

The *c-myc* IRES: structure and mechanism

by

John P.C. Le Quesne M.A. (Cantab)

Thesis submitted for the degree of

Doctor of Philosophy

at the

University of Leicester

January 2000

UMI Number: U127931

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U127931

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

Thanks are due to Simon Engledow for expert technical assistance, and to Eric Westhof for enlightening discussion of RNA structure. All members of the laboratory are thanked for their support and friendship, and congratulated for their forbearance.

I owe a particularly great debt to Anne Willis, whose qualities as a supervisor, a mentor and a friend have transformed my experience of the last three years.

“. . . my work, which I've done for a long time, was not pursued in order to gain the praise I now enjoy, but chiefly from a craving after knowledge, which I notice resides in me more than in most other men. And therewithal, whenever I found out anything remarkable, I have thought it my duty to put down my discovery on paper, so that all ingenious people might be informed thereof.”

Antony van Leeuwenhoek. Letter of June 12, 1716.

“I am the man. I suffered. I was there.”

Walt Whitman. *Leaves of Grass*.

Abstract

The *c-myc* IRES: structure and mechanism

The proto-oncogene *c-myc* is central to the process whereby the cell commits itself to quiescence, differentiation, proliferation or apoptosis, and the expression of Myc protein is controlled at several levels, including translation.

The 5' UTR of *c-myc* has been shown to contain an internal ribosome entry segment (IRES), allowing translation to proceed via an internally initiated mechanism. To determine the secondary structure of the IRES, structural data were obtained by chemical probing of 5' UTR RNA *in vitro*. These data were used as constraints upon the "mFold" RNA secondary structure prediction algorithm, and the model was refined by phylogenetic analysis. The resulting model contains a number of interesting features. There is no detectable structural homology with viral IRESs.

Mutations were introduced to determine the importance of various IRES moieties. Surprisingly, the IRES seemed resistant to relatively gross structural changes, and a number of mutations were seen to significantly activate IRES function, suggesting that the IRES is in a state of constitutive repression.

The point at which the ribosome enters and begins scanning was investigated, revealing that entry occurs in an unstructured region of the IRES, upstream of an inhibitory pseudoknot element that must be disrupted before ribosome entry can occur.

It has previously been noted that the *c-myc* IRES fails to function in RRL *in vitro* translation assays. In order to obtain an *in vitro* assay to aid isolation of specific *trans*-acting factors, several cellular extracts were tested for their ability to stimulate IRES activity *in vitro*. Nevertheless, the IRES was not activated *in vitro*.

From these data, a picture of the *c-myc* IRES that is distinctly different from the viral paradigms has emerged, and a model of the IRES mechanism is presented and discussed.

Contents

Acknowledgements	2
Abstract	3
Contents	4
Abbreviations	7
Chapter 1: Introduction	12
1.1 Translation	12
Phases of translation	12
Initiation of translation	12
Cap-dependent initiation	13
Formation of the ternary complex	13
Cap binding	14
Ribosome scanning	15
Start codon selection	17
Alternative modes of cap-dependent translation initiation	20
1.2 Internal Ribosome Entry	23
The IRES paradigm	23
IRES properties	24
Picornaviruses	24
Flaviviruses	25
Secondary Structure	25
Sequence	28
Start codon selection	30
Trans-acting factors	34
1.3 C-myc	38
The c-myc family	38
Myc protein	38
C-myc is a transcription factor	39
Normal c-myc function	39
C-myc is a proto-oncogene	40
Control of c-myc expression	41
Chapter 2: Materials and Methods	44
2.1 General Reagents	44

2.2	Tissue Culture Techniques	45
	Tissue culture media and supplements	45
	Cell Lines	45
	Maintenance of cell lines	45
	Calcium phosphate-mediated DNA transfection	45
2.3	Bacterial Methods	47
	Culture media and supplements	47
	Bacterial strains	47
	Preparation of competent cells	47
	Transformation of competent cells	48
2.4	Molecular Biology Techniques	49
	Plasmids	49
	Ethanol precipitation of DNA	49
	Phenol/chloroform extraction	50
	Purification of DNA using glassmilk	50
	Agarose gel electrophoresis	50
	Gel isolation of DNA fragments	51
	Synthesis and purification of oligonucleotides	51
	Oligonucleotides	51
	Restriction enzyme digestion	51
	Mung bean nuclease treatment	52
	Alkaline phosphatase treatment of DNA	52
	Ligations	52
	Small scale preparation of plasmid DNA	53
	Large scale preparation of plasmid DNA	53
	Caesium chloride gradient purification of plasmid DNA	54
	Double stranded DNA sequencing	55
	PCR mutagenesis	56
2.5	Biochemical Techniques	57
	Translating cytoplasmic extracts	57
	Ribosomal salt wash	57
	Nuclear extract and hnRNP A1	58
	Nuclear salt wash	58
	High-efficiency transcription extract	59

Preparation of cell lysates from transfected cells	60
Luciferase assays	60
2.6 RNA Methods	61
<i>In vitro</i> run-off transcription	61
Rabbit reticulocyte translation	62
Coupled transcription/translation	62
Cytoplasmic extract translation	62
Chemical structure probing	62
Primer extension	63
2.7 Theoretical Methods	65
Secondary structure prediction	65
Stability calculations	65
Sequence alignment	65
Covariation analysis	65
Structure searching	65
Chapter 3: Structure Determination	66
3.1 Introduction	66
3.2 Theoretical Approaches	68
3.3 Chemical Probing	70
3.4 Structure Modelling	73
3.5 Features of the Secondary Structure	77
3.6 Conservation of Secondary Structure	79
Chapter 4: Mutagenic Analysis	82
4.1 Introduction	82
4.2 Mutagenesis	83
4.3 IRES assay	84
4.4 Identification of the Ribosomal Entry Window	85
4.5 Deletion Mutants	88
4.6 Mutations of Domains 1 and 2	89
4.7 Pseudoknot Mutations	92
4.8 Mutations of the Spacer Region	95
Chapter 5: In Vitro Analysis	96
5.1 Introduction	97
5.2 Translation in Rabbit Reticulocyte Lysate	100

5.3	Translation in HeLa Cell Cytoplasmic Lysate	102
5.4	Wild-type vs. mutant IRES activity <i>in vitro</i>	103
5.5	Translation in GM 2132 Cell Cytoplasmic Lysate	105
5.6	Addition of Specific Protein Factors	106
5.7	Re-folding of IRES RNA	108
5.8	Addition of Nuclear Extracts	109
Chapter 6: Discussion		113
6.1	A Mechanistic Model	113
	IRES architecture	113
	Diffuse function	113
	Ribosome capture	114
	Ribosome entry	115
	Scanning and initiation	116
6.2	The <i>c-myc</i> IRES and Disease	118
	C255U and multiple myeloma	118
	A highly repressed IRES	118
Bibliography		122

Abbreviations

AMV	Avian myeloblastosis virus
ATP	Adenosine 5'-triphosphate
BiP	Immunoglobulin heavy chain binding protein
bp	Base pairs
BMK	Borate, magnesium, potassium buffer
BSA	Bovine serum albumin
BVDV	Bovine viral diarrhoea
CAT	Chloramphenicol acetyltransferase
CIAP	Calf intestinal alkaline phosphatase
CMCT	N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho- <i>p</i> -toluenesulphonate
CMV	Cauliflower mosaic virus
CSFV	Classical swine fever virus
CTP	Cytidine 5'-triphosphate
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
ddNTP	Dideoxynucleoside triphosphate
dGTP	deoxyadenosine 5'-triphosphate
DMS	Dimethyl sulphate
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol

dTTP	Deoxythymidine 5'-triphosphate
<i>E.Coli</i>	<i>Echerischia Coli</i>
EDTA	Ethylenediaminetetra-acetate
EGTA	Ethyleneglycol-bis-(b-amino-ethyl ether)N,N'-tetra-acetic acid
eIF	Eukaryotic initiation factor
EMCV	Encephalomyocarditis virus
FCS	Foetal calf serum
FGF2	Fibroblast growth factor
FMDV	Foot and mouth disease virus
GBV-C	"G.B." virus C (patient initials)
GDP	Guanosine diphosphate
GTP	Guanosine 5'-triphosphate
HAV	Hepatitis A virus
HCV	Hepatitis C virus
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid
HKM	HEPES, potassium, magnesium buffer
HNE	HeLa nuclear extract
hnRNP	Heteronuclear ribonucleoprotein
HRV	Human rhinovirus
IRES	Internal ribosome entry segment
ITC	Isopropanol, TE, caesium
kb	Kilobases
kCal	Kilocalories
kDA	Kilodaltons

LB	Luria-Bertani broth
Met	Methionine
mRNPs	Messenger ribonucleoprotein
NSW	Nuclear salt wash
nt	Nucleotides
ODC	Ornithine decarboxylase
ORF	Open reading frame
PCBP2	Poly(rC) binding protein 2
PCR	Polymerase chain reaction
PDGF2	Platelet derived growth factor 2
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
PTB	Polypyrimidine binding tract
RLU	Relative light units
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RNAasin	Ribonucleic acid hydrolase inhibitor
RRL	Rabbit reticulocyte lysate
rRNA	Ribosomal RNA
RSW	Ribosomal salt wash
SDS	Sodium dodecyl sulphate
SSPB	Standard structure probing buffer
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TE	Tris EDTA

HTE	High-efficiency transcripton extract
TKM	Tris, potassium, magnesium buffer
TMEV	Theiler's murine encephalomyelitis virus
tRNA	Transfer RNA
UNR	Upstream of N- <i>ras</i>
uORF	Upstream open reading frame
UTP	Uridine 5'-triphosphate
UTR	Untranslated region
UV	Ultra-violet
VEGF	Vascular endothelial growth factor

Chapter 1

Introduction

1.1 Translation

Phases of translation

Eukaryotic translation of mRNA into protein is naturally divided into three phases: initiation, elongation and termination. Initiation describes the processes underlying the binding of an 80S ribosome to a start codon. Elongation is an iterative sequence of charged tRNA acquisition, peptide bond formation, ribosome translocation and spent tRNA release. Termination is achieved by the recognition of a stop codon and cleavage of the final peptidyl-tRNA bond, releasing the protein from the ribosome.

