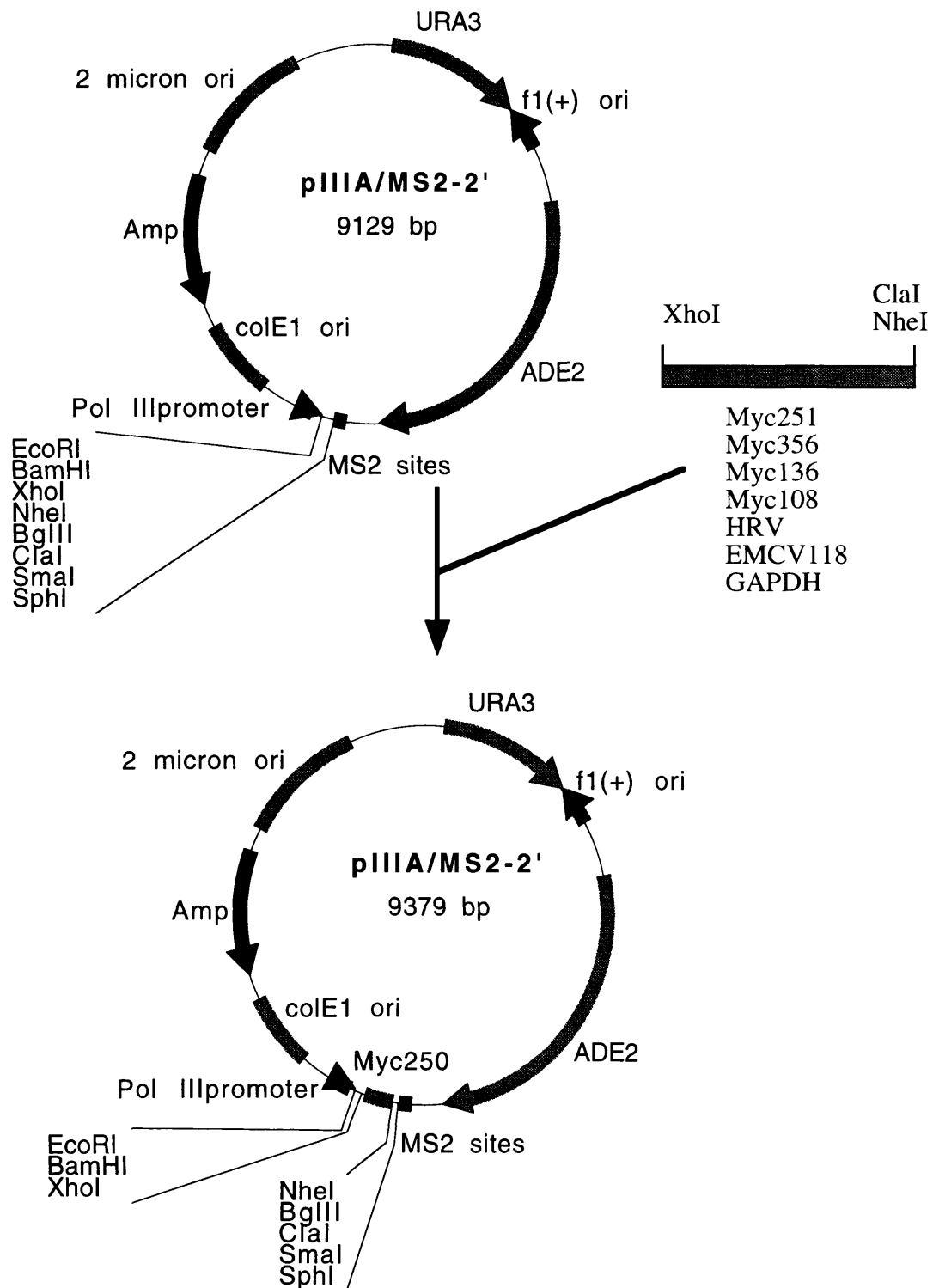
Figure 4.2. Cloning of pIIIA/MS2-2' constructs. Products were PCR amplified with gene specific primers containing the required restriction sites and cloned into the polylinker site of pIIIA/MS2-2' upstream of the tandem MS2 sites

system used. The RNA hybrid is expressed from the yeast RNase P RNA (RPR1) promoter, an RNA polymerase III promoter. This limits the RNA sequences that can be analysed as RNA polymerase III can terminate at runs of four or more U's. It has also been found that a minimal binding site is preferred as sequences over 150-200nt were found to produce lower reporter signals. Therefore domain 1 of the c-*myc* IRES (251nt) was chosen, as this removes a U repeat region at 357-363nt, and represents a known region of structure of the IRES (Le Quesne et al., 2001) (see Figure 1.9). Any region shorter than this may not fold into the correct structure for protein binding and therefore this was considered to be the minimum region that could be used for a library screen. This was cloned into the plasmid pIIIA/MS2-2' to produce an RNA hybrid in which the IRES RNA was upstream of the MS2 binding sites (Figure 4.3).

Other domains of the c-*myc* IRES were also cloned, including domain 1+2 (356nt), domain 2 (136nt) and a section of domain 1(108nt). Several controls were also cloned into the RNA plasmid. These were the HRV IRES known to bind UNR, a fragment of 118nt of the EMCV IRES known to bind PTB (stem loop H) (Witherell et al., 1993), and a negative control of a 151nt region of GAPDH DNA.

c-*myc* fragments of interest were amplified by PCR from pSKML template with gene specific primers containing the required restriction sites. Myc108 was amplified by PCR from the plasmid pMint1L. PCR reactions were carried out under standard conditions and PCR products purified. Restriction digests were then performed before cloning into pIIIA/MS2-2'. Primers used are shown in table 4.1. HRV, and EMCV templates were amplified from pJHRV and pRemcvF respectively, and GAPDH from pSKGAP:E/H. See table 2.2 for primers sequences. All constructs were fully sequenced before use.

To introduce the protein prey component, proteins of interest were cloned into the plasmid pACTII in frame with the Gal4 activation domain (see section 3.3). The *trans*-acting binding proteins UNR, PTB, La, PCBP1, PCBP2, and ITAF45, and the canonical factors eIF4G (457-932), and DAP5 (p97) and (p86) were cloned. The control plasmids expressing the Iron Response Element (IRE), and its interacting protein (IRP) were already available. These produce a known interaction that can be used as a control for the system.

Fragment cloned	Primers used
Myc 251	XHOMF1 + CLAMIR2
Myc 356	XHOMF1 + NHEMR3

GUUUUACGUUUGAGGCCUCGUGGCGCACAUGGUACGCUGU
GGUGCUCGCGGCUGGGAACGAAACUCUGGGAGCUGCGAU
UGGCAGAAUUCCGGCUAGAACUAGUGGAUCCCCCCUCGAGua
auuccagcgagaggcagagggagcgagcgggcggccggcuagggguggaagagccgggcgagc
agagcugcgcugcgggcguccugggaagggagauccggagcgaauaggggggcuucgccucug
gcccagcccucccgcugauccccccagccagcgguccgcaacccuugccgcauccacgaaacuuu
gcccauagcagcgggcgggcacuuugcacuggaacuuacaacacccgagcaaggacgcgacA
UCGAUCCCGGGCAGCUUGCAUGCCUGCAGGUCGACUCUA<u>GAA</u>
<u>AACAUGAGGAUCACCCAUGUCUGCAGGUCGACUCUAGAAAA</u>
<u>CAUGAGGAUCACCCAUGUC</u>UGCAGGUCGACUCUAGAGGAUC
GGAAUUCCCCCAUAUCCAACUUCCAAUUUAAUCUUUCUUUU
U

Figure 4.3. Sequence of the Myc-251 RNA hybrid. The c-*myc* sequence is shown in lowercase, RNase P 5' leader and 3' trailer sequences in bold, and MS2 binding sites are underlined.

Myc 136	XHOMIF3 + NHEMR3
Myc 108	MINTFOR + MINTREV
HRV	HRVFOR + HRVREV
EMCV 118	EMCVFX + EMCVRN
GAPDH	GAPDHFOR + GAPDHREV

Table 4.1. PCR primers used to amplify RNA fragments of interest.

4.3 Confirmation of expression of yeast three-hybrid components

Controls were carried out to ensure that the components of the yeast three-hybrid system were being expressed in the yeast strain L40-coat. Northern blot analysis was performed to confirm expression of the Myc-251 RNA hybrid. The plasmid pIIIA/MS2-2/Myc251 was transformed into yeast by small-scale transformation, and yeast plated onto –Uracil media. Colonies were grown at 30°C for 3-5 days, before a 20ml overnight culture was grown, and RNA extracted. RNA was fractionated by denaturing agarose gel electrophoresis and transferred to nitrocellulose membrane. To prepare the radiolabelled probe pIIIA/MS2-2/Myc251 was digested with *Xho*I and *Sma*I to produce the Myc251 fragment. This was then purified and labelled by random priming in the presence of ^{32}P-labelled CTP. The probe was hybridised to the membrane overnight at 65°C. Northern blotting confirmed expression of a c-*myc* fragment of the size expected (Figure 4.4A). The fragment produced is of the expected 537nt to contain the c-*myc* fragment, MS2 sites and RNase P 5' leader and 3' trailer sequences.

Western Blot analysis was used to confirm expression of the transcriptional activation domain hybrid (Figure 4.4B). Yeast were transformed with pACTII/UNR, and grown on –Leucine Media. A 20ml overnight culture was then grown and protein extracted. Protein extracts were fractionated by SDS-PAGE, and transferred to nitrocellulose membrane. UNR was detected using a specific antibody. Western analysis confirmed expression of a protein of the correct size for the UNR transcriptional activation domain hybrid.

4.4 Interaction of known proteins with the c-*myc* IRES

Work on picornaviral translation has identified a large number of proteins that are involved in IRES initiation. The coding regions of a number of proteins of interest were

A

Size kb

1.0 —

0.5 — ◄── Myc 251 RNA Hybrid

0.25 —

1   2

B

Size kDa

150 —

100 — ◄── UNR AD Hybrid

75 —

1   2
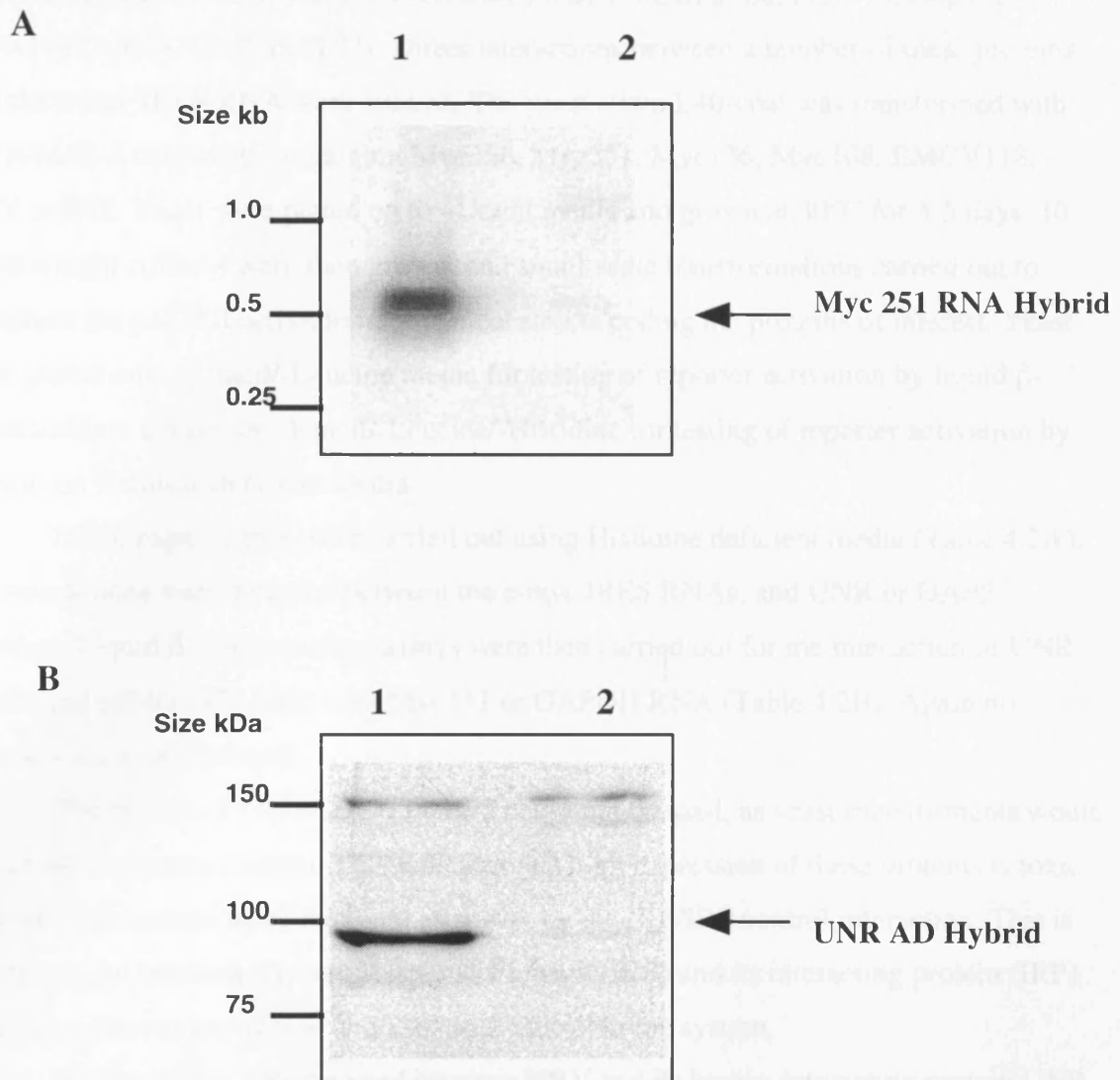
Figure 4.4. Yeast three-hybrid controls. A) Northern blot analysis of pIIIA/MS2-2/MYC251 transformed yeast (1) and un-transformed control (2). A band of the expected 537nt for the Myc-251 RNA hybrid was highly expressed. B) Western blot analysis of pACTII/UNR transformed yeast (1) and un-transformed control (2). A band of the expected size for the UNR activation domain hybrid was expressed.

cloned into the transcriptional activation domain hybrid in frame with the Gal4 activation domain (see section 3.3). Those cloned were PCBP1, PCBP2, La, ITAF45, Dap5 (p97), eIF4G (457-932), UNR and PTB. Direct interactions between a number of these proteins and the c-*myc* IRES RNA were studied. The yeast strain L40-coat was transformed with pIIIA/MS2-2 constructs containing Myc356, Myc251, Myc136, Myc108, EMCV118, HRV or IRE. Yeast were plated on to –Uracil media and grown at 30°C for 3-5 days. 10 ml overnight cultures were then grown, and small-scale transformations carried out to introduce the pACTII activation domain constructs coding the proteins of interest. Yeast were plated onto –Uracil/-Leucine media for testing of reporter activation by liquid β-Galactosidase assays, or –Uracil/-Leucine/-Histidine for testing of reporter activation by growth on histidine deficient media.

Initial experiments were carried out using Histidine deficient media (Table 4.2A). No interactions were detected between the c-*myc* IRES RNAs, and UNR or DAP5 proteins. Liquid β-Galactosidase assays were then carried out for the interaction of UNR, DAP5 and eIF4G (457-932) with Myc251 or GAPDH RNA (Table 4.2B). Again no interactions were detected.

The proteins PTB, PCBP1, PCBP2 could not be used, as yeast transformants would not grow on selective media. This indicates that high expression of these proteins is toxic to yeast. The system was functional as shown by the IRE/IRP control interaction. This is an interaction between the Iron Response Element (IRE) and its interacting protein (IRP) which is a known interaction and used as a control in the system.

No interaction was detected between HRV and its known interacting protein UNR. However, the HRV RNA hybrid used was 546nt and this may explain why no reporter signal was detected. This is longer than recommended for the system, as it is noted RNAs over 150-200nt produce lower signals. The other control of EMCV118 and PTB could not be studied due to being unable to grow PTB transformants. Overall no interaction between any known IRES interacting protein and the c-*myc* IRES RNA, or any control IRES RNA was detected.

4.5. Yeast three-hybrid library screen

A yeast three-hybrid library screen was carried out to identify novel c-*myc* IRES RNA binding proteins. The Myc251 RNA hybrid was chosen for this screen. This construct contains the full domain 1 sequence of the IRES, which may be required for

A

Protein→ RNA↓	UNR	DAP5	IRP
Myc356	−	−	−
Myc251	−	−	−
Myc136	−	−	−
Myc108	−	−	−
HRV	−	−	−
EMCV118	−	−	−
GAPDH	−	−	−
IRE	−	−	**+**

B

Protein→ RNA↓	UNR	DAP5	eIF4G (457-932)	IRP
Myc251	−	−	−	−
GAPDH	−	−	−	−
IRE	−	−	−	**+**

Table 4.2. The study of direct protein:RNA interactions using the yeast three-hybrid system. A) Interactions of UNR and DAP5 with the c-*myc*, HRV, and EMCV IRES RNAs. Reporter activation identified by growth of yeast on histidine deficient media. The IRE/IRP interaction was used as a control. **B)** Interactions of UNR, DAP5, and eIF4G (457-932) with the c-*myc* IRES RNA. Reporter activation identified using liquid β-Galactosidase assays. The IRE/IRP interaction was used as a control.

folding of the correct structure for protein binding. Yeast was transformed with pIIIA/MS2-2/Myc251 by small-scale transformation, and plated onto –Uracil media. Yeast colonies were grown at 30°C for 3-5 days, and then used to grow an overnight culture for large-scale transformation of the cDNA library. A human testis cDNA library was used in the plasmid pACTII. The library was amplified and plasmid DNA prepared. 100μg library DNA was transformed into the yeast by large-scale transformation and yeast were plated onto thirty 10cm –Uracil/-Leucine/-Histidine plates. Plates were supplemented with 2mM 3-aminotriazole (3-AT) to prevent detection of background His3 reporter expression, and left at 30°C for 8 days. Further –Leucine/-Uracil plates were plated with dilutions of the transformation in order to analyse the transformation efficiency obtained. The transformation efficiency was calculated at 1.3×10^4 cfu/μg, and 1.3×10^6 independent clones screened.

From this initial screen 2428 His+ colonies were obtained. A further screen was then carried out to eliminate RNA independent positives. The initial colonies detected were picked onto –Leucine/-Histidine media with 2mM 3-AT. This causes the interactions that do not require the RNA hybrid for His3 activation to lose the RNA plasmid. Those losing the RNA plasmid go pink/red due to the loss of the Ade2 gene encoded by the plasmid pIIIA/MS2-2. The yeast strain L40-coat is an Ade2 mutant, and when levels in the cell become low the yeast attempt to synthesize it accumulating a red purine metabolite and turning the yeast pink/red in colour. Therefore RNA independent positives can be eliminated as they show pink/red. This screen identified 1261 colonies as pink/red and these were eliminated leaving 1139 positives.

The remaining 1139 positives that were dependent on the RNA for His3 reporter activation were then assayed for LacZ activity. This enabled the interactions to be tested with a separate reporter system. A β-galactosidase overlay assay was carried out, and plates left at 30°C for several hours for the blue colour to develop. Those interactions causing expression of the LacZ gene turned the yeast blue due the breakdown of X-Gal. All but 15 colonies turned blue after several hours. This showed that nearly all interactions that were found to activate the His3 reporter were also found to activate the LacZ reporter. This is not surprising as previous reports have shown that this to occur (Bernstein et al., 2002).

A further screen was then used to identify any remaining RNA independent positives. The 1124 interactions were plated onto -Leucine + 5-fluoroorotic acid (5-FOA)

plates. These plates cure the yeast of the RNA plasmid. Yeast expressing the URA3 gene product convert 5-FOA into 5-fluororuracil, which is toxic to the cells. Therefore yeast lose the RNA plasmid to enable them to grow on this media. This could be detected by the slow growth of pink/red colonies. Interactions were then tested again and those still able to activate the LacZ reporter were eliminated as RNA independent positives. Colonies were re-assayed for LacZ reporter activity by β-galactosidase overlay assay, and only 24 colonies were found to turn blue. Therefore 1100 RNA dependent positives remained.

A high number of positives were initially obtained from the screen (Results summarised in Table 4.3). The 5-FOA selection carried out suggested that all RNA independent positives were eliminated. All remaining positives were found not to activate β-Gal after 5-FOA selection, showing that they required the RNA hybrid for reporter activation. The high number of remaining positives could be due to the low stringency of the conditions used. The screen was carried out in the presence of only 2mM 3-AT to eliminate background His3 expression. However this allowed all interactions, even very weak ones, to be detected. This stringency was chosen because the strength of interaction of proteins with IRES elements is not known, and therefore the interaction of any protein of interest may be weak. Also several proteins may be required for strong binding, and binding of individual proteins may be weak. However, the large number of positives obtained meant that all the interactions could not be studied. Further screens to identify only the strongest interactions were carried out.

Protocol	Result
Transformation Efficiency	1.3×10^4 cfu/μg
Independent Clones Screened	1.3×10^6
Initial His+ Colonies	2428
His+ after colony colour screen	1139
His+ after 5-FOA/βGal to eliminate RNA independent positives	1100

Table 4.3. Summary of results from yeast three-hybrid library screen.

4.6. Identification of strong c-*myc* IRES binding proteins

Yeast plates containing varying amounts of 3-AT from 1-25mM were used to determine the level required to identify the strongest interactions. 3-AT is a competitive inhibitor of the His3 encoded protein and therefore higher levels eliminate the detection of weaker reporter signals. The level required was found to be 15mM 3-AT. All 1100 positives obtained were plated onto –Leucine/-Uracil plates containing 15mM 3-AT, and grown for 3-5 days. This identified the 100 strongest interactions. Overnight cultures were grown and plasmid DNA extracted. This DNA was electroporated into HB101 bacterial cells to rescue the pACTII plasmid. HB101 cells are a Leucine minus strain and require the Leucine expressing pACTII plasmid to grow on minimal media. This selects for the protein encoding pACTII containing cells, and not those transformed with the RNA hybrid pIIIA/MS2-2 plasmid. Cells were grown on M9 plates containing ampicillin for 7 days at 37°C, before colonies picked and overnight cultures grown. High quality plasmid DNA was prepared from these cultures by miniprep and used for sequencing.

T-Track sequencing was carried out to identify identical clones. This was done by manual sequencing from the plasmid DNA prepared. Sequencing reactions were performed using a pACTII specific reverse primer (VR1) and labelled with ^{35}S- dATP. Samples were fractionated on 6% polyacrylamide/7M urea gels and results visualised by autoradiography. T-track sequences were compared to identify identical patterns. Of the 100 clones sequenced, 49 clones appeared a number of times and were reduced in number to 10. This left a total of 61 individual clones identified all of which were sent for ABI sequencing with the pACTII forward primer (VF1). The 5' sequence data obtained was used for Blast searches on the NCBI databases (Altschul et al., 1990). Clones were arranged into order of interest, calculated from the number of times they appeared in the screen and their clone identities. Totally unknown sequences that had no known sequence matches or matches to regions with Alu repeat warnings were not pursued. These sequences appeared unlikely to code for genuine interacting proteins. This left 31 proteins identified from the screen. Results of clone identities and frequency of appearance are summarised in table 4.4 (For full sequence data see appendix).

4.7. Elimination of remaining RNA Independent positives

A final list of 31 proteins was obtained from the screen (Table 4.4). These remaining positives were transformed into yeast without the RNA hybrid to eliminate any

Clone Number	Number of times identified	Identity
8	13	ADH9 (Aldehyde dehydrogenase 9)
24	10	Novel (Fibrillarin-Like)
13	6	Methyl-CpG binding domain protein 2
93	5	Ubiquitin specific proteinase 20
33	3	ARMET (Arginine rich protein)
21	3	OATPRP3 (orgainic anion transporter polypeptide-related protein 3)
83	3	Putative protein KIAA1430
36	2	Putative protein KIAA0179 (homology to Nop52)
69	2	Nitrogen regulatory protein
18	2	Novel (homology to PTB associated factor PSF)
48	1	TIAR (Tial-1 RNA binding protein like-1)
100	1	Fibrillarin
4	1	MAD2L1 (Mitotic arrest deficient-like 1 protein)
16	1	Calpain 4 (Calcium dependent protease)
72	1	Ste20 like (Serine/ Threonine protein kinase homolog)
19	1	MMSETII/WHSC1 (SET domain protein)
23	1	Protamine 1
37	1	THOC2 (THO complex 2)
38	1	Ribosomal L29/ Heparin Sulphate Binding protein
43	1	ATP dependent chromatin remodelling protein (ACF1)
51	1	Endoplasmic reticulum lumenal protein 28
56	1	Crystallin zeta-like 1 (quinone reductase)
58	1	Rho GTPase activating protein 5
77	1	DNA J (Hsp40) protein homolog
81	1	Unknown (Identity to glutathione-S-Transferase)
84	1	Serine proteinase inhibitor
86	1	Peroxisomal membrane protein 3
99	1	PCOLCE2 (Procollagen-c-endopeptidase enhancer 2)
76	1	Putative testis protein
28	1	TCOF-1 (Treacher Collins-Franceschetti syndrome 1)
70	1	Putative protein KIAA0339

Table 4.4. 31 clones as identified by Blast search. The list is in order of priority calculated from number of times the clones were identified and interest in the clone identity. 24 unknown clones that did not match anything significant from Blast searches were 11, 20, 27, 30(X2), 53, 54, 55(X3), 59, 61, 63, 65, 67, 73, 74(X2), 78, 90, 91,95(X2), 97. 6 clones failed to sequence on several occasions.

remaining clones that were able to activate the LacZ reporter on their own and were therefore RNA independent. Yeast were transformed with the pACTII plasmids and plated on –Leucine media. Colonies were grown at 30°C for 3-5 days and used to grow overnight cultures. Liquid β-galactosidase assays were then carried out to look at relative levels of LacZ activation (Figure 4.5).

The level of β-galactosidase activity varied considerably. It was found that 14 of the proteins did not appear to activate the reporter greatly on their own, but the remaining clones activated β-galactosidase to some degree. Those that showed either partial or high β-Galactosidase activity were 81, 48, 93, 83, 4, 77, 84, 99, 51, 16, 69, 43, 56, 58, 8. These RNA independent positives should have been eliminated by the use of 5-FOA screening, but the results indicate that this was not efficient.

However this left 14 proteins that did not appear to have β-Galactosidase activity and were studied further. In addition, clones that appeared of interest were not discarded at this stage. Clone 8 (Aldehyde dehydrogenase 9) was of interest due to the fact it contains a known RNA binding domain and was identified a high number of times in the screen, and clone 48 (TIAR) was of interest as a known RNA binding protein.

4.8 Level of interaction with c-*myc* IRES RNA.

The remaining 16 proteins were then transformed into yeast with either Myc-251 IRES RNA or GAPDH RNA negative control. This was to determine the specificity of their interaction with c-*myc* as compared to a non-specific RNA. Yeast were transformed, plated onto –Leucine/-Uracil media for 3-5 days, and overnight cultures grown. Liquid β-Galasctosidase assays were then performed (Figure 4.6). The activities seen show that almost all proteins activated the reporter more in the presence of RNA. Those that didn't were clones 48 and 8 that were both previously found to activate the reporter without RNA, and as no additional activation was observed these were eliminated. For the remaining clones the activation of the reporter was found to be high with both the c-*myc* IRES RNA and with GAPDH RNA. However, there were differences in the activation levels observed with the two RNAs. The majority of proteins showed higher β-Galactosidase activity with the c-*myc* IRES RNA than with GAPDH RNA. Those that showed higher activity with GAPDH RNA were clone 72 (Serine/ Threonine protein kinase homolog), and clone 37 (THO complex 2).
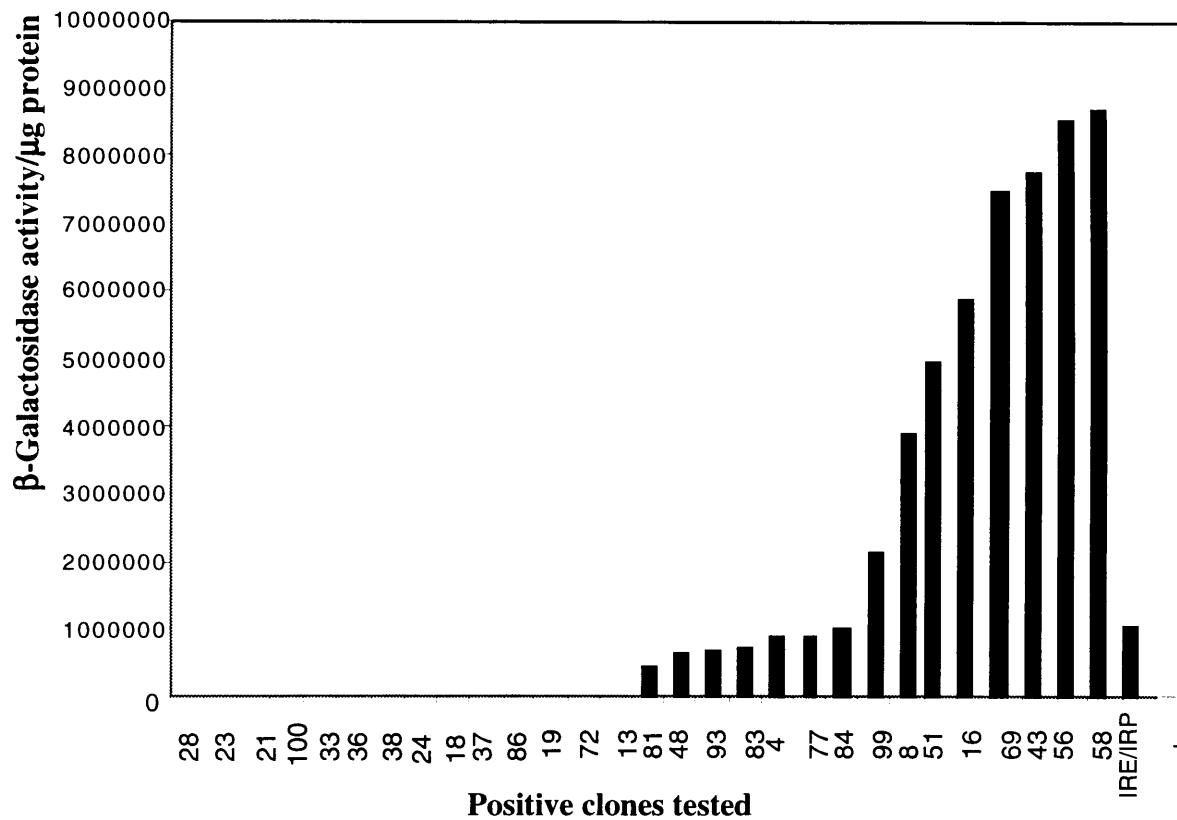
Figure 4.5. Identification of RNA independent positives by liquid β-Galactosidase assay. All positive clones were transformed into yeast without the RNA hybrid and β-Galactosidase activity levels were measured by liquid assay. This allowed positives that were RNA independent to be eliminated. Results show β-Galactosidase activity per μg protein. A respresentative set of data is shown.

Figure 4.6. **Interaction of proteins with c-*myc* or GAPDH RNA in the yeast three-hybrid system by liquid β-galactosidase assay.** β-Galactosidase activity levels were measured by liquid assay. Values for activity are shown for c-*myc* IRES RNA ■ or GAPDH RNA ▨ . Results show β-galactosidase activity per μg protein. A representative set of data is shown.

The significance of this extra activation seen with the c-*myc* IRES RNA is unclear. The chemiluminescent assay performed is very sensitive, and activities observed are only a general guide of strength of interactions. Results do not directly correlate to Kd values of RNA/Protein interactions (Bernstein et al., 2002). In this system shorter RNAs are known to produce stronger reporter signals. The GAPDH RNA used is 140nt compared to the 251nt of the c-*myc* IRES RNA, and may explain why a strong interaction was observed. Therefore the differences in interaction with the c-*myc* IRES RNA and GAPDH RNA may be larger than this assay indicates. Also any c-*myc* IRES interacting proteins are RNA binding proteins, and may therefore show a degree of interaction with other RNAs.

All 14 proteins were studied further. Additional sequencing was carried out to identify those sequences that were not in frame with the Gal4 activations domain, and therefore not coding real proteins. Those found to be out of frame were clone 13 (Methyl-CpG binding domain protein 2), 21 (orgainic anion transporter polypeptide-related protein 3), clone 72 (Serine/ Threonine protein kinase homolog), clone 37 (THO complex 2), clone 86 (Peroxisomal membrane protein 3), clone 28 (Treacher Collins-Franceschetti syndrome 1), and clone 18 (homology to PTB associated factor PSF). These were eliminated leaving 7 proteins.

4.9 Final positives

A final list of the 7 proteins is shown in Table 4.5 (For sequence alignments see appendix). This list includes a number of RNA binding proteins, or sequences with high homology to RNA binding proteins. Fibrillarin is a nucleolar protein involved in rRNA processing and contains an RNA binding domain. The screen also identified a Fibrillarin-like protein 10 times. This has a 74% identity at the protein level to Fibrillarin and is highly homologous in the region containing the fibrillarin RNA binding domain. A clone matching the putative KIAA0179 protein was identified in the screen. KIAA0179 has 56% identity to Nop52 at its N-terminus, which is a protein involved in rRNA processing and which has RNA binding properties. Ribosomal L29 has a large excess of basic residues and as a ribosomal protein may bind RNA, and ARMET is an arginine rich protein, which is a feature common to RNA binding proteins. Protamine and MMSETII are the exceptions as these are DNA binding proteins. Out of the proteins obtained Fibrillarin (100), Protamine (23), and Ribosomal L29 (38) are full length proteins. Fibrillarin-like protein

Clone Reference	Name	Description
24	Novel (Fibrillarin-Like)	Protein with 74% identity to Fibrillarin.
33	ARMET	Arginine rich protein mutated in early stage tumors.
36	Novel (KIAA0179)	KIAA0179 has 56% identity to Nop52 at its N-terminus. Nop52 is a novel nuclear protein thought to be the human homolog of yeast ribosomal RNA processing factor RRP1.
100	Fibrillarin	Human nucleolar protein involved in processing rRNA. Has a RNA binding domain.
19	MMSETII/WHSC1	A SET domain containing protein present in Wolf-Hirschorn syndrome.
23	Protamine 1	Small arginine rich protein involved in condensation of sperm chromatin. Testis specific.
38	Ribosomal Protein L29	Homology to mouse ribosomal protein L29. Also known as HIP - a novel cell surface heparan sulfate binding protein.

Table 4.5. Final 7 clones of interest. These clones coded for proteins of interest, the interactions were not RNA independent, and the clones activated more with c-*myc* IRES RNA than with GAPDH RNA.

(24) and ARMET (33) appear to be missing only the 5' end of the clone, and KIAA0179 (36) and MMSETII (19) are partial clones.

The high number of RNA binding proteins in this list shows that the yeast three-hybrid screen has been successful in identifying RNA binding proteins that bind to the c-*myc* IRES hybrid RNA. The proteins Fibrillarin, Fibrillarin-like and KIAA0179 (Nop-52-like) are interesting due to the recent discovery of Nucleolin as an IRES trans-acting factor (Izumi et al., 2001). Nucleolin and Fibrillarin are nucleolar proteins involved in rRNA processing (Leary and Huang, 2001) and Nop-52 is a homologue of the yeast rRNA processing protein RRP1 (Savino et al., 1999). Fibrillarin has also been found to be involved in pre-mRNA processing (Jang et al., 2002). However, the screen may have identified general RNA binding proteins with no role in IRES translation. Functional tests to investigate effects on c-*myc* IRES translation were carried out.

4.10. Effect of proteins on c-*myc* IRES activity *in vivo*.

Functional tests were performed to determine if the proteins identified in the yeast three-hybrid screen were general RNA binding proteins or had a role in c-*myc* IRES translation. The effect of the proteins on c-*myc* IRES activity *in vivo* was studied. The proteins were cloned into pcDNA3. Clones 100, 23 and 38 were full length and could be excised directly by restriction digestion with *Eco*RI and *Xho*I, and transferred into pcDNA3. Other clones were partial and were PCR amplified to add a start site. PCR was carried out under standard conditions from pACTII templates using primers containing the relevant restriction sites. Fragments were gel purified and restriction digestions performed before cloning into pcDNA3. Clones 33 and 19 were cloned using *Eco*RI and *Xho*I, clone 24 using *Hind*III and *Xho*I, and clone 36 was cloned using *Hind*III and *Bam*HI. Primers used are shown in table 4.6. For primer sequences see table 2.2.

Protein cloned	Primers used
24	FLFOR + FLREV
33	33F + 33R
36	36F + 36R
19	19F + 19R

Table 4.6. Primers used to PCR amplify protein fragments.