Each phase is subject to some variation, but unlike elongation and termination, initiation proceeds by a number of highly distinct mechanisms.

Initiation of translation

In most translation systems, initiation is believed to be the rate-limiting step, making this the stage at which most control is exerted over the overall rate of translation. Two main modes have been described, cap-dependent initiation and internal initiation. Cap-dependent initiation operates on the great majority of cellular messages, and has been the focus of most research. Internal initiation was only much more recently discovered in eukaryotes, and proceeds by a number of distinctively different pathways, reports of which continue to accumulate.

Cap-dependent initiation

The cap-dependent initiation pathway depends upon a complex series of molecular interactions that are broadly the same on all messages, as the same set of canonical eukaryotic initiation factors (eIFs) is globally responsible for cap-dependent initiation. These canonical eIFs are summarised in Table 1.a.

The chain of interactions that results in the ribosome recognising the authentic start codon and beginning peptide synthesis will be considered, beginning at the end of the previous round of translation. In many cases the precise *in vivo* order of these associations and disassociations has not been determined, and the events described below have been incorporated into a number of slightly differing models by different researchers. A general model of cap-dependent initiation is represented in Figure 1.i.

Formation of the ternary complex

Ribosomal subunits that have recently terminated translation and dissociated from mRNA will, at physiological salt concentrations, tend to associate with each other. Their dissociation is encouraged by eIF3, which is thought to act by competing with the 60S subunit for a binding site on the 40S subunit (Goumans *et al.*, 1980; Hannig, 1995).

This stable eIF3/40S complex is competent to bind the ternary complex, consisting of initiator Met-tRNA, eIF2, and GTP. The ternary complex is able to bind the 40S subunit in the absence of eIF3 *in vitro* when purified proteins are used, but the interaction is apparently stabilized by eIF3 (Trachsel and Staehelin, 1979). The resulting eIF3-40S-tRNA_i^{met}-eIF2-GTP complex has a sedimentation coefficient of 43S.

There is some evidence that eIF1A plays a part in the assembly of this, the 43S complex (Chaudhuri *et al.*, 1997), but recent research has revealed a more definite role for this factor (Pestova *et al.*, 1998a).

Factor	Function	Subunits (kDa)
eIF-1	Scanning complex assembly	14
eIF-1A	Scanning complex assembly: 40S/eIF-3 complex formation?	17
eIF-2	GTP/GDP binding: tRNA _i ^{met} delivery and attachment to the 40S subunit: start codon selection?	36, 38, 52
eIF-2B	Guanine nucleotide exchange factor: drives exchange of GDP for GTP on eIF-2	81, 71, 58, 43, 34
eIF-3	Dissociation of ribosomal subunits: stabilization of 40S/eIF-2 complex: interaction with eIF-4F	170, 116, 110, 66, 48, 47, 44, 40, 36, 35, 28
eIF-4A	RNA helicase	46
eIF-4B	Stimulates helicase activity of eIF-4A	69
eIF-4E	Cap-binding protein	25
eIF-4F	Cap-binding complex, comprising eIFs 4A, 4E and 4G	46, 25, 220
eIF-4G	Bridging protein joining 4E and 4A	220
eIF-5	Hydrolysis of GTP, with eIF2: triggers eIF dissociation and subunit association: start codon selection?	58
eIF-5B	Involved in subunit association: 60S subunit delivery?	175

Table 1.a. Canonical eukaryotic initiation factors.

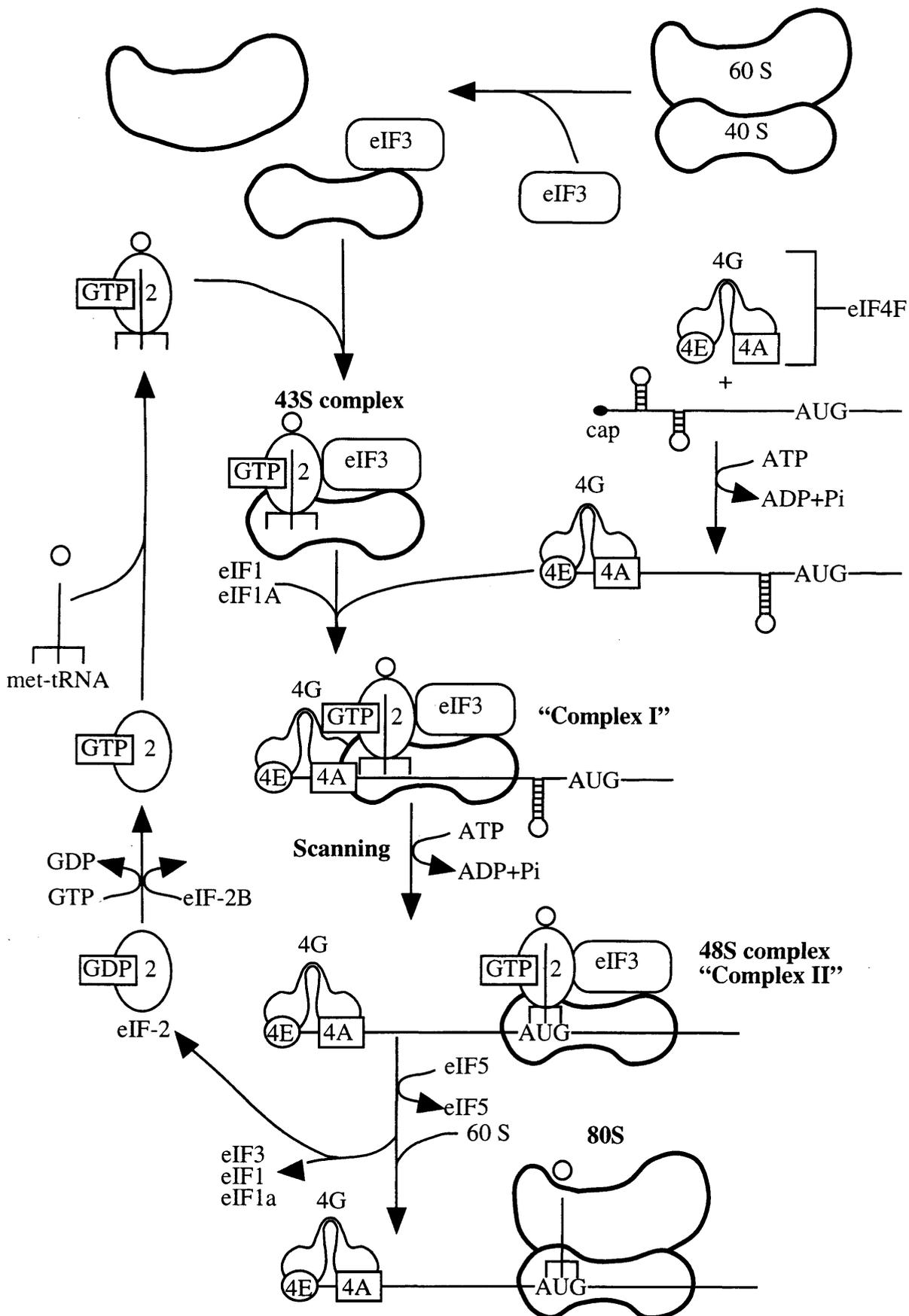


Figure 1.i. Cap-dependent initiation of translation. The post-termination ribosome is separated by the action of eIF3. The eIF4F trimer binds the 5' cap of the mRNA, and subsequently recruits the 43S ternary complex. Upon locating the authentic start codon, the 80S ribosome assembles and polypeptide synthesis proceeds.

Cap binding

The m⁷G cap at the 5' end of the mRNA is recognised and bound by eIF4E (Marcotrigiano *et al.*, 1997), which is able to form a trimeric complex with eIF4A and eIF4G. EIF4G acts as a bridge joining eIF4E and eIF4A, and the three factors are collectively known as eIF4F.

This cap-binding complex performs two essential functions. Firstly, it establishes a stretch of single-stranded RNA close to the cap upon which the 43S complex can “land”. Thus, *in vitro*, purified eIF4F unwinds secondary structures adjacent to the cap 15 base pairs in length (Lawson *et al.*, 1986). This is due to the action of eIF4A. Indeed, eIF4A is the archetypal DEAD box protein, a family of RNA-activated ATPases proposed to act as RNA helicases. The mechanism of structure disruption remains unclear, but observed RNA- and ATP-binding kinetics and conformational changes of the protein as it binds these substrates are consistent with an active unwinding process (Lorsch and Herschlag, 1998a; Lorsch and Herschlag, 1998b). The addition of eIF4B significantly enhances the activity (Lawson *et al.*, 1989), though once again the precise role of the factor is unclear. It has been proposed that it might act as a recycling factor, displacing eIF4A from the eIF4F complex thus permitting the released eIF4E/eIF4G component to recruit further molecules of eIF4A (Ray *et al.*, 1986). However, the enhancing effect is also seen in the absence of eIF4B (Lawson *et al.*, 1989), suggesting a functional role as part of a higher-order helicase complex, perhaps as a “transmission”, acting as an adapter protein between the eIF4A “motor” and the RNA substrate. It has been shown in rabbit reticulocyte lysate that the deliberate insertion of stable secondary structure motifs very close to the cap inhibits initiation (Kozak, 1989), suggesting both that the unwinding activity is limited, and that the mRNA binding track in the 40S subunit is only capable of accommodating fully unwound RNA.