The pcDNA3 constructs were co-transfected into HeLa cells along with the dicistronic c-myc IRES plasmid pRMF (Figure 4.7A). pRMF contains the c-myc IRES between the Renilla luciferase and Firefly luciferase genes, allowing the proteins effect on c-myc IRES translation to be studied. Translation levels of Firefly and Renilla luciferase were determined using a dual luciferase assay. All transfections were carried out with the addition of pJLacZ, and results shown relative to β-galactosidase activity to adjust for changes in transfection efficiency. The results show no clear activation of c-myc IRES activity in HeLa cells for any of the proteins studied (Figure 4.7B). Slight activations were observed with some proteins, but results were not consistent and some proteins appeared to inhibit c-myc IRES activity. The results observed may be due to the protein fragments all exhibiting c-myc IRES RNA binding ability. The addition of some proteins may have caused inhibition of IRES activity due to this binding ability for the RNA. Also only clones 100, 23 and 38 were full length and other proteins fragments may contain RNA binding domains but not other domains required for full protein activity.

4.11 Effect of proteins on c-myc IRES activity in vitro.

Proteins were studied to see the effect of the proteins on c-myc IRES activity in vitro. This was done using the rabbit reticulocyte lysate system. To look at the effect of proteins in this system recombinant proteins must be added. Clone 24 (Fibrillarin-like) and clone 100 (Fibrillarin) were chosen due to the high number of times the Fibrillarin-like protein was identified in the initial screen. The protein fragments were amplified by PCR using primers containing the relevant restriction sites. For Fibrillarin-like protein the primers FLFOR and PETFLREV were used, and for Fibrillarin the primers YH100F and YH100R. Both were cloned into pET21a in frame with the His-Tag extension using EcoRI and XhoI. The pET21a constructs were transformed into BL21 cells and protein expression induced at 37°C for 4 hr. Proteins were purified using nickel affinity columns.

In vitro translations were carried out using the rabbit reticulocyte lysate system and the effect of the addition of protein factors analysed. RNA transcripts were produced by in vitro run-off transcription from the dicistronic plasmid pRMF linearised with HpaI or pRF linearised with BamHI (Figure 4.8A). Transcripts were added to rabbit reticulocyte lysate with 150ng protein factors and incubated at 30°C for 1.5 hours. Reactions were then assayed for luciferase activity.
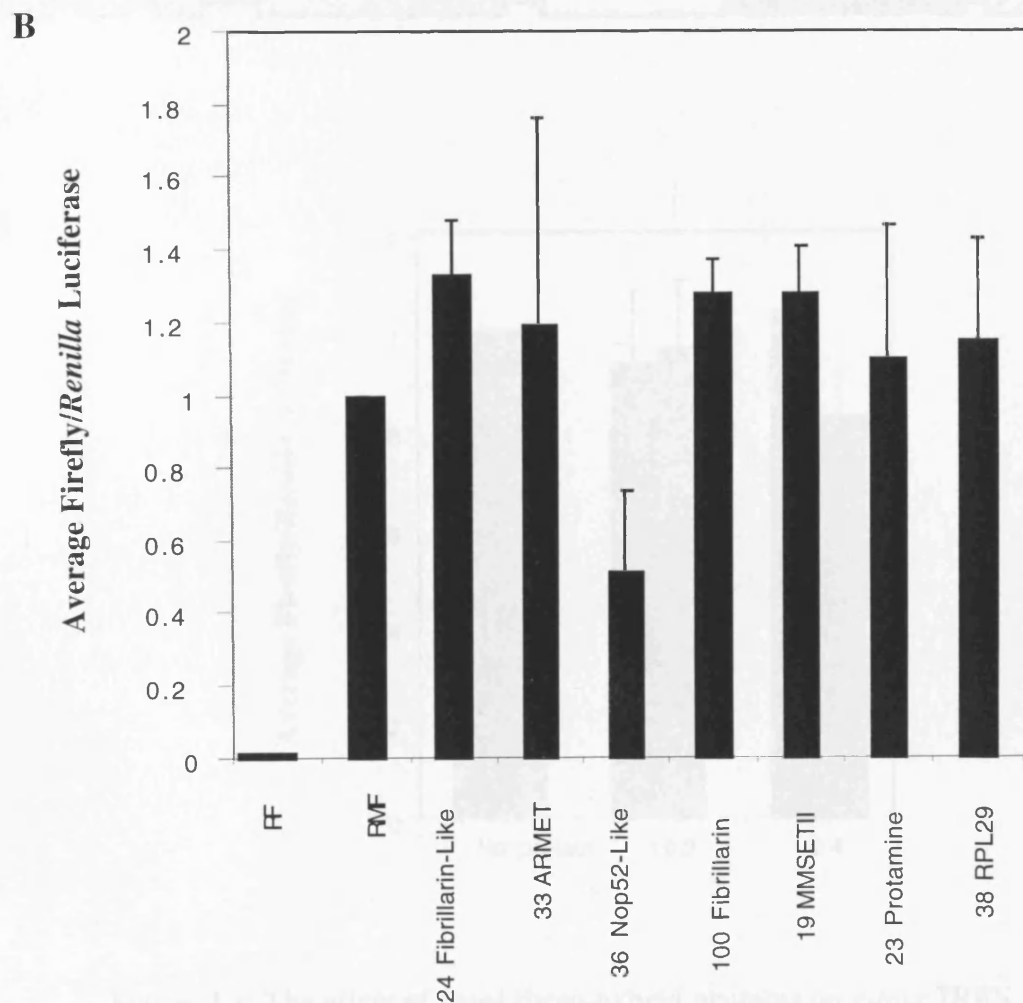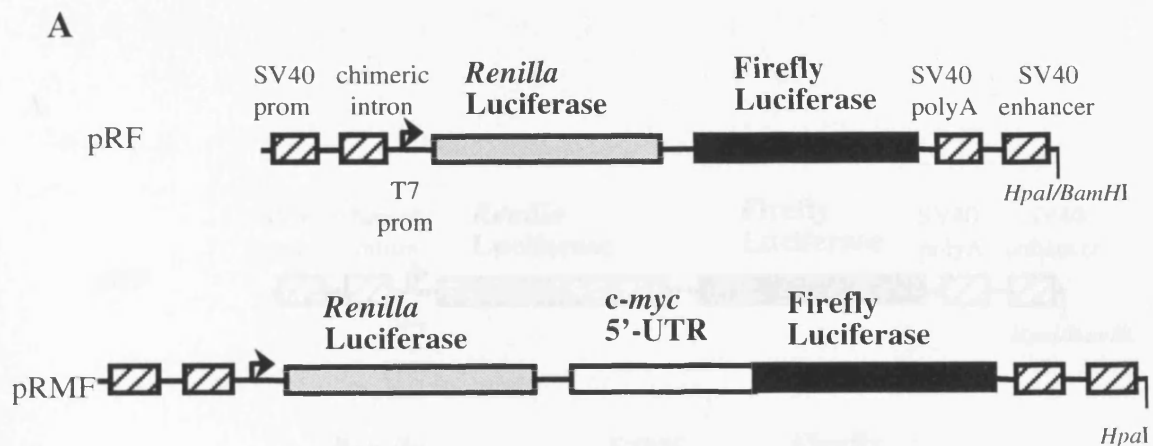
Figure 4.7. The effect of yeast three-hybrid proteins of c-*myc* IRES activity in *vivo*. A) pRMF dicistronic construct containing the c-*myc* IRES between the *Renilla* and Firefly luciferase genes. B) The effect of Y3H positive proteins of c-*myc* IRES activity *in vivo*. Results show co-transfection of 0.5μg pcDNA3 constructs and 1μg pRMF into HeLa cells. Results are average Firefly/*Renilla* luciferase normalised to pRMF. Error bars show standard deviation.

A

SV40 chimeric **Renilla** **Firefly** SV40 SV40
prom intron **Luciferase** **Luciferase** polyA enhancer

pRF

T7
prom

HpaI/BamHI

Renilla **c-myc** **Firefly**
Luciferase **5'-UTR** **Luciferase**

pRMF

HpaI
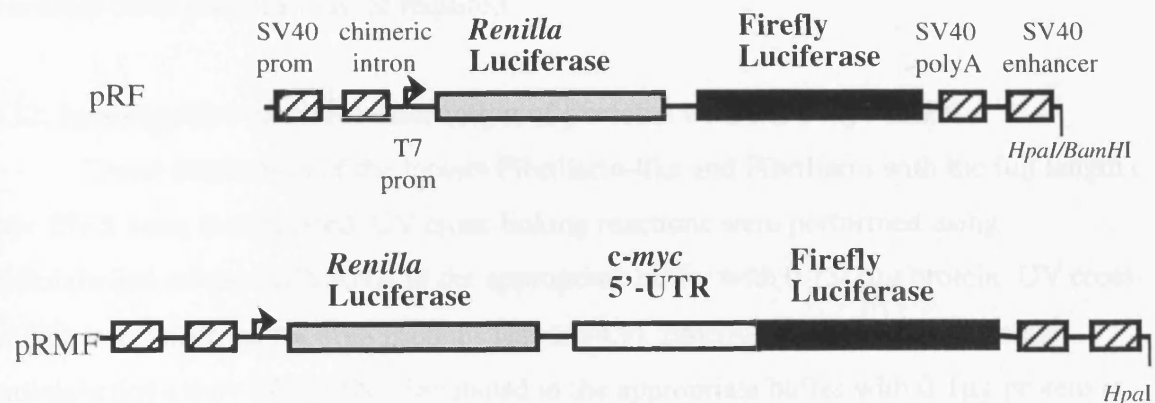
B
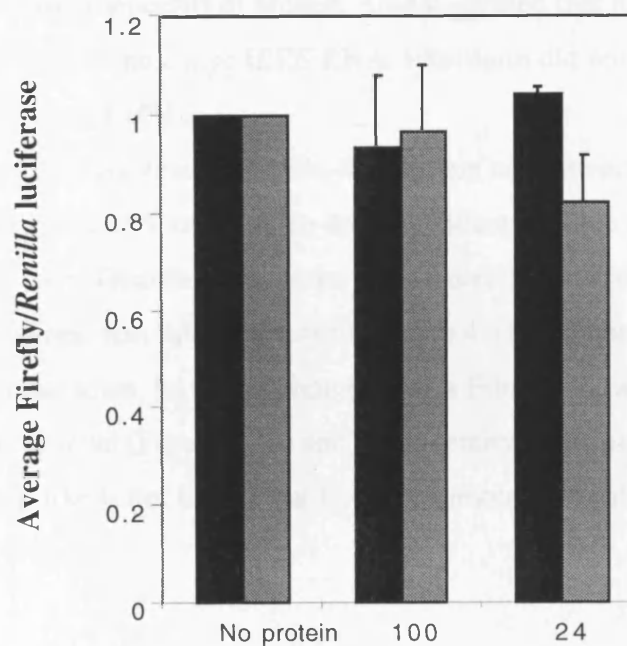
Figure 4.8. The effect of yeast three-hybrid proteins on c-*myc* IRES activity in *vitro*. A) Transcripts were produced by *in vitro* run-off transcription from pRF or pRMF linearised plasmids. B) The effect of yeast three-hybrid proteins on c-*myc* IRES activity in *vitro*. Results are shown as normalised average Firefly/*Renilla* luciferase readings. ■ pRF transcript ▨ pRMF transcript.

No effect on c-*myc* IRES activity was observed with the addition of the protein factors Fibrillarin or Fibrillarin-like (Figure 4.8B). This could be due to these individual factors not being limiting in rabbit reticulocyte lysate. Other work investigating known *trans*-acting factors identified an effect only when proteins were added in combination and therefore other proteins may be required.

4.12. Investigation of direct interaction of proteins with the c-*myc* IRES.

Direct interaction of the factors Fibrillarin-like and Fibrillarin with the full length c-*myc* IRES were investigated. UV cross-linking reactions were performed using radiolabelled c-*myc* IRES RNA in the appropriate buffer with 0.75-1µg protein. UV cross-linking was observed for both proteins (Figure 4.9). EMSAs were carried out using radiolabelled c-*myc* IRES RNA incubated in the appropriate buffer with 0-1µg protein at room temperature. Reactions were then run on 5% polyacrylamide/TBE gels. A shift was observed for Fibrillarin-like protein (Figure 4.10A). This produced two shifts in the gel with the addition of increasing amounts of protein. This suggested that more than one molecule of protein can bind to the c-*myc* IRES RNA. Fibrillarin did not to shift the c-*myc* IRES RNA by EMSA (Figure 4.10B).

The results obtained show that Fibrillarin-like protein can interact with the full length c-*myc* IRES RNA by EMSA and was the most abundant positive to be isolated from the yeast three-hybrid screen. Therefore this clone was studied in more detail. The Fibrillarin-like clone obtained was fully sequenced (Figure 4.11). It contained an open reading frame of 304 amino acids. Its nearest homolog was Fibrillarin, with which it had 74% identity at the protein level (Figure 4.12) and 72% identity at the nucleotide level. The start codon for Fibrillarin-like is not known but from its homology to Fibrillarin it appears to be missing the most 5'end.

4.13 Discussion

The yeast three-hybrid system was used to study the interactions between proteins and the c-*myc* IRES. Initial work investigating proteins known to have a role in IRES translation did not identify any interactions. A three-hybrid library screen was performed to identify novel c-*myc* IRES binding proteins. The screen identified seven proteins that may have a role in c-*myc* IRES translation. These were Fibrillarin-like protein, ARMET, KIAA0179 protein (Nop52-like), Fibrillarin, MMSETII, Protamine-1 and Ribosomal
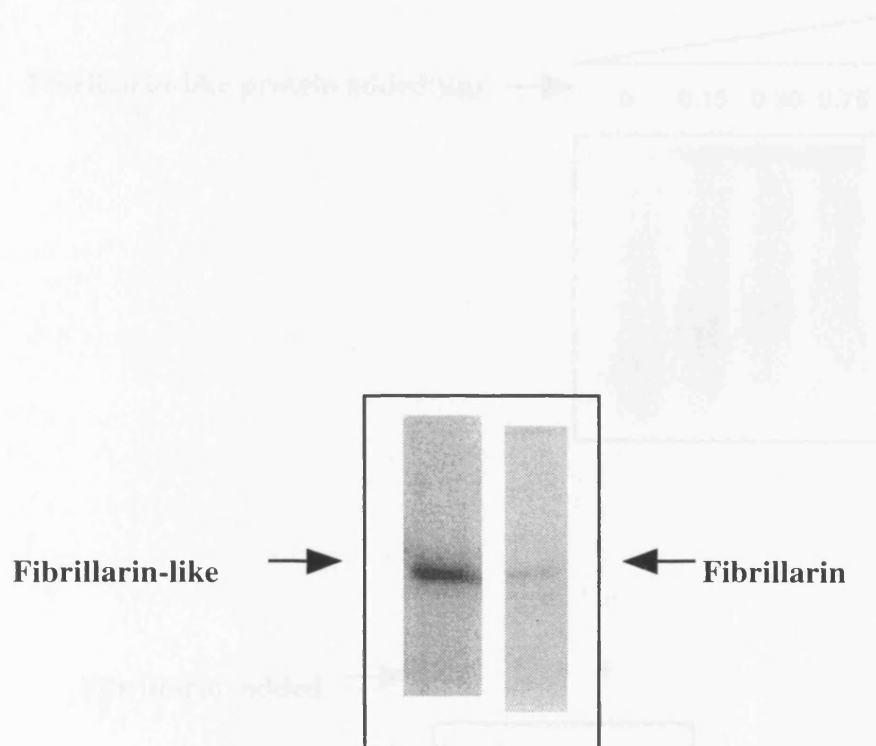
Figure 4.9. UV crosslinking of Fibrillarin-like (24) and Fibrillarin (100) proteins to the c-*myc* IRES RNA. 0.75µg Fibrillarin-like and 1µg Fibrillarin were used.
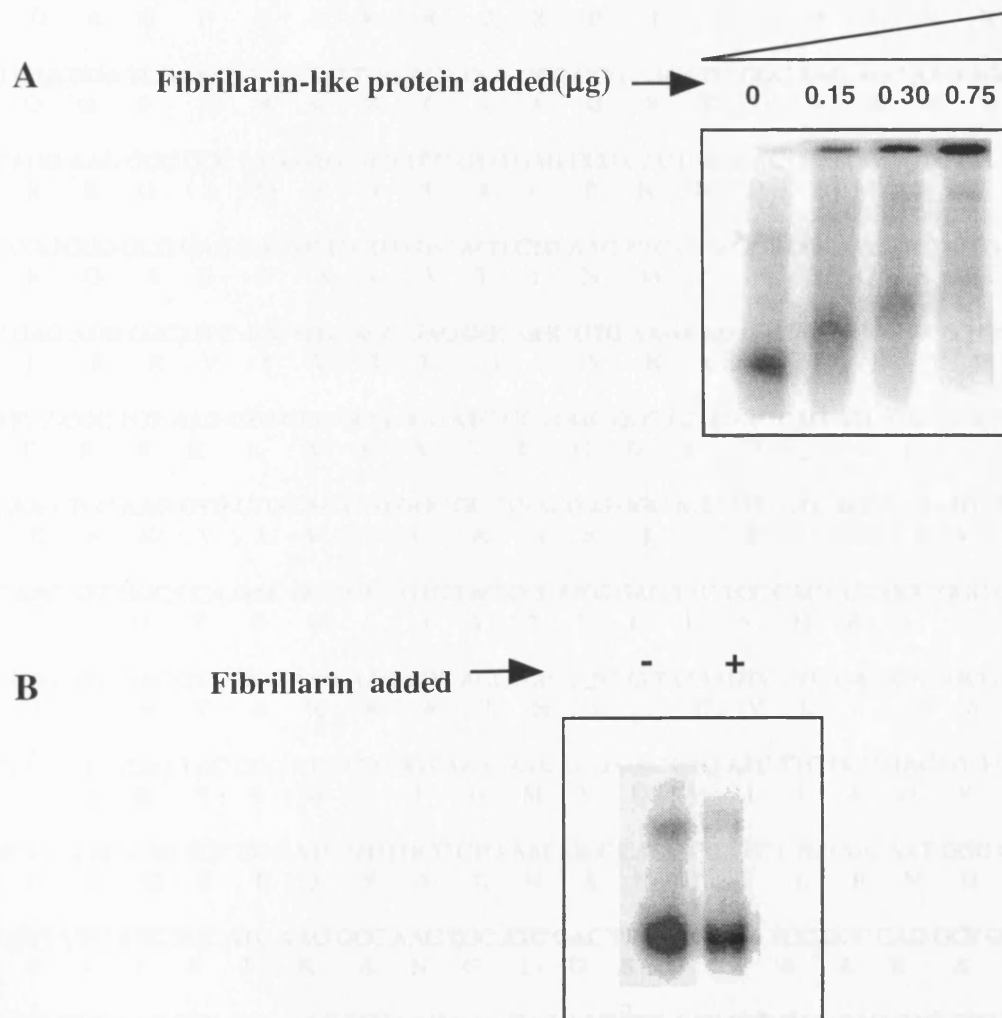
Figure 4.10. EMSA of the c-*myc* IRES RNA with Fibrillarin-like and Fibrillarin proteins. A) EMSA of the c-*myc* IRES RNA with Fibrillarin-like protein. Increasing amounts of protein were added from 0-0.75 μg. B) EMSA of the c-*myc* IRES RNA with 1μg Fibrillarin.

GGC GGC GCG GGC AAG GGC GGC GGG GGC GAC TGG CGG CGG CCA GGG GGG CAA GGG CGG CTT
 G   G   A   G   K   G   G   G   G   D   W   R   R   P   G   G   Q   G   R   L

CGG GGC GCG GGC CCG GGG CTT CGG CGG GGG CGG CCG GGG CCG GGG GCG CGG CGG CGG CGA
 R   G   A   G   P   G   L   R   R   G   R   P   G   P   G   A   R   R   R   R

AGG CAA GGA TCG CGG CGG CGG TGG ACA GCG GCG GGG CGG CGT GGC AAG AGC AAG AGC CGC
 R   Q   G   S   R   R   R   W   T   A   A   G   R   R   G   K   S   K   S   R

CGC AGG AAG GGC GCC ATG GTG GTG TCG GTG GAG CCG CAC CGG CAC GAG GGC GTC TTC ATC
 R   R   K   G   A   M   V   V   S   V   E   P   H   R   H   E   G   V   F   I

TAC CGC GGG GCG GAG GAC GCG CTG GTC ACG CTG AAC ATG GTG CCG GGC CAG TCT GTG TAC
 Y   R   G   A   E   D   A   L   V   T   L   N   M   V   P   G   Q   S   V   Y

GGC GAG AGG CGC GTC ACG GTG ACC GAG GGC GGC GTG AAG CAG GAG TAC CGC ACG TGG AAC
 G   E   R   R   V   T   V   T   E   G   G   V   K   Q   E   Y   R   T   W   N

CCG TTC CGC TCT AAG CTG GCC GCG GCC ATC CTG GGC GGG GTG GAC CAG ATC CAC ATC AAG
 P   F   R   S   K   L   A   A   A   I   L   G   G   V   D   Q   I   H   I   K

CCC AAG TCC AAG GTG CTG TAC CTG GGC GCC GCG TCG GGC ACC ACC GTC TCC CAT GTC TTC
 P   K   S   K   V   L   Y   L   G   A   A   S   G   T   T   V   S   H   V   F

GAC ATC ATT GGC CCA GAC GGC CTG GTC TAC GCC GTC GAG TTC TCC CAC CGC GCC GGC CGC
 D   I   I   G   P   D   G   L   V   Y   A   V   E   F   S   H   R   A   G   R

GAT CTG GTC AAC GTG GCC AAG AAG CGC ACC AAC ATC ATT CCG GTC CTG GAG GAC GCG CGG
 D   L   V   N   V   A   K   K   R   T   N   I   I   P   V   L   E   D   A   R

CAC CCG CTC AAG TAC CGC ATG CTC ATC GGG ATG GTG GAC GTG ATC TTC GCC GAC GTG GCC
 H   P   L   K   Y   R   M   L   I   G   M   V   D   V   I   F   A   D   V   A

CAG CCG GAC CAG TCC CGC ATC GTG GCC CTG AAC GCC CAC ACC TTC CTG CGC AAT GGG GGC
 Q   P   D   Q   S   R   I   V   A   L   N   A   H   T   F   L   R   N   G   G

CAC TTT CTC ATC TCC ATC AAG GCC AAC TGC ATC GAC TCC ACC GCA TCC GCC GAG GCT GTG
 H   F   L   I   S   I   K   A   N   C   I   D   S   T   A   S   A   E   A   V

TTT GCT TCT GAG GTG AGG AAG TTG CAG CAG GAG AAC TTG AAG CCT CAA GAG CAG CTG ACC
 F   A   S   E   V   R   K   L   Q   Q   E   N   L   K   P   Q   E   Q   L   T

CTG GAG CCC TAT GAG CGG GAC CAC GCT GTG GTG GTC GGG GTC TAC CGA CCT CTT CCC AAG
 L   E   P   Y   E   R   D   H   A   V   V   V   G   V   Y   R   P   L   P   K

AGC AGC AGC AAA TAG CAC CCA GCT CAG GCT CGC CTG CCA TCT CCC CAA GGC TGC GTT GTG
 S   S   S   K   *   H   P   A   Q   A   R   L   P   S   P   Q   G   C   V   V

TTT GCT ATT ATT TTC TGT GTG TTT TCT TTG TGA GTG TTT TGT TTT GTT GTT TTT CTA TTA
 F   A   I   I   F   C   V   F   S   L   *   V   F   C   F   V   V   F   L   L

AAC TGC ATA AAG AAA CGG CAA AAA AAA AAA AAA AAA AAA AA
 N   C   I   K   K   R   Q   K   K   K   K   K

Figure 4.11. Sequence of Fibrillarin-Like clone. Sequence
showing protein translation

```
Fibrillarin-Like:   1   GGAGKGGGGDWRRPGGQGRL---RGAGPGLR-RGRPGPGARRRRRQGSRRR--WTAAG
                        G  G+GG GD      GG+G    RG G G R RGR G G       G R    +  +  G
Fibrillarin       :  11 GFGGRGGFGDRGGRGGRGGFGGGRGRGGGFRGRGRGGGGGGGGGGGGGRGGGGFHSGG

Fibrillarin-Like:  55   GKSKSRRR--KGAMVVSVEPHRHEGVFIYRGAEDALVTLNMVPGQSVYGERRVTVTEG
                        G+ +   +R   +      V VEPHRHEGVFI RG EDALVT N+VPG+SVYGE+RV+++EG
Fibrillarin       :  71 GRGRGGKRGNQSGKNVMVEPHRHEGVFICRGKEDALVTKNLVPGESVYGEKRVSISEG

Fibrillarin-Like: 113   KQEYRTWNPFRSKLAAAILGGVDQIHIKPKSKVLYLGAASGTTVSHVFDIIGPDGLVY
                        K EYR WNPFRSKLAAAILGGVDQIHIKP +KVLYLGAASGTTVSHV DI+GPDGLVY
Fibrillarin       : 131 KIEYRAWNPFRSKLAAAILGGVDQIHIKPGAKVLYLGAASGTTVSHVSDIVGPDGLVY

Fibrillarin-Like: 173   EFSHRAGRDLVNVAKKRTNIIPVLEDARHPLKYRMLIGMVDVIFADVAQPDQSRIVAL
                        EFSHR+GRDL+N+AKKRTNIIPV+EDARHP KYRMLI MVDVIFADVAQPDQ+RIVAL
Fibrillarin       : 191 EFSHRSGRDLINLAKKRTNIIPVIEDARHPHKYRMLIAMVDVIFADVAQPDQTRIVAL

Fibrillarin-Like: 233   HTFLRNGGHFLISIKANCIDSTASAEAVFASEVRKLQQENLKPQEQLTLEPYERDHAV
                        HTFLRNGGHF+ISIKANCIDSTASAEAVFASEV+K+QQEN+KPQEQLTLEPYERDHAV
Fibrillarin       : 251 HTFLRNGGHFVISIKANCIDSTASAEAVFASEVKKMQQENMKPQEQLTLEPYERDHAV

Fibrillarin-Like: 293   GVYRPLPK 300
                        GVYRP PK
Fibrillarin       : 311 GVYRPPPK 318
```

Figure 4.12. Sequence alignment of Fibrillarin-like protein with Fibrillarin.
There is 74% identity at the protein level.

Protein L29. A number of these positives were known RNA binding proteins or highly homologous to known RNA binding proteins. As discussed Fibrillarin is a nucleolar protein involved in rRNA processing, and KIAA0179 has homology to Nop-52, a homologue of the yeast ribosomal RNA processing protein RRP1. Therefore Fibrillarin-like and KIAA0179 (Nop52-like) proteins are likely to bind RNA. The protein Ribosomal L29 has a large excess of basic residues and as a ribosomal protein may bind RNA, and ARMET is an arginine rich protein, which is a feature common to RNA binding proteins. Only Protamine-1 and MMSETII do not fall into this category, as they are known DNA binding proteins.

The relevance of the interaction of these proteins with c-*myc* IRES RNA has not been determined. Functional tests on their effect on IRES activity in *vivo* and *vitro* were non conclusive. The proteins identified may be general RNA binding proteins, and have no functions in c-*myc* IRES translation. The yeast three-hybrid system requires that the RNA of interest folds into the correct structure for protein binding. The c-*myc* IRES fragment used was large and therefore it is unknown if it would fold the correct structure *in vivo*. In addition the screen used identified the strongest interactions, which may not be relevant when identifying IRES interacting proteins. It is thought that several proteins bind to the IRES structure for efficient IRES translation and therefore the interaction of several proteins at once may be required for efficient binding. This has been shown for the Apaf-1 IRES to which PTB binds only in the presence of UNR (Mitchell et al., 2001). However, the fact these proteins can bind the c-*myc* IRES RNA strongly, especially for Fibrillarin-like protein which interacted with the full length IRES by EMSA, suggests these interactions may not be random. The reason that the proteins showed no effect on c-*myc* IRES activity may be because some proteins were not full length and may therefore not have been functional.

Several of the proteins identified were nuclear proteins or showed high homology to nuclear proteins. It has been found previously that the c-*myc* IRES requires a nuclear event for translation to occur (Stoneley et al., 2000b) and therefore a nuclear protein may be required for efficient IRES translation. In is also interesting that Fibrillarin, Fibrillarin-Like and Nop52-like proteins were identified, as Nucleolin, another protein involved in rRNA processing, has been found to have a role in IRES translation from HCV and PV (Izumi et al., 2001). In addition yeast Fibrillarin has been shown to be involved in pre-mRNA

processing by stabilising the uvi15+ pre-mRNA (Jang et al., 2002). Further work needs to be carried out to determine if these proteins have any role in c-*myc* IRES translation.

Chapter 5

Internal ribosome entry in the c-*myb* 5'UTR

5.1 Introduction

c-*myb* is a member of the Myb family of transcription factors and is involved in the control and differentiation of haematopoietic cells. Like c-*myc* its expression is tightly regulated in the cell, including at the level of transcription. The gene has multiple transcriptional start sites and the protein has also been found to act on its own transcription. The main transcription start site was found to be 200bp upstream of the AUG codon in human transcripts, producing a transcript with a 73% GC rich 5'UTR. This would be predicted to form a strong secondary structure and suggests that the gene may be controlled at the level of translation. c-*myb* is involved in control of cell growth and is an oncogene, and several genes with these features are known to be controlled by internal ribosome entry (Hellen and Sarnow, 2001). Therefore the effect of the c-*myb* 5'UTR on translation was investigated.

5.2 Cloning of the c-*myb* 5'UTR

The c-*myb* 5'UTR was obtained by PCR from human Lymphoblast cDNA. Primers were designed to amplify the region of 200nt from the main transcription start site to the ATG start codon (Figure 5.1). The cDNA was produced from GM01953 cell Poly A+ RNA by reverse transcription. The fragment was amplified using the primers MybECO and MybNCO, under standard PCR conditions with the addition of 0.5M Betaine and 5% DMSO. A fragment of the expected 200bp was obtained and gel purified. A restriction digest was performed using *Eco*RI and *Nco*I and the fragment cloned into the dicistronic plasmid pRF in the intercistronic region between the *Renilla* and Firefly luciferase genes (Figure 5.2). The 200bp insert was sequenced and found to exactly match that of the published sequence.

5.3 Effect of the c-*myb* 5'UTR on translation from pRF.

In order to test the effect of the c-*myb* 5'UTR on translation the construct pRmybF was transfected into HeLa cells. pRmybF contains the c-*myb* 5'UTR between the *Renilla* and Firefly luciferase genes (Figure 5.3A), and allows the effect of the 5'UTR on translation from these cistrons to be studied. Translation levels of Firefly and *Renilla*

ATATCAACCT GTTTCCTCCT CCTCCTTCTC CTCCTCCTCC
GTGACCTCCT CCTCCTCTTT CTCCTGAGAA ACTTCGCCCC
AGCGGTGCGG AGCGCCCTGC GCAGCCGGGG AGGGACGCAG
GCAGGCGGCG GGCAGCGGGA GGCGGCAGCC CGGTCGGTCC
CCGCGGCTCT CGCGGAGCCC CGCCGCCCGC CGCGCCATGG

Figure 5.1. The c-*myb* 5'UTR sequence. The sequence is 200nt in length
and 73% GC rich.

SV40 promoter

upstream poly(A)   Intron   EcoRV 502

fl origin

Rluc

Amp

pRF
6499 bp

SpeI 1490
PvuII 1498
EcoRI 1502
NcoI 1521

Fluc

Enhancer   SV40 poly(A)

EcoRI   NcoI

c-*myb* 5'UTR

SV40 promoter

upstream poly(A)   Intron   EcoRV 502

fl origin

Rluc

SpeI 1490
PvuII 1498
EcoRI 1502

Amp

pRmybF
6700 bp

c-myb 5'UTR

NcoI 1703

Fluc

Enhancer   SV40 poly(A)
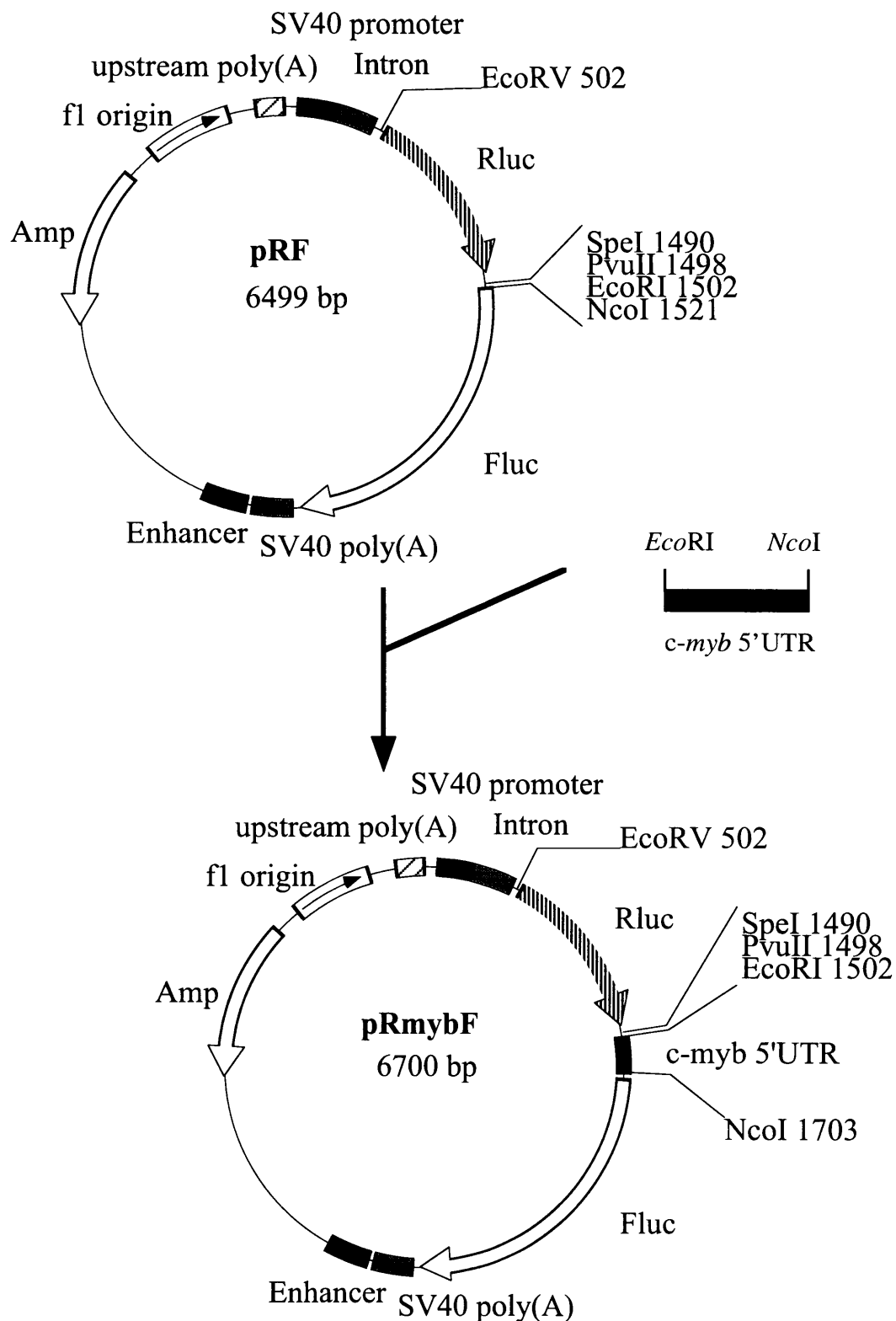
Figure 5.2. Cloning of pRmybF. The 5'UTR of c-*myb* was PCR amplified with gene specific primers containing the required restriction sites and cloned into the plasmid pRF between the *Renilla* and Firefly genes.

A

pRF

SV40   chimeric   *Renilla*   **Firefly**   SV40   SV40
prom   intron   Luciferase   Luciferase   polyA   enhancer

                    T7   *Hpa*I/*Bam*HI
                    prom

pRmybF   *Renilla*   c-*myb*   **Firefly**
                Luciferase   5'-UTR   Luciferase

                                                                          *Hpa*I
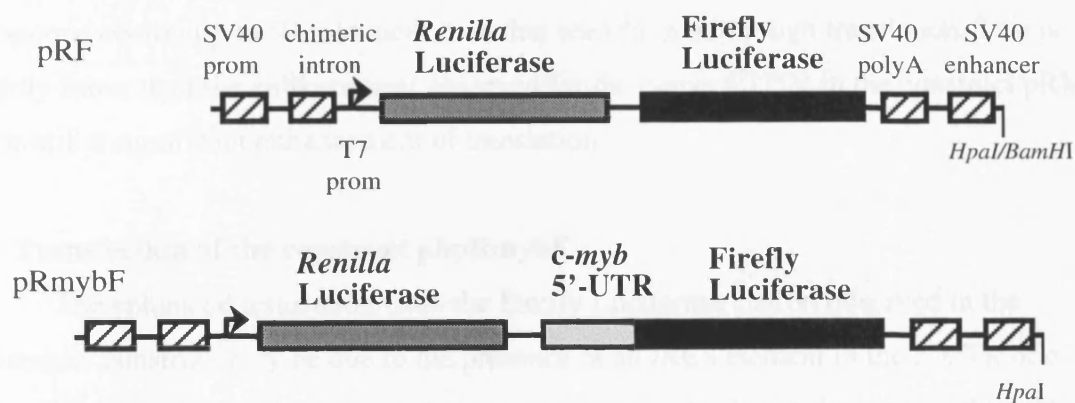
B
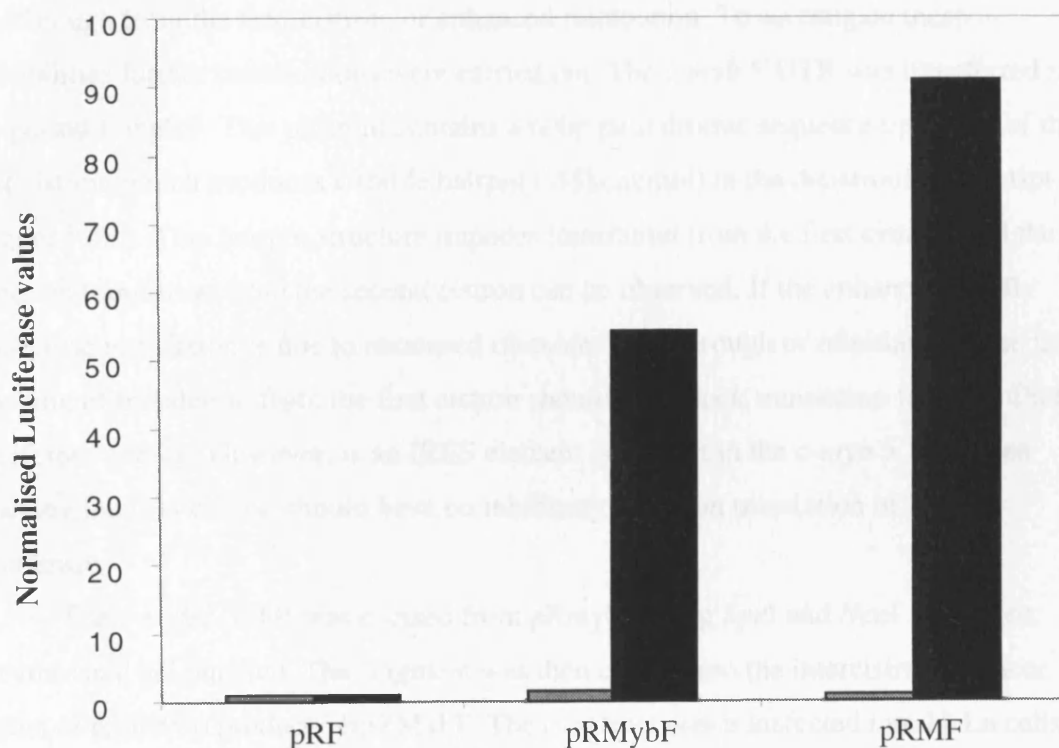
Figure 5.3. Transient transfection of pRF, pRmybF and pRMF into HeLa cells. A) pRmybF dicistronic construct containing the c-*myb* IRES between the *Renilla* and Firefly luciferase genes. B) Transient transfection of pRF, pRmybF and pRMF into HeLa cells. Results show normalised *Renilla* ⬜ and Firefly ⬛ luciferase readings and are adjusted to relative β-galactosidase activity. A representative set of data is shown .

luciferase were determined using a dual luciferase assay. The plasmids pRF, and pRMF were transfected in parallel, and all transfections were carried out with the addition of pJLacZ to enable results to be adjusted for changes in transfection efficiency.

It was found that the c-*myb* 5'UTR enhanced translation from the second cistron of the dicistronic construct (Figure 5.3B). The 5'UTR was found to increase translation from the second cistron up to 50 fold more than that seen for readthrough translation. This is slightly lower than the enhancement observed for the c-*myc* 5'UTR in the construct pRMF, but is still a significant enhancement of translation.

5.4 Transfection of the construct phpRmybF

The enhanced translation from the Firefly Luciferase cistron observed in the dicistronic construct may be due to the presence of an IRES element in the 5'UTR of c-*myb*. However, the increase in translation seen may also be due to the increased length of the intercistronic region. This increase in length could cause enhanced ribosomal readthrough from the first cistron, or enhanced reinitiation. To investigate these possibilities further transfections were carried out. The c-*myb* 5'UTR was transferred into the plasmid phpRF. This plasmid contains a 60bp palindromic sequence upstream of the first cistron, which produces a stable hairpin (-55kcal/mol) in the dicistronic transcript (Figure 5.4A). This hairpin structure impedes translation from the first cistron, and the effect on translation from the second cistron can be observed. If the enhanced Firefly Luciferase translation is due to increased ribosomal readthrough or reinitiation, then the blocking of translation from the first cistron should also block translation from the Firefly Luciferase cistron. However, is an IRES element is present in the c-*myb* 5'UTR then blocking the first cistron should have no inhibitory effect on translation of Firefly Luciferase.

The c-*myb* 5'UTR was excised from pRmybF using *Spe*I and *Nco*I restriction enzymes and gel purified. The fragment was then cloned into the intercistronic spacer region of phpRF to produce phpRMybF. The construct was transfected into HeLa cells in parallel with the constructs pRF, pRMybF. Again transfections were carried out with the addition of pJLacZ to enable results to be adjusted for changes in transfection efficiency. Results showed that translation from *Renilla* luciferase was decreased greatly with the insertion of the stable hairpin structure (Figure 5.4B). Translation from the Firefly luciferase cistron was not decreased, showing that its translation was independent from that
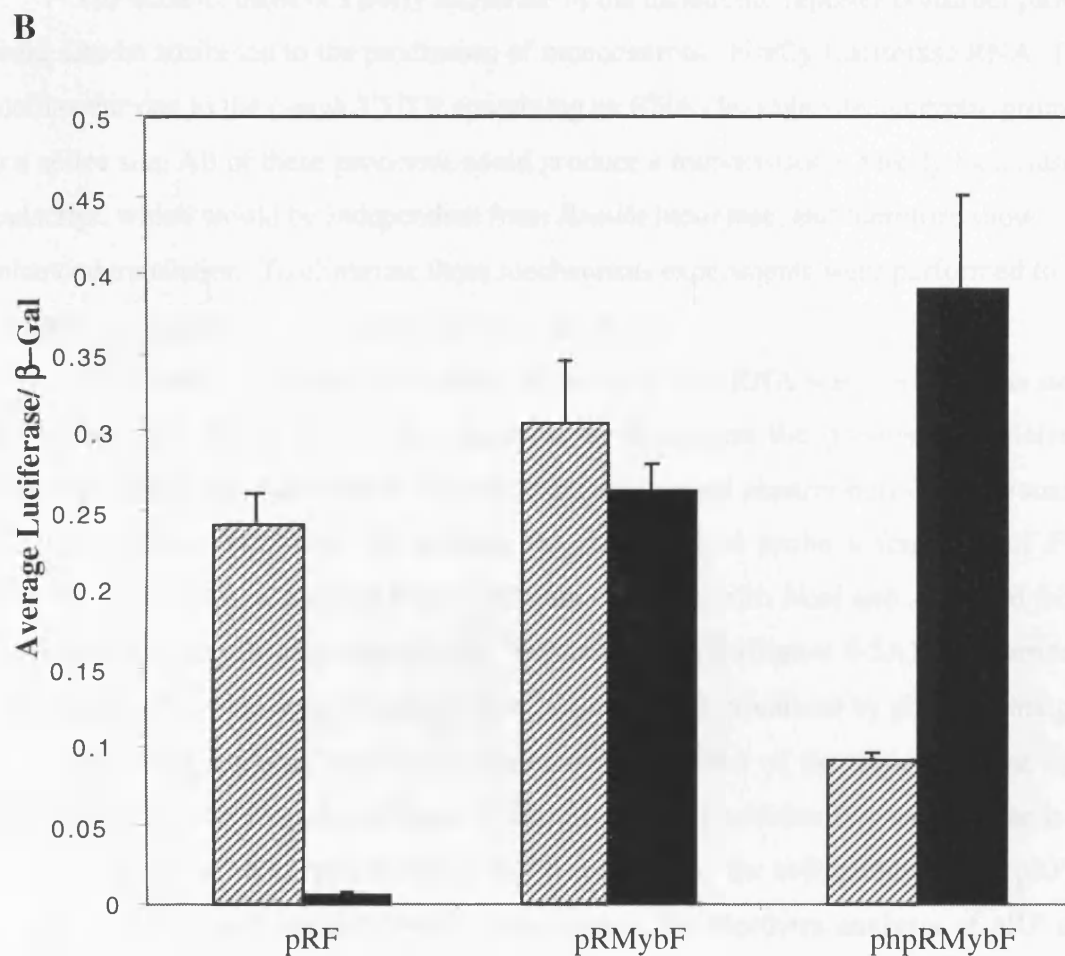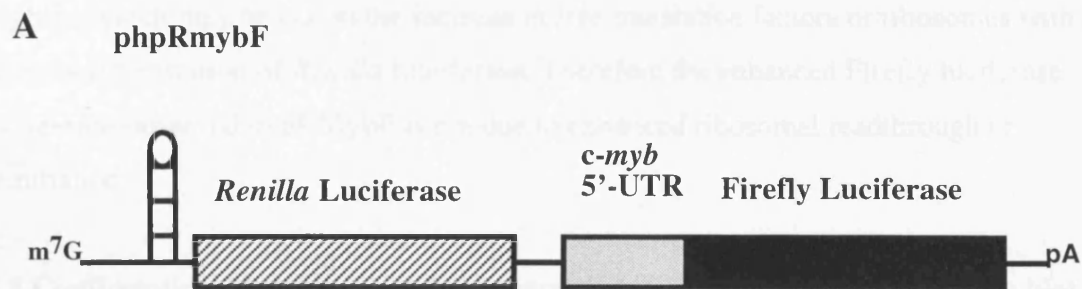
Figure 5.4. Transient transfection of pRF, pRmybF and phpRmybF into HeLa cells. A) phpRmybF dicistronic construct containing the c-*myb* IRES between the *Renilla* and Firefly luciferase genes. B) Transient transfection of pRF, pRmybF and phpRmybF into HeLa cells. Results show average *Renilla* ▨ and Firefly ■ luciferase readings and are adjusted to relative β-galactosidase activity. A representative set of data is shown .

of the first cistron. The translation of Firefly Luciferase was in fact found to increase slightly, which may be due to the increase in free translation factors or ribosomes with the decreased translation of *Renilla* Luciferase. Therefore the enhanced Firefly luciferase expression observed in pRMybF is not due to enhanced ribosomal readthrough or reinitiation.

5.5 Confirmation of expression of a dicistronic transcript *in vivo* by northern blot analysis

The enhancement of Firefly luciferase in the dicistronic reporter construct pRMybF could also be attributed to the production of monocistronic Firefly Luciferase RNA. This could occur due to the c-*myb* 5'UTR containing an RNA cleavage site, a cryptic promoter, or a splice site. All of these processes could produce a monocistronic Firefly luciferase transcript, which would be independent from *Renilla* luciferase, and therefore show enhanced translation. To eliminate these mechanisms experiments were performed to confirm the expression of a dicistronic message in *vivo*.

Cells were transfected with pRMybF, and poly A+ RNA was isolated. The isolated RNA was used for Northern Blot analysis to determine the presence of dicistronic transcript. RNA was fractionated by denaturing agarose gel electrophoresis and transferred to nitrocellulose membrane. To prepare the radiolabelled probe a fragment of Firefly luciferase was excised from pRF by restriction digestion with *Nco*I and *Ava*I and labelled by random priming in the presence of ^{32}P-labelled dCTP (Figure 5.5A). The probe was hybridised to the membrane overnight at 65°C and results visualised by phosphorimager.

Northern analysis confirmed expression of a band of the expected size for the dicistronic pRMybF transcript (Figure 5.5B). However, in addition two smaller bands were also detected. A band of approximately 1.3kb was present for both pRMybF and pRF. The presence of this band has previously been reported for Northern analysis of pRF and is thought to be an abnormally processed luciferase transcript, with no luciferase function (Coldwell et al., 2000). More importantly, a smaller band of approximately 1.8kb was also detected for the pRMybF sample only. The full-length monocistronic firefly luciferase transcript is 2kb and therefore this band could encode the full firefly luciferase coding region. However, this band is less abundant than the dicistronic band and is unlikely to be responsible for the large increase in firefly luciferase activity observed *in vivo*. In addition a similar sized band has previously been reported for Northern analysis of RNA from

A
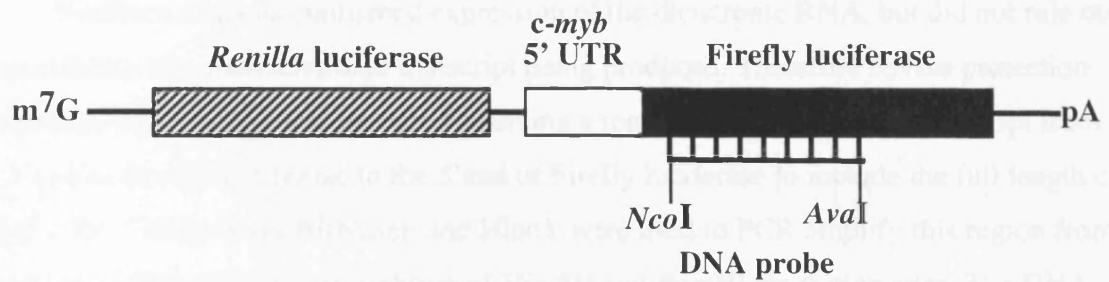
B
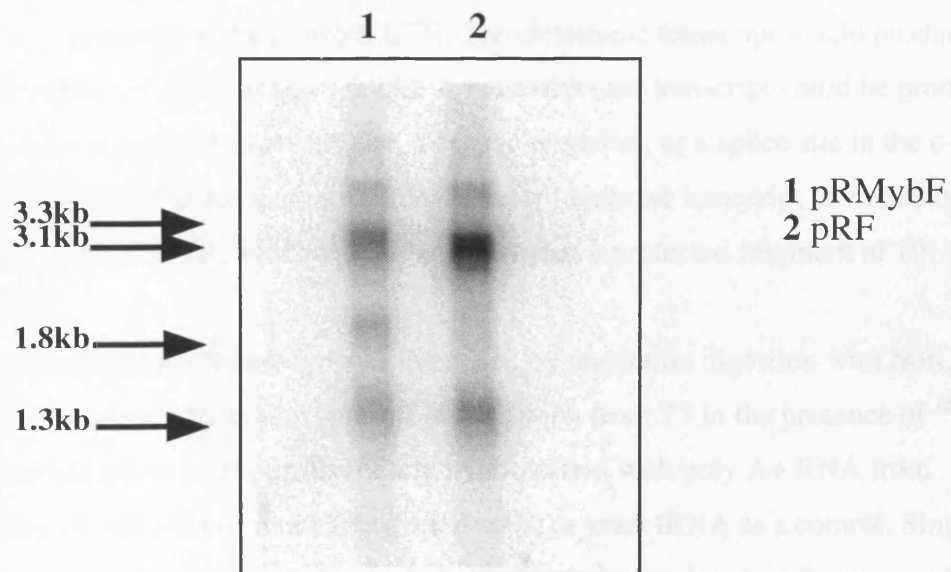
Figure 5.5. Northern Blot analysis to confirm expression of a dicistronic pRMybF transcript. A) Position of Northern probe on dicistronic transcript. B) Results of northern blots on polyA+ RNA from transfected cells. 1)pRMybF 2)pRF.

pRMF, pRBF, and pRNF transfected cells and may represent an artefact (M. Stoneley, M.J. Coldwell, Catherine Jopling-personal communication). Due to the potential for the 1.8kB band encoding the full-length firefly luciferase open-reading frame further analysis of the samples was carried out.

5.6 Confirmation of expression of a dicistronic transcript *in vivo* by RNase protection

Northern analysis confirmed expression of the dicistronic RNA, but did not rule out the possibility of a monocistronic transcript being produced. Therefore RNase protection was performed. A probe was prepared spanning a region of the dicistronic transcript from the 3'end of *Renilla* luciferase to the 5'end of Firefly luciferase to include the full length c-*myb* 5'UTR. The primers BRNaseF and Hluc3' were used to PCR amplify this region from the plasmid pRMybF with the addition of *Hind*III and *Bam*HI restriction sites. The DNA fragment was gel purified and restriction digestion performed, before cloning downstream of the T7 site in the plasmid pSK+Bluescript (Figure 5.6). This produced the plasmid pSKRNase/Myb, which could be used to produce an antisense riboprobe.

The riboprobe hybridises to dicistronic transcripts, and monocistronic transcripts produced by processing in the c-*myb* 5'UTR. The dicistronic transcript would produce a protected fragment of 428nt (Figure 5.7A). A monocistronic transcript could be produced by the presence of an RNA cleavage site, a cryptic promoter, or a splice site in the c-*myb* 5'UTR. This would produce a monocistronic Firefly luciferase transcript preceded by 1-200nt of the c-*myb* 5'UTR, which would be detected as a protected fragment of 101-301nt (Figure 5.7B).

The plasmid pSKRNase/Myb was linearised by restriction digestion with *Not*I, and the riboprobe produced by *in vitro* run-off transcription from T7 in the presence of ^{32}P-CTP. The riboprobe was gel purified before hybridisation with poly A+ RNA from pRMybF transfected cells or mock transfected cells, or yeast tRNA as a control. Single stranded RNA was digested, and reactions purified before samples were fractionated by electrophoresis on a 4% polyacrylamide/ 7M Urea gel. A protected fragment of approximately 428nt corresponding to the size of a dicistronic message was present for the pRMybF transfected sample (Figure 5.8- lane 3). No clear lower band corresponding to the size of a monocistronic transcript was observed. Several faint bands were detected but are unlikely to account for the large increase in Firefly Luciferase activity observed in
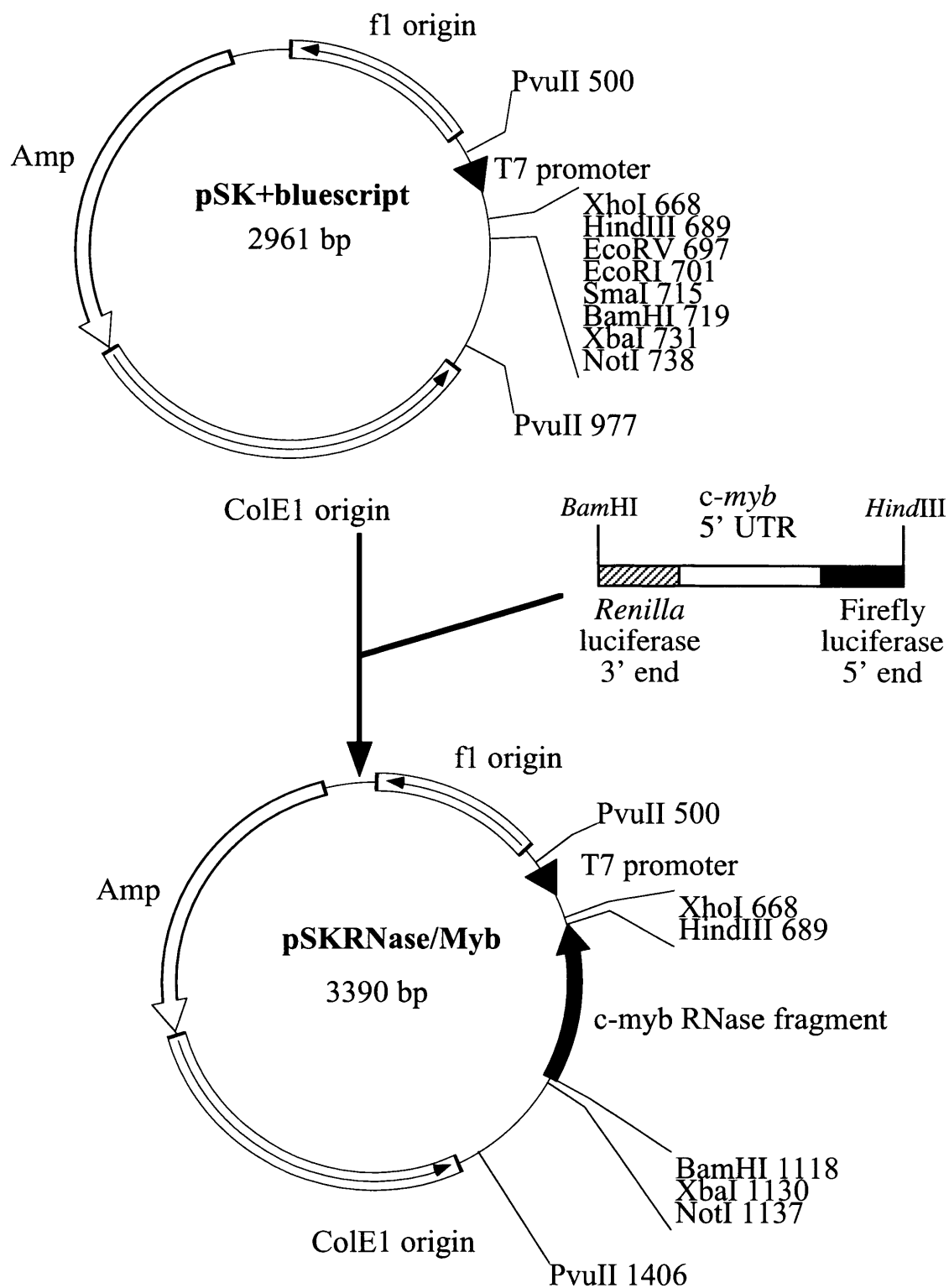
Figure 5.6. Cloning of pSKRNase/Myb. A fragment from the 3'end of *Renilla* luciferase to the 5' end of firefly luciferase and including all the c-*myb* 5'UTR was amplified by PCR and cloned into pSK+Bluescript. The fragment was cloned between *Hind*III and *Bam*HI to enable transcription of an antisense riboprobe.

A Dicistronic message

74nt   200nt   **101nt**

Renilla **luciferase**   c-*myb* 5'UTR   **Firefly luciferase**

m7G   pA

3'   **5'** **Riboprobe 527nt**

428nt protected

B Monocistronic message

101nt

1-200nt   **Firefly luciferase**
c-*myb* 5'UTR

                                         pA

3'   **5' Riboprobe 528nt**

101-301nt protected
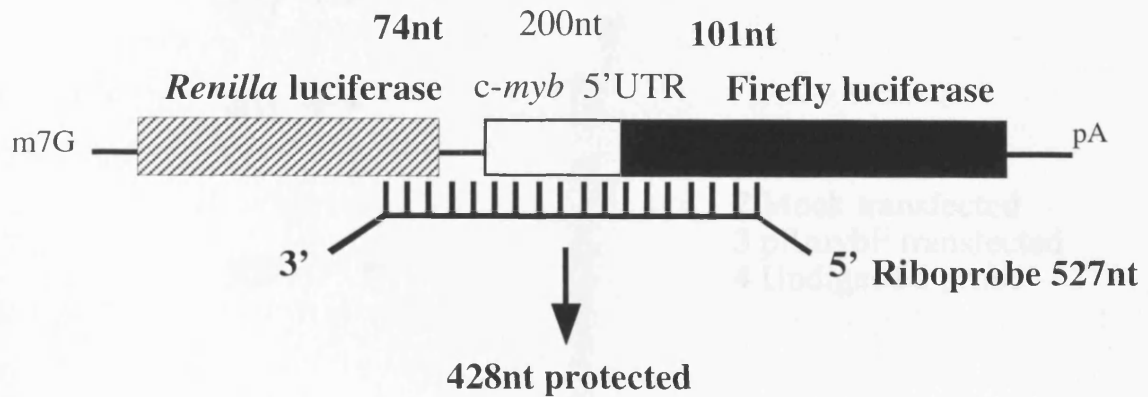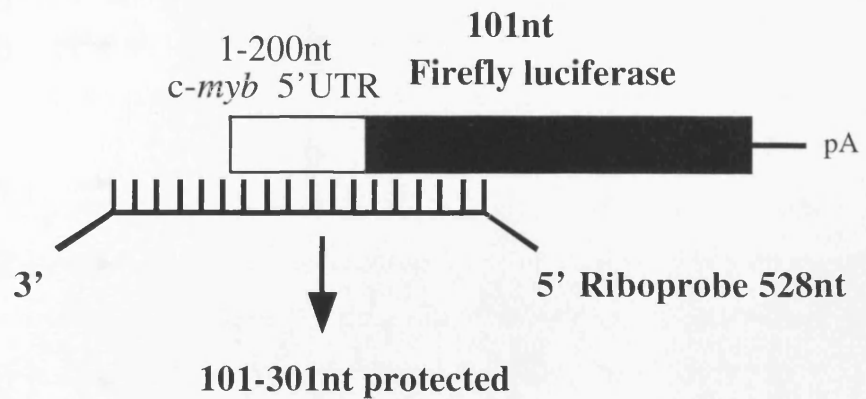
Figure 5.7 RNase protection diagram. RNase protection to confirm the presence of dicistronic message. The full dicistronic message will produce a protected fragment of 428nt (A). A monocistronic message can be produced by the presence of an RNA cleavage site, a cryptic promoter, or a splice site in the c-*myb* 5'UTR. This would produce a protected fragment of 101-301nt (B).

RNA size nt

1 2 3 4

622 →
527 →
404 →
309 →
242 →
238 →
217 →
201
190 →
180
160 →
147 →
122 →
110 →
90 →

1 tRNA
2 Mock transfected
3 pRmybF transfected
4 Undigested probe
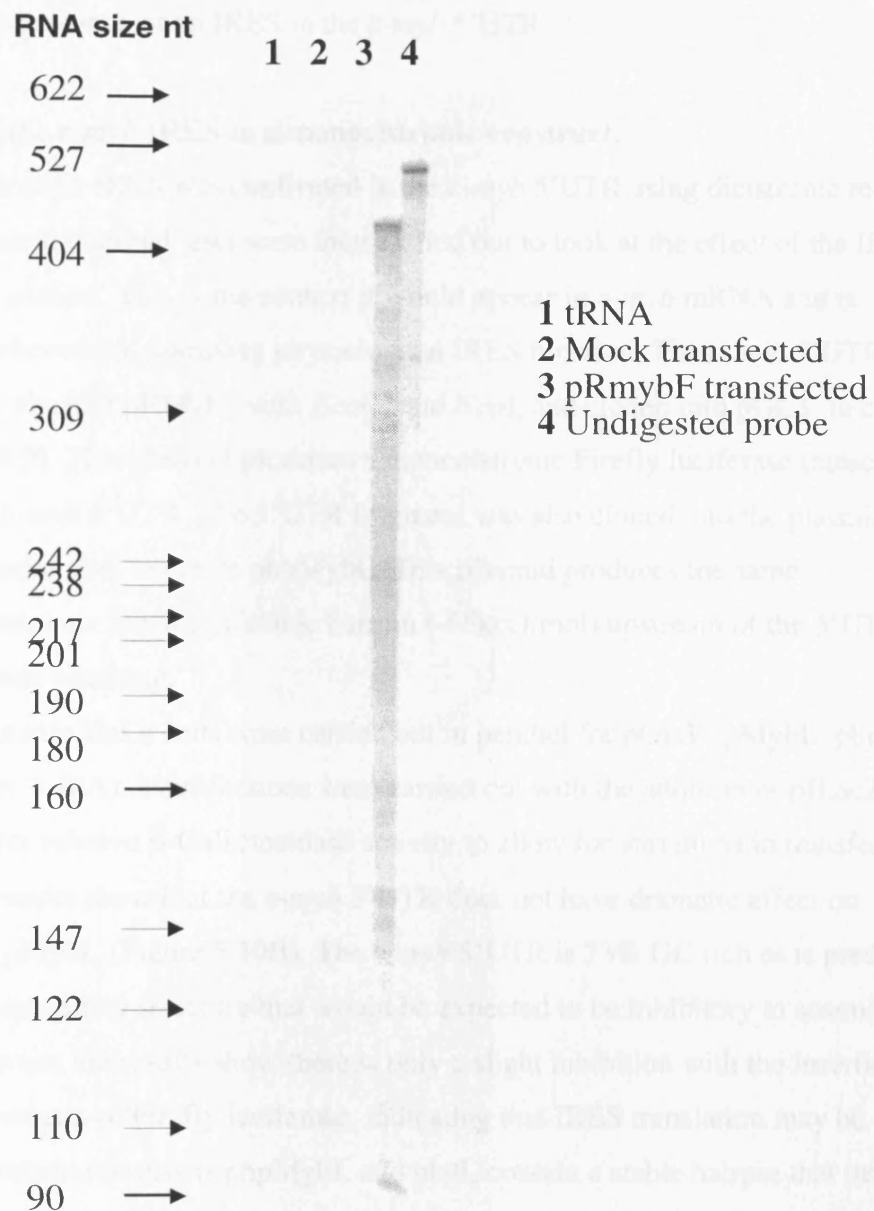
Figure 5.8 RNase protection to confirm the presence of a dicistronic pRMybF transcript. RNase protection of RNA from pRMybF transfected, mock transfected cells, or yeast tRNA. Samples were fractionated by electrophoresis in parallel with pBR322/*Hpa*II markers.

pRMybF transfected cells. Therefore the increase in Firefly luciferase activity observed *in vivo* is due to the presence of an IRES in the c-*myb* 5'UTR.

5.7 Function of the c-*myb* IRES in a monocistronic construct.

The presence of an IRES was confirmed in the c-*myb* 5'UTR using dicistronic reporter constructs. Further functional tests were then carried out to look at the effect of the IRES in a monocistronic context. This is the context it would appear in c-*myb* mRNA and is therefore more relevant for assessing physiological IRES function. The c-*myb* 5'UTR was excised from the plasmid pRMybF with *Eco*RI and *Nco*I, and cloned into pGL3' to create pMybL (Figure 5.9). This plasmid produces a monocistronic Firefly luciferase transcript preceded by the c-*myb* 5'UTR. The 5'UTR fragment was also cloned into the plasmid phpL using the same sites to create phpMybL. This plasmid produces the same monocistronic construct but with a stable hairpin (-55kcal/mol) upstream of the 5'UTR sequence to impede scanning.

Transfections into HeLa cells were carried out in parallel for pGL3', pMybL, phpL and phpMybL (Figure 5.10A). Transfections were carried out with the addition of pJLacZ and results adjusted for relative β-Galactosidase activity to allow for variations in transfection efficiency. The results show that the c-*myb* 5'UTR does not have dramatic effect on translation from pMybL (Figure 5.10B). The c-*myb* 5'UTR is 73% GC rich as is predicted to form a stable secondary structure that would be expected to be inhibitory to scanning of ribosomes. However, the results show there is only a slight inhibition with the insertion of this sequence upstream of Firefly luciferase, indicating that IRES translation may be occurring. The hairpin constructs phpMybL and phpL contain a stable hairpin that impedes scanning. The results show that in the presence of the c-*myb* 5'UTR there is still some translation of Firefly Luciferase, above that seen for the construct phpL. This translation is not due to scanning, and shows that internal ribosome entry is occurring. The observed IRES translation from phpMybL is lower that that seen for pMybL indicating that both scanning and IRES translation occur in a monocistronic context, and that IRES translation may account for approximately 50% of the translation observed. This agrees with results for the c-*myc* IRES, which has also been found to be translated by both cap-dependent and IRES translation (Stoneley et al., 2000b).
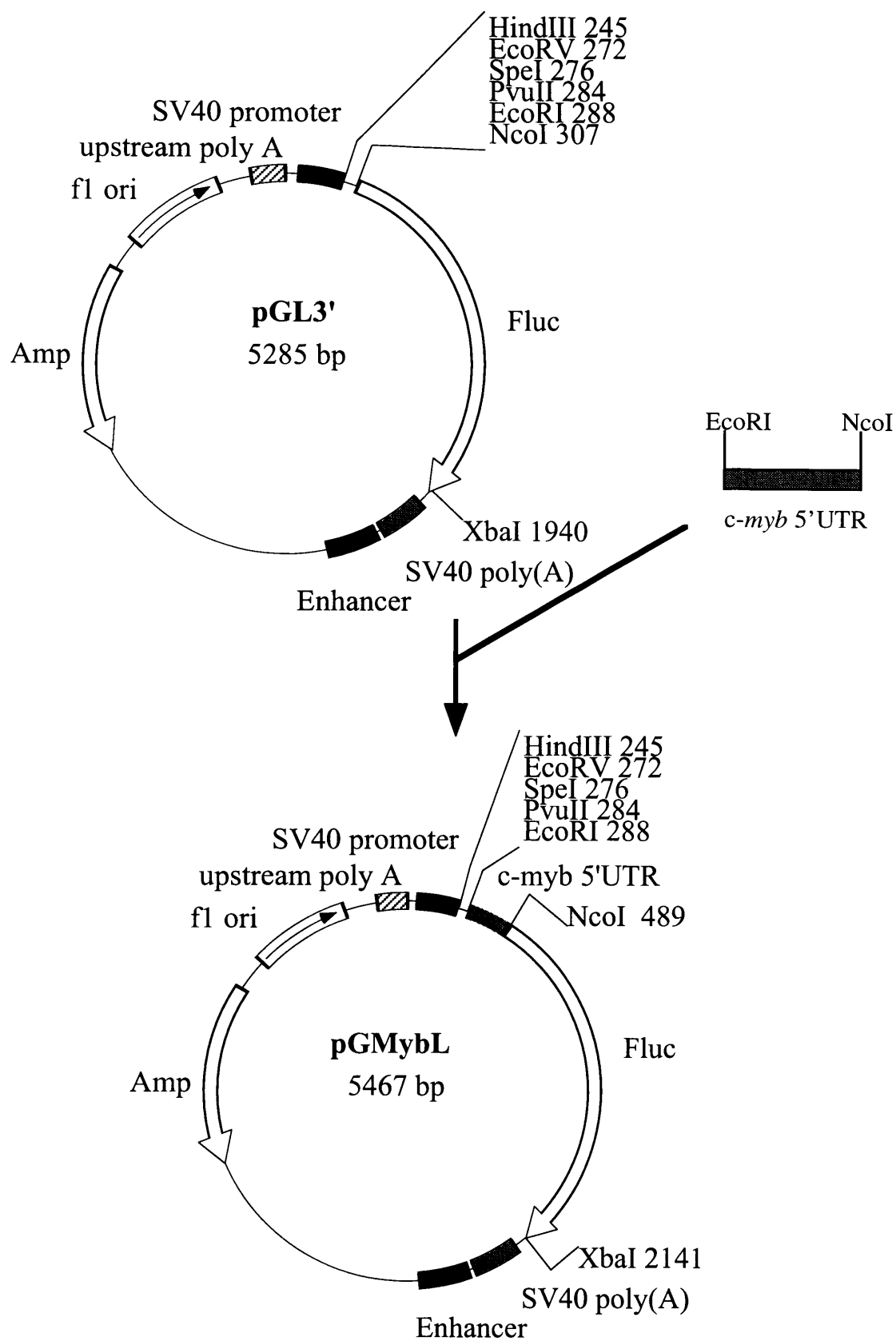
Figure 5.9. Cloning of pGMybL. The c-*myb* 5'UTR was excised from pRmybF using *Eco*RI and *Nco*I and cloned upstream of firefly luciferase in the plasmid pGL3'.
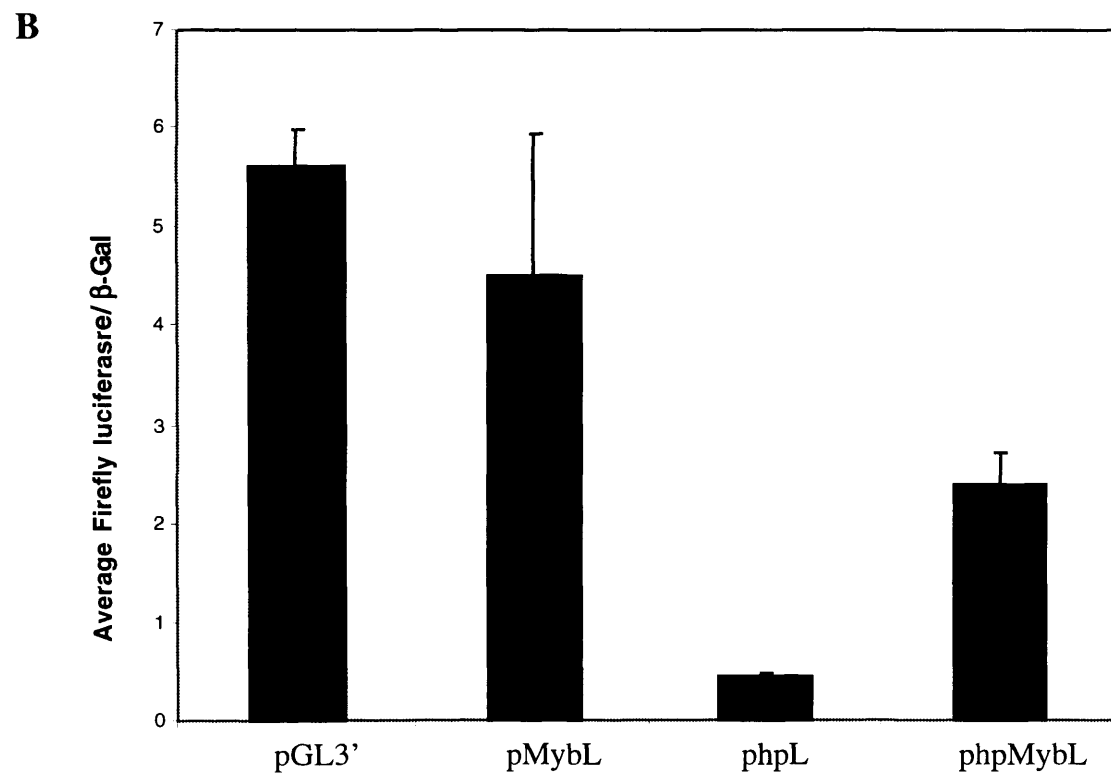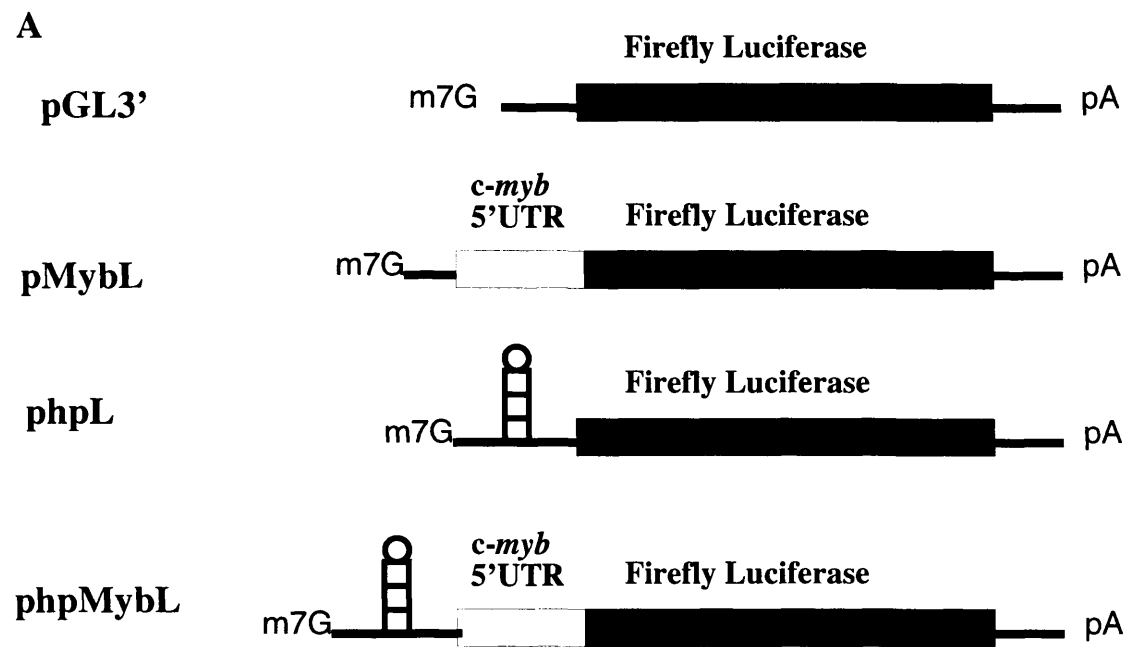
Figure 5.10. Transfection of monocistronic constructs into HeLa cells. A) Diagram of monocistronic constructs produced. B) Transfection of monocistronic constructs into HeLa cells. Results show average Firefly Luciferase Readings normalised to β-Galactosidase activity. Error bars show standard deviation.