Secondly, the cap-binding complex recruits the 43S complex to the mRNA. The precise interactions that underlie this step have yet to be characterised, although interactions between eIF4G and the 40S ribosomal subunit (Lamphear *et al.*, 1995), between eIF4G and eIF3 (Mader *et al.*, 1995), between eIF4B and eIF3 (Hershey, 1999), between eIF4B and the 40S ribosomal subunit (Methot *et al.*, 1996) and, of course, between the 40S subunit and mRNA have all been described.

Ribosome scanning

When the ribosome has acquired the mRNA into its binding track, it must locate a start codon. It is generally accepted that this process occurs via a scanning mechanism, in which the 40S subunit proceeds unidirectionally from the cap, one base at a time, until conditions are met such that protein synthesis can begin.

In vitro reconstitution experiments using purified components of the translation machinery have revealed roles for eIF1 and eIF1A in scanning (Pestova *et al.*, 1998a). In this system, in the absence of eIF1 and eIF1A, 43S complexes were able to bind at the 5' extremity of mRNA, but did not scan. This was described as "complex I". Addition of eIF1 alone reduced the formation of complex I, and enabled some of the ribosomes to locate the authentic AUG by scanning, where they stalled due to the absence of post-initiation factors; this was described as "complex II". Addition of eIF1A alone enhances formation of complex I, without the formation of complex II. Both complexes are required for the efficient formation of complex II. Furthermore, eIF1/1A-dependent conversion of complex I to complex II seems to depend upon the disassembly of complex I and subsequent re-assembly in the presence of these factors.

This is interpreted as a mechanism improving the accuracy of start codon selection, which is supported by the observation that mutations in the yeast eIF1 homologue Sui1 allow

aberrant initiation (Yoon and Donahue, 1992), although it is not clear that this function is exercised *in vivo*. That a particular factor is required for the fidelity of a given process does not quite mean that the function of that factor is to maintain fidelity, although the distinction is a fine one. It is also suggested that eIF1 and eIF1A might complete the assembly of a competent scanning complex by sealing the RNA-binding tunnel in the 40S subunit.

That the scanning process is unidirectional is in accordance with the observation that longer 5'UTRs are not disproportionately slowly traversed (Kozak, 1998), as would be expected if the ribosomes were proceeding by a random walk. The energy-consuming component of the scanning machinery is believed to be the helicase activity of eIF4A, which is not itself processive (Rogers *et al.*, 1999), so presumably some form of molecular ratchet is at work to ensure that the ribosome does not backtrack.

It seems likely that both eIF4A and eIF4B are involved in RNA unwinding both before 43S complex binding and during scanning, although some data suggest that the majority of initiation-dependent ATP consumption occurs during RNA unwinding before 43S complex binding (Jackson, 1991). Certainly the unwinding activity seen before and after ribosome entry is quantitatively different, as the pre-40S-binding activity is less competent to unwind structure than the activity observed during scanning (Kozak, 1989). This study showed that a hairpin loop containing 13 base pairs with a calculated stability of -30 kCal/mol prevented 40S subunits binding to mRNA when positioned 12 nucleotides from the 5' cap, but was readily scanned through when positioned 52 nucleotides from the cap. This requires explanation if the same factors are responsible for the unwinding, namely eIF4A and eIF4B. It is possible that via the observed interaction between eIF4B and 18S rRNA (Methot *et al.*, 1996) the helicase activity is transferred to the 40S subunit, and that the enhanced helicase activity is due to the sequestration of single strands, recently unwound from helical elements,

into the mRNA binding tunnel of the 40S subunit. This would be in contrast to the non-processive cap-bound helicase activity, which without any such mechanism to maintain RNA in a single-stranded form would be confined to repeatedly unwinding the same, short stretch of RNA. This model also provides a satisfactory explanation for the directionality of the scanning process, by proposing that helicase activity is confined to the 3' side of the ribosome, and that the ratchet effect is due to the preference of the ribosome for the newly unwound RNA to its 3' side rather than the randomly structured RNA on the 5' side.

Start codon selection

The great majority of proteins are translated from the most 5' AUG start codon in good context. Recognition of the start codon is accompanied by transient association of the release factor eIF5 with eIF2, hydrolysis of the eIF2 bound GTP and release of eIFs. This leaves the way open for the binding of the 60S ribosomal subunit, accommodation of the next aminoacyl-tRNA into the A-site, and the beginning of the elongation phase.

Experiments have been performed in yeast, in which the anticodon loop of the initiator Met-tRNA is engineered from 3'-UAC-5' to 3'-UCC-5', which show efficient initiation from an AGG codon (Cigan *et al.*, 1988), demonstrating the primacy of the mRNA-Met tRNA interaction in start site selection.

When two AUG codons are very closely spaced, initiation can be just as efficient from the second as the first (Kozak, 1995; Williams and Lamb, 1989). This is variously interpreted as due to "leaky scanning", that is poor recognition of the upstream codon by the traversing 40S subunit due to other peculiarities of the RNA, or alternatively a generally less than rigorously stepwise and systematic search by the scanning machinery. It is certain that leaky scanning can occur, but there is insufficient evidence to be certain that it is solely responsible for the omission of the upstream AUG in all cases. In any case, it is clear that the ribosome

does not under all circumstances recognise an AUG, even when in good context, that passes through the recognition centre.

The next most important determinant of AUG initiation is the identity of neighbouring bases. The “ideal” context was first identified by a mass alignment of vertebrate mRNA sequences (Kozak, 1987), and the particular importance of a G at position +4 (where AUG is +1 to +3) and a purine at position -3 emerged afterward (Kozak, 1997). Context effects are presumably mediated by electrostatic interactions between the bases in question and rRNA/protein components of the mRNA binding track on the scanning complex. It is interesting that the contextually important bases are confined to a region that is much shorter than the mRNA binding track, suggesting that only this segment of the message is bound sufficiently intimately for inspection during scanning. It may well be that contextual effects have their origin in the requirement of the scanning machinery to sufficiently structurally constrain the mRNA surrounding the putative start codon to allow a rigorous test of the codon-anticodon interaction.

The presence of secondary structure can also affect AUG recognition. If the AUG is situated within or downstream of a hairpin-loop structure that is beyond the unwinding ability of the scanning-associated helicase activity, then no recognition can occur since scanning is halted (Kozak, 1989). If, however, a small hairpin (-19 kCal/mol) which does not ordinarily impede scanning is positioned a critical distance downstream of an AUG in poor context, then initiation is enhanced (Kozak, 1990). The effect is strong when the spacing between the AUG and the hairpin is 14 nucleotides, and may be stronger still at another distance somewhere between 8 and 32 nucleotides. The optimal spacing seems likely to coincide with the 3' boundary of the 40S ribosome: 12-15 nucleotides when determined by nuclease protection (Kozak, 1977), or 15-17 nucleotides according to primer extension “toeprinting” experiments

(Pestova *et al.*, 1998a). It is interesting to speculate that a similar structure placed a critical distance 5' of a start codon might impair codon recognition, energetically favouring the progress of the scanning complex by “winding up” as it emerges from the 5' opening of the mRNA binding track.

Since codon-anticodon recognition, context and downstream structure have cumulative effects upon the efficiency of codon recognition, it is unsurprising that initiation can occur at codons other than AUG if either or both of the other criteria are satisfied. Thus, a functional CUG upstream of the AUG codon is a relatively common feature of growth factor genes and proto-oncogenes, including *c-myc* (Hann *et al.*, 1988). In the case of *c-myc*, this CUG codon is in good context, lacks any predicted stable downstream structures, yet accounts for a significant fraction of cellular Myc protein. Interestingly, experiments using engineered *c-myc* constructs translated in RRL showed that as well as the “classical” context, the identity of the codons at positions +5 and +6 were also crucial to CUG usage (Boeck and Kolakofsky, 1994), although it was not demonstrated that this effect was genuinely due to altered requirements of the codon-recognising centre rather than being a general effect undetected in studies of AUG codons due to a perfect codon-anticodon interaction and good context saturating the recognition criteria. Likewise, the insertion of an eight base-pair long helical element with a predicted stability of -19 kCal/mol downstream of a GUG was seen to enhance initiation from this codon three-fold (Kozak, 1990).

Thus three highly individual criteria affect start codon recognition: the degree of codon-anticodon complementarity, the degree of similarity between start codon context and the ideal context, and the presence or absence of crucially spaced downstream structures. All three criteria affect the thermodynamics of the ribosome/mRNA interaction. The mechanistic consequence that these criteria seem to hold in common is temporal in nature; all might be

expected to cause the scanning ribosome to pause. So, hypothetically, for start codon recognition to occur, the scanning complex must hesitate over the start codon for long enough to allow another step to occur.

The best candidate for this step is the hydrolysis of eIF2-bound GTP, which is dependent upon an interaction with eIF5 (Chaudhuri *et al.*, 1994). GTP hydrolysis sets in train the release of initiation factors (Hershey, 1999) and joining of the 60S subunit, a process which also involves eIF5B (Pestova, 1999). It has been shown in yeast that mutations in eIF5 or eIF2 that allow faster hydrolysis of GTP allow initiation at a normally unused UUG codon (Huang *et al.*, 1997). If start codon selection truly depends on the kinetics of this process, it is only likely to be crucial in the selection of codons which only partially fulfil the three criteria described above, since in a reconstituted system the scanning complex will halt indefinitely at an AUG codon in good context in the absence of eIF5 (Pestova *et al.*, 1998a).