5.8 Activity of the c-myb IRES in different cell lines

Cellular IRESs, such as that for c-myc, have been found to have differing activities between cell lines. The c-myc IRES was found to be most active in HeLa cells and MRC5 cells, with decreased activity in a range of other cell lines. This is thought to the presence of varying amounts of required trans-acting factors in these cell lines (Stoneley et al., 2000b). The activity of the c-myb IRES was determined in a range of cell types, including lymphocyte cell lines where the c-myb protein may be expressed (Bender and Kuehl, 1986).

A range of cell types were transfected with the plasmid pRMybF using Fugene-6 or Electroporation. Translation levels of Firefly and *Renilla* luciferase were determined using a dual luciferase assay. All transfections were carried out with the addition of pJLacZ to enable results to be adjusted for changes in transfection efficiency. It was found that the activity of the c-myb IRES varied greatly between cell lines (Figure 5.11). Activity was found to be highest in HeLa cells, followed by MRC5. These cell lines has previously been shown to have high activity for the c-myc IRES. Other cell lines showed low activity for the IRES, including lymphocyte cell lines such as Jurkats and Raji cells. c-myb expression is known to be very limited and is found to be highly expressed only in immature lymphocyte cells. Therefore no cell line with known high expression of c-myb was tested. The c-myb IRES was found to have some activity in porcine (PAE), and bovine (BAE) cells, but have almost no activity in rat (H9c2) or mouse (Balbc/3T3) cells. Overall the c-myb IRES was found to show high activity in HeLa cells and low in other cell lines tested. The IRES activity observed may represent varying amounts of required trans-acting factors present in these cell lines.

5.9 Effect of protein factors on translation from the c-myb IRES.

As with the c-myc IRES, the activity of the c-myb IRES was found to differ greatly between cell lines suggesting a need for protein factors that are present in these cell lines in varying amounts. Therefore the effect of known IRES interacting proteins on c-myb IRES activity was studied, including both IRES trans-acting proteins and canonical factors. Those proteins studied were PCBP1, PCBP2, HnRNPK, UNR, UNRIP, ITAF45, La, DAP5 (p97 and p86) and eIF4G (457-932). All proteins were cloned into the plasmid pcDNA3 for transfection (see section 3.3). Co-transfections of pRMybF and pcDNA3 constructs were carried out in HeLa cells. Translation levels of Firefly and *Renilla*
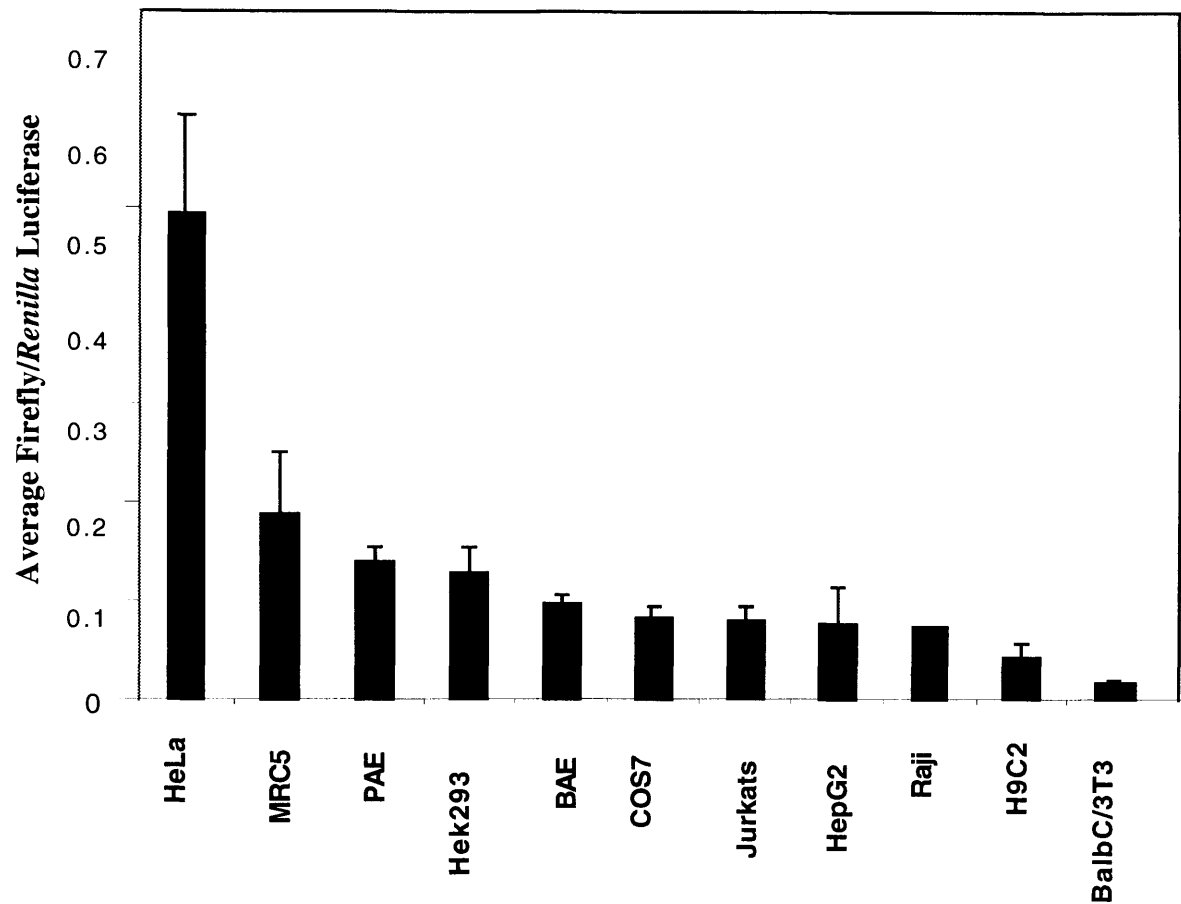
Figure 5.11. Activity of the c-*myb* IRES in different cell lines. Results show average Firefly/*Renilla* Luciferase readings. Error bars show standard deviations.

luciferase were determined using a dual luciferase assay. All transfections were carried out with the addition of pJLacZ to enable results to be adjusted for changes in transfection efficiency.

PCBP2 and ITAF45 were found to slightly stimulate c-*myb* IRES activity to approximately 1.2 fold, and HnRNPK to 1.3 fold (Figure 5.12). The factors eIF4G (457-932) and DAP5 (p97 and p86) were found to inhibit IRES translation slightly. It is interesting that PCBP2 and HnRNPK were found to stimulate c-*myb* IRES activity as these factors have been found to effect c-*myc* IRES activity, suggesting a common mechanism may occur for both genes. However, the activations observed are small and are not statistically significant. Further tests *in vivo* would need to be carried out, along with assessing the direct interaction of these factors with the c-*myb* IRES.

5.10. Discussion.

The c-*myb* 5'UTR was found to contain an IRES in the 200nt upstream of the AUG codon. The presence of an IRES was confirmed by functional tests, showing that the increase in Firefly Luciferase was not due to enhanced ribosomal readthrough, and unlikely to be due to the production of a monocistronic Firefly Luciferase transcript. The c-*myb* IRES was found to enhance Firefly expression to 50 fold over readthrough initiation. This is high activity when compared to other reported cellular IRESs such as Apaf-1, which stimulates Firefly Luciferase 10 fold over readthrough (Coldwell et al., 2000), and Bag-1, which stimulates Firefly Luciferase 17 fold over readthrough (Coldwell et al., 2001). In addition the activity is higher than that observed for the viral IRESs EMCV and HRV (Coldwell et al., 2001).

Further work showed that translation from a monocistronic transcript preceded by the c-*myb* IRES can occur by both cap-dependent scanning and IRES initiation. This would occur under physiological conditions for the c-*myb* mRNA. IRES translation had been found to maintain the expression of genes under cellular conditions when cap-dependent translation is inhibited. For example the c-*myc* IRES maintains expression of the protein during apoptosis (Stoneley et al., 2000a) and following DNA damage (Subkhankulova et al., 2001). Therefore the c-*myb* IRES may be utilised under certain conditions to ensure continued expression of the gene.

Cellular IRESes have been shown to have differing activities between cell types. This was observed for the c-*myb* IRES, where activity varied greatly among cell lines.
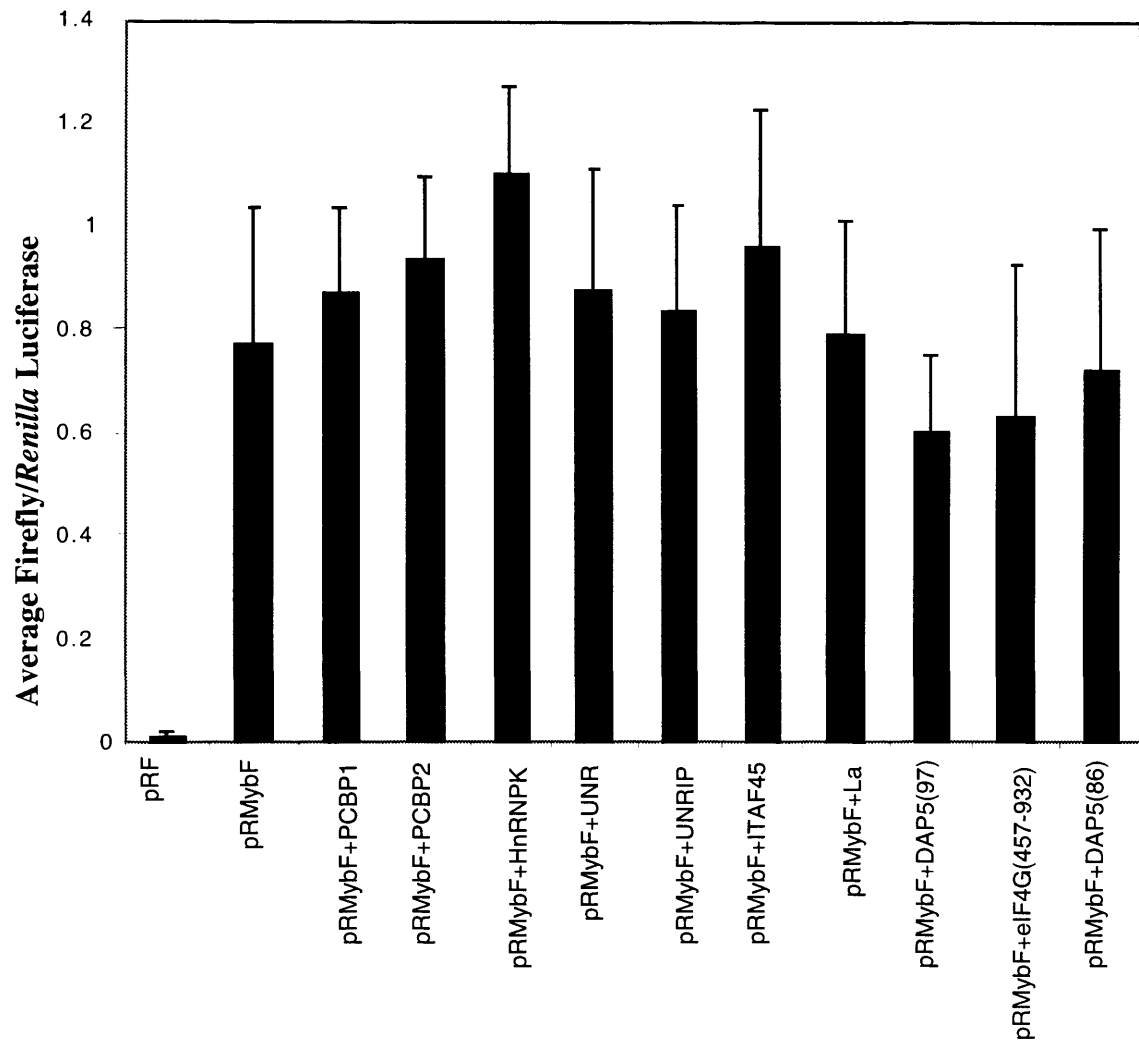
**Figure 5.12. Co-transfections of protein factors and pRMybF into
HeLa cells.** Results show average Firefly/*Renilla* Luciferase readings and
error bars show standard deviation.

Activity was found to be high in HeLa cells. This cell line has been shown previously to have high activity for the c-*myc*, N-*myc* (Jopling and Willis, 2001), and Apaf-1 IRESs suggesting certain cellular factors required for IRES translation may be abundant in this cell type. Activity was found to be low in other cell lines tested. Co-transfection of HeLa cells with protein factors known to effect IRES translation showed slight stimulation of the c-*myb* IRES activity with PCBP2, HNRNPK and ITAF45. Activations observed were low and may not represent significant effects. However, it is interesting to note that PCBP2 and HnRNPK may have an effect on c-*myb* IRES activity as these proteins were found to have an effect on c-*myc* IRES activity, and may represent a common mechanism for the two IRESs.

Chapter 6

Discussion

6.1. Internal initiation from the cellular c-*myc* IRES involves *trans*-acting factors

The majority of investigation into protein factors that are required for IRES-mediated translation has concentrated on the factors required for viral IRESs. It was found that internal ribosome entry required almost the same initiation factors as cap-dependent translation (Pestova et al., 1996a). In addition type 1 picornaviral IRESs, and other viral IRESs such as HCV, were found to require additional *trans*-acting factors for efficient translation. This study has examined the factors required for internal initiation from the cellular c-*myc* IRES. The investigation of protein factors known to be involved in IRES translation identified the proteins PCBP1, PCBP2, HnRNPK, UNR and UNRIP as having a role in c-*myc* IRES translation. UV cross-linking experiments and EMSAs showed the direct interaction of PCBP1, PCBP2, HnRNPK and UNR with the c-*myc* IRES RNA, and competition UV cross-linking experiments showed specificity of the interaction with PCBP1, PCBP2 and HnRNPK.

The use of EMSAs highlighted differences in the interaction of the c-*myc* IRES RNA with these proteins. The shifted position of the bands observed suggested that PCBP1 was binding the RNA as a dimer, and PCBP2 as a monomer. The factor HnRNPK also appeared to be able to bind the RNA as a dimer. These results were not unexpected as these proteins are all members of the poly (C) binding protein family, which have been shown to bind RNA as both homo and heterodimers (Gamarnik and Andino, 1997; Makeyev and Liebhaber, 2002). In addition the proteins were found to bind to different regions of the c-*myc* IRES RNA. PCBP2 appeared to bind in the region up to 98 nucleotides, and PCBP1 and HnRNPK in the region up to 211 nucleotides. All three proteins bind in domain 1 of the c-*myc* IRES that is highly structured and contains a pseudoknot structure (Le Quesne et al., 2001). The proteins are likely to bind in this highly folded region to maintain secondary and tertiary structure of the IRES. The ribosome entry window for the c-*myc* IRES was determined between 177-194 nt and falls within the 211nt that bind PCBP1, PCBP2 and HnRNPK. Therefore this suggests a role of the factors in 40S ribosome recruitment or binding. PCBP1 binds in the region of pseudoknot helix α close to the ribosome entry window, and may act to open up the IRES structure for

ribosome recruitment. In agreement with this model mutations show this region of structure is inhibitory to c-*myc* IRES function (Le Quesne et al., 2001).

All the factors found to interact with the c-*myc* IRES contain multiple RNA binding domains that are likely to contact the RNA at multiple points to maintain the correct RNA structure for efficient internal initiation. The proteins PCBP1, PCBP2, HnRNPK all contain 3 KH RNA binding domains, and UNR contains 5 cold shock RNA binding domains. UNRIP was discovered as a UNR interacting protein and contains WD-40 repeats (Hunt et al., 1999). Proteins with these repeats are likely to be involved in protein complexes, and therefore this protein may be involved in the formation of a protein complex with other *trans*-acting or canonical factors.

Investigation of the proteins effect on c-*myc* IRES activity showed that PCBP1 and PCBP2 stimulated IRES activity *in vivo*. HnRNPK was found to have a slight stimulatory effect *in vivo* although further work is required to prove a direct role in c-*myc* IRES translation. The levels of HnRNPK in the cell lines were not determined as no antibody was available, and therefore the low levels of activation seen may represent the fact that high levels are present in the cell lines tested. UNR or UNRIP were not found to simulate IRES activity *in vivo*, which again may be due to the high levels found in HeLa cells. UNR and UNRIP stimulated IRES activity *in vitro* in combination with PCBP1, and PCBP2 showing that a combination of factors may be required. Many IRESs have been shown to require several factors for activity. PTB has been found to only bind the cellular Apaf-1 IRES in the presence of UNR (Mitchell et al., 2001). Viral IRESs such as EMCV have also shown a requirement for several factors.

Other proteins studied were not found to interact with the c-*myc* IRES. No interaction was identified between the canonical initiation factors eIF4G (457-932) and DAP5 (p97) with the c-*myc* IRES RNA. eIF4G(457-932) may only interact with the IRES in conjunction with other factors, or may require additional domains over and above that required to support picornaviral IRES translation.

The factors PCBP1, PCBP2 and UNR are IRES *trans*-acting factors (ITAFs) known to effect translation from viral IRESs, and UNRIP was identified as a UNR interacting protein in HRV IRES translation. The factor HnRNPK is a member of the poly (C) binding protein family that includes PCBP1 and 2, and has also been found to have role in transcription from the c-*myc* gene (Michelotti et al., 1996). Therefore its potential as a factor involved in the translation of c-*myc* as well as its transcription is of interest.

Internal initiation from c-*myc* has similarities with other cellular IRESs, which have been shown to require ITAFs for efficient translation. For example the XIAP IRES has been found to require La, and the Apaf-1 IRES requires both PTB and UNR for efficient translation (Holcik and Korneluk, 2000; Mitchell et al., 2001). The c-*myc* IRES has also been shown to require a nuclear event for activity, and it is possible that a nuclear protein may be involved in its translation (Stoneley et al., 2000b). The factors PCBP1, PCBP2, and HnRNPK are all known to shuttle between the nucleus and the cytoplasm (Makeyev and Liebhaber, 2002).

6.2 The identification of novel factors that interact with the c-*myc* IRES

A large number of known viral IRES *trans*-acting proteins have been reported, several of which have been shown to have a role in translation from more than one IRES element. However, novel factors have also been discovered such as ITAF45, which interacts with the FMDV IRES (Pilipenko et al., 2000). Therefore the translation of the c-*myc* IRES may involve novel ITAFs. To identify potentially novel ITAFs a yeast three-hybrid library screen was conducted. Many RNA-Protein interactions have been discovered using this system, however the system has not previously been used to identify IRES interacting proteins. The strength of interaction of IRES *trans*-acting factors and IRES elements are not known, and may be weak or require the interactions of several proteins for tight binding.

The library screen identified a high number of positives due to the stringency of the conditions used. The use of 2mM 3-AT to remove background His3 expression allowed the detection of even very weak positives, and further rounds of screening had to be carried out to identify the strongest interacting proteins. This may have eliminated the detection of IRES-ITAF interactions. The screen identified 7 positives including 5 potential RNA binding proteins and 2 DNA binding proteins. These results show that the screen was successful in identifying c-*myc* IRES hybrid RNA interacting proteins. These 7 proteins all interacted strongly to the IRES hybrid RNA, and interacted with it more than with GAPDH hybrid RNA showing specificity of the interaction. It is not known if the proteins identified were general RNA binding proteins or if these factors were involved in c-*myc* IRES translation. It was found that both Fibrillarin and Fibrillarin–like protein interacted with the full-length c-myc IRES by UV cross-linking, and that the Fibrillarin-like protein interacted by EMSA showing a strong interaction was occurring. The protein Fibrillarin-like is of

interest due to the discovery of Fibrillarins role in pre-mRNA stabilisation, and the discovery of another rRNA processing factor, Nucleolin, in IRES translation of HRV and PV. Fibrillarin is also a nuclear protein and as discussed it is possible that a nuclear protein may be involved in translation of the c-*myc* IRES as it requires a nuclear event for activity. No direct role of these proteins in c-*myc* IRES translation was determined, and cloning of the full-length transcripts of these proteins and investigation of their effect on c-*myc* activity is required.

6.3 The c-*myb* IRES

A number of cellular IRESs have been reported, and appear to be present in a group of genes involved in cell regulation and growth. 5'UTRs containing IRESs are highly GC rich and predicted to contain strong secondary structure elements that would impede binding and scanning of ribosomes. The c-*myb* IRES 5'UTR is highly GC rich and when placed into the dicistronic reporter system was found to cause enhanced translation of the downstream cistron. This enhanced translation was found not to be due to increased ribosome read through or reinitiation, and was found to be unlikely to be due to the production of a monocistronic message. Therefore it was concluded that the c-*myb* 5'UTR contained an IRES.

c-*myb* is involved in the proliferation, cell cycle regulation and differentiation of haematopoietic cells. Like c-*myc,* its expression and activity in the cell is highly controlled. The c-*myb* 5'UTR is translated by both cap-dependent and cap-independent translation. This agrees with the results found for c-*myc* and N-*myc* 5'UTRs (Jopling and Willis, 2001; Stoneley et al., 2000b). Cellular IRESs have been found to maintain expression of the proteins in situations of reduced cap-dependent translation. The c-*myc* 5'UTR was found be utilized in situations of reduced cap-dependent translation such as in apoptosis, and following DNA damage (Stoneley et al., 2000a; Subkhankulova et al., 2001). Therefore the c-*myb* IRES is likely to maintain translation of the protein in situations of reduced cap-dependent translation where the protein is required. c-Myb is required in the cell cycle progression in late G1 or early S phase, and in cell proliferation. The protein is also required in embryogenesis, as the homozygous knockout was found to result in embryonic lethality (Mucenski et al., 1991). The c-*myb* IRES may therefore have a role in developmental control of translation. The IRESs of c-*myc* and FGF2 were found to be involved in developmental control (Creancier et al., 2001; Creancier et al., 2000).

The expression of c-*myb* is mainly limited to immature haematopoietic cells. The investigation of IRES activity between cell lines showed high activity in HeLa cells and low activity in other cell lines tested. This low activity observed may be due to the low expression of c-*myb* in these cell lines and higher activity may be found in cell lines with higher RNA levels. This has been found with the N-*myc* IRES, which has high activity in neuronal cell lines that express N-Myc (Jopling and Willis, 2001). The differences observed in the c-*myb* IRES activity between cell lines suggests that *trans*-acting factors required for efficient translation are present in these cell lines in varying amounts. The activity was found to be high in HeLa cells, which also show high activity for other cellular IRESs such as c-*myc*, suggesting that protein factors required for internal initiation are abundant in this cell line. The study of the effect of known protein factors on c-*myb* IRES translation showed a slight stimulatory effect for PCBP2, ITAF45 and HnRNPK. The proteins PCBP2 and HnRNPK were also found to effect c-*myc* IRES translation and may represent *trans*-acting factors common to the two genes. Further investigation of these proteins would determine if a common mechanism is occurring.

6.4 The protein factor requirements of cellular IRESs.

This work has shown that internal initiation from the cellular c-*myc* IRES requires specific *trans*-acting factors for efficient translation. In addition the differing activity of the c-*myb* IRES between cell lines suggests a requirement for *trans*-acting factors, and the factors PCBP2 and HnRNPK, along with ITAF45 may have a role in it translation. These *trans*-acting factors are likely to maintain the correct secondary and tertiary structure of the IRESs for 40S ribosomal subunit recruitment and binding. Further investigation of the exact binding regions for these proteins on the RNA and the interactions between them would help to elucidate the mechanism involved. Other cellular IRESs such as Apaf-1, XIAP, c-sis, and BiP have all been shown to require specific *trans*-acting factors for efficient internal initiation. The *trans*-acting proteins are mainly proteins previously shown to be involved in the internal initiation from viral IRESs, suggesting the viruses have utilized proteins already present in the cell with this role. Cellular IRESs are likely to function by a common mechanism using a number of these proteins and canonical initiation factors for maintenance of the correct structure for efficient recruitment of the 40S ribosomal subunit.

Appendix

A) The effect of protein factors on c-*myc* IRES activity *in vivo*.

Extra data for figure 3.12 showing separate firefly and renilla luciferase readings, and standard deviations.

	Average Renilla Luciferase	Average Firefly Luciferaese	Standard Deviation Renilla	Standard Deviation Firefly
pRF	0.19031434	0.00401177	0.01467047	0.00258067
pRMF	0.26239073	0.59028941	0.10687742	0.1935948
pRMF+UNR	0.23576515	0.49164734	0.11932819	0.20120659
pRMF+ITAF	0.18932749	0.43298851	0.10900464	0.23471989
PRMF+PCBP1	0.21799551	0.8687426	0.07307328	0.31612782
pRMF+PCBP2	0.16097369	0.50761298	0.08175312	0.23293818
pRMF+LA	0.66458893	0.9068956	0.64244663	0.63453787
pRMF+4G	0.25426047	0.51321442	0.07448822	0.06368825
pRMF+DAP5	0.20471931	0.32466318	0.11115385	0.09279597
pRMF+HnRNPK	0.31197656	0.7848422	0.12631233	0.25132244

B) Sequences of positive clones from yeast three-hybrid screen, and Blast alignments showing closest match.

Sequence of positive clones:

Clone 8 sequence:
GAGCTGCGTCCGCCACTCATGTTTCTCCGAGCAGGCCTGGCCGCGCTCTCCCCG
CTTCTTCGCAGTCTTCGGCCCTCTCCTGTCGCCGCCATGAGCACTGGCACCTTC
GTCGTGTCGCAGCCGCTCAATTACCGCGGCGGGGCCCGCGTGGAGCCGGCGGA
CGCCTCCGGTACCGAGAAAGCTTTCAGCCAGCAACCGGCCGAGTGATAGCTAC
TTTCACATGTTCAGGAGAAAAGGAAGTAAATTTGGCTGTTCAAAATGCAAAGG
CTGCTTTTAAAATATGGAGTCAAAAATCTGGCATGGAGCGTTGCCGAATCCTTT
TGGAGGCTGCCAGGATAATAAGGGAACGGGAGGATGAAATTGCTACTATGGA
GTGCATCAACAATGGCAAGTCATCTTTGAGGCCGCTTGGACATCGACATTTCCT
GGCAGTCCTGGAGTATTATGCGGGCTTGGCTGCATTCATGGCTGGTGAACACA
TCACTCCCAGGTGGATCGTT

Clone 24 sequence: see below

Clone 13 sequence:

CCGGGGCCCGGCGTCCCCCGAGTGGCGGCGGGCCTTTGGCGGCGACGCCGGCG
CTTCGGCGGCGGCGGCGGGTTGGCGGCGGCCACCCGGCGGGAGTCACCTTTCC
CGTCGGGAGCGCGGGCCCGGGGCAAGGGGACCCCGGCCACGGAGAGCGGGAA
GAGGATGGATTGCCCCGCCCTTTCCCCCCGGATGGAAAGAAGGAGGAAGTGAT
CCCGAAAATTTTGCCTTAAGTCTGGCAAGAGCCATGTCTTACTTACTTCAGTCC
AAGTGGTAAAGAGGTTAAAAAGCAAACCTAATTTGGCAAAGGCCCTGGAAAA
TACTTTTTGGATCTTAGCAGGTTTGACTTTAGAACTTGGGAAAGAGATGCCTTG
TAATTACGGAAGGAACAAACAGGGACTGCGAAACGATCTTAATAAAATAAGG
GTAAACAAGACTTGGATACAAACATTTGCAATTAGGACAAACAGGCATAATTT
TTAAACAACCGGACCAAAGTCACA

Clone 93 sequence:

CATCTTCATGGGAAATTTGCTGAGCCCCCACCAGGGAACCGGATGATGAAACA
GGGATACCTCACAGCTTGGCCATTTGAGGCAAAGGCAGCTTCCCGAGCTGATG
CTAAAGAAGACAGACTTTCCCTTCCTCCCAGCAGCAGCAGTGCAGAGCCCGCC
TGGAGGGATGTGGGGGCTGTGCAGGGTGCAGCGCTCAGGTGGATCCTGGGAA
GCAGCCTCTGGATGCTGAGTGGAGGGAGCCACTGAGCACAGCAAGGCACCAA
AGCCCCTGGAGAAACCGCAGGGCGAGGTGCGACCATCATCAGGATCAAAGCA
GACGGGGCGTGGGTGGGGAAGGGGCTCTGGGACCAGACCCCCCACACTACTG
CGTCTTTGTTTCTATCAGTCTTTGTAGAAGCAGGTGGTGGTGGAAATTCCAGCA
GGTGGGTCCGCAAAGGCCCTGAGGCCTCACTTTTCGGATCTTCTGTCCCAGATC
TGCTCCCTTCCTGCTGAGCCTGGGG

Clone 33 sequence: see below

Clone 21 sequence:

GAGGAGCTGTGCCTTCCACCTCTCCAGCCCCGGCAGGACGGGGGCGGCCGCGC
GAACCCGGGGCGGGGACAGCACGCAGCCTCGAGGCGCGCACCCCCGCCCGGC
AGCGGCCCCGACACCCGGGGCGAGCGGGAAAGCGGCAGCGGCGGCGGCGGCG
GCGGCGGCGGGGGAAGGATGCAGGGGAAGAAGCCGGGCGGTTCGTCGGGCGG
CGGCCGGAGCGGCGAGCTGCAGGGGGACGAGGCGCAGAGGAACAAGAAAAA
GAAAAAGAAGGTGTCCTGCTTTTCAACATCAAGATCTTCCTGGTGTCCGAGTG
CGCCCTGATGCTGGCGCAAGGGCACGGTGGGCCGCCTACCTGGTGAGCGTCCT
GACCACCCTGGAGCGTAGGTTCAACCTGCAGAGCGCTGACGTGGGTGTGATCG
CTAGCAGCTTCGAGATCGGGAACCTTGGCGCTCATCCTCTTCGTGAGCTACTTC
GGGGCACGCGGGCACCGGCCGCGCCTG

Clone 83 sequence:

CGGGAGTGCGCCAGCGCCGCGCGTGGGGCTGTGGTGGCCCGCGGCTCTCAGAT
ATATTTTTGCCATCATGGATCAGTTTGGAGATATATTAGAAGGTGAAGTGGAC
CATTCTTTCTTTGACAGTGACTTTGAAGAAGGAAAGAAATGTGAAACTAACTC
AGTTTTTGACAAGCAAATGATGACCCAAAGGAAAGAATAGATAAGATACAAA
AATGTAAATTCGAACACTGGAATGCAAACAACAGAAAATTATCTTACTGAGAA
GGGAAATGAAAGAAACGTGAAATTTCCCCCAGAACACCCCGTGAGAATGATG

TTACACAAACTGTAAGTTCTTTCTCATTGGCAGCCTCTTCAAGATCAAAAAAT
TGTGTGATGTTTCAACAGGACTTAAAATACACGTGTCCATTCCAAATAGAATTC
CCAAAATTGTAAAAGAAGGTGAAGATGATTACTACACAATGGAGAGGAAAGC
GTGATGATGGGAAAGAAATACCAT

Clone 36 sequence: see below

Clone 69 sequence:

CTTGACTGGGGCTGGAAGATCTGCTTCCAGCTCACATGGATGTTGACCAGAGA
CCTTAGCTCCTTGCCACATGGGTTCCTCCAGTGACCTCATGACAAAGCAGCTGG
CTTCCCCCAGAGCAAGTGATCCAAGAGAGAGACCAAGGAAGAGTCTGCAGTG
CCTTTTATGACCCAGTCGTAGGTCACACCTCTCACTTCCACCATATTCTCCTTAT
TAGATCAGAGTCACTAATCCAGCCCATACTCAGGGGAGGGGCACTGAGCTCCA
CTTCTTGAGAAAGGACTAGCAAAGAATCTAGGGGCATATGTTAAAACCGCCTG
GCCTCACCAAAAGTTAATGTTGAAGAAAGTGACGGTTTTCCAGGAACCGGTGA
AAGGAACCAAGGATGTTGGGATTCGGGGAATTCAGTGATCCACACTCTAGTGA
GGGGAGCCATGGTTATCAAGAGGAGGTATTGGGCACATGTTTCCAAGTTTCTG
AGGTTGGTTTCTGGGCATGGG

Clone 18 sequence:

AGGAGGAGGAGGAGGAGGAACAGGAGGAGGAGGAGGAACAGGAGGAGGAG
GAGGAGGAGGCAGCGGCGGCACTGTGGGTCTTCCACAGGGTGTGCCCGGAGG
AGGTCAAGACTCCACTGCCCTGAAGCAAACAAGGCGTCCAGCCCGAGCCAGG
GTCACACGCCCAACAGGTAGGATCCAGCAACCCCTGCGCAGCAGCCCCCGCCA
TGCGAGGGAGGACCCGGAGTTGAGCGTGGCCCTGCCTGACATCTGACCCCGCC
GTCAGGAGCTCCACCAGGTCAGGGCTGCTCAGCTGACCCCACAGCCCCACCAG
GCATGGTCAGCACGGTGCCGAATGCAGGCCTGGAAAGGGAATAAAGGAGGCC
CTGGGGAGGGGCCTTGGTGCTTCATCCCAAGCCAACCCCGAAAGTGGAAGCCA
CAGGGCCAGGGGGACGTCGTGGCTGAGATGCGAGCAGGGTGCCGGGCACCCC
CAAGGCAAGCACTGCTGTCCCTGAGAACAG

Clone 48 sequence:

AAAAATTGTACTGTGTACTGTGGAGGAATTGCGTCTGGGTTAACAGATCAGCT
TATGAGACAGACATTCTCACCATTTGGACAAATTATGGAAATAAGAGTTTTCC
CAGAAAAGGGCTATTCATTTGTCAGATTTTCAACCCATGAAAGTGCAGCCCAT
GCCATTGTTTCGGTGAACGGTACTACGATTGAAGGACATGTGGTTAAATGCTA
TTGGGGTAAAGAATCTCCTGATATGACTAAAAACTTCCAACAGGTAATTCGAT
TTTTCATAGCATTCTTTAAGGTTTCCATCTTACATGTCACATAAAAGCTTTGGA
AACTCTGTAAAATGAACAATAAAATAAAGCATATAGCTACTTTCAGTTGGATT
GGATTTTCAAAATTGATTATTTGCGGTATTAACTGAATCTTAATACTTTCTTTTC
ACAGGTGTCACCACCCCAATAAACTTAGACAATGATAAATAACAGAGTCCTAT
TCACATAAGGGGTGCTTACT

Clone 100 sequence: see below

Clone 4 sequence:

CTGGAGCCGCTGTGGTTGCTGTCCGCGGAGTGGAAGCGCGTGCTTTTGTTTGTG
TCCCTGGCCATGGCGCTGCAGCTCTCCCGGGAGCAGGGAATCACCCTGCGCGG
GAGCGCCGAAATCGTGGCCGAGTTCTTCTCATTCGGCATCAACAGCATTTTATA
TCAGCGTGGCATATATCCATCTGAAACCTTTACTCGAGTGCAGAAATACGGAC
TCACCTTGCTTGTAACTACTGATCTTGAGCTCATAAAATACCTAAATAATGTGG
TGGAACAACTGAAAGATTGGTTATACAAAGTGTTCAGTTCAGAAACTGGTTGT
AGTTATCTCAAATATTGAAAGTGGTGAGGTCCTGGAAAGATGGCAGTTTGATA
TTGAGTGTGACAAGACTGCAAAAGATGACAGTGCACCCAGAGAAAAGTCTCA
GAAAGCTATCCAGGATGAAATCCGTTCAGTGATCAGACAGATCACAAGCTACC
GGTGACATTCTGCACTGTTG

Clone 16 sequence:

AATCTTTTTTTTTTTGCCGTCACTATGCTCTCACCCCTCATTTGTTGACCAATT
TTTTTCTTGGTATCTATGCTGTAACGTCCTTCCCATTTCTAGTCTGGCAGTCATC
GCATCAGCATTGCTGCTCAGTTCAACCCGAGCCCCCCCTCTCAACATTACTCCA
ACTTGAGCCAACGAGAGTGAAGGTCGGCAGTTCCGGAACTTTTTGCCCACTGG
CTGGAATGACATGGAGGTCATCCCCAGAACTCATGAACATTTCAATAGGTTGT
GACACGACACCTGATCTGAAGACTGATGGTTTGGCATTTGACACATGTCGCAC
ATGGTGGCCCGTGATGGATAGCGACACCACAGGCAAGCTGGGCTTTGAGGAAT
TCAAGTACTTGTGGAACAACATAAAGGTGGCAGGCCTATACAAACAATTCGA
CACTTGACCGATCAGGGACCAATTTGCAGTAGTGACTCCCAGGTGCCTTTG
AGGCAGGGGTTCCACCTGGA

Clone 72 sequence:

CCGGCGCTGCTCCTCCCTCCGCCCAGCTCGGCTCACTTACTCTCCCCTTGCTAG
CTGGCTTCGGCGGGACCCGGCTGGCATCACCGCAATCCGCCGCCGTTGCGGAT
GAGGCCTGCGCGCCTGAGAAGCTGGCGGCAGCAAAAGGAAGTCTTCAAAGTG
GAAAGGTCGTCCCGTTGGACGACGCGCCTCCTTGCTTTTCCTTCCAAATTATGG
CGGGCTAAAAAGGGCTTTCCCCCGTCACTGCAAGATTCCAAGCAAAGGACCCT
GTTCAAAATTGAGAAAGATCCCTGCGGCGGATAAAACCCCTGACCTGGGGTT
CCAGGGCAAAGCTCTACTGGGCCCATGCTTTAGGCAGGCAAGGGTGTAGGGTT
GGCTTTTGAAGGACCGTCCAAGCTTACACGGTGTATCTGCAGGAGAGCCAACT
CTTGGTGGGTTAAAGGCATCCCGTGCAAAACCAGGCAAAAAACGTTACAAAT
AGAAACGGAAACAACTTGGGAA

Clone 19 sequence: see below

Clone 23 sequence: see below

Clone 37 sequence:

ACCCTGGTAAAGAAAAACCAAAGGAAGAGCGGCCAAATAAAGATGAAAAAG
CAAGAGAGACCAAGGAAAGAACGCCGAAGTCTGACAAAGAGAAAGAAAAT
TCAAGAAGGAAGAAAAAGCTAAAGATGAGAAATTTAAGACCACTGTCCCCAA
CGCAGAATCAAAATCAACTCAAGAAAGGGAAAGAGAGAAGGAGCCATCCAGA
GAAAGAGATATAGCAAAGGAAATGAAATCAAAGGAAAATGTTAAAGGAGGA

GAAAAAACACCAGTTTCTGGGTCCTTGAAATCACCTGTTCCCAGATCAGATAT
TCCAGAGCCTGAAAGGGAACAAAAACGCCGCAAAATTGATACTCACCCTTCTC
CATCACATTCCTCCACAGTAAAGGTTACAGCCATACTTCCAAAGTTCCTCTGGG
TTCTGAGAACTATGCCAGCTCACCTGTCATCTCCATTCATTTTCTACAGGACAG
TCTCATCGAACTCAAGGAATCTTCAGCAA

Clone 38 sequence: see below

Clone 43 sequence:

GTCGACGCGGCCGCGTCGTTTGAGAGGACAATGTTGCTAATAAACAGACTCTT
GCAAGTTATAGGAGCAAAGCTACTAAAGAAAGAGATAAACTTTTGAAACAAG
AAGAAATGAAGTCACTGGCTTTTGAAAGGCTAAATTAAAAAGAGAAAAAGC
AGATGCCCTAGAAGCGAAGAAAAAGAAAAAGAAGATAAAGAGAAAAAGAG
GGAAGAATTGAAAAAATTGTTGAAGAAGAGAGACTAAAGAAAAAAGAAGA
AAAAGAGAGGCTTAAAGTAGAAAGAGAAAAGGAAAGAGAGAAGTTACGTGA
AGAAAAGCGAAAGTATGTGGAATACTTAAAACAGTGGAGTAAACCTAGAGAA
GATATGGAATGTGATGACCTTAATGCATTTGGGGAACTTTTTGATCTTCAAGAT
GAGTTTCCTGATGGAGTAACCCTAGAAGTATTAGAGGAAGCTCTTGTAGGAAA
TGACCAGTGAAGGCCCACTGTGTGAATTGCT

Clone 51 sequence:

CAGGAACCCGGCGATATGGCTGCCGCTGTGCCCCGCGCCGCATTTCTCTCCCC
GCTGCTTCCCCTTCTCCTGGGCTTCCTGCTCCTCTCCGCTCCGCATGGCGGCAG
CGGCCTGCACACCAAGGGCGCCCTTCCCCTGGATACGGTCACTTTCTACAAGG
TCATTCCCAAAAGCAAGTTCGTCTTGGTGAAGTTCGACACCCAGTACCCCTAC
GGTGAGAAGCAGGATGAGTTCAAGCGTCTTGCTGAAAACTCGGCTTCCAGCGA
TGATCTCTTGGTGGCAGAGGTGGGGATCTCAGATTATGGTGACAAGCTGAACA
TGGAGCTGAGTGAGAAATACAAGCTGGACAAAGAGAGCTACCCAGTCTTTTAC
CTCTTTCGGGATGGGGACTTTGAGAACCCAGTCCCATACACTGGGGCAGTTAA
GGGTGGAGCCATCCAGCGCTGGCTGAAAGGGGCAAGGGGTCTACCTAGGTAT
GCCTGGGTGCCTGCCTGTATACG

Clone 56 sequence:

CATGGACGGAAGCAGCAGGAAGCATTCGGGATGGAGTGCGTGCCTATACAGC
TCTGCATTATCTTTCTCATCTCTCTCCTGGAAAATCAGTGCTGATAATGGATGG
AGCAAGTGCATTTGGTACAATAGCTATTCAGTTAGCACATCATAGAGGAGCCA
AAGTGATTTCAACAGCATGCAGCCTTGAAGATAAGCAGTGCCTTGAAAGATTC
AGACCTCCCATAGAGAACAGTACAATTGCCTGATTTCCTGTCCTTTACAGCCCG
AGTGATTGATGTATCTAATGGGAAAGTTCATGTTGCTGAAAGCTGTTTTGGAA
GAAACAGGTGGCCTGGGAGTAGATATTGTCCTAGATGCTGGAGTGAGATTATA
TAGTAAAGATGATGAACCAGCTGTAAAACTCAACTACTACCACATAAACATGA
TATCATCACACTTCTTGGTGTTGGAGGCCACTGGGTACAACAGAAGAAAACCT
TCAGTTGGATCCTCCAGATAGC

Clone 58 sequence:

GAGGAGGCGGCGGCGGCGGGAGCGGTCCCCAGGAATGTCGCTGCCGCCGCCA
CCGCCGGGGCCGCTGCCGTTGAGGAGGAGACGGAGGAGACCGACGTTGTTAG
GAAGATGATCCCTATGATCTTGAAGATGTTTCTGCACAGAAATGAGGGAAATA
CAAAGAACCAAATACAGTTCTGAAATTTGGGATCTGTATTTTGAGATGATTTTA
TTTTCAGAATGAGAAGCATATCTGGTTACCTTTATGAATGTAGAGACATGAGA
AGAGAGTTATGATGGCAAAAAACAAAGAGCCTCGTCCCCCATCCTATACCATC
AGTATAGTTGGACTCTCTGGGACTGAAAAGACAAAGGTAACTGTGGAGTTGG
AAAGTCTTGTTGTGCAATAGATTTTGTACGCTCAAAAGCAGATGAATATTATCC
AGAGCATACTTCTGTGCTTAGCACCATTGACTTTGGAGGACGAGTAGTAAACA
ATGATCACTTTTTGTCTGGGGTG

Clone 77 sequence:

CGCGCCCCAGCTTTCGTCCCGCTGTGCGGAGATGCCACCCAGCTCCACCGCAG
GACCGCGGAGCGAGAGGCGGCAGACGCGCTCCGGGGGCCGCCGGTCAAGAGG
AAGATGAAGGGCGCGACGACTTCGCGGTTTCCAGAATGGGCTCCTTCTGGCCC
AACAATAAGTGCCATTTGGAATCTTCTCGGCCTTCTCTGTGCCTTCCGCGCGTC
TGGGCTCCCCGGAGAGAAAGGAAAGCGCGAGGAGCCGCCGCCACCACCAGCG
CAGCAGTCCTGGAGCTGTGAGGAGATTCGGGCCGTCACCCTGCCTCCCCTGCG
TCCCGCACCGGCCGCTTCTGTCCTCGGACCCATTCCAACAATCTCGTAAACAT
GGTGGATTACTATGAAGTTCTAGGCGTGCAGAGACATGCCTCACCCGAGGATA
TTAAAAAGGCATATCGGAAACTGGCACTGAAGTGGCATCCAGATAAAAATCCT
GAGAATAAAGAAAGAAGCAGAG

Clone 81 sequence:

CGCGGCCGCGTCGACTTTTCTAGTAATAATTTTAAAATAGCATTACTTTTTATT
ATCACAATATTAAAAAATAGGAAAGTAATTGTGATGTCAAAATACTCAACACA
TGGTATTTTTATTTAATTAAAATATGTATTTTGGGGAATAGCTGATGATTATTA
CTATTTATCCTCCTCCTTACCAATATACAAATATAATTTTTTAGCCTTCAATAAT
ATAATAAGACTATATTAATTCCTTAATTTAGAAAGAGGCTATTTCCTTTTCTAG
GAAAAGGGCATTGAAGTGATGTTTCTCCCCACCCTTGCCCTACTTGGACTCTT
GATTGGAATGTTCTCCCTGCAGCAGTCAGCCCAAAGGAAGGTGAGTTTTCTTTT
TTCTTTAAGTTTTATTCATACTTTGGATACTGCTTCAGATATGTCTTCTAATTCT
CTAAATCTCCCTTTGGCTTTATCTAATATTCTGATTATACTATCAGTTGAGTTTT
CATTTTATTTTT

Clone 84 sequence:

AGAGCGCTGTCAAGAAGACCCACACGCCCCCCTCCAGCAGCTGAATTCCTGCA
GCTCAGCAGCCGCCGCCAGAGCAGGACGAACCGCCAATCGCAAGGCACCTCT
GAGAACTTCAGGATGCAGATGTCTCCAGCCCTCACCTGCCTAGTCCTGGGCCT
GGCCCTTGTCTTTGGTGAAGGGTCTGCTGTGCACCATCCCCCATCCTACGTGGC
CCACCTGGCCTCAGACTTCGGGGTGAGGGTGTTTCAGCAGGTGGCGCAGGCCT
CAAGGACCGCAACGTGGTTTTCTCACCCTATGGGGTGGCCTCGGTGTTGGCCA
TGCTCCAGCTGACAACAGGAGGAGAAACCCAGCAGCAGATTCAAGCAGCTAT
GGGATTCAAGATTGATGACAAGGGCATGGCCCCCGCCCTCGGCATCTGTACAA

GGAGCTCATGGGGCCATGGAACAAGGATGAGATCAGCACCACAGACGCGATC
TTTGTCAGCGGGATCTGAAGCTGGT

Clone 86 sequence:

CCTCGGGACTATTCAAAAGACAGCTTTCCAGAGAATGTTTCAAGAAAGTTACG
TGGAGCGGGGCGTTTCGCAACTCCTAAGGGTCTGGAACCTAACCTGCAGTGTC
TTGAGCTTCTGGCAGTTGAGGCAGGATTTCACAAAAACTTGGAGCAAAAATTC
CCACTGTCATAGTATAGGCTGCCAGTCTTGAGGGAGCTGCAGGACAGGAAAAA
GAGACCTTCAGAGAAGACAGGTTCCAGAAAAGAGAATGCGAAGAGGCAAACA
GAGGCTTAAGAATAAACCAGTTGGATGCACTTTGAACTAAACAAGGCCTGGAC
AGTAGTTTGGTCCCAGTTTACTAGTGCTTTTCATGGATTTAAACCTGGGCTGTT
AGCTCGCTTTTGAGCCAAGGGAAAGCGTCTTATGGGTTTTTTGGGAGATTCCCA
TCTATCCAAAAAATGCCACAGGGGACAGTCAGTTTTGAATATTAAGTTCAAAA
ATGATTTTTCCCCTAACCTGGG

Clone 99 sequence:

CTGGAGAGTCTGGATTTATTGGCAGTGAAGGTTTTCCTGGAGTGTACCCTCCAA
ATAGCAAATGTACTTGGAAAATCACAGTTCCCGAAGGAAAAGTAGTCGTTCTC
AATTTCCGATTCATAGACCTCGAGAGTGACAACCTGTGCCGCTATGACTTTGTG
GATGTGTACAATGGCCATGCCAATGGCCAGCGCATTGGCCGCTTCTGTGGCAC
TTTCCGGCCTGGAGCCCTTGTGTCCAGTGGCAACAAGATGATGGTGCAGATGA
TTTCTGATGCCAACACAGCTGGCAATGGCTTCATGGCCATGTTCTCCGCTGCTG
AACCAAACGAAGAGGGGATCAGTATTGTGGAGGACTCCTTGACAGACCTTCC
GGCTCTTTTAAAACCCCCAACTGGCCAGACCGGGATTACCCTGCAGGAGTCAC
TTGTGTGTGGCACATTGTAGCCCCAAAGAATCAGCTTATAGAATTAAAGTTTG
AGAAGTTTGATGTGGAGCGA

Clone 76 sequence:

CAGCAGGCCATGGAACTCCAGGAGGCTGCCAAGAAGAATTATGCAAACAAGG
CCCCGGGAGAGCAACAGAAATTGCAAGGAAATCCAACAAAGAAAACCAAATC
AAAAAGACCAGATCCCCTCAAAGGACAGAAGGTTATTGCAAGATGTGATGAA
AATGGCTTTTATTTTCCAGGGGTTGTGAAGAAGTGTGTGAGCCGCACCCAAGC
ACTGGTGGGCTTCAGTTACGGAGACACCAAGGTCGTGTCCACCTCCTTCATCA
CGCCTGTGGGGGGCGCCATGCCCTGCCCGCTGCTCCAGGTTGGAGATTATGTG
TTTGCCAAAATTGTGATACCCAAAGGATTTGACTTCTATGTCCCTGCCATTGTC
ATAGCACTTCCCAATAAGCATGTGGCCACAGAAAAATTCTACACAGTTTTGAA
GTGTAACAACCGGAGAGAATTTTGCCCTCGGAGTGCACTTATTAAGATCAGCC
AAAACAAGTATGCGCTCTCTTTGCT

Clone 28 sequence:

AAAAGAAGAAGAAAGCAAAAAAGGCCTCAACCAAAGATTCTGAGTCACCGTC
CCAGAAGAAAAAGAAGAAAAAGAAGAAGACAGCAGAGCAGACTGTATGACG
AGCACCAGCACCAGGCACAGGGATTTCCTAGCCGAGCAGTGGCCATCCCCATG
CCTCTGACCTCCACCGACCTCTGCCCACCATGGGTTGGAACTAAACTGTTACCT

TCCCTCGCTCCACAGAAGAAGACAGCCAGCTTCAGGGGTCCCTGTGCTGGCCA
AGCCAGTGAGCCTGCGGGGAGGCTGGTCCAAGGAGAAAGTGGACCAGCTCCC
ATGACCTCACCCCACTCCCCCAACACAGGACGCTTCATATAGATGTGTACAGT
ATATGTATTTTTTTAAGTGACCTCCTCTCCTTCCACAGACCCCACATGCCCAAA
GGCCTCGGGACTTCCCACCACCTTGCTCCACAGATCCAGCTAGGCCTGACCTGT
GCCTCATCCCGTGCCGCTCGGTCT

Clone 70 sequence:

NNTNATTTGTTTGANATNCCATNCGCTCGCACGGGTGGGCATATGGCCATGGA
NGCCCCGGGGGATCCGAATNCGCGGCCGCGTCGACCNTGAGGCGGAGGAGCC
NAAGCCGAAGCAACTGCANCAGCAGCNGNAGGAGGGCGAANAGGAANGGGG
AGGAAGAGGNNTGANGAANAGNAGGAGGANTANTNTTNACAGCANNTGGCA
TNNAANTATNGGNAGGGCNCCCTCACNNANNCGCNANNTCCNATTCCACNCC
NAGNCGCNNCTNTCCATTNGANAANANTAACTTTNNCCTACTCTATNCTCTAT
CNTTNCTAAGTACTAAAATTTTTNGAATNGAACTAANNACANTTATTATTTNTA
NTGACAATTCNCGTAC

Final 7 clones of interest:

Clone 24

Fibrillarin-Like: See Figures 4.11 and 4.12.

Clone 33
ARMET

Sequence:
GGGTGCGGTTCAGTCGGTCGGCGGCGGCAGCGGAGGAGGAGGAGGAGGAGGA
GGATGAGGAGGATGAGGAGGATGTGGGCCACGCAGGGGCTGGCGGTGGCGCT
GGCTCTNANCGTGCTGCCGGGCAGCCGGGCGCTGCGGCCGGGCGACTGCTAAG
TTTGTATTTCTTATCTGGGAAGATTTTACCAGGACCTCAAAGACAGAGATGTCA
CATTCTCACCAGCCACTATTGAAAACGAACTTATAAAGTTCTGCCGGGGAAGC
AAGAGGCAACAAAAATCGGTTGTGCTNCTNTATCNGGGCCACAGATGATGCA
GCCACCAAAATCATNAATGAGGTATCAAAGCCTCTGGCCC

Alignment with ARMET (NM_006010) Homo sapiens arginine-rich, mutated in early
stage tumors (ARMET),mRNA.

```
Score =  600 bits (312), Expect = e-169
Identities = 341/356 (95%), Gaps = 1/356 (0%)

Query: 1    GGGTGCGGTTCAGTCGGTCGGCGGCGGCAGCGGAGGAGGAGGAGGAGGAGGAGGATGAGG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 226  GGGTGCGGTTCAGTCGGTCGGCGGCGGCAGCGGAGGAGGAGGAGGAGGAGGAGGATGAGG
ARMET  32   G  C  G  S  V  G  R  R  R  Q  R  R  R  R  R  R  R  M  R
```

```
Query:  61  AGGATGAGGAGGATGTGGGCCACGCAGGGGCTGGCGGTGGCGCTGGCTCTNANCGTGCTG
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |   |  | | | | | | |
Sbjct: 286  AGGATGAGGAGGATGTGGGCCACGCAGGGGCTGGCGGTGGCGCGTGGCTCTGAGCGTGCTG
ARMET  52   R   M   R   R   M   W   A   T   Q   G   L   A   V   R   V   A   L   S   V   L


Query: 121  CCGGGCAGCCGGGCGCTGCGGCCGGGCGACTGCTAAGTTTGTATTTCTTATCTGGGAAGA
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 346  CCGGGCAGCCGGGCGCTGCGGCCGGGCGACTGCGAAGTTTGTATTTCTTATCTGGGAAGA
ARMET  72   P   G   S   R   A   L   R   P   G   D   C   E   V   C   I   S   Y   L   G   R


Query: 181  TTTTACCAGGACCTCAAAGACAGAGATGTCACATTCTCACCAGCCACTATTGAAAACGAA
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 406  TTTTACCAGGACCTCAAAGACAGAGATGTCACATTCTCACCAGCCACTATTGAAAACGAA
ARMET  92   F   Y   Q   D   L   K   D   R   D   V   T   F   S   P   A   T   I   E   N   E


Query: 241  CTTATAAAGTTCTGCCGGGGAAGCAAGAGGCAACAAAAATCGGTTGTGCTNCTNTATCNG
            | | | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | |   |   | | | | | | | | | | | | |   | |   | | | |   |
Sbjct: 466  CTTATAAAGTTCTGCC-GGGAAGCAAGAGGCAAAGAGAATCGGTTGTGCTACTATATCGG
ARMET  112  L   I   K   F   C   R   E   A   R   G   K   E   N   R   L   C   Y   Y   I   G


Query: 301  GGCCACAGATGATGCAGCCACCAAAATCATNAATGAGGTATCAAAGCCTCTGGCCC  356
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |   | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 525  GGCCACAGATGATGCAGCCACCAAAATCATCAATGAGGTATCAAAGCCTCTGGCCC  580
ARMET  132  A   T   D   D   A   A   T   K   I   I   N   E   V   S   K   P   L   A
```

Clone 36

KIAA0179 protein

Sequence:
CACAATAAAAGGAAACGGCCACGGAAGAAGAGCCCGAGGGCCCACAGGGAA
ATGTTGGAATCAGCAGTGTTGCCCCCAGAGGACATGTCTCAGAGTGGCCCGAG
TGGCAGTCATCCTCAGGGACCTAGAGGGTCCCCGACAGGTGGAGCCCAACTCC
TAAAAAGGAAGCGGAAACTTGGAGTTGTGCCCGTCAATGGCAGTGGCCTGTCC
ACGCCGGCCTGGCCTCCATTGCAGCAGGAAGGCCCTCCCACAGGCCCCGCAGA
GGGGGCGAACAGCCACACCACGCTGCCCCAGCGCAGGAGGCTGCAGAAAAAG
AAGGCAGGGCCCGGCAGCCTGGAGCTCTGTGGCCTGCCCAGCCAGAAAACAG
CAAGTTTGAAAAAGAGGAAGAAAATGAGAGTGATGTCAAACTTTGGTGGAGC
ACAACGGGGTGCTGGAGTCCGAAGCTGGGCAACCCCAGGCTCTGGGAAGCAG
TGGGACTTTGCAGTTCCCTGAAGAAGCAGAAGCTGAGGGCAGAGAGCGACTTT
GTGAAAGTTTGACACCCCCTTCTTACCAAAGCCCCTTGTTCTTTAGAAGAGCCA
AGAGCAGCACTGGCACCCACCCTTCANGCCCTTGNCGTNCAAGTTNAACAAGA
CACCCATCCAGCTTCCAAGAAAGTCACCTTTTGGGTTGACAAGAAACATTGAC
TGGCCGATTTCANGAAGACAGGACAAGAGTTTTTTTTGGTTAGTTCCCACGGG
CCCNTTTTTGAGTGGGCTTNGACCTTGAACAGAAACCCCT

Alignment with (XM_035973) Homo sapiens KIAA0179 protein (KIAA0179), mRNA.

```
Score = 1131 bits (588), Expect = 0.0
Identities = 735/777 (94%), Gaps = 16/777 (2%)

Query:  1     CACAATAAAAGGAAACGGCCACGGAAGAAGAGCCCGAGGGCCCACAGGGAAATGTTG
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1519  CACAATAAAAGGAAACGGCCACGGAAGAAGAGCCCGAGGGCCCACAGGGAAATGTTG
KIAA0179 469  H   N   K   R   K   R   P   R   K   K   S   P   R   A   H   R   E   M   L


Query:  58    GAATCAGCAGTGTTGCCCCCAGAGGACATGTCTCAGAGTGGCCCGAGTGGCAGTCAT
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1577  GAATCAGCAGTGTTGCCCCCAGAGGACATGTCTCAGAGTGGCCCGAGTGGCAGTCAT
KIAA0179 488  E   S   A   V   L   P   P   E   D   M   S   Q   S   G   P   S   G   S   H


Query:  115   CCTCAGGGACCTAGAGGGTCCCCGACAGGTGGAGCCCAACTCCTAAAAAGGAAGCGG
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1634  CCTCAGGGACCTAGAGGGTCCCCGACAGGTGGAGCCCAACTCCTAAAAAGGAAGCGG
KIAA0179 507  P   Q   G   P   R   G   S   P   T   G   G   A   Q   L   L   K   R   K   R


Query:  172   AAACTTGGAGTTGTGCCCGTCAATGGCAGTGGCCTGTCCACGCCGGCCTGGCCTCCA
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1691  AAACTTGGAGTTGTGCCCGTCAATGGCAGTGGCCTGTCCACGCCGGCCTGGCCTCCA
KIAA0179 526  K   L   G   V   V   P   V   N   G   S   G   L   S   T   P   A   W   P   P


Query:  229   TTGCAGCAGGAAGGCCCTCCCACAGGCCCCGCAGAGGGGGCGAACAGCCACACCACG
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1748  TTGCAGCAGGAAGGCCCTCCCACAGGCCCCGCAGAGGGGGCGAACAGCCACACCACG
KIAA0179 545  L   Q   Q   E   G   P   P   T   G   P   A   E   G   A   N   S   H   T   T


Query:  286   CTGCCCCAGCGCAGGAGGCTGCAGAAAAAGAAGGCAGGGCCCGGCAGCCTGGAGCTC
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1805  CTGCCCCAGCGCAGGAGGCTGCAGAAAAAGAAGGCAGGGCCCGGCAGCCTGGAGCTC
KIAA0179 564  L   P   Q   R   R   R   L   Q   K   K   K   A   G   P   G   S   L   E   L


Query:  343   TGTGGCCTGCCCAGCCAGAAAACAGCAAGTTTGAAAAAGAGGAAGAAAATGAGAGTG
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1862  TGTGGCCTGCCCAGCCAGAAAACAGCAAGTTTGAAAAAGAGGAAGAAAATGAGAGTG
KIAA0179 583  C   G   L   P   S   Q   K   T   A   S   L   K   K   R   K   K   M   R   V


Query:  400   ATGTCAAACTTTGGTGGAGCACAACGGGGTGCTGGAGTCCGAAGCTGGGCAACCCCA
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1919  ATGTCAAAC-TTGGTGGAGCACAACGGGGTGCTGGAGTCCGAAGCTGGGCAACCCCA
KIAA0179 602  M   S   N   L   V   E   H   N   G   V   L   E   S   E   A   G   Q   P   Q


QUERY:  457   GGCTCTGGGAAGCAGTGGGACTTTGCAGTTCCCTGAAGAAGCAGAAGCTGAGGGCAG
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
SBJCT:  1975  GGCTCTGGGAAGCAGTGGGAC-TTGCAGTTCCCTGAAGAAGCAGAAGCTGAGGGCAG
KIAA0179 621  A   L   G   S   S   G   T       C   S   S   L   K   K   Q   K   L   R   A   E
```

```
Query:  514   AGAGCGACTTTGTGAAAGTTTGACACCCCCTTCTTACCAAAGCCCCTTGTTCTTTAG
              |||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
SBJCT: 2031   AGAGCGACTTTGTG-AAGTTTGACACCCCCTTCTTACCAAAGCCCC-TGTTCTTCAG
KIAA0179 640    S   D   F   V   K   F   D   T   P   F   L   P   K   P   L     F   F   R


Query:  571   AAGAGCCAAGAGCAGCACTGGCACCCACCCTTCANGCCCTTGNCGTNCAAGTTNAAC
              |||||||||||||||||||||||||||||||||||||||| |||| || ||| | |||| |||
Sbjct: 2086   AAGAGCCAAGAGCAGCACTGCCACCCACCCTCCAGGCCC-TGCCGTCC-AGCTAAAC
KIAA0179 659   R   A   K   S   S   T   A   T   H   P   P   G   P   A   V   Q   L   N


Query:  628   AAGACACCCATCCAGCTTCCAAGAAAGTCACCTTTTGGGTTGACAAGAAACATTGA
              |||||| ||||||||| |||||||||||||| |||||||||||||||||||||||| |||
Sbjct: 2141   AAGACA-CCATCCAGC-TCCAAGAAAGTCACC-TTTGGGCTGAACAGAAACA-TGA
KIAA0179 677   K   T    P   S   S    S   K   K   V   T    F   G   L   N   R   N   M    T


Query:  685   CTGGCCGATTTCANGAAGACAGGACAAGAGNNNNNNNNGGTTAGTTCCCACGGGCCC
              || |||||||||| ||||||| ||||||||               ||| || |||||||||||||
Sbjct: 2194   CT-GCCGAATTCAAGAAGACA-GACAAGAG--TATCTTGGTCAG-TCCCACGGGCCC
KIAA0179 695    A   E   F   K   K   T   D   K   S     I   L   V   S    P   T   G   P


Query:  742   NTTTTTGAGTGGGCTTNGACCTTGAACAGAAACCCCT 778
              || | |||||| ||| |||| ||||||||| |||||
Sbjct: 2246   -TTCTCGAGTGGCCTTCGACCCTGAACAGAAGCCCCT 2281
KIAA0179 711   S   R   V   A   F   D   P   E   Q   K   P   L
```

Alignment of (XP_035973) Homo sapiens KIAA0179 protein (KIAA0179) with (NP_003674) Homo sapiens NNP-1/Nop52, novel nuclear protein 1

```
Score =  372 bits (956), Expect = e-101
Identities = 186/332 (56%), Positives = 244/332 (73%), Gaps = 3/332 (0%)


Query:    7   PAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFYCMWVQD
              P EIQ AQRLA +E+  RDRAV+KLR+YI  +TQR GGF+ +ELLK+WKGLFYCMW+QD
Sbjct:    8   PPEIQLAQRLAGNEQVTRDRAVRKLRKYIVARTQRAAGGFTHDELLKVWKGLFYCMWMQD


Query:   67   EPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQTMNREWKGIDRLRLDKYYMLIRLVLR
              +PLLQEEL  TI+QLVHA  + AQHLF+Q FWQTMNREW GIDRLRLDK+YML+R+VL
Sbjct:   68   KPLLQEELGRTISQLVHAFQTTEAQHLFLQAFWQTMNREWTGIDRLRLDKFYMLMRMVLN


Query:  127   QSFEVLKRNGWEESRIKVFLDVLMKEVLCPESQSPNGVRFHFIDIYLDELSKVGGKELLA
              +S +VLK  GWEE +I+  L++LM E+L P SQ+PNGV+ HFI+I+L+EL+KVG +EL A
Sbjct:  128   ESLKVLKMQGWEERQIEELLELLMTEILHPSSQAPNGVKSHFIEIFLEELTKVGAEELTA


Query:  187   DQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVGDGDLSA
              DQNLKFIDPFC+IAA+TKD ++  I RG+FE IV+Q+P  E+ + E T+  D ++++
Sbjct:  188   DQNLKFIDPFCRIAARTKDSLVLNNITRGIFETIVEQAPLAIEDLLNELDTQ--DEEVAS


Query:  247   EEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDT-GPLLQFDYKAVAD
              +    +E  R    +K +            E G + G   D+ GP+LQFDY+AVA+
Sbjct:  246   DSDESSEGGERGDALSQKRSEKPPAGSICRAEPEAGEEQAGDDRDSGGPVLQFDYEAVAN


Query:  306   RLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEG 337
              RL EM SR++TP  NRKRL K+I+K QDL+ G
Sbjct:  306   RLFEMASRQSTPSQNRKRLYKVIRKLQDLAGG 337
```

Clone 100

Fibrillarin

Sequence:
CCGGAGCCGCACAAACCAGGGCTCGCCATGAAGCCAGGATTCAGTCCCCGTGG
GGGTGGCTTTGGCGGCCGAGGGGGCTTTGGTGACCGTGGTGGTCGTGGAGGCC
GAGGGGGCTTTGGCGGGGGCCGAGGTCGAGGCGGAGGCTTTAGAGGTCGTGG
ACGAGGAGGAGGTGGAGGCGGCGGCGGCGGTGGAGGAGGAGGAAGAGGTGG
TGGAGGCTTCCATTCTGGTGGCAACCGGGGTCGTGGTCGGGGAGGAAAAAGA
GGAAACCAGTCGGGGAAGAATGTGATGGTGGAGCCGCATCGGCATGAGGGTG
TCTTCATTTGTCGAGGAAAGGAANATGCACTTGGTCACCAAAGAACCTGNTTC
CTGGGGAATCAGTTATGGAGAGAAGAGAGTCTTGATTTNGGAAGGAGATGAC
AAAATTGAGTACCGAGCCTGGAACCCCTTTCGCTCCAAGCTAGCANCANCAAT
CCTGGGTGGTGTGGACCAGATCCACATCAAACCGGGGGCTAANGTTTTNTACC
TNGGGGCTGNCTTCGGGCACCACGGTTTCCATNTTTNTTACATCGNTGGTCCGG
ATGGTNTAGTTTATGCAGTTGANTTNTNCCACCNTTTTGCCGNNACCTCATTAA
CTTGNCCAAAAAAAGGACCAANAT

Alignment with (NM_001436) Homo sapiens fibrillarin (FBL), mRNA.