Alternative modes of cap-dependent translation initiation

While the great majority of eukaryotic messages are translated by the mechanism described above, a number of variations are known to operate under certain circumstances. The mechanisms of re-initiation and ribosome shunting will be briefly discussed, and internal ribosome entry via IRESs is dealt with separately in the next section.

Short upstream open reading frames (uORFs) are present in a number of genes, including several growth factors and proto-oncogenes (Willis, 1999), yeast genes (Hinnebusch, 1996), and viral genes (Moustakas *et al.*, 1993). These may be negotiated in two ways: the uORF may be bypassed altogether by leaky scanning (see above), or the ribosome can recognise the upstream codon, translate the uORF, and then re-initiate at the second cistron. This re-initiation mechanism is poorly understood, but a number of observations have yielded some insight.

Re-initiation operates downstream of a uORF with a maximum length of about 30 codons (Luukkonen *et al.*, 1995), leading to the temporal suggestion that the time taken to traverse the uORF is critical; perhaps re-initiation depends upon the presence of eIFs that only slowly dissociate after major subunit association (Kozak, 1999). The minimum spacing between ORFs is also crucial, apparently to give sufficient opportunity for the 40S subunit to re-acquire a tRNA_i^{met} (Hinnebusch, 1997).

Initiation is also observed from start codons that are just upstream of the termination codon of the uORF, suggesting that post-termination 40S subunits, unlike their scanning counterparts, are capable of backtracking (Peabody and Berg, 1986). This is consistent with models in which unidirectional scanning is a consequence of the helicase activity of associated eIF4 factors, which are presumably dislodged by 60S subunit joining.

Finally, the nature of the peptide encoded by the uORF is also sometimes important. A hexapeptide with the sequence MAGDIS, encoded by the uORF upstream of the S-adenosylmethionine decarboxylase ORF, seems to stall the ribosome as it is synthesized (Geballe and Morris, 1994). This repression may be mediated by the intracellular concentration of polyamines (Ruan *et al.*, 1996).

A variation of the scanning model known variously as discontinuous scanning, ribosome shunting or ribosome hopping has been described in plant translation systems. This process is characterised by the ability of scanning ribosomes to bypass regions of the 5' UTR, apparently leaving and re-entering the message. Requirements for shunting vary between systems, the paradigm being a short uORF closely followed by a highly structured shunted region, as found in CMV RNAs (Futterer and Hohn, 1996; Pooggin *et al.*, 1999). Shunting has also been described operating upon the cytopathic late adenoviral RNAs, requiring a

structured region, a critically-sized spacer and non-canonical *trans*-acting factors to operate (Huang and Schneider, 1991; Schneider, 1999).

1.2 Internal Ribosome Entry

The IRES paradigm

In eukaryotes, internal initiation was first observed to act upon RNA genomes of the poliovirus (Dorner *et al.*, 1984), a member of the family of picornaviridae (Jackson and Kaminski, 1995). It was known that infection with poliovirus rapidly caused host cell protein synthesis to be shut down, but that expression of viral proteins continued unabated. The 2A viral protease cleaves host cell eIF4G between the domains that interact with eIF4E and eIF4A, separating the cap-binding and helicase activities of eIF4F (Lamphear *et al.*, 1995; Lamphear *et al.*, 1993). Thus eIF4A is no longer recruited to the cap, the crucial cap-ternary complex contacts are lost, and the cap-dependent mechanism is curtailed. Poliovirus RNA does not require intact eIF4G for initiation (Pestova *et al.*, 1996) and is translated efficiently at the expense of the cell.

Internal initiation was subsequently shown to depend upon a discrete region of the viral 5' UTR, named the internal ribosome entry site (or segment), or IRES (Jang *et al.*, 1988; Pelletier and Sonenberg, 1988).

The classic test for IRES activity involves the use of bicistronic constructs (Figure 1.ii). A suspected IRES sequence is inserted between two reporter gene ORFs. If the sequence contains no IRES, then translation from the 5' cistron will proceed as normal, with at most a very low level of expression from the 3' cistron, probably due to re-initiation by post-termination ribosomes that remain associated with the mRNA. If there is a functional IRES between the cistrons, then expression from the downstream cistron will be greatly enhanced, due to internal initiation.

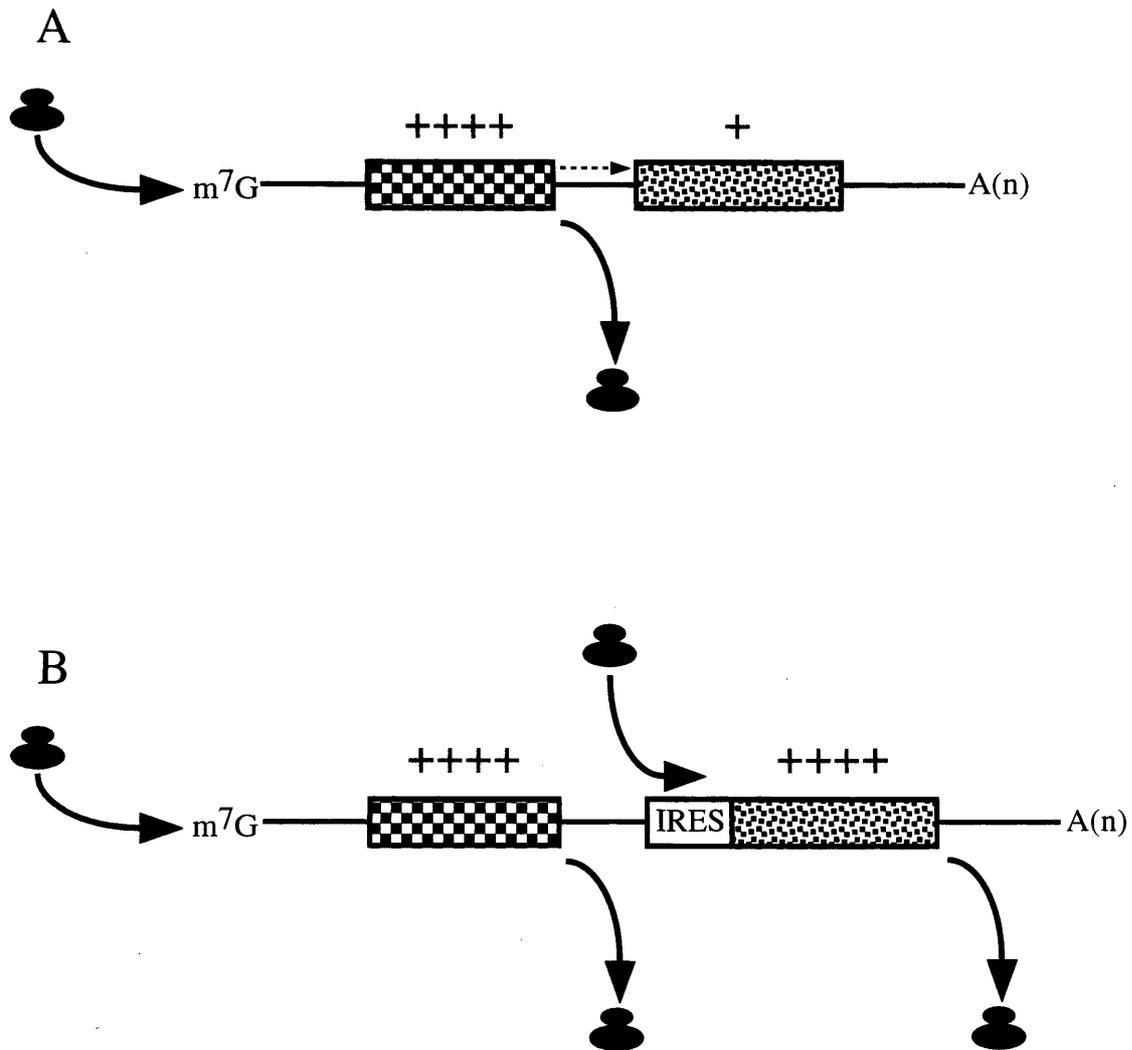


Figure 1.ii. The bicistronic assay for IRES function. A: The upstream reporter is efficiently translated by the cap-dependent mechanism. Slight expression of the downstream reporter is performed by ribosomes that re-initiate (“readthrough”). B: The downstream reporter gene is also efficiently translated, via an IRES-mediated cap-independent ribosomal entry mechanism.

Subsequent studies have identified IRESs in several other viruses, and more recently in the 5' UTRs of a number of cellular mRNAs.

IRES properties

Although the molecular details underlying IRES function are as yet incompletely determined, there are sufficient data to classify IRESs in number of ways, such as their secondary structure, mode of initiation, and their requirements for *trans*-acting factors. A description of these various categories and a brief treatment of the properties of IRESs that fit them is a more fitting approach to this enormous body of data than a mechanical listing and description of a number of IRESs. Such a thematic approach fosters consideration of a novel IRES without imposing conformity to any particularly well-understood paradigm, and also enables the discussion of interesting properties of IRESs that an exhaustive list might obscure.

As picornavirus and flavivirus IRESs will form a large part of the discussion, it is necessary to briefly describe these families.

Picornaviruses

The picornaviruses are a family of positive-stranded RNA viruses. Their IRES elements are around 450 nucleotides in length, contain multiple AUG codons and form elaborate and extensive secondary structures (Jackson and Kaminski, 1995). The IRESs have been divided into three classes on the basis of their taxonomy, secondary structures and accompanying mechanistic peculiarities.

Class I contains the entero- and rhinoviruses, including poliovirus, human coxsackie viruses and human rhinovirus (HRV), responsible for the common cold.

Class II contains the cardio- and aphthoviruses, including the foot and mouth disease virus (FMDV), the encephalomyocarditis virus (EMCV) and Theiler's murine encephalomyelitis virus (TMEV).

Class III consists solely of the hepatitis A virus (HAV).

Flaviviruses

The only IRESs to have received a degree of scrutiny comparable to those of the picornaviridae are found in the family flaviviridae (Lemon and Honda, 1997). The most infamous member is the hepatitis C virus (HCV), but the family includes other economically important viruses including the pestiviruses bovine viral diarrhoea virus (BVDV) and classical swine fever virus (CSFV). GB virus C (GBV-C), responsible for hepatitis G, is a distantly related flavivirus that has not as yet been assigned to a genus (Karayiannis *et al.*, 1998).

Secondary Structure

A glance at an IRES sequence immediately reveals the potential for extensive secondary structure formation, as they are without exception long and G/C-rich. The importance of secondary structure is highlighted by the high frequency of base covariance in aligned viral IRES sequences, which facilitates structural modelling. The secondary structures of a number of viral IRESs have been determined, largely by this method, but as yet the published models of cellular IRESs rely almost entirely upon computer modelling of individual sequences, a method which cannot be relied upon in isolation. For this reason, this section will concentrate upon viral IRES structures.

The class I IRESs comprise six domains (I-VI), separated by short unstructured regions (5-14 nucleotides), as shown in Figure 1.iii. Individual domains are generally simple helices

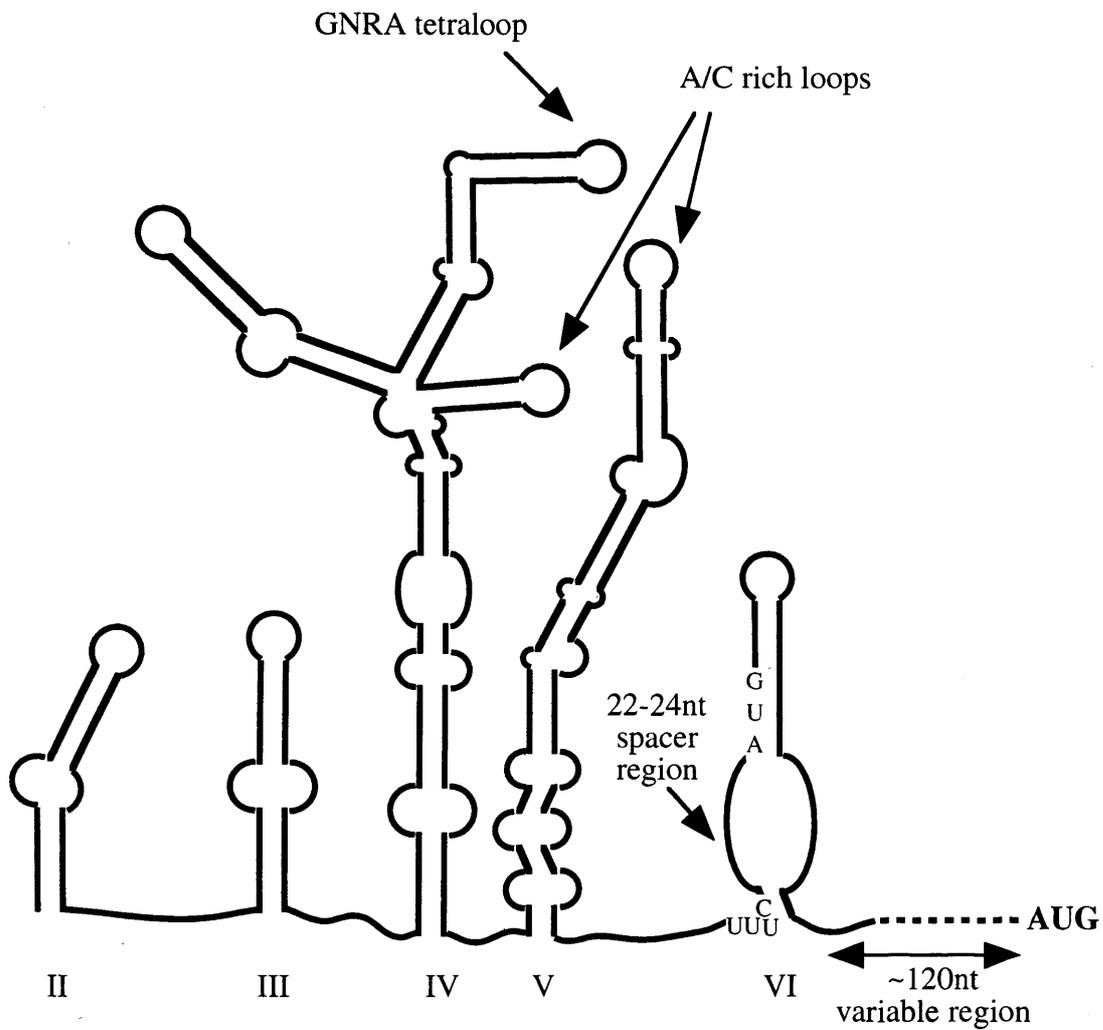


Figure 1.iii. Secondary structure of a class I picornavirus IRES. The authentic initiation codon is in bold type, and the structure is about 620nt long in total. The model was derived by computer modelling and chemical/enzyme probing, and is adapted from Stewart and Semler, 1997.

interrupted by small internal bulges and bulged loops, but the largest domain, domain IV, is more complex, including a four-way junction.

Class II IRESs have 12 domains (A-L) (Figure 1.iv); domain I includes two four-way junctions, domains J and K are the terminal elements of a higher-order three-way junction.

The hepatitis C virus IRES is relatively simple (Figure 1.v), containing three domains only, but including a pseudoknot structure 12 nucleotides upstream of the start codon. If this pseudoknot is disrupted by mutagenesis, the IRES is rendered inactive (Wang *et al.*, 1995). The pseudoknot is also a feature of pestivirus IRESs, where once again its presence is essential for IRES function (Rijnbrand *et al.*, 1997).

The distantly related flavivirus GBV-C also contains a pseudoknotted domain (Figure 1.vi), although in contrast to the hepatitis C virus it is situated far upstream of the start codon. Like HCV, this IRES contains structural elements downstream of the authentic AUG start codon.

Thus both picornavirus and flavivirus IRESs comprise more than one highly distinctive secondary structural form, apparently with few features in common. The picornavirus elements have been proposed to share a structural core more closely resembling Figure 1.iv than Figure 1.iii, with a branched 3' terminal structure, and also incorporating a hepacivirus IRES-like pseudoknot at the 3' boundary of the IRES (Le and Maizel, 1998). This model is derived from a mass alignment of picornavirus IRESs of all three classes. It is as yet unsupported by structure probing or energy minimization studies, and the proposed pseudoknot is erratically conserved and poorly delineated. The undiminished efficiency of chimeric IRES elements having non-complementary 5' and 3' "pseudoknot" elements from IRESs of different classes clearly demonstrates that there is no functional requirement for pseudoknot formation (Ohlmann and Jackson, 1999).

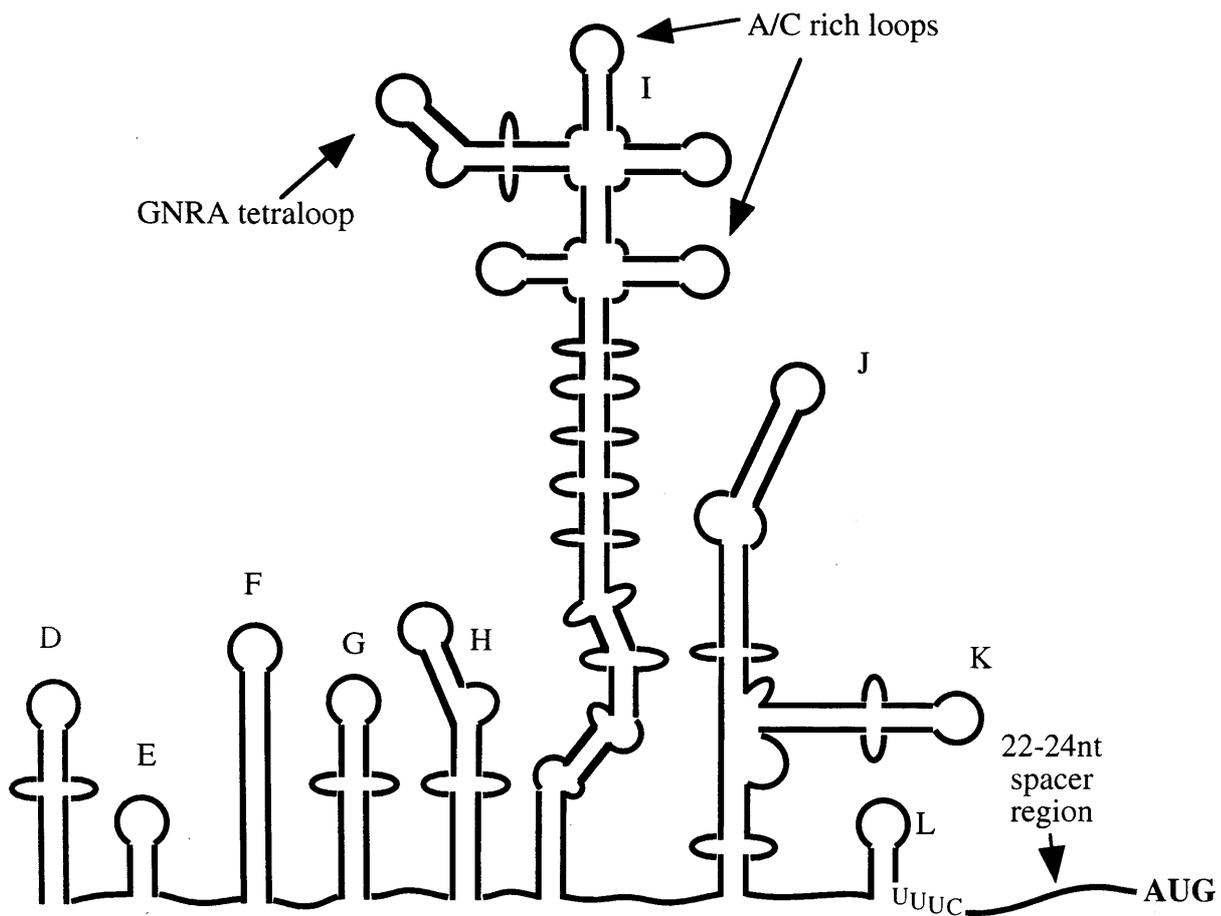


Figure 1.iv. Secondary structure of a class II picornavirus IRES. The authentic initiation codon is in bold type, and the structure is about 550nt long in total. The model was derived by computer modelling and chemical/enzyme probing, and is adapted from Stewart and Semler, 1997.