Score = 1019 bits (530), Expect = 0.0
Identities = 619/659 (93%), Gaps = 6/659 (0%)
 Strand = Plus / Plus

Query: 1 CCGGAGCCGCACAAACCAGGGCTCGCCATGAAGCCAGGATTCAGTCCCCGTGGGGGTGGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 33 CCGGAGCCGCACAAACCAGGGCTCGCCATGAAGCCAGGATTCAGTCCCCGTGGGGGTGGC
FBL   1   M   K   P   G   F   S   P   R   G   G   G

Query: 61 TTTGGCGGCCGAGGGGGCTTTGGTGACCGTGGTGGTCGTGGAGGCCGAGGGGGCTTTGGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 93 TTTGGCGGCCGAGGGGGCTTTGGTGACCGTGGTGGTCGTGGAGGCCGAGGGGGCTTTGGC
FBL   12 F   G   G   R   G   G   F   G   D   R   G   G   R   G   G   R   G   G   F   G

Query: 121 GGGGGCCGAGGTCGAGGCGGAGGCTTTAGAGGTCGTGGACGAGGAGGAGGTGGAGGCGGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 153 GGGGGCCGAGGTCGAGGCGGAGGCTTTAGAGGTCGTGGACGAGGAGGAGGTGGAGGCGGC
FBL   32 G   G   R   G   R   G   G   G   F   R   G   R   G   R   G   G   G   G   G   G

Query: 181 GGCGGCGGTGGAGGAGGAGGAAGAGGTGGTGGAGGCTTCCATTCTGGTGGCAACCGGGGT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 213 GGCGGCGGTGGAGGAGGAGGAAGAGGTGGTGGAGGCTTCCATTCTGGTGGCAACCGGGGT
FBL   52 G   G   G   G   G   G   G   R   G   G   G   G   F   H   S   G   G   N   R   G

Query: 241 CGTGGTCGGGGAGGAAAAAGAGGAAACCAGTCGGGGAAGAATGTGATGGTGGAGCCGCAT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 273 CGTGGTCGGGGAGGAAAAAGAGGAAACCAGTCGGGGAAGAATGTGATGGTGGAGCCGCAT
FBL   72 R   G   R   G   G   K   R   G   N   Q   S   G   K   N   V   M   V   E   P   H

```
Query:  301  CGGCATGAGGGTGTCTTCATTTGTCGAGGAAAGGAANATGCACTTGGTCACCAAAGAACC
             || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | | | | |  | | | | | | | |  | | | | | |
Sbjct:  333  CGGCATGAGGGTGTCTTCATTTGTCGAGGAAAGGAAGATGCAC-TGGTCACC-AAGAACC
FBL      92  R   H   E   G   V   F   I   C   R   G   K   E   D   A   L   V   T   K   N   L


Query:  361  TGNTTCCTGGGGAATCAG-TTATGGAGAGAAGAGAGTCTTGATTTNGGAAGGAGATGACA
             | |  |  | | | | | | | | | | | | |  | | | | | | | | | | | | | | | | | |  | | | | |  | | | | | | | | | | | | |
Sbjct:  391  TGGTCCCTGGGGAATCAGTTTATGGAGAGAAGAGAGTCTCGATTTCGGAAGGAGATGACA
FBL     111  V   P   G   E   S   V   Y   G   E   K   R   V   S   I   S   E   G   D   D   K


Query:  420  AAATTGAGTACCGAGCCTGGAACCCCTTTCGCTCCAAGCTAGCANCANCAATCCTGGGTG
             | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | | | | | | | | | | | | |  | |  | | | | | | | | | | | |
Sbjct:  451  AAATTGAGTACCGAGCCTGGAACCCCTTCCGCTCCAAGCTAGCAGCAGCAATCCTGGGTG
FBL     131  I   E   Y   R   A   W   N   P   F   R   S   K   L   A   A   A   I   L   G   G


Query:  480  GTGTGGACCAGATCCACATCAAACCGGGGGCTAANGTTTTNTACCTNGGGGCTGNCTTCG
             | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | |  |  | | | | |  | | | | | | |  |  | | |
SBJCT:  511  GTGTGGACCAGATCCACATCAAACCGGGGGCTAAGGTTCTCTACCTCGGGGCTGCC-TCG
FBL     151  V   D   Q   I   H   I   K   P   G   A   K   V   L   Y   L   G   A   A   S


Query:  540  GGCACCACGGT-TTCCATNTTTNTTACATCGNTGGTCCGGATGGTNTAGTTTATGCAGTT
             | | | | | | | | | | |  |  | | | |  |  |  |  | | | | | |  | | | | | | | | | | |  | | | |  | | | | | | | |
Sbjct:  570  GGCACCACGGTCTCCCATGTCTCTGACATCGTTGGTCCGGATGGTCTAGTCTATGCAGTC
FBL     171  G   T   T   V   S   H   V   S   D   I   V   G   P   D   G   L   V   Y   A   V


Query:  599  GANTTNTNCCACC-NTTTTGCCGNNACCTCATTAACTTGNCCAAAAAAAGGACCAANAT
             | |  | |  |  | | | | |  |  |  | | | |  | | | | | | | | | | | | | |  | | | |  | |  | | | | | | | |  | |
Sbjct:  630  GAGTTCTCCCACCGCTCTGGCCGTGACCTCATTAACTTGGCCAAGAAGAGGACCAACAT
FBL     191  E   F   S   H   R   S   G   R   D   L   I   N   L   A   K   K   R   T   N   I
```

Clone 19

MMSETII/WHSC1

Sequence:
CCTTACGGGAAAGTCCAGATCTACACAGCGGATATTTCAGAAATCCCTAAGTG
CNACTGCAAGCCCACAGATGAGAATCCTTGTGGCTTTGATTCGGAGTGTCTGA
ACAGGNTGCTGATGTNNGNGTGCCACCCCNAAGTGGGTTCCGNGCGCGANNN
NNGNCACACCCAGTGCTTCACCAANCGCCAGTACCCAGAGACCAAGATCATCA
AGACAGATGGCANANGGTGGGGCCTGGTCGCCCANAGGNACATTATNAAGGG
AGAATTTGTTAANGANTACGTTTGGGAGNTGATCGACGAGGAGGAGTGCATG
GCGAGAATCAAGCACGCACACCGAGAACGACATCACCCACTTCTACATGCTCA
CTATAGACAAGGACCGTATNATAGACGCTGGCCCCAAAGGAAACTACTCTCGA
TTTATGAATCACAGCTGCCAGCCCAACTGTGAGACCCTCAAGGTGGACAGTGA
ATGGGGACACTCGTGTTGGGCCTGTTTGGCGTCTGTGANATTTCTGCNNGGGA
CGGGAGCTGACTTTTTAACTACAACCT

Alignment with (NM_133336) Homo sapiens Wolf-Hirschhorn syndrome candidate 1 (WHSC1), MMSETII.

```
Score =  802 bits (417), Expect = 0.0
Identities = 510/554 (92%), Gaps = 6/554 (1%)
 Strand = Plus / Plus


Query:1      CCTTACGGGAAAGTCCAGATCTACACAGCGGATATTTCAGAAATCCCTAAGTGCNACTGC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct:1612   CCTTACGGGAAAGTCCAGATCTACACAGCGGATATTTCAGAAATCCCTAAGTGCAACTGC
WHSC1 218    P   Y   G   K   V   Q   I   Y   T   A   D   I   S   E   I   P   K   C   N   C


Query:61     AAGCCCACAGATGAGAATCCTTGTGGCTTTGATTCGGAGTGTCTGAACAGGNTGCTGATG
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct:1672   AAGCCCACAGATGAGAATCCTTGTGGCTTTGATTCGGAGTGTCTGAACAGGATGCTGATG
WHSC1 238    K   P   T   D   E   N   P   C   G   F   D   S   E   C   L   N   R   M   L   M


Query:121    TNNGNGTGCCACCCCNAAGTGGGTTCCGNGCGCGANNNNNGNCACACCCAGTGCTTCACC
             |   |  |||||||||   |  |||  ||  |||  |  ||||         |  ||  | |||||||||||||
Sbjct:1732   TTTGAGTGCCACCCGCAGGTGTGTCCCGCGGGCGAGTTCTGCCAGAACCAGTGCTTCACC
WHSC1 258    F   E   C   H   P   Q   V   C   P   A   G   E   F   C   Q   N   Q   C   F   T


Query:181    AANCGCCAGTACCCAGAGACCAAGATCATCAAGACAGATGGCANANGGTGGGGCCTGGTC
             ||  ||||||||||||||||||||||||||||||||||||||||||| |  |||||||||||||
Sbjct:1792   AAGCGCCAGTACCCAGAGACCAAGATCATCAAGACAGATGGCAAAGGGTGGGGCCTGGTC
WHSC1 278    K   R   Q   Y   P   E   T   K   I   I   K   T   D   G   K   G   W   G   L   V


Query:241    GCCCANAGGNACATTATNAAGGGAGAATTTGTTAANGANTACGTTTGGGAGNTGATCGAC
             |||  |  |||  ||||  |   |||||||||||||||||| ||  ||||||  |||||  ||||||||
Sbjct:1852   GCCAAGAGGGACATCAGAAAGGGAGAATTTGTTAACGAGTACGTTGGGGAGCTGATCGAC
WHSC1 298    A   K   R   D   I   R   K   G   E   F   V   N   E   Y   V   G   E   L   I   D


Query:301    GAGGAGGAGTGCATGGCGAGAATCAAGCACGCACACCGAGAACGACATCACCCACTTCTA
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct:1912   GAGGAGGAGTGCATGGCGAGAATCAAGCACGCACA-CGAGAACGACATCACCCACTTCTA
WHSC1 318    E   E   E   C   M   A   R   I   K   H   A   H   E   N   D   I   T   H   F   Y


Query:361    CATGCTCACTATAGACAAGGACCGTATNATAGACGCTGGCCCCAAAGGAAACTACTCTCG
             ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Sbjct:1971   CATGCTCACTATAGACAAGGACCGTATAATAGACGCTGGCCCCAAAGGAAACTACTCTCG
WHSC1 338    M   L   T   I   D   K   D   R   I   I   D   A   G   P   K   G   N   Y   S   R


Query:421    ATTTATGAATCACAGCTGCCAGCCCAACTGTGAGACCCTCAAGGTGGACAGTGAATGGGG
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct:2031   ATTTATGAATCACAGCTGCCAGCCCAACTGTGAGACCCTCAA-GTGGACAGTGAATGGGG
WHSC1 358    F   M   N   H   S   C   Q   P   N   C   E   T   L   K   W   T   V   N   G   D
```

```
Query:481  ACACTCGTGTTGGGCCTGTTTGGCGTCTGTGANATTTCTGCNNGGGACGGGAGCTGACTT
           ||||||||| ||||||||||| ||||||||| ||| ||||  ||||| ||||||||| |
Sbjct:2090 ACACTCGTG-TGGGCCTGTTTGCCGTCTGTGACATTCCTGC-AGGGAC-GGAGCTGAC-T
WHSC1 377     T   R   V   G   L   F   A   V   C   D   I   P   A   G   T   E   L   T
```

```
Query:541  TTTAACTACAACCT 554
           ||||||||||||||
Sbjct:2146 TTTAACTACAACCT 2159
WHSC1 396  F   N   Y   N   L
```

Clone 23

Protamine

Sequence:
AACCNGAGTTCCACCTGCTCACAGGTTGGCTGGCTCAGCCAAGGNGGTGCCCT
GCTNTGAGCNTTCAGGCCAAGCCCATCCTGCACCATGGCCAGGTNCAGATGCT
GTCGCAGCCAGAGCCGGAGCAGATNTTACCGCCAGAGACAAAGAAGTCGCAG
ACGAAGGAGGCGGAGCTGCCANACNCGGAGGAGAGCCATGAGGTGCTGCCGC
CCCAGGTNCAGACCGCGATGTANAANACNCTANTTGCNCAAAATAGCACATCC
ACCAAACTCNTGCCTGAGAATGTTACCAGACTTCAAGATCCTNTTGCCACATCT
TGAAAATGCCACCNTCCAATAAAAATCAGGAGCCTGNTAAGGAACAATGCCG
CCTGTCAATAAATGTTGAAAGTCAAAAAAAAAAAAAAAAAAAAA

Alignment with (BC003673) Homo sapiens, protamine 1, clone MGC:12307
IMAGE:3935638, mRNA, complete cds.

```
Score =  710 bits (369), Expect = 0.0
Identities = 390/409 (95%)
 Strand = Plus / Plus
```

```
Query:6    GAGTTCCACCTGCTCACAGGTTGGCTGGCTCAGCCAAGGNGGTGCCCTGCTNTGAGCNTT
           ||||||||||||||||||||||||||||||||||||||| ||||||||||| ||||| ||
Sbjct:9    GAGTTCCACCTGCTCACAGGTTGGCTGGCTCAGCCAAGGTGGTGCCCTGCTCTGAGCATT
```

```
Query:66   CAGGCCAAGCCCATCCTGCACCATGGCCAGGTNCAGATGCTGTCGCAGCCAGAGCCGGAG
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Sbjct:69   CAGGCCAAGCCCATCCTGCACCATGGCCAGGTACAGATGCTGTCGCAGCCAGAGCCGGAG
Prot.1 1                           M   A   R   Y   R   C   C   R   S   Q   S   R   S
```

```
Query:126  CAGATNTTACCGCCAGAGACAAAGAAGTCGCAGACGAAGGAGGCGGAGCTGCCANACNCG
           |||||  |||||||||||||||||||||||||||||||||||||||||||||||| || ||
Sbjct:129  CAGATATTACCGCCAGAGACAAAGAAGTCGCAGACGAAGGAGGCGGAGCTGCCAGACACG
Prot.1 14  R   Y   Y   R   Q   R   Q   R   S   R   R   R   R   R   R   S   C   Q   T   R
```

```
Query:186  GAGGAGAGCCATGAGGTGCTGCCGCCCCAGGTNCAGACCGCGATGTANAANACNCTANTT
           |||||||||||||||||||||||||||||||| ||||||| |||||| || || ||| ||
Sbjct:189  GAGGAGAGCCATGAGGTGCTGCCGCCCCAGGTACAGACCGAGATGTAGAAGACACTAATT
Prot.1 34  R   R   A   M   R   C   C   R   P   R   Y   R   P   R   C   R   R   H   ^^^
```

114

```
Query:246 GCNCAAAATAGCACATCCACCAAACTCNTGCCTGAGAATGTTACCAGACTTCAAGATCCT
           ||  ||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Sbjct:249 GCACAAAATAGCACATCCACCAAACTCCTGCCTGAGAATGTTACCAGACTTCAAGATCCT


Query:306 NTTGCCACATCTTGAAAATGCCACCNTCCAATAAAAATCAGGAGCCTGNTAAGGAACAAT
           |||||||||||||||||||||||||| ||||||||||||||||||||| |||||||||||
Sbjct:309 CTTGCCACATCTTGAAAATGCCACCATCCAATAAAAATCAGGAGCCTGTTAAGGAACAAT


Query:366 GCCGCCTGTCAATAAATGTTGAAAAGTCAAAAAAAAAAAAAAAAAAAA 414
           |||||||||||||||||||||||||||| |||||||||||||||||||
Sbjct:369 GCCGCCTGTCAATAAATGTTGAAAAGCCAAAAAAAAAAAAAAAAAAAA 417
```

Clone 38

Ribosomal Protein L29

Sequence:
TCGGGAGCCGCGGCTTATGGTGCAGACATGGCCAAGTCCAAGAACCACACCAC
ACACAACCAGTCCCGAAAATGGCACAGAAATGGTATCAAGAAACCCCGATCA
CAAAGATACGAATCTCTTAAGGGGGTGGACCCCAAGTTCCTGAGGAACATGCG
CTTTGCCAAGAAGCACAACAAAAAGGGCCTAAAGAAGATGCAGGCCAACAAT
GCCAAGGCCATGAGTGCACGTGCCGAGGCTATCAAGGCCCTCGTAAAGCCCAA
GGAGGTTAAGCCCAAGATCCCAAAGGGTGTCAGCCGCAAGCTCGATCGACTTG
CCTACATTGCCACCCCAAGCTTGGGAAGCGTGCTCGTGCCCGTATTGCCAAGG
GGCTCAGGCTGTGCCGGCAAAGGCCAAGGCCAAGGCCAAGGCCAAGGATCAA
ACCAAGGCCCAGGCTGCAGCCCCAGCTTNAGTTCCAGCTCAGCTTCCAAACGT
ACCCAGGCCCCTACAAAGGCTTCAGAGTAGATATCTCTGCCAACATGAGGACA
GAAAGGACTGGTGCGACCCCCCACCCCCGCCCTTGGCTACCATNTGCATGGGG
GCTTGGGGTNCTCCTGTGCTATTTGTACAAATTAACCTGAGGCCGGGAAAAAA
A

Alignment with (NM_000992) Homo sapiens ribosomal protein L29 (RPL29), mRNA.

```
Score = 1083 bits (563), Expect = 0.0
Identities = 622/638 (97%), Gaps = 7/638 (1%)
 Strand = Plus / Plus


Query:1    TCGGGAGCCGCGGCTTATGGTGCAGACATGGCCAAGTCCAAGAACCACACCACACACAAC
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:68   TCGGGAGCCGCGGCTTATGGTGCAGACATGGCCAAGTCCAAGAACCACACCACACACAAC
RPL29 1                                M   A   K   S   K   N   H   T   T   H   N


Query:61   CAGTCCCGAAAATGGCACAGAAATGGTATCAAGAAACCCCGATCACAAAGATACGAATCT
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:128  CAGTCCCGAAAATGGCACAGAAATGGTATCAAGAAACCCCGATCACAAAGATACGAATCT
RPL29 12   Q   S   R   K   W   H   R   N   G   I   K   K   P   R   S   Q   R   Y   E   S
```

```
Query:121  CTTAAGGGGGTGGACCCCAAGTTCCTGAGGAACATGCGCTTTGCCAAGAAGCACAACAAA
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:188  CTTAAGGGGGTGGACCCCAAGTTCCTGAGGAACATGCGCTTTGCCAAGAAGCACAACAAA
RPL29 32   L   K   G   V   D   P   K   F   L   R   N   M   R   F   A   K   K   H   N   K


Query:181  AAGGGCCTAAAGAAGATGCAGGCCAACAATGCCAAGGCCATGAGTGCACGTGCCGAGGCT
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:248  AAGGGCCTAAAGAAGATGCAGGCCAACAATGCCAAGGCCATGAGTGCACGTGCCGAGGCT
RPL29 52   K   G   L   K   K   M   Q   A   N   N   A   K   A   M   S   A   R   A   E   A


Query:241  ATCAAGGCCCTCGTAAAGCCCAAGGAGGTTAAGCCCAAGATCCCAAAGGGTGTCAGCCGC
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:308  ATCAAGGCCCTCGTAAAGCCCAAGGAGGTTAAGCCCAAGATCCCAAAGGGTGTCAGCCGC
RPL29 72   I   K   A   L   V   K   P   K   E   V   K   P   K   I   P   K   G   V   S   R


Query:301  AAGCTCGATCGACTTGCCTACATTG-CCACCCCAAGCTTGGGAAGCGTGCTCGTGCCCGT
           |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Sbjct: 68  AAGCTCGATCGACTTGCCTACATTGCCCACCCCAAGCTTGGGAAGCGTGCTCGTGCCCGT
RPL29 92   K   L   D   R   L   A   Y   I   A   H   P   K   L   G   K   R   A   R   A   R


Query:360  ATTGCCAAGGGGCTCAGGCTGTGCCGG-CAAAGGCCAAGGCCAAGGCCAAGGCCAAGGAT
           |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Sbjct:428  ATTGCCAAGGGGCTCAGGCTGTGCCGGCCAAAGGCCAAGGCCAAGGCCAAGGCCAAGGAT
RPL29 112  I   A   K   G   L   R   L   C   R   P   K   A   K   A   K   A   K   A   K   D


Query:419  CAAACCAAGGCCCAGGCTGCAGCCCCAGCTTNAGTTCCAGCTCA-GCTTCCAAACGTACC
           |||||||||||||||||||||||||||||||  |||||||||||  |||  ||||||||||
Sbjct:488  CAAACCAAGGCCCAGGCTGCAGCCCCAGCTTCAGTTCCAGCTCAGGCTCCCAAACGTACC
RPL29 132  Q   T   K   A   Q   A   A   A   P   A   S   V   P   A   Q   A   P   K   R   T


Query:478  CAGGCCCCTACAAAGGCTTCAGAGTAGATATCTCTGCCAACATGAGGACAGAAAGGACTG
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct:548  CAGGCCCCTACAAAGGCTTCAGAGTAGATATCTCTGCCAACATGAGGACAG-AAGGACTG
RPL29 152  Q   A   P   T   K   A   S   E   ^^^


Query:538  GTGCGACCCCCCACCCCCG-CCCTTGGCTACCATNTGCATGGGGGGCTTGGGGTNCTCCTG
           |||||||||||||||||||  ||||  ||||||||||  ||||||||||    ||||||  ||||||
Sbjct:607  GTGCGACCCCCCACCCCCGCCCCTGGGCTACCATCTGCATGGGG--CTGGGGTCCTCCTG


Query:597  TGCTATTTGTACAAATTAACCTGAGGCCGGGAAAAAAA  634
           ||||||||||||||||  ||||||||||  ||  |||||||
Sbjct:665  TGCTATTTGTACAAATAAACCTGAGGCAGGAAAAAAA  702
```

116

References

Adams, J. M., Harris, A. W., Pinkert, C. A., Corcoran, L. M., Alexander, W. S., Cory, S., Palmiter, R. D., and Brinster, R. L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature *318*, 533-538.

Afroze, T., and Husain, M. (1999). c-*myb* binding sites mediate G1/S associated repression of the plasma mambrane Ca2+-ATPase promoter. The Journal of Biological Chemistry *275*, 9062-9069.

Ali, I. K., McKendrick, L., Morley, S. J., and Jackson, R. J. (2001). Activity of the Hepatitis A Virus IRES Requires Association between the Cap-Binding Translation Initiation Factor (eIF4E) and eIF4G. Journal of Virology *75*, 7854-7863.

Ali, N., Pruijn, G. J. M., Kenan, D. J., Keene, J. D., and Siddiqui, A. (2000). Human La antigen is required for the Hepatitis C virus internal ribosome entry site-mediated translation. The Journal of Biological Chemistry *275*, 27531-27540.

Ali, N., and Siddiqui, A. (1997). The La antigen binds 5' noncoding region of the hepatitis C virus RNA in the context of the initiator AUG codon and stimulates internal ribosome entry site-mediated translation. Proc Natl Acad Sci USA *94*, 2249-2254.

Altschul, Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403-410.

Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I., and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature *359*, 423-426.

Asano, K., Clayton, J., Shalev, A., and Hinnebusch, A. G. (2000). A multifactor complex of eukayotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation intermediate *in vivo*. Genes and Development *14*, 2534-2546.

Belsham, G. J., and Sonenberg, N. (1996). RNA-Protein Interactions in Regulation of Picornavirus RNA Translation. Microbiology Reviews *60*, 499-511.

Bender, T. P., and Kuehl, M. (1986). Murine myb protooncogene mRNA:cDNA sequence and evidence for 5'heterogeneity. Proceedings of the National Academy of Science U S A *83*, 3204-3208.

Bentley, D. L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-*myc* in differentiated HL60 cells. Nature *321*, 702-706.

Bernstein, D. S., Buter, N., Stumpf, C., and Wickens, M. (2002). Analyzing mRNA-protein complexes using a yeast three-hybrid system. Methods *26*, 123-141.

Bernstein, J., Sella, O., Le, S., and Elroy-Stein, O. (1997). PDGF/c-sis mRNA leader contains a differentiation-linked internal ribosomal entry site. J Biol Chem *272*, 9356-9362.

Biedenkapp, H., Borgmeyer, U., Sippel, A. E., and Klempnauer, K. H. (1988). Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature *335*, 835-837.

Bommer, U. A., Lutsch, G., Stahl, J., and Bielka, H. (1991). Eukayotic initiation factors eIF-2 and eIF-3: interactions, structure and localisation in ribosomal intiation complexes. Biochimie *73*, 1007-1019.

Borman, A., Howell, M. T., Patton, J. G., and Jackson, R. J. (1993). The involvement of a spliceosome component in internal initiation of human rhinovirus RNA translation. J Gen Virol *74*, 1775-1788.

Borman, A. M., and Kean, K. M. (1997). Intact eukaryotic initiation factor 4G is required for hepatitis A virus internal initiation of translation. Virololgy *237*, 129-136.

Borman, A. M., Michel, M. Y., and Kean, K. M. (2001). Detailed analysis of the requirements of Hepatitis A virus internal ribosome entry segment for the Eukaryotic initiation factor comple eIF4F. Journal of Virology *75*, 7864-7871.

Burk, O., Mink, S., Ringwald, M., and Klempnauer, K. H. (1993). Synergistic activation of the chicken mim-1 gene by v-myb and c/EBP transcription factors. EMBO J *12*, 2027-2038.

Bushell, M., McKendrick, L., Janicke, R. U., Clemens, M. J., and Morley, S. J. (1999). Caspase-3 is necessary and sufficient for cleavage of protein synthesis eukaryotic initiation factor 4G during apoptosis. FEBS letters *451*, 332-336.

Bushell, M., Wood, W., Clemens, M. J., and Morley, S. J. (2000a). Changes in integrity and association of eukaryotic protein synthesis initiation factors during apoptosis. European Journal of Biochemistry *267*, 1083-1091.

Bushell, M., Wood, W., Clemens, M. J., and Morley, S. J. (2000b). Cleavage of polypeptide chain initiation factor eIF4GI during apoptosis in lymphoma cells: characterisation of an internal fragment generated by caspase-3-mediated cleavage. Cell Death and Differentiation *7*, 628-636.

Byrd, M. P., Zamora, M., and Lloyd, R. E. (2002). Generation of multiple isoforms of eukaryotic translation initiation factor 4GI by use of alternative translation initiation codons. Molecular and Cellular Biology 22, 4499-4511.

Carter, M. S., and Sarnow, P. (2000). Distinct mRNAs That Encode La Autoantigen Are Differentially Expressed and Contain Internal Ribosome Entry Sites. J Biol Chem 275, 28301-28307.

Chakrabarti, A., and Maitra, U. (1991). Function of Eukaryotic initiation factor 5 in the formation of an 80S ribosomal polypeptide chain initiation complex. The Journal of Biological Chemistry 266, 14039-14045.

Chappell, S. A., Edelman, G. M., and Mauro, V. P. (2000a). A 9-nt segment of a cellular mRNA can function as internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. Proc Natl Acad Sci USA 97, 1536-1541.

Chappell, S. A., Le Quesne, J. P. C., Paulin, F. E. M., deSchoolmeester, M. L., Stoneley, M., Soutar, R. L., Ralston, S. H., Helfrich, m. H., and Willis, A. E. (2000b). A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: A novel emchanism of oncogene de-regulation. Oncogene 19, 4437-4440.

Chappell, S. A., Owens, G. C., and Mauro, V. C. (2001). A 5' Leader of Rbm3, a Cold Stress-induced mRNA, Mediates Internal Initiation of Translation with Increased Efficiency under Conditions of Mild Hypothermia. Journal of Biology.

Chaudhuri, J., chowdhury, D., and Maitra, U. (1999). Distinct functions of Eukaryotic translation initiation factors eIF1A and eIF3 in the formation of the 40S ribosomal preinitiation complex. The Journal of Biological Chemistry 25, 17975-17980.

Chen, C., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular mRNAs. Science 268, 415-417.

Chkheidze, A. N., Lyakhov, D. L., Makeyev, A. V., Morales, J., Kong, J., and Liebhaber, S. A. (1999). Assembly of the alpha-globin mRNA stability complex reflects binary interaction between the pyrimidine-rich 3 ' untranslated region determinant and poly(C) binding protein alpha CP. Molecular and Cellular Biology 19, 4572-4581.

Coldwell, M., deSchoolmeester, M. L., Fraser, G., Pickering, B. M., Packham, G., and Willis, A. E. (2001). The p36 isoform of BAG-1 is translated by internal ribosome entry following heat shock. Oncogene 20, 4095-4100.

Coldwell, M., Mitchell, S., Stoneley, M., MacFarlane, M., and Willis, A. (2000). Initiation of Apaf-1 by internal ribosome entry. Oncogene *19*.

Cornelis, S., Bruynooghe, Y., Denecker, G., Huffel, S. V., Tinton, S., and Beyaert, R. (2000). Identification and Characterization of a Novel Cell Cycle-Regulated Internal Ribosome Entry Site. Molecular Cell *5*, 597-605.

Craig, A. W., Haghighat, A., Yu, A. T. K., and Sonenberg, N. (1998). Interaction of polyadenylate-binding protein with the eIF4G homoloue PAIP enhances translation. Nature *392*, 520-523.

Creancier, L., Mercier, P., Prats, A., and Morello, D. (2001). c-myc Internal Ribosome Entry Site Activity Is Developmentally Controlled and Subjected to a Strong Tanslational Repression in Adult Transgenic Mice. Molecular and Cellular Biology *21*, 1833-1840.

Creancier, L., Morello, D., Mercier, P., and Prats, A.-C. (2000). Fibroblast growth factor 2 IRES activity ex vivo and in transgenic mice reveals a stringent tissue-specific regulation. J Cell Biol *150*, 275-281.

Cuesta, R., Laroia, G., and Schneider, R. J. (2000). Chaperone Hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initation complexes. Genes Dev *14*, 1460-1470.

Dang, C. (1999). C-Myc target genes involved in cell growth, apoptosis and metabolism. Molecular and Cellular Biology *19*, 1-11.

Dang, C. V., Resar, L. M. S., Emison, E., Kim, S., Li, Q., Prescott, J. E., Wonsey, D., and Zeller, K. (1999). Function of the c-*myc* oncogenic transcription factor. Experimental Cell Research *253*, 63-77.

Dash, A. B., Orrico, F. C., and Ness, S. A. (1996). The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb. Genes and Development *10*, 1858-1869.

Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R., and Bradley, A. (1993). A null c-*myc* mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes and Development *7*, 671-682.

Day, D. A., and Tuite, M. F. (1998). Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. Journal of Endocrinology *157*, 361-371.

Dean, M., Levine, R. A., Ran, W., Kindy, M. S., Sonenshein, G. E., and Campisi, J. (1986). Regulation of *c-myc* transcription and mRNA abundance by serum growth factors and cell contact. Journal of Biological Chemistry *261*, 9161-9166.

Dmitrovsky, E., Kuehl, M., Hollis, G. F., Kirsch, I. R., Bender, T. P., and Segal, S. (1986). Expression of a transfected humna c-*myc* oncogene inhibits differentiation of a mouse erythroleukaemia cell line. Nature *322*, 748-750.

Dudek, H., Tantravahi, R. V., Rao, V. N., Reddy, E. S. P., and Reddy, E. P. (1992). Myb and Ets proteins cooperate in transcriptional activation of the mim-1 promoter. Proc Natl Acad Sci U S A *89*, 1291-1295.

Eisenman, R. N. (2001). Deconstructing Myc. Genes and Development *15*, 2023-2030.

Evan, G. I., Wyllie, A. H., Gilbert, C. S., Littlewood, T. D., Land, H., Brooks, M., Waters, C. M., Penn, L. Z., and Hancock, D. C. (1992). Induction of Apoptosis in Fibroblasts by c-myc Protein. Cell *69*, 119-128.

Feigenblum, D., and Schneider, R. J. (1996). Cap-binding protein (Eukaryotic factor 4E) and 4E-Inactivating protein BP-1 independently regulate cap-dependent translation. Molecular and Cellular Biology *16*, 5450-5457.

Fernandez, J., Bode, B., Koromilas, A., Diehl, J. A., Krukovets, I., Snider, M. D., and Hatzoglou, M. (2002a). Translation mediated by the internal ribosome entry site of the cat-1 mRNA is regulated by glucose availability in a PERK kinase-dependent manner. The Journal of Biological Chemistry *277*, 11780-11787.

Fernandez, J., Yaman, I., Merrick, W. C., Koromilas, A., Wek, R. C., Sood, R., Hensold, J., and Hatzoglou, M. (2002b). Regulation of internal ribosome entry site-mediated translation by eukaryotic initiation factor-2alpha phosphorylation and translation of a small upstream open reading frame. The Journal of Biological Chemistry *277*, 2050-2058.

Fernandez, J., Yaman, I., Mishra, R., Merrick, W. C., Snider, M. D., Lamers, W. H., and Hatzoglou, M. (2001). Internal Ribosome Entry Site-mediated Translation of a Mammalian mRNA Is Regulated by Amino Acid Availability. Journal of Biological Chemistry *276*, 12288-12291.

Futterer, J., Kiss-Laszlo, Z., and Hohn, T. (1993). Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. Cell *73*, 789-802.

Galaktionov, K., Chen, X., and Beach, D. (1996). Cdc25 cell-cycle phosphatase as a target of c-*myc*. Nature *382*, 511-117.

Gamarnik, A. V., and Andino, R. (1997). Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. RNA *3*, 882-892.

Geballe, A. P., and Morris, D. R. (1994). Initiation codons within 5'-leaders of mRNAs as regulators of translation. TIBS *19*, 159-164.

Gewirtz, A. M., Anfossi, G., Venturelli, D., Valpreda, S., Sims, R., and Calabretta, B. (1989). G1/S transition in normal human T-lymphocytes requires the nuclear protein encoded by c-myb. Science *245*, 180-183.

Gewirtz, A. M., and Calabretta, B. (1988). A c-*myb* antisense oligodeoxynucleotide inhibits normal human hematopoiesis in vitro. Science *242*, 1303-1306.

Gingras, A. C., Raught, B., Gygi, S. P., Niedzwiecka, A., Miron, M., Burley, S. K., Polakiewicz, R. D., Wyslouch-Cieszynska, A., Aebersold, R., and Sonenberg, N. (2001). Hierarchical phosphorylation of the translation inhibitor 4e-BP1. Genes and Development *15*, 2852-2864.

Gingras, A. C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. Annual Reviews in Biochemistry *68*, 913-963.

Gingras, A. C., Svitkin, Y. V., Belsham, G. J., Pause, A., and Sonenberg, N. (1996). Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. PNAS *93*, 5578-5583.

Gradi, A., Imataka, H., Svitkin, Y. V., Rom, E., Raught, B., Morino, S., and Sonenberg, N. (1998a). A novel functional human eukaryotic translation initiation factor 4G. Molecular and Cellular Biology *18*, 334-342.

Gradi, A., Svitkin, Y. V., Imataka, H., and Sonenberg, N. (1998b). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with thr shutoff of host protein synthesis after poliovirus infection. PNAS *95*, 11089-11094.

Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000). The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior. Annuual Reviews in Cellular and Developmental Biology *16*, 653-699.

Gray, N. K., and Hentze, M. W. (1994). Regulation of protein synthesis by mRNA structure. Molecular Biology Reports *19*, 195-200.

Gray, N. K., and Wickens, M. (1998). Control of translation initiation in animals. Annuual Reviews in Cellular and Developmental Biology *14*, 399-458.

Guo, J., Hui, D. J., and Merrick, W. C. (2000). A new pathway of translational regulation mediated by eukaryotic initiation factor 3. The EMBO Journal *19*, 6892-6899.

Haghighat, A., Mader, S., Pause, A., and Sonenberg, N. (1995). Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. EMBO J *14*, 5701-5709.

Hahm, B., Kim, Y. K., Kim, J. H., Kim, T. Y., and Jang, S. K. (1998). Heterogeneous Nuclear Ribonuclearprotein L Interacts with the 3' Border of the Internal Ribosome Entry Site of the Hepatitis C Virus. Journal of Virology *72*, 8782-8788.

Hayashi, S., Nishimura, K., Fukuchi-Shimogori, T., Kashiwagi, K., and Igarashi, K. (2000). Increase in cap- and IRES-dependent protein synthesis by overproduction of translation initiation factor eIF4G. Biochemical and Biophysical Research Communications *277*, 117-123.

Hellen, C. U., Witherell, G. W., Schmid, M., Shin, S. H., Pestova, T. V., Gil, A., and Wimmer, E. (1993). A cytoplasmic 57-kDa protein that is required for translation of picornavirus RNA by internal ribosome entry is identical to the nuclear pyrimidine tract-binding protein. PNAS *90*, 7642-7646.

Hellen, C. U. T., and Sarnow, P. (2001). Internal Ribosome entry sites in eukaryotic mRNA molecules. Genes and Development *15*, 1593-1612.

Hemmings-Meiszczak, M., and Hohn, T. (1999). A stable hairpin preceded by a short open reading frame promotes non linear ribosome migration on a synthetic mRNA leader. RNA *5*, 1149-1157.

Henis-Korenblit, S. (2000). A novel form of Dap5 protein accumulates in apoptotic cells as a result of caspase cleavage and internal ribosome entry site mediated translation. Molecular and Cellular Biology *20*, 496-506.

Henis-Korenblit, S., Shani, G., Marash, L., Shohat, G., and Kimchi, A. (2002). The caspase-cleaved DAP5 protein supports internal ribosome entry site-mediated translation of death proteins. PNAS *99*, 5400-5405.

Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B., and Klausner, R. D. (1987). Identification of the Iron-Responsive element for the translational regulation of Human Ferritin mRNA. Science *238*, 1570-1573.

Hermeking, H., and Eick, D. (1994). Mediation of c-myc-induced apoptosis by p53. Science *265*, 2091-2092.

Hirst, S. K., and Grandori, C. (2000). Differential activity of conditional MYC and its variant MYC-S in human mortal fibroblasts. Oncogene *19*, 5189-5197.

Holcik, M., and Korneluk, R. G. (2000). Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: Role of La autoantigen in XIAP translation. Mol Cell Biol *20*, 4648-4657.

Holcik, M., Lefebvre, C., Yeh, C., Chow, T., and Korneluk, R. G. (1999). A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection. Nature Cell Biol *1*, 190-192.

Holcik, M., and Liebhaber, S. A. (1997). Four highly stable eukaryotic mRNAs assemble 3' untranslated region RNA-protein complexes sharing cis and trans components. Proceedings of the National Academy of Science U S A *94*, 2410-2414.

Holcik, M., Yeh, C., Korneluk, R. G., and Chow, T. (2000). Translational upregulation of X-linked inhibitor of apoptosis (XIAP) increases resistance to radiation induced cell death. Oncogene *19*, 4174-4177.

Hunt, S., and Jackson, R. (1999). PTB is necessary but not sufficient for efficient internal initiation of translation of human rhino virus 2 RNA. RNA *5*, 344-359.

Hunt, S. L., Hsuan, J. J., Totty, N., and Jackson, R. J. (1999). unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of human rhinovirus RNA. Genes and Development *13*, 437-448.

Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. The EMBO Journal *17*, 7480-7489.

Imataka, H., Olsen, H. S., and Sonenberg, N. (1997). A new translational regulator with homology to eukaryotic translation initiation factor 4G. The EMBO Journal *16*, 817-825.

Izumi, R. E., Valdez, B., Banerjee, R., Srivastava, M., and Dasgupta, A. (2001). Nucleolin stimulates viral internal ribosome entry site-mediated translation. Virus Research *76*, 17-29.

Jackson, R. J., Howell, M. T., and Kaminski, A. (1990). The novel mechanism of Initiation of picornavirus RNA translation. TIBS *15*, 477-511.

Jackson, R. J., Hunt, S. L., Reynolds, J. E., and Kaminski, A. (1995). Cap-
Dependent and Cap-Independent Translation: Operational Distinctions and Mechanistic
Interpretations. Curr Top Microbiol Immunol *203*, 1-29.

Jang, S. K., Krausslich, H., Nicklin, M. J. H., Duke, G. M., Palmenberg, A. C., and
Wimmer, E. (1988). A segement of the 5' Nontranslated region of encephalomyocarditis
virus RNA directs internal entry of ribosomes during in vitro translation. Journal of
Virology *62*, 2636-2643.

Jang, S. K., and Wimmer, E. (1990). Cap-independent translation of
encephalomyocarditis virus RNA; structural requirements of the internal ribosomal entry
site and involvement of a cellular 57 kD RNA binding protein. Genes & Dev *4*, 1560-1572.

Jang, Y. K., Kim, M., and Park, S. D. (2002). Fibrillarin binds to a cis-regulatory
element in pre-mRNA of uvi15+ infission yeast. Biochemical and Biophysical Research
Communications *294*, 1184-1190.

Janike, R. U., Lee, F. H. H., and Porter, A. G. (1994). Nuclear c-myc plays an
important role in the cytotoxicity of tumor- necrosis-factor-alpha in tumor-cells. Molecular
and Cellular Biology *14*, 5661-5670.

Johannes, G., and Sarnow, P. (1998). Cap-independent polysomal association of
natural mRNAs encoding c-myc, Bip, and eIF4G conferred by internal ribosome entry
sites. RNA *4*, 1500-1513.

Jopling, C. L., and Willis, A. E. (2001). N-myc translation is initiated via an
internal ribosome entry segment that displays enhanced activity in neuronal cells.
Oncogene *20*, 2664-2670.

Jordan, M., Schallhorn, A., and Wurm, F. M. (1996). Transfecting mammalian
cells: Optimization of critical parameters affecting calcium-phosphate precipitate
formation. Nucleic Acids Res *24*, 596-601.

Juin, P., Heuber, A., Littlewood, T., and Evan, G. (1999). c-Myc-induced
senstization to apoptosis is mediated through cytochrome C release. Genes and Dev *13*,
1367-1381.

Juin, P., Hunt, A., Littlewood, T., Griffiths, B., Swigart, L. B., Korsmeyer, S., and
Evan, G. (2002). c-Myc functionally cooperates with Bax to induce apoptosis. Molecular
and Cellular Biology *22*, 6158-6169.

Kaspar, R.L., Rychlik, W., White, M.W., Rhoads, R.E., and Morris, D.R. (1990).
Simultaneous cytoplasmic redistribution of ribosomal protein L32 mRNA and

phosphorylation of eukaryotic initiation factor 4E after mitogenic stimulation of swiss 3T3 cells. The Journal of Biological Chemistry *265*, 3619-3622.

Kato, G. J., Lee, W. M. F., Chen, L., and Dang, C. V. (1992). Max: functional domains and interaction with c-myc. Genes and Development *6*, 81-92.

Khaleghpour, K., Svitkin, Y. V., Craig, A. W., DeMaria, C. T., Deo, R. C., Burley, S. K., and Sonenberg, N. (2001). Translational repression by a novel partner of human poly(A) binding protein, Paip2. Molecular Cell *7*, 205-216.

Kim, Y. H. (1999). La protein is required for efficient translation driven by encephalomyocarditis virus internal ribosome entry site. Journal of General Virology *80*, 3159-3166.

Kim, Y. K., Back, S. H., Rho, J., Lee, S. H., and Jang, S. K. (2001). La autoantigen enhances translation of BiP mRNA. Nucleic Acids Res *29*, 5009-5016.

Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. The International Journal of Biochemistry and Cell Biology *31*, 25-29.

Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Leibig, H. D., Sommergruber, W., Sobrino, F., Blaas, D., Rhoads, R. E., and Skern, T. (1994). Foot-and-mouth disease virus leader proteinase: Purification of the Lb form and determination of its cleavage site on eIf-4G. Journal of Virology *68*, 5677-5684.

Klempnauer, K. H., Gonda, T. J., and Bishop, J. M. (1982). Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell *31*, 453-463.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Research *15*, 8125-8148.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. The Journal of Cell Biology *115*, 887-903.

Kozak, M. (1996). Interpreting cDNA sequences: some insights from studies on translation. Mammalian Genome *7*, 563-574.

Kozak, M. (2001). Constraints on reinitiation in mammals. Nucleic Acids Research *29*, 5226-5232.

Krishnamoorthy, T., Pavitt, G. D., Zhang, F., Dever, T. E., and Hinnsbusch, A. G. (2001). Tight binding of the phosphorylated a subunit of initiation factor 2 (eIF2a) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initation. Molecular and Cellular Biology *21*, 5018-5030.

Lavu, S., and Reddy, E. P. (1986). Structural organization and nucleotide sequence of mouse c-myb oncogene: activation in ABPL tumors is due to viral integration in an intron which results in the deletion of the 5' coding sequences. Nucleic Acids Res *14*, 5309-5320.

Le Quesne, J. P. C., Stoneley, M., Fraser, G. A., and Willis, A. E. (2001). Derivation of a structural Model for the c-myc IRES. Journal of Molecular Biology *310*, 111-126.

Le, S.-Y., and Maizel Jr, J. V. (1997). A common RNA structural motif involved in the internal initiation of cellular mRNAs. Nucleic Acids Research *25*, 362-369.

Leary, D. J., and Huang, S. (2001). Regulation of ribosome biosynthesis within the nucleolus. Febs Letters *509*, 145-150.

Leffers, H., Dejgaard, K., and Celis, J. E. (1995). Characterisation of two cellular poly(rC)-binding proteins, each containing three K-homologous(KH) domains. European Journal of Biochemistry *230*, 447-453.

Li, Q., Imataka, H., Morino, S., Rogers, G. W., Richter-Cook, N. J., Merrick, W. C., and Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. Molecular and Cellular Biology *19*, 7336-7346.

Lomakin, I. B., Hellen, C. U., and Pestova, T. V. (2000). Physical association of eukaryotic initiation factor 4G (eIF4G) with eIF4A strongly enhances binding of eIF4G to the internal ribosome entry site of encephalomyocarditis virus and is required for internal initiation of translation. Molecular and Cellular Biology *20*, 6019-6029.

Lorsch, J. R., and Herschlag, D. (1998). The DEAD Box proein eIF4A. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. Biochemistry *37*, 2180-2193.

Macejak, D. G., and Sarnow, P. (1991). Internal initiation of translation mediated by the 5' leader of a cellular mRNA. Nature *353*, 90-94.

Makeyev, A. V., Chkheidze, A. N., and Liebhaber, S. A. (1999). A set of highly conserved RNA-binding proteins, alpha CP-1 and alpha CP-2, implicated in mRNA stabilization, are coexpressed from an intronless gene and its intron-containing paralog. Journal of Biological Chemistry *274*, 24849-24857.

Makeyev, A. V., and Liebhaber, S. A. (2002). The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. RNA *8*, 265-278.

Mateyak, M. K., Obaya, A. J., Adachi, S., and Sedivy, J. M. (1997). Phenotypes of c-myc-deficient Rat Fibroblasts isolated by targeted homologous recombination. Cell Growth and differentiation *8*, 1039-1048.

McClinton, D., Stafford, J., Brents, L., Bender, T. P., and Kuehl, W. M. (1990). Differentiation of mouse erythroleukemia cells is blocked by late up-regulation of a c-myb transgene. Mol Cell Biol *10*, 705-710.

McKendrick, L., Pain, V. M., and Morley, S. J. (1999). Translation initiation factor 4E. The International Journal of Biochemistry and Cell Biology *31*, 31-35.

McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J., and Lewis, J. D. (2001). Interaction of Eukaryotic Translation Factor 4G with the Nuclear Cap-Binding Complex Provides a Link between Nuclear and Cytoplasmic Functions of the m7 Guanosine Cap. Molecular and Cellular biology *21*, 3632-3641.

Meerovitch, K., Svitkin, Y. V., Lee, H. S., Lejbkowicz, F., Kenan, D. J., Chan, E. K. L., Agol, V. I., Keene, J. D., and Sonenberg, N. (1993). La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. J Virol *67*, 3798-3807.

Michel, Y. M., Poncet, D., Piron, M., Kean, K. M., and Borman, A. M. (2000). Cap-Poly(A) synergy in mammalian cell-free extracts. The Journal of Biological Chemistry *275*, 32268-32276.

Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I., and Levens, D. (1996). Heterogeneous nuclear ribonucleoprotein K is a transcription factor. Molecular and Cellular Biology *16*, 2350-2360.

Minich, W. B., Balasta, M. L., Goss, D. J., and Rhoads, R. E. (1994). Chromatographic resolution of in vivo phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF4E: in creased cap affinity of the phosphorylated form. Proceedings of the National Academy of Science U S A *91*, 7668-7672.

Mitchell, S., Brown, E. C., Coldwell, M., Jackson, R. J., and Willis, A. E. (2001). Protein factor requirement of the Apaf-1 internal ribosome entry segment: Roles of polypyrimidine tract binding protein and upstream of N-ras. Molecular and Cellular Biology *21*, 3364-3374.

Morris, D. R., and Geballe, A. P. (2000). Upstream open reading frames as regulators of mRNA translation. Molecular and Cellular Biology *20*, 8635-8642.

Mucenski, M. L., Mclain, K., Kier, A. B., Swerdlow, S. H., Schreiner, C. M., Miller, T. A., Pietryga, D. W., Scott, W. J., and Potter, S. S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell 65, 677-689.

Nagy, E., and Rigby, W. F. C. (1995). Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD+-binding region (Rossmann Fold). The Journal of Biological Chemistry 270, 2755-2763.

Nanbru, C., Lafon, I., Audiger, S., Gensac, G., Vagner S., Huez, G., and Prats, A.-C. (1997). Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. Journal of Biological Chemistry 272, 32061-32066.

Nasi, S., Ciarapica, R., Junker.R., Rosati, J., and Soucel, L. (2001). Making decisions through Myc. Febs Letters 490, 153-162.

Nesbit, C. D., Tersak, J. M., and Prochownik, E. V. (1999). MYC oncogenes and human neoplastic disease. Oncogene 18, 3004-3016.

Ness, S. A. (1999). Myb binding proteins: regulators and cohorts in transformation. Oncogene 18, 3039-3046.

Obaya, A. J., Mateyak, M. K., and Sedivy, J. M. (1999). Mysterious liaisons: the relationship between c-Myc and the cell cycle. Oncogene 18, 2934-2941.

Oelgeschlager, M., Krieg, J., Luscher-Firzlaff, J. M., and Luscher, B. (1995). Casein Kinase II phosphorylation site mutations in c-Myb affect DNA binding and transcriptional cooperativity with NF-M. Molecular and Cellular Biology 15, 5966-5974.

Oh, I. H., and Reddy (1999). Myb gene family in cell growth, differentiation, and apoptosis. Oncogene 18, 3017-3033.

Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M., and Hentze, M. W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3'end. Cell 89, 597-606.

Pain, V. M. (1996). Initiation of protein synthesis in eukaryotic cells. European Journal of Biochemistry 236, 747-771.

Paulin, F. E. M., Chappell, S. A., and Willis, A. E. (1998). A single nucleotide change in the c-myc internal ribosome entry segment leads to enhanced binding of a group of protein factors. Nucleic Acids Research 26, 3097-3103.

Pause, A., Belsham, G. J., Gingras, A. C., Donze, O., Lin, T., Laurence, J. C., and Sonenberg, N. (1994). Insulin-dependent stimulation of protien synthesis by phosphorylation of a regulator of 5' cap function. Nature 371, 762-767.

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320-325.

Perez-Roger, I., Solomon, D. L., Sewing, L., and Land, H. (1997). Myc activation of cyclin E/Cdk2 kinase involves induction of cyclin E gene transcription and inhibition of p27(Kip1) binding to newly formed complexes. Oncogene *14*, 2373-2381.

Pestova, T. V., Borukhov, S. I., and Hellen, C. U. T. (1998a). Eukaryotic ribosomes require initiation factors 1 and 1A to locate inititation codons. Nature *394*, 854-859.

Pestova, T. V., Hellen, C. U. T., and Shatsky, I. N. (1996a). Canonical Eukaryotic Initiation Factors Determine Initiation of Translation by Internal Ribosomal Entry. Molecular and Cellular Biology *16*, 6859-6869.

Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. (2001). Molecular mechanisms of translation initiation in eukaryotes. Proceedings of the National Academy of Science U S A *98*, 7029-7036.

Pestova, T. V., Lomakin, I. B., Lee, J. H., Choi, S. K., Dever, T. E., and Hellen, C. U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. Nature *403*, 332-335.

Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J., and Hellen, C. U. (1998b). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. Genes and Development *12*, 67-83.

Pestova, T. V., Shatsky, I. N., and Hellen, C. U. T. (1996b). Functional Dissection of Eukaryotic Initiation Factor 4F: the 4A Subunit and the Central Domain of the 4G Subunit Are Sufficient To Mediate Internal Entry of 43S Preinitiation Complexes. Molecular and Cellular Biology *16*, 6870-6878.

Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I., and Hellen, C. U. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. Genes and Development *14*, 2028-2045.

Pinkstaff, J. K., Chappell, S. A., Mauro, V. P., Edelman, G. M., and Krushel, L. A. (2001). Internal initiation of translation of five dendritically localized neuronal mRNAs. PNAS *98*, 2770-2775.

Poulin, F., Gingras, A. C., Olsen, H., Chevalier, S., and Sonenberg, N. (1998). 4E-BP3, a new member of the Eukaryotic initiation factor 4E-binding protein family. The Journal of Biological Chemistry *273*, 14002-14007.

Prendergast, G. C. (1999). Mechanisms of apoptosis by c-myc. Oncogene *18*, 2967-2987.

Pyronnet, S., Dostis, J., and Sonenberg, N. (2001). Suppression of cap-dependent translation in mitosis. Genes and Development *15*, 2083-2093.

Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. (1999). Human eukaryotic translation initiation factor 4G(eIF4G) recruits Mnk1 to phoshorylate eIF4E. The EMBO Journal *18*, 270-279.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). A Cell Cycle-Dependent Internal Ribosome Entry Site. Molecular Cell *5*, 607-616.

Rabbitts, P. H., Watson, J. V., Lamond, A., Forster, A., Stinson, M. A., Evan, G., Fischer, W., Atherton, E., Sheppard, R., and Rabbitts, T. H. (1985). Metabolism of c-myc gene products: c-myc mRNA and protein expression in the cell cycle. The EMBO Journal *4*, 2009-2015.

Ramos, R., and Martinez-Salas, E. (1999). Long-range RNA interactions between structural domains of the aphthovirus internal ribosome entry site (IRES). RNA *5*, 1374-1383.

Ramsay, R. G., Ikeda, K., Rifkind, R. A., and Marks, P. A. (1986). Changes in gene expression associated with induced differentiation of erythroleukemia: protooncogenes, globin genes, and cell division. Proc Natl Acad Sci U S A *83*, 6849-6853.

Reynolds, J. E., Kaminski, A., Kettinen, H. J., Grace, K., Clarke, B. E., Carroll, A. R., Rowlands, D. J., and Jackson, R. J. (1995). Unique features of internal initiation of hepatitis C virus RNA translation. EMBO J *14*, 6010-6020.

Rhoads, R. E. (1999). Signal transduction pathways that regulate Eukaryotic protein synthesis. The Journal of Biological Chemistry *274*, 30337-30340.

Richter-Cook, N. J., Dever, T. E., Hensold, J. O., and Merrick, W. C. (1998). Purification and characterisation of a new Eukaryotic protein translation factor. The Journal of Biological Chemistry *273*, 7579-7587.

Rijnbrand, R., Bredenbeek, P. J., Haasnoot, P. C., Kieft, J. S., and Spaan, W. J. M. (2001). The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C and other flaviviral RNAs. RNA *7*, 585-597.

Rijnbrand, R., vanderStraaten, T., vanRijin, P., Spaan, W., and Bredenbeek, P. (1997). Internal ribosome entry of ribosomes is directed by the 5'non-coding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. J Virol *71*, 451-457.

Robertson, M. E. M., Seamons, R. A., and Belsham, G. J. (1999). A selection system for functional internal ribosome entry site (IRES) elements: Analysis of the requirement for a conserved GNRA tetraloop in the encephalomyocarditis virus IRES. RNA *5*, 1167-1179.

Ryan, K. M., and Birnie, G. D. (1996). Myc oncogenes: the enigmatic family. Biochemical Journal *314*, 713-721.

Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Tonegawa, S. (1983). Activation of the *c-myc* gene by translocation: a model for translational control. Proceedings of the National Academy of Sciences, USA *80*, 7476-7460.

Sakamuro, D., and Prendergast, G. C. (1999). New Myc-interacting proteins: a second Myc network emerges. Oncogene *18*, 2942-2954.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989). Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. Proc Natl Acad Sci U S A *86*, 5758-5762.

Sarnow, P. (1989). Translation of glucose-regulated protein 78/immunoglobulin heavy-chain binding protein mRNA is increased in poliovirus-infected cells at a time when cap-dependent translation of cellular mRNA is inhibited. Proc Natl Acad Sci USA *86*, 5795-5799.

Sasaki, J., and Nakashima, N. (1999). Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus *in vitro*. J Virol *73*, 1219-1266.

Sasaki, J., and Nakashima, N. (2000). Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. PNAS *97*, 1512-1515.

Savino, T. M., Bastos, R., Jansen, E., and Hernandez-Verdun, D. (1999). The nucleolar antigen Nop52, the human homologue of the yeast ribosomal RNA processing RRP1, is recruited at late stages of nucleologenesis. Journal of cell science *112*, 1889-1900.

Scheper, G. C., Morrice, N., Kleijn, M., and Proud, C. G. (2001). The mitogen-activated protein kinase Mnk2 is a eukaryotic initiation factor 4E kinase with high levels of basal activity in mammalian cells. Molecular and Cellular Biology *21*, 743-754.

Scheper, G. C., van Kollenburg, B., Hu, J., Luo, Y., Goss, D. J., and Proud, C. G. (2002). Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA. The Journal of Biological Chemistry *277*, 3303-3309.

Schmidt, E. (1999). The role of c-myc in cellular growth control. Oncogene *18*, 2988-2996.

Schultz, D. E., Hardin, C. C., and Lemon, S. M. (1996). Specific interaction of glyceraldehyde 3-phosphate dehydrogenase with the 5'-nontranslated RNA of Hepatitis A virus. The Journal of Biological Chemistry *271*, 14134-14142.

Sella, O., Gerlitz, G., Le, S., and Orna, E. (1999). Differentiation-Induced Internal Translation of c-sis mRNA: Analysis of the cis Elements and Their Differentistion-Linked Binding to the hnRNP C Protein. Molecular and Cellular Biology *19*, 5429-5440.

Shatkin, A. J. (1976). Capping of Eucaryotic mRNAs. Cell *9*, 645-653.

Si, K., Chaudhuri, J., Chevesich, J., and Maitra, U. (1997). Molecular cloning and functional expression of a human cDNA encoding translation initiation factor 6. Proceedings of the National Academy of Science U S A *94*, 14285-14290.

Silvera, D., Gamarnik, A. V., and Andino, R. (1999). The N-terminal K homology domain of the Poly(rC)-binding protein is a major determinant for binding to the poliovirus 5'-untranslated region and acts as an inhibitor of viral translation. Journal of Biological Chemistry *274*, 38163-38170.

Simons, M., Edelman, G. M., DeKeyser, J., Langer, R., and Rosenberg, R. D. (1992). Antisense c-myb oligonucleotides inhibit intimal arterial smooth muscle cell accumulation in vivo. Nature *359*, 67-70.

Simons, M., Morgan, K. G., Parker, C., Collins, E., and Rosenberg, R. D. (1993). The proto-oncogene cmyb mediates an intracellular calcium rise during the late G1 phase of the cell cycle. The Journal of Biological Chemistry *268*, 627-632.

Spahn, C. M. T., Kieft, J. S., Grassucci, R. A., Penczek, P. A., Zhou, K., Doudna, J. A., and Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in conformation of the 40S ribosomal subunit. Science *291*, 1959-1962.

Spangberg, K., and Schwartz, S. (1999). Poly(C)-binding prtoein interacts with the hepatitis C virus 5' untranslated region. Journal of General Virology *80*, 1371-1376.

Spotts, G. D., Patel, S. V., Xiao, Q., and Hann, S. R. (1997). Identification of downstream-initiated c-myc proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. Molecular and Cellular Biology *17*, 1459-1468.

Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z., and Keshet, E. (1998). Translation of vascular endothelial growth factor mRNA by internal ribosome entry: Implications for translation under hypoxia. Molecular and Cellular Biology *18*, 3112-3119.

Stewart, S., and Semler, B. L. (1997). RNA determinants of picornavirus cap-independent translation initiation. Seminars in Virology *8*, 242-255.

Stoneley, M., Chappell, S., Jopling, C., Dickens, M., MacFarlane, M., and Willis, A. (2000a). c-Myc protein synthesis is initiated from the internal ribosome entry segment during apotosis. Molecular and Cellular Biology *20*, 1162-1169.

Stoneley, M., Paulin, F. E. M., Le Quesne, J. P. C., Chappell, S. A., and Willis, A. E. (1998). C-myc 5' untranslated region contains an internal ribosome entry segment. Oncogene *16*, 423-428.

Stoneley, M., Subkhankulova, T., Le Quesne, J., Coldwell, M., Jopling, C., Belsham, G., and Willis, A. (2000b). Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. Nucleic Acids Research *28*, 687-694.

Subkhankulova, T., Mitchell, S. A., and Willis, A. E. (2001). Internal ribosome entry segment-mediated initiation of c-myc protein synthesis following genotoxic stress. Journal of Biochemistry *359*, 183-192.

Tanikawa, J., Yasukawa, t., Enari, M., Ogata, K., Nishimura, y., Ishii, S., and Sarai, A. (1993). Recognition of specific dna-sequences by the c-myb protooncogene product - role of 3 repeat units in the dna-binding domain. Proc Natl Acad Sci U S A *90*, 9320-9324.

Thisted, T., Lyakhov, D. L., and Liebhaber, S. A. (2001). Optimized RNA targets of two closely related triple KH domain proteins,heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. Journal of Biological Chemistry *276*, 17484-17496.

Vagner, S., Galy, B., and Pyronnet, S. (2001). Irresistable IRES. EMBO Reports *2*, 893-898.

Vennstrom, B., Sheiness, D., Zabielski, J., and Bishop, J. M. (1982). Isolation and Characterization of c-*myc*, a Cellular Homolog of the Oncogene (v-*myc*) of Avian Myelocytomatosis Virus Strain 29. J Virol *42*, 773-779.

Waggoner, S., and Sarnow, P. (1998). Viral ribonucleoprotein complex formation and nucleolar-cytplasmic relocalization of Nucleolin in poliovirus-infected cells. Journal of Virology *72*, 6699-6709.

Walker, B. L. (1999). Differential utilization of poly (rC) binding protein 2 in translation driven by picornavirus IRES elements. RNA *5*.

Wang, C., Le, S., Ali, N., and Siddqui, A. (1995). A RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' non-coding region. RNA *1*, 526-537.

Wang, X., Paulin, F. E. M., Campbell, L. E., Gomez, E., O'Brien, K., Morrice, N., and Proud, C. G. (2001). Eukaryotic initiation factor 2B: Identification of multiple phosphorylation sites in the E-subunit and their functions in vivo. The EMBO Journal *20*, 4349-4359.

Waskiewicz, A. J., Johnson, J. C., Penn, B., Mahalingam, M., Kimball, S. R., and Cooper, J. A. (1999). Phosphorylation of the cap-binding protein Eukaryotic translation initiation factor 4E by protein kinase Mnk1 in vivo. Molecular and Cellular Biology *19*, 1871-1880.

Weinstein, Y., Ihle, J. N., Lavu, S., and Reddy, E. P. (1986). Truncation of the c-myb gene by a retroviral integration in an interleukin 3-dependent myeloid leukemia cell line. Proc Natl Acad Sci U S A *83*, 5010-5014.

Wilson, J. E., Pestova, T. V., Hellen, C. U. T., and Sarnow, P. (2000a). Initiation of protein synthesis from the A site of the ribosome. Cell *102*, 511-520.

Wilson, J. E., Powell, M. J., Hoover, S. E., and Sarnow, P. (2000b). Naturally occuring dicistronic Cricket Paralysis Virus RNA is regulated by two internal ribosome entry sites. Mol Cell Biol *20*, 4990-4999.

Witherell, G. W., Gil, A., and Wimmer, E. (1993). Interaction of Polypyrimidine Tract Binding protein with the Encephalomyocarditis Virus mRNA Internal Ribosome Entry Site. Biochemistry *32*, 8268-8275.

Woo, C. H., Sopchak, L., and Lipsick, J. S. (1998). Overexpression of an alternatively spliced form of c-Myb results in increases in transactivation and transforms avian myelomonoblasts. Journal of Virology *72*, 6813-6821.

Yamanaka, S., Zhang, X., Maeda, M., Miura, K., Wang, S., Farese Jr, R. V., Iwao, H., and Innerarity, T. L. (2000). Essential role of NAT1/p97/DAP5 in embryonic differentiation and the retinoic acid pathway. The EMBO Journal *19*, 5533-5541.

Yang, G., and Sarnow, P. (1997). Location of the internal ribosome entry site in the 5' non-coding region of the immunoglobulin heavy-chain binding protein (BiP) mRNA: Evidence for specific RNA-protein interactions. Nucleic Acids Research *25*, 2800-2807.

Ye, X., Fong, P., Iizuka, N., Choate, D., and Cavener, D. R. (1997). Ultrabithorax and Antennapedia 5' untranslated regions promote developmentally regulated internal translation. Mol Cell Biol *17*, 1714-1721.

Yi, M., Schultz, D. E., and Lemon, S. M. (2000). Functional Significance of the Interaction of Hepatitis A Virus RNA with Glyceraldehyde 3-Phoshate Dehydrogenase (GAPDH): Opposing Effects of GAPDH and Polypyrimidine Tract Binding Protein on Internal Ribosome Entry Site Function. Journal of Virology *74*, 6459-6468.

Yuan, X., Davydova, N., Conte, M. R., Curry, S., and Matthews, S. (2002). Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. Nucleic Acids Research *30*, 456-462.

Yueh, A., and Schneider, R. J. (2000). Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. Genes and Development *14*, 414-421.

Publication

Pickering, B.M., Mitchell, S.A., Evans, J.R., Willis, A.E. (2003). Polypyrimidine tract binding protein and poly (rC) binding protein 1 interact with the BAG-1 IRES and stimulate its activity *in vitro* and *in vivo*. Nucleic.Acids.Res.31(2), 1-8. 639-646.

Polypyrimidine tract binding protein and poly r(C) binding protein 1 interact with the BAG-1 IRES and stimulate its activity *in vitro* and *in vivo*

Becky M. Pickering, Sally A. Mitchell, Joanne R. Evans and Anne E. Willis*

Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK

ABSTRACT

The 5′-untranslated region of Bag-1 mRNA contains an internal ribosome entry segment (IRES) and the translation of Bag-1 protein can be initiated by both cap-dependent and cap-independent mechanisms. In general, cellular IRESs require non-canonical *trans*-acting factors for their activity, however, very few of the proteins that act on cellular IRESs have been identified. Proteins that interact with viral IRESs have also been shown to stimulate the activity of cellular IRESs and therefore the ability of a range of known viral *trans*-acting factors to stimulate the Bag-1 IRES was tested. Two proteins, poly r(C) binding protein 1 (PCBP1) and polypyrimidine tract binding protein (PTB), were found to increase the activity of the Bag-1 IRES *in vitro* and *in vivo*. The regions of the Bag-1 IRES RNA to which they bind have been determined, and it was shown that PCBP1 binds to a short 66 nt section of RNA, whilst PTB interacts with a number of sites over a larger area. The minimum section of the RNA that still retained activity was determined and both PCBP1 and PTB interacted with this region suggesting that these proteins are essential for Bag-1 IRES function.

INTRODUCTION

The human Bag-1 gene (Bcl-2 associated athanogene) encodes three major isoforms generated from a single transcript, p50, p46 and p36 (BAG-1L, BAG-1M and BAG-1S respectively) that differ at their N-termini. BAG-1L, which initiates from a non-canonical CUG codon, contains an SV40-like nuclear localisation signal (NLS) at its N-terminus, and this is thought to be responsible for the mainly nuclear distribution of this isoform (1). The AUG-initiated isoforms, BAG-1M and BAG-1S, lack this NLS and as such are localised predominantly cytoplasmically (1,2). The variation in the N-termini of the Bag-1 proteins also leads to different protein binding specificities and the isoforms have diverse cellular roles. Bag-1 was originally identified as RAP46 (receptor associated protein) through its interaction with the glucocorticoid receptor (3) and has been found to

associate with numerous other members of the steroid hormone receptor superfamily. It was also documented as HAP46 (Hsc70/Hsp70 associated protein) through its interaction with the 70 kDa heat-shock proteins (4,5) where it has a role as a co-chaperone in the protein folding response (6–8). The murine homologue, Bag-1, was identified through its interaction with Bcl-2, an anti-apoptotic gene (9). Bag-1 has been shown to interact with Bcl-2 and to promote the anti-apoptotic properties of this protein, blocking a step in the apoptotic pathway (10).

It was originally suggested that the three main isoforms of Bag-1, in addition to a minor isoform, p29, are translated by leaky scanning (1), however, expression of Bag-1 *in vivo* does not support this hypothesis (1,2,11). Recently we have demonstrated that the p36 isoform of Bag-1 can be translated by internal ribosome entry through use of an internal ribosome entry segment (IRES) in addition to the cap-dependent scanning mechanism (11). IRES elements are used to initiate translation under conditions where cap-dependent scanning is compromised (12) and our data suggest that the IRES of Bag-1 is required to maintain translation of the p36 isoform following heat shock (11).

The internal ribosome entry mechanism of translation was first identified in picornaviruses and these have been widely studied in terms of structure, mechanisms and *trans*-acting factor requirements (13). Viral IRESs vary widely in their dependence on *trans*-acting factors. For example, some viral IRESs such as the encephalomyocarditis virus (14,15), in contrast, other viral IRESs such as polio virus or the human rhino virus (HRV) require the addition of extracts derived from HeLa cells to *in vitro* systems before they are active (16–18). In the case of HRV, two proteins that are required have been identified, upstream of N-ras (unr) (19), and polypyrimidine tract binding protein (PTB) (20). A number of additional viral IRES binding/ activating proteins have been identified, including the La autoantigen which is used by polio virus IRES (21), and poly r(C) binding protein 2 (PCBP2) which binds to polio virus IRES (22) and has been shown to activate entero/rhino virus IRESs *in vitro* (23). These proteins are thought to act as RNA chaperones to either maintain or aid the RNA to form a structure that is competent for ribosome recruitment.

Recently, a large number of cellular IRESs have been identified but the mechanisms by which they initiate translation are poorly understood (12). The protein factor

*To whom correspondence should be addressed. Tel: +44 116 2523362; Fax: +44 116 2523369; Email: aew5@le.ac.uk

requirements for cellular IRESs are much less well defined although data produced thus far would suggest that each IRES has a requirement for a specific set of *trans*-acting factors. First, cellular IRESs show considerable cell tropism and this would suggest that levels of endogenous *trans*-acting factors vary between cell lines (24–26). Secondly, the proteins that interact with two cellular IRESs studied so far are different, thus the Apaf-1 IRES requires PTB and unr for function *in vitro* and *in vivo* (27), whilst the XIAP IRES has a requirement for La (28).

In this study the protein factor requirements for efficient Bag-1 IRES activity has been investigated. We demonstrate a direct and specific interaction of PTB and poly r(C) binding protein 1 (PCBP1) with the Bag-1 IRES, which stimulates IRES function both *in vitro* and *in vivo*.

MATERIALS AND METHODS

Plasmid constructs

The plasmids pRF and pRBF harbouring deletion segments of Bag-1 are described in Figures 1A, 2A and 4A. The plasmid pSKL is based upon the vector pSK+bluescript (Stratagene); the Bag-1 5'-untranslated region (5'-UTR) or deletions were cloned into this vector in frame with the firefly luciferase gene (Figs 3A and 5A). Deletion fragments were generated by PCR using specific primers to the regions required. For expression of proteins used, the cDNAs were present in PET28a vectors, enabling expression of protein in *Escherichia coli* and subsequent purification of the protein. For expression in tissue culture cells, the cDNAs were subcloned into pCDNA3.1 and for expression in insect cells (for purification of protein) subcloned into pBlueBac4 (Invitrogen).

Protein expression

Proteins were overexpressed in *E.coli* from the pET28a vector by the addition of isopropyl-β-D-thiogalactopyranoside to the growth medium. The proteins that contained a His tag were purified using a nickel affinity column (Qiagen). Alternatively, unr was purified from cultures of Sf9 cells that had been infected with a recombinant baculovirus expressing unr-His (Invitrogen). Cells were harvested and lysed in phosphate-buffered saline containing 0.1% Triton X-100, and the tagged protein purified on a nickel affinity column.

Cell culture and transient transfections

Cells were typically grown in Dulbecco's modified Eagle's medium (Gibco-BRL) containing 10% fetal calf serum, under humidified atmosphere containing 5% CO_2. The cell lines CALU1, CAL51 and CAMA1 were a kind gift from Dr G. Packham (CRUK Unit, Southampton, UK). MCF7 and HeLa were originally purchased from ATCC. Cells were transfected using FuGene 6 (Roche) as specified by the manufacturer. Alternatively, calcium phosphate-mediated transfections were performed as described, with minor modifications (29). Lysates were prepared from transfected cells using 1× passive lysis buffer. Firefly and *Renilla* luciferase activities were measured using the 'Stop and glo' dual-luciferase reporter assay system (Promega) according to the manufacturer's instructions with the exception that only 25 µl

of each reagent was used. Light emission was measured over 10 s using an OPTOCOMP I luminometer. Activity of the β-galactosidase transfection control was measured using a Galactolight Plus assay system (Tropix). All transfections were carried out in triplicate on at least three independent occasions.

In vitro transcription reactions

Vector DNA was linearised by restriction digestion using a site downstream of the region of interest (*Hpa*I for dicistronic, *Nco*I for monocistronic); transcripts were synthesised in a reaction mixture containing 1× transcription buffer [40 mM HEPES–KOH (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol (DTT), 10 mM NaCl], 40 U RNAguard or RNasin, 1 mM ATP, 1 mM UTP, 1 mM CTP, 0.5 mM GTP, 1 µM $m^7G(5')ppp(5')G$, 1 µg of DNA template and 20 U T7 or T3 RNA polymerase to a final volume of 50 µl. For radiolabelled RNAs, CTP was replaced with 50 µCi [α-^{32}P]CTP. The reaction mixture was incubated at 37°C for 1.5 h and the RNA purified.

In vitro translation reactions

The Promega rabbit reticulocyte flexi-lysate *in vitro* translation system was primed with 5 ng/µl RNA and used according to the manufacturer's instructions. The reaction was performed in a final volume of 12.5 µl and 0.1 µg of each protein was added where appropriate. Luciferase activities were assayed as described above, and the firefly and *Renilla* values expressed relative to the control plasmid pRF, which was assigned a value of 1. All experiments were performed in triplicate on at least three independent occasions.

UV-crosslinking analysis

Radiolabelled transcript was generated from pSKBL linearised with *Nco*I. Approximately 2.5 pmol per reaction was incubated with 0.25 µg of protein in 1× UV-crosslinking buffer [10 mM HEPES (pH 7.4), 3 mM $MgCl_2$, 100 mM KCl, 5 mM creatine phosphate, 1 mM DTT, 1 mM ATP, 6% glycerol, 0.1 µg/µl tRNA] for 15 min at room temperature. For competition assays, unlabelled competitor RNAs were added with labelled RNA. The reaction mixtures were UV irradiated using a 305 nm UV light source for 30 min on ice. RNase A and RNase V1 (0.2 mg/ml) were added to the mixture to degrade any unprotected RNA by incubation at 37°C for 30 min. Sample buffer was added and the samples separated on a 10% polyacrylamide gel by SDS–PAGE. Gels were dried at 80°C under vacuum for 2 h and analysed on a Molecular Dynamics Phosphorimager.