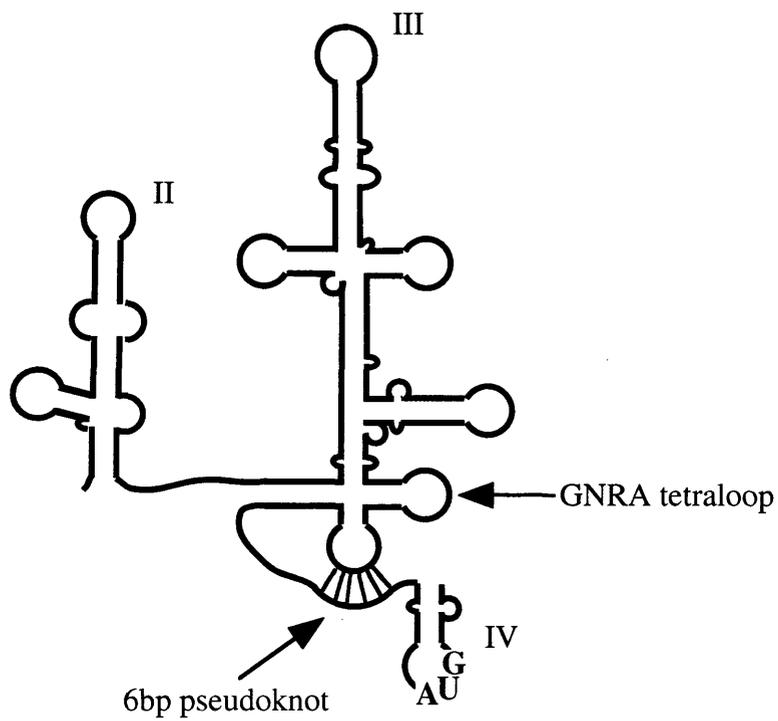


Figure 1.v. Secondary structure of the HCV IRES. The authentic initiation codon is in bold type, and the structure shown is 492nt long in total. The model is adapted from Lemon and Honda, 1997.

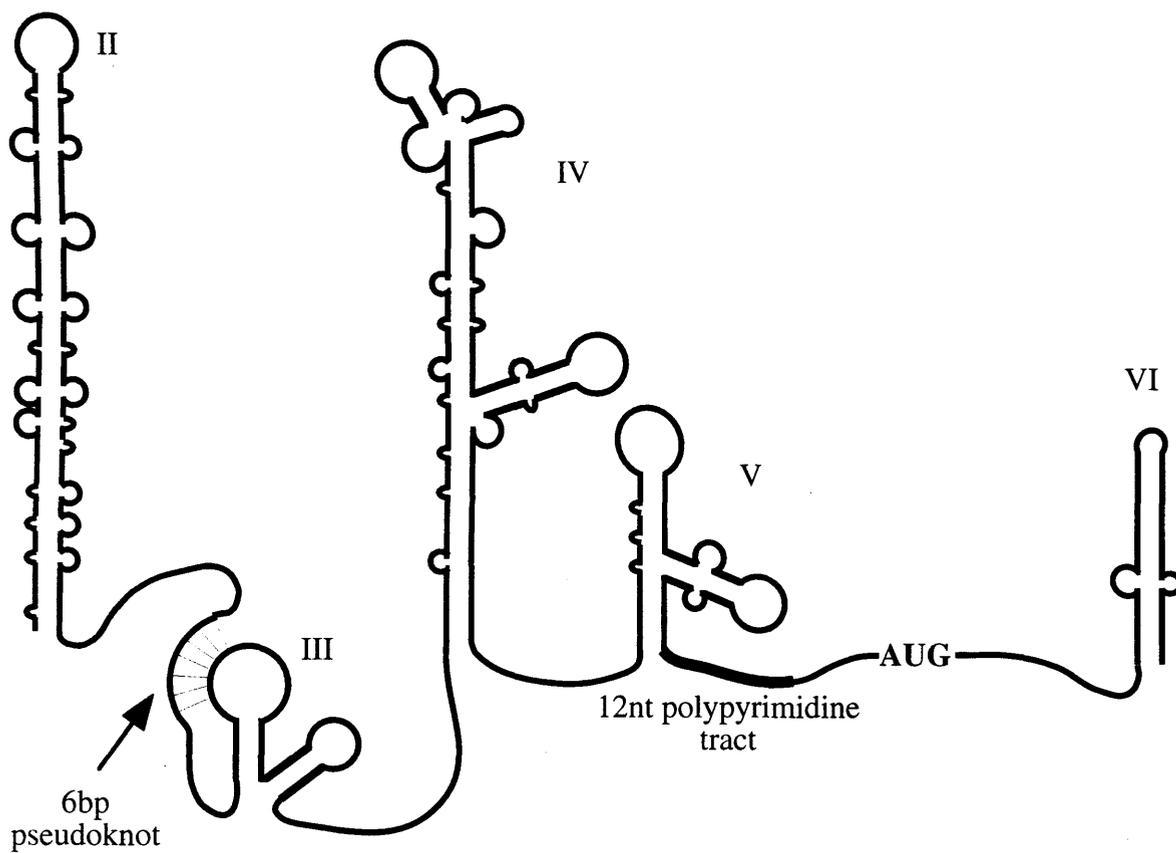


Figure 1.vi. Secondary structure of the GBV-C IRES. The authentic initiation codon is in bold type, and the structure shown is 550nt long. The model is based largely upon sequence analysis, and is adapted from Lemon and Honda, 1997.

Attempts have also been made to unify the picornavirus, hepacivirus and pestivirus IRES structures by once again postulating a common structural core (Le *et al.*, 1996). This depends upon the same pseudoknot motif as the above model, and suffers from the same criticism. Furthermore, the gross differences in mechanism between HCV and picornavirus IRESs are consistent with divergent structures, and the unclassified flavivirus IRES sequences, such as the GBV-C IRES, are wholly incompatible with this model.

Finally, a similar study sought to determine a common secondary structure for a range of cellular IRESs, including BiP, FGF2 and antennapedia, of which the main determinant is a “Y-shaped motif” (Le and Maizel, 1997). Again, in the absence of physical data it is impossible to confirm or eliminate this model, and there is insufficient mechanistic data regarding these IRESs to support the speculation that they rely upon similar intermolecular contacts.

Too much significance can be attributed to secondary structure models. There are bound to be many combinations of primary and secondary structures that will satisfy the need to present particular moieties in particular positions in three-dimensional space, thus eliminating any absolute theoretical need for congruence of secondary structure of taxonomically distant IRESs, even if they do rely upon similar intermolecular contacts. The danger of wholly theoretical phylogenetic studies is that in the search for principles governing IRES secondary structure, “rules” will emerge that merely reflect the coincidental ability of a number of sequences to share a common theoretical structure. The more relaxed these rules are, the more likely it is that the sequences under study will comply, and for any given set of sequences, a common structure will be found if the rules are made sufficiently lax.

Sequence

Even within a species, viral IRESs can differ considerably at the sequence level. HCV IRES sequences, for example, diverge by up to 15% (Bukh *et al.*, 1992), and about 50% of loci within the IRES have been seen to vary. The sequences of the structurally similar class I entero- and rhinovirus IRESs vary by as much as 36% (Rivera *et al.*, 1988). This variation is largely silent, and is typically accommodated in helical portions of the IRES in the form of covariant base-pairs. This is consistent with a scaffolding role for much of the IRES RNA, serving to present generally unpaired bases to the translational machinery in a suitable three-dimensional conformation.

Conserved terminal loops and internal bulges are strong candidates for such sequence-dependent contact points, and many mutations in such areas have been found to negatively impact IRES function. This might reflect loss of binding of rRNA or sequence-specific proteins. Alternatively, “unpaired” bases may be directly involved in tertiary contacts with other parts of the IRES, via base triples, tetraloop/receptor interactions and other, base/backbone contacts. Thus the observation that a mutant RNA has a perturbed protein-binding profile does not, unfortunately, prove that the mutant locus is a primary *trans*-acting factor recognition determinant.

A number of small conserved sequence elements have been identified. Picornaviral class I and II IRESs both contain conserved GNRA tetraloops, in domains IV and I respectively. The sequence of this loop is crucial for EMCV (Robertson *et al.*, 1999) and FMDV (deQuinto and MartinezSalas, 1997) IRES function, although it remains unclear whether it interacts intra- or inter-molecularly. The current picornaviruses IRES models do not contain classic type I ribozyme-like (Cate *et al.*, 1996) tetraloop receptors, however.