Electrophoretic mobility shift assays (EMSAs)

Approximately 20 pmol of RNA was incubated with protein as appropriate in a buffer mix containing 40 U RNAguard, 2 µl of 5× transcription buffer [200 mM Tris—HCl (pH 8.0), 40 mM $MgCl_2$, 10 mM spermidine, 250 mM NaCl, 50 mM DTT, 15 µg tRNA], and 2 µl of 10 mM ATP in a reaction volume of 15 µl for 10 min at room temperature. DNA loading dye was added (50% sucrose, 0.1 M EDTA, 0.2% bromophenol blue) and samples loaded onto 0.7× Tris-borate-EDTA (TBE) agarose gels. Samples were electrophoresed at 100 V for ~3 h in 1× TBE loading buffer. All buffers and loading dyes were
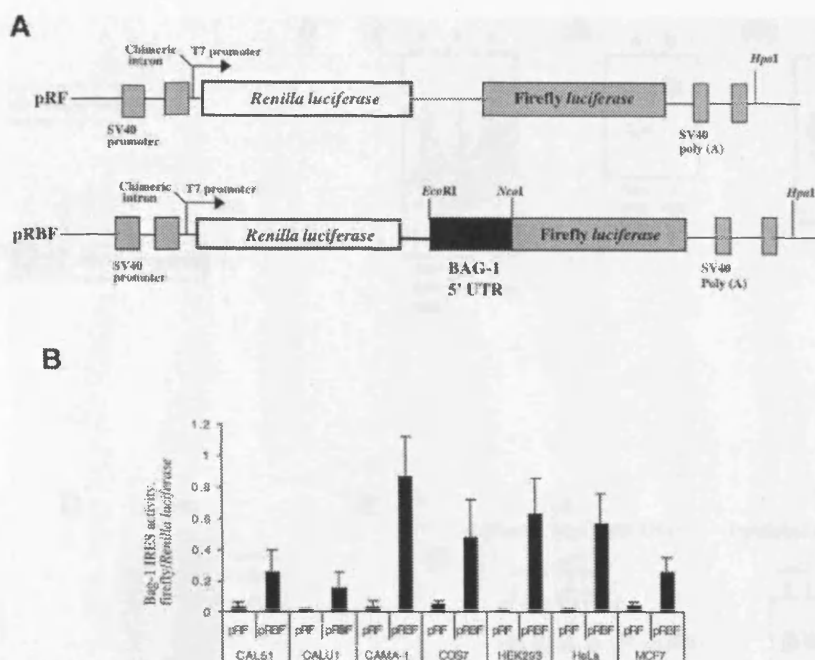
Figure 1. A comparison of the efficiency of Bag-1 IRES-mediated translation in cell lines of different origin. (**A**) Schematic representation of the dicistronic reporter constructs pRF and pRBF where pRBF contains the Bag-1 5′-UTR inserted into the vector pRF and fused in-frame with the firefly luciferase gene. (**B**) The plasmids pRF and pRBF were transfected into the cell lines indicated. CAL51, CAMA-1 and MCF7s are of human breast carcinoma origin; CALU1, human lung cancer; COS7, a monkey epithelial cell line (CV-1) immortalised with SV40 DNA; HEK293, a human embryonic kidney cell line immortalised with adenovirus; and HeLa S3, of human cervical epitheloid carcinoma origin. IRES activity was expressed as the ratio of downstream cistron expression to upstream cistron expression (firefly/*Renilla* luciferase), with any differences in transfection efficiencies corrected for using the β-galactosidase transfection control (PJ7*lac*Z). Error bars indicate standard deviations as determined from at least three independent experiments performed in triplicate.

filter-sterilised. The gels were dried under vacuum at 60°C for 2 h and exposed on a phosphoimager.

RESULTS

Comparison of Bag-1 IRES-mediated internal initiation between cell types

The majority of cellular IRESs studied thus far show considerable cell tropism in that they do not work efficiently in all cell types (24–26). This is presumably because the expression of specific IRES *trans*-acting factors varies between cell lines (24–26). The ability of the Bag-1 IRES to function in a variety of cell lines was tested by transfecting HeLa, COS7, HEK293, MCF7, CAL51, CALU1 and CAMA1 cell lines with the dicistronic construct pRBF or the control construct pRF. The expression from both *Renilla* and firefly luciferase cistrons was assayed and normalised to the transfection control β-galactosidase. The efficiency of the IRES is represented as a ratio of firefly luciferase to *Renilla* luciferase expression from pRBF (Fig. 1A). As expected, the Bag-1 IRES showed a wide range of activities in different cell types and firefly luciferase activity was found to vary considerably according to cell line (Fig. 1B). The Bag-1 IRES was highly active in CAMA-1 and active in HeLa cells, COS7 and HEK293 cells, yet relatively inactive in MCF7, CAL51 and CALU1 cells (Fig. 1B). The expression of the Bag-1 IRES in these cell lines differed from that observed with

the c-*myc* IRES (25) and the Apaf-1 IRES (30), which would again suggest that cellular IRESs have different requirements for *trans*-acting factors.

PCBP1 and PTB bind to the Bag-1 IRES

EMSAs were performed to identify putative Bag-1 IRES *trans*-acting factors by using a range of known IRES interacting proteins including PCBP1 and PCBP2 (22), PTB (20), DAP5 (31,32), La (28) and unrip (19). Radiolabelled Bag-1 IRES RNA was generated from *in vitro* transcription reactions primed with DNA derived from the monocistronic constructs (Fig. 2A); the resulting RNA was incubated with protein and the products separated on 0.5% TBE agarose gels. Only when BAG-1 IRES RNA was incubated with PTB or PCBP1 was a decrease in the mobility of the RNA observed (Fig. 2B and B.M.Pickering and A.E.Willis, unpublished data), suggesting both of these proteins bind the IRES directly. No difference in mobility of Bag-1 IRES RNA was observed with the other proteins tested, for example La (Fig. 2B, i, lane b). To test the specificity of this interaction, the proteins were also incubated with a non-specific RNA segment from glyceraldehyde-3-phosphate dehydrogenase (G3PDH) of approximately the same size. No alterations in mobility of this RNA were observed with any of the proteins tested (Fig. 2B, iii). To confirm this interaction, UV-crosslinking analysis was performed. Thus, radiolabelled Bag-1 IRES RNA was incubated with PCBP1 and/or PTB, samples were
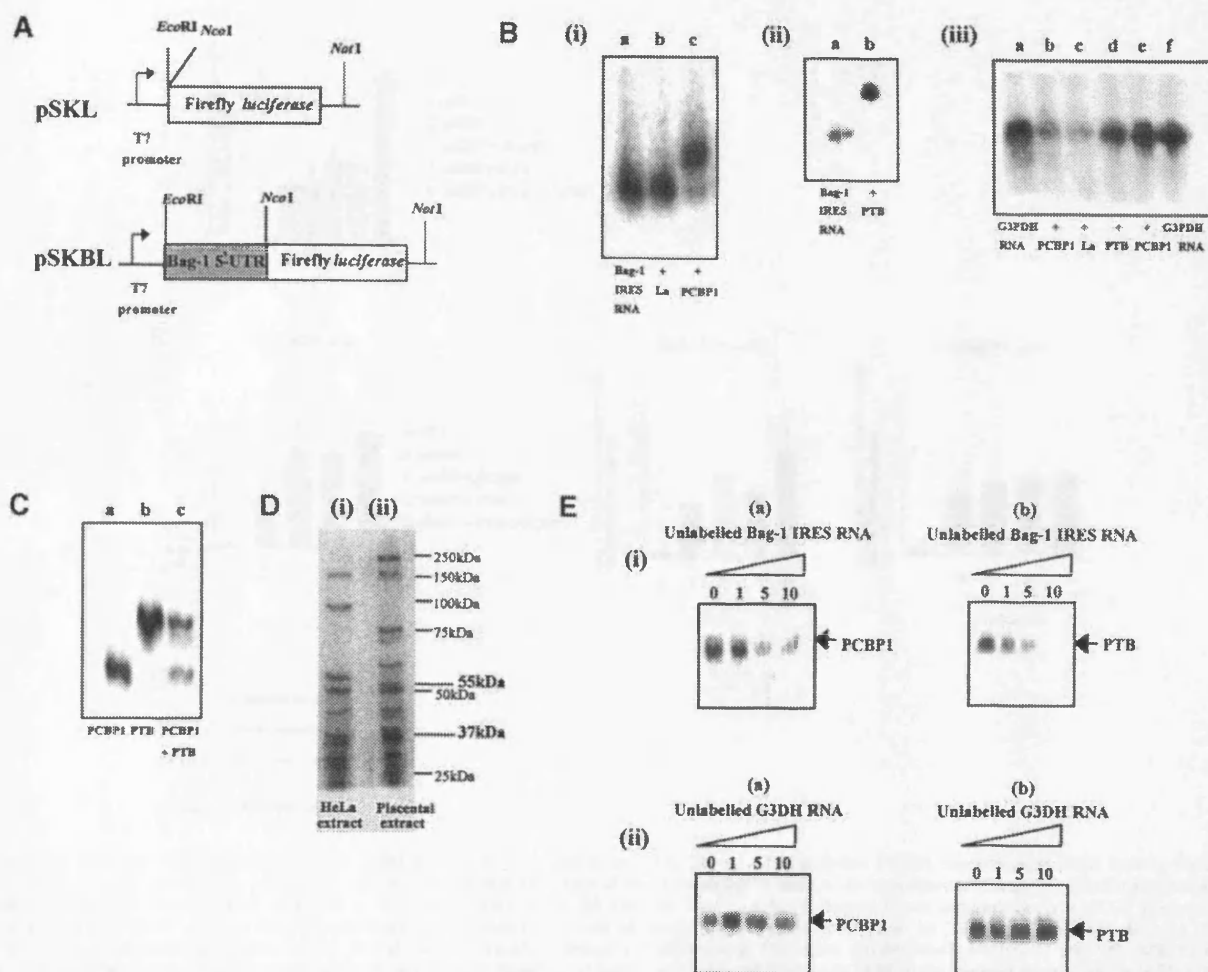
Figure 2. PTB and PCBP1 bind directly to Bag-1 IRES RNA. (**A**) Schematic diagram of the monocistronic plasmids pSKL and pSKBL where pSKBL contains the Bag-1 5′-UTR fused in-frame with the luciferase gene. (**B**) (i and ii) EMSAs of radiolabelled Bag-1 IRES RNA alone (i and ii, lane a) and in combination with 0.2 µg of protein (i, lane b), La (i, lane c PCBP1). A gel retardation is observed with PTB and PCBP1, suggesting complex formation, but not with La. (iii) Control EMSAs were carried out with non-specific radiolabelled RNA of the same size from G3PDH (lanes a and f, G3PDH RNA alone). No gel retardation is observed with 0.2 µg of any protein tested (lanes b–e). (**C**) UV-crosslinking assay of radiolabelled Bag-1 IRES RNA in combination with 0.2 µg of PTB and/or PCBP1, again showing both proteins binding to Bag-1 RNA. (**D**) UV-crosslinking assay of radiolabelled Bag-1 IRES RNA with cell extracts. (Lane i) HeLa extract and (lane ii) placental extract. Bands corresponding in size to PTB (55 kDa) and PCBP1 (37 kDa) can be observed in both lanes, marked in bold type. Sizes and position of markers are indicated in plain type. There is also a marker at 37 kDa. (**E**) (i) UV-crosslinking assay of radiolabelled Bag-1 IRES RNA in combination with (a) 0.2 µg of PCBP1, where binding is competed by the addition of a 5× molar excess of unlabelled Bag-1 IRES RNA and (b) 0.2 µg of PTB, where binding is competed by the addition of an equimolar quantity of unlabelled Bag-1 IRES RNA. (ii) UV-crosslinking assay of radiolabelled Bag-1 RNA in combination with (a) 0.2 µg of PCBP1 or (b) 0.2 µg of PTB, where there is no competition observed with a 10× molar excess of unlabelled G3DH RNA.

exposed to UV light, any RNA not bound to protein digested with RNases and the products separated by PAGE. Both PTB and PCBP1 either singly or in combination interacted with Bag-1 IRES RNA (Fig. 2C). Crosslinking analysis was also undertaken with cell extracts made from HeLa or placenta. Thus, radiolabelled Bag-1 IRES RNA was incubated with HeLa or placental extracts, exposed to UV light and unbound RNA was digested with RNases. The products were separated by PAGE. A number of proteins that interact with the Bag-1 IRES were identified in both extracts with sizes of approximately 150, 55, 45, 37, 30 and 29 kDa in addition to some

extract-specific proteins (Fig. 2D). The 55 and 37 kDa proteins are the same size as PTB and PCBP1, respectively. The other proteins remain to be identified.

To determine the specificity of the interaction between the Bag-1 IRES and these proteins, UV-crosslinking experiments were performed in the presence of excess unlabelled Bag-1 IRES RNA or G3PDH mRNA (Fig. 2E). Both proteins bind specifically to the Bag-1 IRES. Hence, there was a reduction in the binding of protein to the radiolabelled transcripts with a 1-fold molar excess of unlabelled BAG-1 IRES RNA, but not with a 10-fold molar excess of G3PDH RNA (Fig. 2E, i and ii).
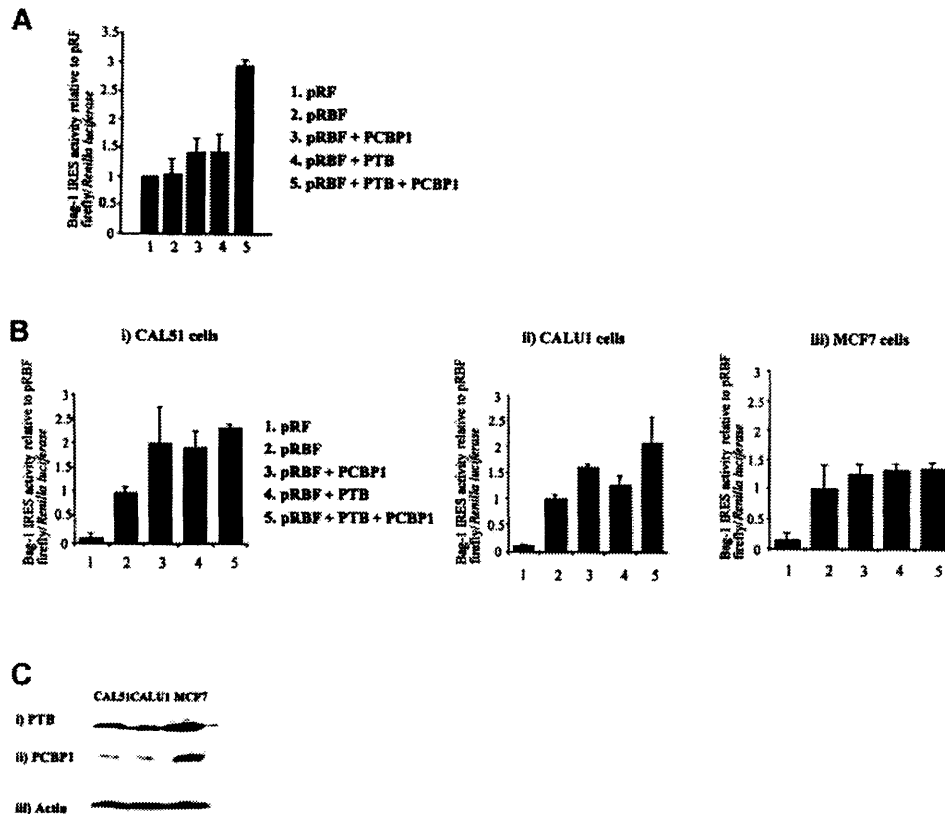
Figure 3. PTB and PCBP1 stimulate Bag-1 IRES activity *in vitro* and *in vivo*. (**A**) 200 ng of PTB and/or PCBP1 can stimulate IRES activity from the dicistronic plasmid pRBF (Fig. 1A). IRES activity is expressed as a ratio of the downstream cistron to the upstream cistron (firefly/*Renilla* luciferase), in rabbit reticulocyte lysates primed with 100 ng of capped pRBF RNA. (**B**) The cell lines (i) CAL51 (human breast carcinoma), (ii) CALU1 (human lung cancer) and (iii) MCF7 (human breast carcinoma) were co-transfected with plasmids containing the pRBF (Fig. 1A) and/or PCBP1/PTB. Activity of the Bag-1 IRES is expressed using the ratio of downstream cistron expression to upstream cistron expression (firefly/*Renilla* luciferase) with any differences in transfection efficiencies corrected for using the β-galactosidase transfection control and expressed relative to pRBF alone to show the increase in IRES activity produced by each *trans*-acting factor. Error bars indicate standard deviations as determined from at least three independent experiments performed in triplicate. (**C**) Western blots of cell lysates for endogenous protein levels with (i) anti-PTB antibody, (ii) anti-PCBP1 antibody and (iii) anti-actin antibody as a loading control show a correlation with the level of activation shown by transfection of each protein in each cell line.

Crosslinking analysis was also performed with La, unrip, DAP5 and PCBP2 but no protected RNA was detected (data not shown).

PCBP1 and PTB stimulate the Bag-1 IRES *in vitro* and *in vivo*

In general, cellular IRESs work very inefficiently (if at all) *in vitro*, but we have shown that it is possible to stimulate certain cellular IRESs by the addition of known viral *trans*-acting factors (27). The activity of the Bag-1 IRES in a dicistronic RNA (generated from pRBF and pRF; Fig. 1A) was tested in the rabbit reticulocyte lysate system with the addition of PCBP1 and PTB. The Bag-1 IRES functioned very inefficiently *in vitro* and no appreciable luciferase activity was detected over that produced from RNA derived from the vector pRF which does not contain an IRES (Fig. 1A). However, activation was observed with PTB and PCBP1, when added singly, each stimulating IRES activity to 1.4-fold (Fig. 3A). Moreover, addition of PCBP1 and PTB had an

additive effect in combination producing 3-fold stimulation of the IRES (Fig. 3A).

To test whether these proteins could stimulate the function of the Bag-1 IRES *in vivo*, co-transfections were carried out using pRBF with plasmids expressing PCBP1 or PTB either singly or in combination into the cell lines which showed low Bag-1 IRES activity (Fig. 1). Thus, transfection of CAL51 and CALU1 with either PTB or PCBP1 alone had a stimulatory effect, which was additive when the plasmids containing DNA encoding these proteins were transfected in combination (Fig. 3C, i and ii). In contrast, transfection of MCF7s with PTB and/or PCBP1 did not significantly stimulate Bag-1 IRES activity (Fig. 3C, iii). To test whether there was a correlation between the expression of PTB/PCBP1 and IRES function western analysis was performed and cell lysates were immunoblotted for PTB, PCBP1 or actin as a loading control (Fig. 3D). There is a good correlation between endogenous protein levels and activation of the IRES by these proteins. For example, MCF7s which are not stimulated by co-transfection
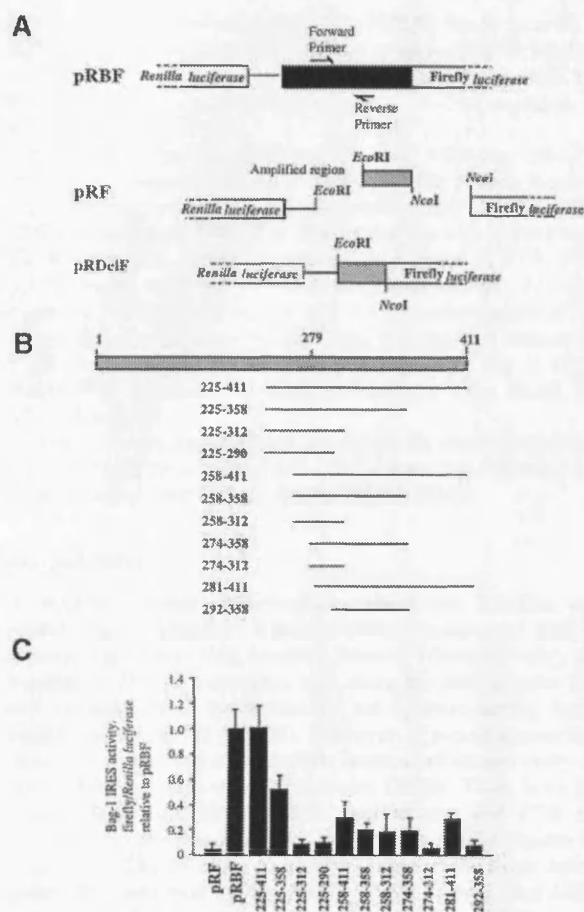
Figure 4. PCBP1 and PTB bind to the minimal active element of the Bag-1 IRES. (**A**) Schematic diagram showing construction of deletion constructs from the discistronic plasmid pRBF (Fig. 1A) by PCR where delF indicates the forward primer, which introduces an *Eco*RI site and delR the reverse primer, which introduces an *Nco*I site, primer sequences are described in Materials and Methods. (**B**) Representation of the sections of the Bag-1 5′-UTR amplified in comparison with the full-length 5′-UTR. (**C**) Relative IRES activity of the deletion constructs in HeLa cells taken as a ratio of firefly/*Renilla* luciferase, normalised to a β-galactosidase transfection control and expressed relative to pRBF, which is assigned a value of 1.

with PTB and/or PCBP1 have comparatively high levels of these proteins. In contrast, CALU1 or CAL51 cell lines which are stimulated by these proteins have lower expression of PTB and PCBP1 than MCF7 cells (Fig. 2D).

A comprehensive deletion analysis was then performed to identify the minimal element that harboured IRES activity to determine whether PTB or PCBP1 were essential for Bag-1 IRES function.

The minimum active fragment of the Bag-1 IRES is 186 nt

Regions of the Bag-1 5′-UTR DNA were obtained by PCR and subcloned into the dicistronic reporter vector pRF (Fig. 4A and B). Deletions were transfected into HeLa cells and cell lysates assayed for luciferase activity (Fig. 4C). A region of
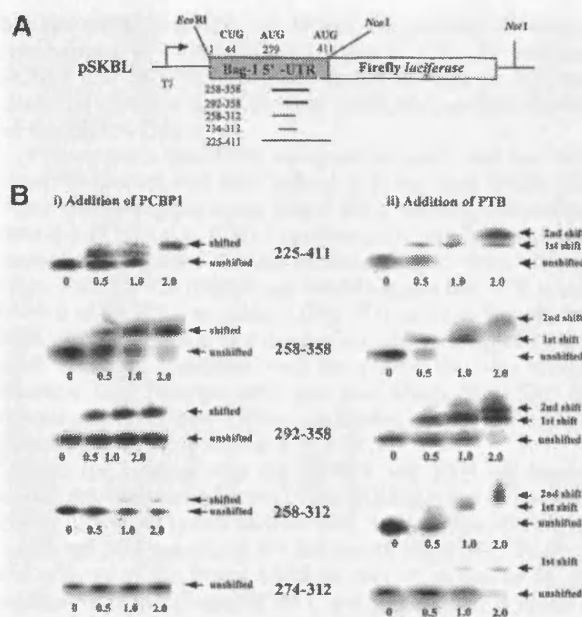
Figure 5. PCBP1 and PTB bind to specific fragments of the Bag-1 5′-UTR. (**A**) Schematic diagram of the monocistronic constructs pSKL and pSKBL showing sites for run-off transcription. (**B**) EMSAs of radiolabelled segments of the Bag-1 5′-UTR with the addition of 2.5, 5 or 10 μg/μl (0.5, 1 or 2 μl total per reaction, respectively) of (i) PCBP1 or (ii) PTB. Arrows indicate positions of protein–RNA complexes (shifted) or RNA alone (unshifted).

218 nt from 225 to 411 was identified as the minimal element since this region of RNA retained 100% IRES activity. The 5′ end of this section of RNA must be critical for function of the IRES since deleting a further 25 nt at the 5′ end reduced the activity of the 258–441 fragment to 25% (Fig. 4C). It is possible to delete an additional 107 nt from the 3′ end and retain 50% of the IRES function with the 225–358 fragment, but the removal of a further 46 nt to generate the 225–312 fragment resulted in an inactive IRES (Fig. 4C). These data would suggest that the first 225 nt are not required for an active IRES, but this section of RNA may bind other, perhaps regulatory, proteins.

PCBP1 and PTB bind to the minimal active fragment

EMSAs were then performed to identify the regions of Bag-1 IRES RNA that could be bound by PCBP1 and PTB. Fragments of the DNA encoding Bag-1 IRES RNA were obtained by PCR and inserted into the vector pSKL (Fig. 5A). Radiolabelled RNAs were generated from these plasmids by *in vitro* transcription reactions and these were incubated with PTB or PCBP1 (Fig. 5B, i). The minimum active fragment 225–411 binds both PTB and PCBP1 as efficiently as the full-length RNA (Figs 5B and 2B). To refine the binding sites further, RNA was generated from additional deletion constructs (Fig. 5A) and these were used in EMSAs. Thus, incubating the 258–358 segment of radiolabelled Bag-1 IRES RNA with increasing amounts of PCBP1 showed that PCBP1 bound directly to this region of the IRES. A decrease in mobility of the radiolabelled 292–358 fragment was also

observed and this would imply that PCBP1 binds to a 66 nt fragment. The 3′ end of this fragment is important for binding since no protein–RNA complexes were observed with the RNA generated from either 258–312 or 274–312 segments of the 5′-UTR (Fig. 5B, i).

Incubating increasing amounts of PTB with the 258–358 and 292–358 segments of RNA shows that this protein binds to these RNA segments (Fig. 5B, ii). Interestingly, addition of a 2-fold molar excess of PTB to RNA gave rise to a shifted band of a position that would correspond to a dimer of PTB. PTB also interacts with the 258–312 fragment and the 274–312 fragment, but this only occurs at a 2-fold molar excess of the protein and may be non-specific. Thus, this analysis refined the PCBP1 binding site to the 292–312 region of the 5′-UTR, whilst PTB appeared to bind at multiple sites along the 258–358 region.

Thus, both proteins interact strongly with the fragment that retains 50% of the activity (225–358) suggesting that these are indeed *trans*-acting factors for the Bag-1 IRES.

DISCUSSION

To obtain a better understanding about the function and mechanisms of action of cellular IRESs it is essential that the *trans*-acting factors that mediate internal ribosome entry are elucidated. This is a complex task since the data suggest that each cellular IRES has a distinct set of *trans*-acting factor requirements (Fig. 1) (24–28). However, it would appear that some of the proteins that mediate internal ribosome entry on viral IRESs are also used by cellular IRESs. Thus, both the Apaf-1 IRES and HRV IRESs require unr and PTB for function (27) and the XIAP and polio virus IRESs require La for activity (28). In order to identify some of the *trans*-acting factors that are used by the Bag-1 IRES, known viral IRES *trans*-acting factors were tested to determine which could interact with the Bag-1 IRES RNA. Of the seven proteins that were tested, only two, PTB and PCBP1, could bind in EMSAs (Fig. 2B). UV-crosslinking studies showed that both PCBP1 and PTB interact directly and specifically with Bag-1 IRES RNA (Fig. 2C and D). Interestingly, these proteins would appear to work in concert since a combination of these two proteins stimulated the Bag-1 IRES 3-fold *in vitro* (Fig. 3B). Moreover, it was possible to increase the activity of the Bag-1 IRES in the cell lines that had low Bag-1 IRES activity by co-transfection of pRBF with plasmids containing cDNAs encoding these proteins (Fig. 3C, i and ii).

PCBP1, which is a member of the KH domain family of single-stranded nucleic acid binding proteins (33), was found to bind strongly to a 66 nt fragment (Fig. 5). The proteins in the KH domain family generally bind to cytidine-rich sequences (33), however, there is not a cytidine-rich stretch in the linear sequence to which PCBP1 binds. It is likely, therefore, that this protein is recognising secondary or tertiary structural motifs in the Bag-1 IRES. Computer predictions have been carried out using the mfold program, although, in the absence of experimentally derived data it is difficult to predict the structure of RNA that is being recognised. It is of interest to note that PCBP2, which has 90% amino acid similarity to PCBP1, does not bind to Bag-1 IRES RNA (data not shown). Therefore, these proteins must recognise distinct structural motifs. In this regard, it has been shown that PCBP2,

but not PCBP1, is able to restore the activity of entero/rhinoviruses in PCBP-depleted extracts (23). In addition, PCBP1 and PCBP2 have been found to bind to different structural motifs of the polio virus IRES and mediate stability of the mRNA (34).

PTB contains four RNA recognition motifs and has been shown to interact with both cellular (35) and viral IRESs (13). The PTB binding site on the Bag-1 IRES was less well defined than the PCBP1 site (Fig. 5), although it was found to bind strongly to segments of the minimal Bag-1 IRES element (Figs 4 and 5). The EMSA data would suggest that PTB is able to bind to the RNA as a dimer (Fig. 5B) and it is thought that PTB exists in solution as a dimer (36). The binding of PTB to viral IRESs is consistent with the notion that this protein interacts with multiple structures and brings them into the correct conformation for ribosome binding (13) and it is likely that the same is true for the Bag-1 IRES.

Both the binding sites for PCBP1 and PTB are located within the minimum element of the IRES that has full activity (from 225 to 411) and indeed both bind within the region which has 50% activity of the full-length IRES (Fig. 5). Since the activity of the Bag-1 IRES *in vitro* is enhanced by the addition of both PTB and PCBP1, and given that PTB seems to have multiple contact points on the RNA, we suggest that these proteins act as RNA chaperones and allow the Bag-1 IRES to attain the correct structure that is competent for 40S ribosomal subunit entry. To test this theory, work is being carried out to obtain a secondary structural model for the Bag-1 IRES in the presence of these proteins.

ACKNOWLEDGEMENTS

REFERENCES

1. Yang,X., Chernenko,G., Hao,Y., Ding,Z., Pater,M.M., Pater,A. and Tang,S.-C. (1998) Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells. *Oncogene*, **17**, 981–989.
2. Packham,G., Brimmell,M. and Cleveland,J.L. (1997) Mammalian cells express two differently localised Bag-1 isoforms generated by alternative translation initiation. *Biochem. J.*, **328**, 807–813.
3. Zeiner,M. and Gehring,U. (1995) A protein that interacts with members of the nuclear hormone receptor family: identification and cDNA cloning *Proc. Natl Acad. Sci. USA*, **92**, 11465–11469.
4. Zeiner,M., Niyaz,Y. and Gehring,U. (1999) The hsp70-associating protein Hap46 binds to DNA and stimulates transcription. *Proc. Natl Acad. Sci. USA*, **96**, 10194–10199.
5. Stuart,J.K., Myszka,D.G., Joss,L., Mitchell,R.S., McDonald,S.M., Xie,Z., Takayama,S., Reed,J.C. and Ely,K.R. (1998) Characterization of interactions between the anti-apoptotic protein BAG-1 and Hsc70 molecular chaperones. *J. Biol. Chem.*, **273**, 22506–22514.
6. Nollen,E.A.A., Kabakov,A.E., Brunstung,J.F., Kanon,B., Hohfeld,J. and Kampinga,H.H. (2001) Modulation of *in vivo* HSP70 chaperone activity by Hip and Bag-1. *J. Biol. Chem.*, **276**, 4677–4682.
7. Gassler,C.S., Wiederkehr,T., Brehmer,D., Bukau,B. and Mayer,M.P. (2001) Bag-1M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration-dependent as positive and negative cofactor. *J. Biol. Chem.*, **276**, 32538–32544.
8. King,F.W., Wawrzynow,A., Hohfeld,J. and Zylicz,M. (2001) Co-chaperones Bag-1, Hop and Hsp40 regulate Hsc70 and Hsp90 interactions with wild-type or mutant p53. *EMBO J.*, **20**, 6297–6305.

9. Takayama,S., Sato,T., Krajewski,S., Kochel,K., Irie,S., Millan,J.A. and Reed,J.C. (1995) Expression and location of Hsp70/Hsc-binding anti-apoptotic protein BAG-1 and its variants in normal tissues and tumor cell lines. *Cell*, **80**, 279–284.

10. Takayama,S., Krajewski,S., Krajewska,M., Kitada,S., Zapata,J.M., Kochel,K., Knee,D., Scudiero,D., Tudor,G., Miller,G.J. *et al.* (1998) Cloning and functional analysis of BAG-1: a novel Bcl-2-binding protein with anti-cell death activity. *Cancer Res.*, **58**, 3116–3131.

11. Coldwell,M.J., deSchoolmeester,M.L., Fraser,C.A., Pickering,B.M., Packham,G. and Willis,A.E. (2001) The p36 isoform of BAG-1 is translated by internal ribosome entry following heat shock. *Oncogene*, **20**, 4095–4100.

12. Hellen,C.U.T. and Sarnow,P. (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.*, **15**, 1593–1612.

13. Belsham,G.J. and Jackson,R.J. (2000) Translation initiation on picornavirus RNA. In Sonenberg,N., Hershey,J.W.B. and Mathews,M.B. (eds), *Translational Control of Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 869–900.

14. Pestova,T.V., Shatsky,I.N. and Hellen,C.U.T. (1996) Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol. Cell. Biol.*, **16**, 6870–6878.

15. Pestova,T.V., Hellen,C.U.T. and Shatsky,I.N. (1996) Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.*, **16**, 6859–6869.

16. Brown,B. and Ehrenfeld,E. (1979) Translation of poliovirus RNA *in vitro*: changes in cleavage pattern and initiation sites by ribosomal salt wash. *Virology*, **97**, 396–405.

17. Borman,A., Howell,M.T., Patton,J.G. and Jackson,R.J. (1993) The involvement of a spliceosome component in internal initiation of human rhinovirus RNA translation. *J. Gen. Virol.*, **74**, 1775–1788.

18. Borman,A.M., Bailly,J.-L., Girard,M. and Kean,K.M. (1995) Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation *in vitro*. *Nucleic Acids Res.*, **23**, 3656–3663.

19. Hunt,S.L., Hsuan,J.J., Totty,N. and Jackson,R.J. (1999) unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of human rhinovirus RNA. *Genes Dev.*, **13**, 437–448.

20. Hunt,S.L. and Jackson,R.J. (1999) Polypyrimidine tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. *RNA*, **5**, 344–359.

21. Svitkin,Y.V., Meerovitch,K., Lee,H.S., Dholakia,J.N., Kenan,D.J., Agol,V.I. and Sonenberg,N. (1994) Internal translation initiation on poliovirus RNA; further characterisation of La function in poliovirus translation *in vitro*. *J. Virol.*, **68**, 1544–1550.

22. Gamarnik,A.V. and Adino,R. (1997) Two functional complexes formed by KH domain containing proteins with the 5′ noncoding region of poliovirus RNA. *RNA*, **3**, 882–892.

23. Walter,B.L., Nguyen,J.H.C., Ehrenfeld,E. and Semler,B.L. (1999) Differential ultilization of poly (rC) binding protein 2 directed by picornavirus IRES elements. *RNA*, **5**, 1570–1585.

24. Jopling,C.L. and Willis,A.E. (2001) N-*myc* translation is initiated via an internal ribosome entry segment that displays enhanced activity in neuronal cells. *Oncogene*, **20**, 2664–2670.

25. Stoneley,M., Subkhankulova,T., Le Quesne,J.P.C., Coldwell,M.J., Jopling,C.L., Belsham,G.J. and Willis,A.E. (2000) Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. *Nucleic Acids Res.*, **28**, 687–694.

26. Creancier,L., Mercier,P., Prats,A.-C. and Morello,D. (2001) c-myc internal ribosome entry site activity is developmentally controlled and subjected to a strong translational repression in adult transgenic mice. *Mol. Cell. Biol.*, **21**, 1833–1840.

27. Mitchell,S.A., Brown,E.C., Coldwell,M.J., Jackson,R.J. and Willis,A.E. (2001) Protein factor requirements of the Apaf-1 internal ribosome entry segment: roles of polypyrimidine tract binding protein and upstream of N-ras. *Mol. Cell. Biol.*, **21**, 3364–3374.

28. Holcik,M. and Korneluk,R.G. (2000) Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Mol. Cell. Biol.*, **20**, 4648–4657.

29. Stoneley,M. (1998) Functional analysis of the 5′ untranslated region of the c-myc proto-oncogene. PhD Thesis, University of Leicester, UK.

30. Coldwell,M.J., Mitchell,S.A., Stoneley,M., MacFarlane,M. and Willis,A.E. (2000) Initiation of Apaf-1 translation by internal ribosome entry. *Oncogene*, **19**, 899–905.

31. Henis-Korenblit,S., Levy Strumpf,N., Goldstaub,D. and Kimchi,A. (2000) A novel form of DAP5 protein accumulates in apoptotic cells as a result of caspase cleavage and internal ribosome entry site-mediated translation. *Mol. Cell. Biol.*, **20**, 496–506.

32. Henis-Korenblit,S., Shani,G., Sines,T., Marash,L., Shohat,G. and Kimchi,A. (2002) The caspase-cleaved DAP5 protein supports internal ribosome entry site-mediated translation of death proteins. *Proc. Natl Acad. Sci. USA*, **99**, 5400–5405.

33. Makeyev,A.V. and Liebhaber,S.A. (2002) The poly (C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA*, **8**, 265–278.

34. Murray,K.E., Roberts A.W. and Barton,D.J. (2001) Poly r(C) binding proteins mediate polio virus mRNA stability. *RNA*, **7**, 1126–1141.

35. Kim,Y.K., Hahm,B. and Jang,S.K. (2000) Polypyrimidine tract-binding protein inhibits translation of Bip mRNA. *J. Mol. Biol.*, **304**, 119–133.

36. Perez,I., McAfee,J.G. and Patton,J.G. (1997) Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry*, **36**, 11881–11890.