Class I and II IRESs also contain A/C rich terminal loops, which are less well conserved than the GNRA loop, but nonetheless contribute to FMDV IRES function in a sequence-dependent manner (deQuinto and MartinezSalas, 1997).

It is tempting to speculate that particular conserved bases in these and other, smaller, single-stranded loops might be involved in Watson-Crick base-pairing with 18S rRNA. But, because the single-stranded sections are so short, it is not possible to identify regions of significant complementarity within rRNA sequences.

All picornavirus IRES elements also contain an absolutely conserved UUUC sequence, followed by a G-poor “spacer” 17-25 nucleotides long, followed by an AUG. This motif lies at the 3' extremity of the IRES. In class I picornavirus elements, the AUG is believed to lie in a helical portion of domain VI, and there is some potential for a short helical segment involving part of the spacer region, although this is poorly conserved. In class II elements, the motif just overlaps with the 3' proximal structural domain. The AUG component of the motif is crucial; it is the only AUG codon within the IRES whose integrity is essential for a competent poliovirus phenotype (Pelletier *et al.*, 1988), and is probably the authentic site of ribosome entry. The UUUC sequence is often important too, as demonstrated by numerous observations that deletions and point substitutions negatively affect IRES function (Gmyl *et al.*, 1993; Haller and Semler, 1992; Kuge and Nomoto, 1987; Kuhn *et al.*, 1990). Finally, the spacing between the UUUC and AUG sequences is also influential (Jang *et al.*, 1990; Pilipenko *et al.*, 1992). One model proposes that the UUUC sequence acts by hybridizing to a complementary GAAG sequence very close to the 3' terminus of the 18S rRNA, and that the AUG hybridizes with a CGU sequence 14 bases further upstream (Figure 1.vii). This model is attractive, being reminiscent of the Shine-Dalgarno interaction involved in prokaryotic start codon selection. However, the presence of highly conserved wobble pairings in both complementary boxes is

18S rRNA	3'	AUUACUAGGAAGGCGUCCAACUGGAUGCCUU	5'
		** ***** *****	
Type I IRES RNA	5'	GUGUUUCc(20-25) gcUUAUGG	3'

Figure 1.vii. Possible hybridization of the 3' terminal region of 18S rRNA and picornaviral IRES RNA. Lower case indicates incomplete conservation among available type I IRES sequences. Pairing potential is indicated by asterisks. Adapted from Pilipenko *et al.*, 1992.

surprising, especially in the case of the AUG. Observations that in class II picornavirus IRESs this AUG is the authentic start codon suggest that a role in rRNA contact is at best a secondary one. Conceivably, the AUG might sequentially interact with the rRNA (via a wobble interaction), and subsequently, precisely, with the anticodon. Neither proposed interaction has as yet been tested by chemical probing of rRNA in the presence and absence of IRES RNA.

There are numerous nucleotides that are invariant in hepacivirus and pestivirus IRESs (Hellen and Pestova, 1999), including a GNRA tetraloop, but there are as yet few indications as to precisely which roles these segments are playing. Interestingly, UV cross-linking of stable HCV IRES/40S subunit binary complexes only revealed one contact with ribosomal protein, apparently between ribosomal protein S9 and a region of the IRES downstream of the pseudoknot (Pestova *et al.*, 1998b), suggesting that the body of the HCV IRES might solely be contacting 18S rRNA.

None of these sequence elements have been reliably identified in cellular IRESs, and the general lack of homologous cellular IRES sequences has hampered identification of IRES sequence elements by alignment analysis.

The *c-myc* IRES has one arresting feature at the sequence level, namely a complete absence of AUG codons in any of the sequences available. One expects non-functional start codons to appear in IRES sequences by genetic drift. This suggests that the *c-myc* 5' UTR is not under all circumstances highly structured, but is at times scanned through like any other leader sequence.

Start codon selection

IRESs fall into two classes: those that position the 40S ribosomal subunit accurately over the authentic initiation codon, and those that set the ribosome scanning toward the first

AUG in good context. Thus far, all cellular IRESs that have been adequately investigated seem to fall into the latter category.

The class II picornaviruses EMCV and TMEV, HCV and pestivirus IRESs typify the “precise placement” model. In these cases, initiation fails if the spacing between the IRES element and the AUG is either increased or reduced.

In HCV the spacing between the pseudoknot element and the AUG is especially critical, as AUGs as little as eight bases closer or further away are not utilised (Reynolds *et al.*, 1996; Rijnbrand *et al.*, 1997). This is consistent with a model of the HCV IRES as a rigid structure, releasing the ribosomal subunit only when the codon-anticodon interaction has been established, perhaps concomitantly with large subunit joining. The pseudoknot in HCV and pestivirus IRESs seems likely to act as a “backstop” for the small ribosomal subunit, as the spacing between the pseudoknot and the AUG (12-13 nucleotides) correlates closely with the estimated length of the portion of the mRNA binding track from the 5' edge of the subunit to the P-site.

The picornaviruses have a greater spectrum of activities. In the class I picornavirus poliovirus IRES, infectivity is dependent upon the presence of the AUG associated with the spacer region (Pelletier *et al.*, 1988), yet practically no initiation takes place at this codon, but instead at the next AUG, 155 nucleotides downstream. Insertion of extraneous AUGs between these two positions reduces utilization of the downstream AUG by an amount somewhat less than would be expected if a mechanism of scanning identical to that seen in cap-dependent initiation was operating (Hellen *et al.*, 1994). Possibly the interaction between poliovirus IRES and the 40S subunit is a “sticky” one, and AUGs further downstream are favoured because it takes time for this association to be broken, permitting 60S subunit joining.

In TMEV, a class II IRES, AUG utilization was shown to be confined to a “window” 12 nt long, situated 16 nt downstream of the UUUC element (Pilipenko *et al.*, 1994), reminiscent of hepacivirus. The relatively large size of the window perhaps belies a less rigid structure, consistent with the larger size and greater scope for articulation between domains in the picornavirus IRES as compared to HCV. In the case of EMCV, another class II picornavirus, nearly all initiation takes place at the spacer-associated AUG (Kaminski *et al.*, 1994). The obvious hypothesis, that this difference in AUG utilization between class I and II IRESs is mediated by divergence in the structural and sequential context of the spacer-associated AUGs has been disproved. Experiments in which chimeric type I/II IRESs, joining at the CUUU element, were constructed and assayed *in vitro*, show that start codon selection is overwhelmingly influenced by the class origin of the 5' portion of the IRES (Ohlmann and Jackson, 1999). Thus it is the “body” of the IRES that determines whether subunit joining occurs at once, or whether the spacer-associated AUG is not utilised and prolonged scanning takes place.

How can an AUG act as a necessary determinant of 40S subunit entry, yet not be used as an initiation codon? The most closely analogous situation is “leaky scanning”, the failure of the scanning ribosome to recognise an AUG during cap-dependent initiation. According to the kinetic model of start codon selection, AUG recognition depends upon the 40S subunit pausing for long enough for eIF5-stimulated hydrolysis of eIF2 bound GTP. One could speculate that the class I IRES body (and associated protein factors) interferes with the action of eIF5 and/or eIF5B by, for example, sterically blocking the eIF5B binding site on the 40S subunit. Thus the AUG would be important for accommodation of the mRNA into the binding track on the ribosome, but would not be recognised as a start codon. Scanning would take place, but AUG recognition would depend upon disassociation of the IRES from the 40S

subunit, which itself might be a “kinetic” process, as mentioned above. Alternatively, in the class I IRESs, the AUG might not be making a codon-anticodon interaction at all, but might merely be involved in contacting 18s rRNA. This is consistent with the observation that a mutilated class I IRES can functionally revert to an ACG codon (Pilipenko *et al.*, 1992), which is more perfectly complementary. Reversion to AUG is much more common, however. The former model is preferable since it predicts less divergent roles for the conserved AUG codons of class I and class II IRESs.

There is no rule that IRES-driven translation must start at an AUG codon, as demonstrated by the recent discovery that the insect picorna-like virus PSIV has an IRES that drives translation from a CUU codon (Sasaki and Nakashima, 1999). In monocistronic constructs, translated by scanning through a 3' fragment of the IRES, the CUU is not used, indicating that the selection of this codon is mediated by upstream portions of the IRES via a scanning-independent mechanism.

The IRESs present in the 5' UTRs of the cellular genes *c-myc* and VEGF have both been shown to contain IRES elements which, when bisected, show IRES activity resident to either the truncated 5' or 3' portions. This is variously interpreted as evidence for a diffuse distribution of IRES function along the length of the molecule (Stoneley *et al.*, 1998), or for “two IRESs” (Huez *et al.*, 1998). In the case of *c-myc*, the halves have a synergistically cumulative effect; in VEGF, the effect is certainly cumulative, although in the absence of numerical data it is impossible to determine whether the 5' and 3' portions contribute to IRES function synergistically. In any case, if two truly independent IRESs were present on the same message, it is hard to imagine how, under any circumstances, internal initiation could be more efficient than that driven by the strongest of the two.

Certainly the IRES elements (whatever their number) contained within the 5' UTRs of *c-myc* and VEGF do not conform to the "precise placement" model of start codon selection as typified by the hepaciviruses and class II picornaviruses. This is clear from the observation that IRES activity tolerates substantial deletions at the 3' end of the 5' UTR, effectively altering the relative positioning of start codon and vital IRES elements. This leaves two possibilities: a mechanism that proceeds by a more-or-less normal scanning process, or a continued influence of IRES complex components in start codon selection, though in a less structurally rigid fashion than that observed for HCV and class II picornavirus IRESs. In *c-myc*, two closely positioned start codons are used, an AUG and an upstream CUG, resulting in two forms of Myc protein.

Trans-acting factors

In order to support initiation, an IRES must functionally replace the cap-binding complex. One extreme view suggests that an IRES might merely mimic the m7G cap, and recruit a full set of canonical initiation factors to the interior of the message. Alternatively, the IRES might have reduced factor requirements, the RNA structure functionally replacing or obviating the need for some or all cap-associated factors. Finally, a particular IRES might require additional protein factors, either to act as scaffolding proteins, facilitating IRES RNA-translation machinery interactions, or to act as bridges, interacting simultaneously with IRES RNA and the translation machinery.

Class II IRESs generally function strongly in unsupplemented reticulocyte lysates (Borman *et al.*, 1995). Furthermore, stable 48S complexes can be assembled upon EMCV IRES RNA *in vitro*, using purified eIFs 1, 1a, 2, 3, 4A, 4B and 4F (Pestova *et al.*, 1998a). The requirement for eIF4F is less than absolute, as eIF4E can be omitted without negative impact (Pestova *et al.*, 1996), and both the N-terminal and C-terminal extremities of eIF4G,

responsible for the interaction with eIF4E and eIF4A respectively, can be disposed of with impunity (Ohlmann *et al.*, 1996; Pestova *et al.*, 1996). UV cross-linking studies suggest that eIF4G and eIF4B both directly interact with class II IRESs (Meyer *et al.*, 1995; Pestova *et al.*, 1996). Both proteins have been reported to interact with 43S complexes, and it is probable that ribosome recruitment is at least partially mediated in this way. Under normal circumstances, the EMCV IRES does not seem to require any additional, non-canonical, initiation factors.

The HCV IRES has a greatly reduced requirement for initiation factors (Hellen and Pestova, 1999). IRES function is unaffected by dominant negative forms of eIF4A (Pestova *et al.*, 1998b), shows no requirement for eIFs 4B, 4E, 4G and only requires eIF3 at the late stage of eIF5/5A stimulated subunit joining. Initiation factors 1 and 1a have a barely perceptible effect upon 48S complex formation on the closely related CSFV IRES RNA, as monitored by reverse transcriptase toeprinting (Pestova *et al.*, 1998a). No non-canonical factors are implicated, resulting in a model highly reminiscent of prokaryotic internal entry (Pestova *et al.*, 1998b), mediated solely by the RNA, 40S subunit and eIFs 2 and 3. HCV IRES RNA is even capable of forming stable binary complexes with 40S subunits alone, successfully positioning the initiation codon very close to the ribosomal P-site.

Class I picornaviral IRESs are considerably more demanding. It was known from early on that poliovirus initiation in rabbit reticulocyte lysate is significantly enhanced and rendered more faithful by the addition of cytoplasmic extracts of HeLa cells (Brown and Ehrenfeld, 1979; Dorner *et al.*, 1984), demonstrating that they require factors that are more abundant in such extracts than in RRL for efficient function. Numerous HeLa cell proteins have been shown to directly interact with class I IRES RNAs, but they have proved difficult to purify and identify.

The first to be identified was the La autoantigen (Meerovitch *et al.*, 1993), which improves poliovirus translation fidelity in RRL, albeit at concentrations orders of magnitude greater than those found *in vivo*. Furthermore, adding back recombinant La to immunodepleted HeLa translating extracts does not restore function (Stewart and Semler, 1997).

A more convincing case is made for the functional involvement of the 57 kDa pyrimidine-tract binding protein (PTB) (Borman *et al.*, 1993; Hellen *et al.*, 1993; Pestova *et al.*, 1991). PTB is a predominantly nuclear protein, involved in alternative splice site selection (Perez *et al.*, 1997; Valcarcel and Gebauer, 1997; Zhang *et al.*, 1999). The addition of recombinant PTB to RRL greatly stimulates translation from entero- and especially rhinovirus IRESs (Hunt and Jackson, 1999). This activity is not limited to the class I picornavirus IRESs; FMDV and EMCV IRESs are also stimulated by PTB *in vitro*, albeit to a much lesser extent (Kaminski and Jackson, 1998; Niepmann *et al.*, 1997). A serendipitously generated mutant form of the EMCV IRES bearing an additional A residue in the A-rich bulge was discovered to be far more dependent upon PTB for efficient initiation than the wild-type (Kaminski and Jackson, 1998; Niepmann *et al.*, 1997). Thus PTB seems able to compensate for the mutation, suggesting a role as scaffolding, maintaining three-dimensional IRES integrity, rather than bridging between the IRES and the translational machinery. PTB has also been shown to bind the IRES-containing 5' UTR of the gene VEGF (Vascular Endothelial Growth Factor) (Huez *et al.*, 1998), but binding to mutant forms of the IRES *in vitro* does not correlate with their activities *in vivo*.

A second HRV IRES-stimulating activity was identified from ion-exchange chromatography of HeLa cytoplasmic lysate (Hunt and Jackson, 1999). The active component of this fraction is the 97 kDa protein UNR (Upstream of N-Ras) (Hunt *et al.*, 1999). UNR and

PTB synergistically enhance initiation from the HRV IRES in RRL, suggesting a model in which both proteins are required for the assembly of a competent IRES superstructure. UNR is predicted to have no less than five RNA-binding cold-shock domains, all of which are required for the stimulating activity (Jackson, 1999). Again, this suggests a role in maintaining higher-order IRES structure. Surprisingly, UNR has no significant enhancing effect upon the PTB-dependent stimulation of poliovirus IRES function (Hunt *et al.*, 1999).

The human cellular protein PCBP2 (poly(rC) binding protein 2) was found to interact with wild-type and mutant forms of the poliovirus IRES domain IV in a way that correlated with IRES activity in HeLa lysate enriched RRL (Blyn *et al.*, 1995; Blyn *et al.*, 1997). HeLa lysates affinity-depleted of PCBP2 are defective with respect to poliovirus IRES translation, and function is restored by addition of the recombinant protein (Blyn *et al.*, 1997). It remains to be seen how closely analogous is the role played by PCBP2 in poliovirus translation to that played by UNR in HRV translation.

No non-canonical initiation factors have as yet been identified that act upon cellular IRESs. Numerous proteins have been shown to bind cellular IRESs including BiP (Yang and Sarnow, 1997), *c-myc* (Paulin *et al.*, 1998) and VEGF (Huez *et al.*, 1998). Unfortunately, until such factors are purified there can be no meaningful investigation of the roles, if any, that they play in internal initiation.

1.3 C-*myc*

The c-*myc* family

C-*myc* was originally identified as a cellular homologue of v-*myc*, an oncogene from the avian myelocytomatosis virus (Vennstrom *et al.*, 1982). It has since been cloned from a number of species, in which it retains a high degree of sequence identity. The *myc* family of genes has emerged with the discovery of L-*myc*, overexpressed in small cell lung cancer (Nau *et al.*, 1985), and N-*myc*, overexpressed in neuroblastoma (Schwab *et al.*, 1983). The three genes share regions of high conservation and a three-exon organisation (Figure 1.viii). Two other peripheral family members have been identified, named S-*myc* and B-*myc* (Marcu *et al.*, 1992).

Myc protein

The Myc protein exists in two isoforms of mass 67 and 64 kDa, named Myc 1 and Myc 2. The size difference arises from the utilization of two translation initiation codons, a canonical AUG in exon 2 and a non-canonical CUG codon in exon 1, 45 nucleotides upstream in the mature transcript. Under normal conditions, Myc 2 (the AUG-initiated form) represents the majority of cellular Myc protein. Some evidence suggests that Myc 1 is antagonistic to Myc 2 (Hann *et al.*, 1994), and is up-regulated as cell density rises as a response to nutrient deprivation (Hann *et al.*, 1992).

The protein has been shown to contain several functional domains: a proline/glutamine rich transcriptional activation domain, a basic region, a helix-loop helix motif, a leucine zipper, and a nuclear localization signal (Kato and Dang, 1992) (Figure 1.ix). Three phosphorylation sites have also been identified.

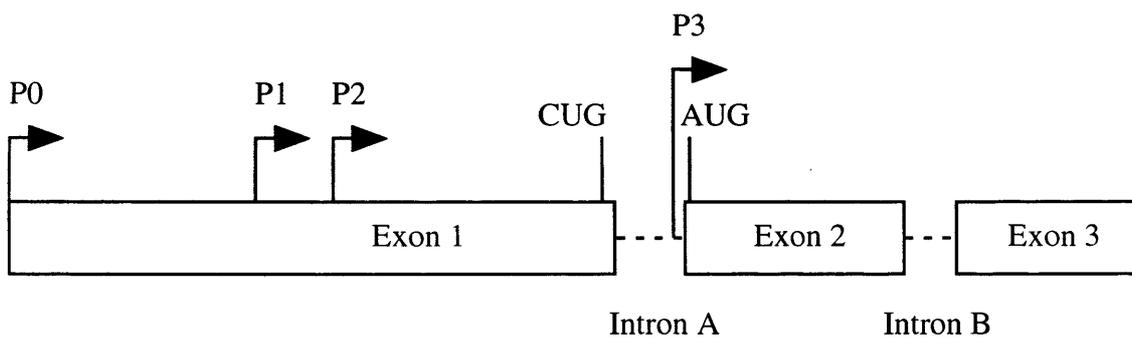


Figure 1.viii. Exon structure of the *c-myc* gene. The length of exon 1 is variable, as a number of promoters are used *in vivo*, resulting in UTRs with lengths of ≈ 1000 , 554, 394 and 20 nt respectively.

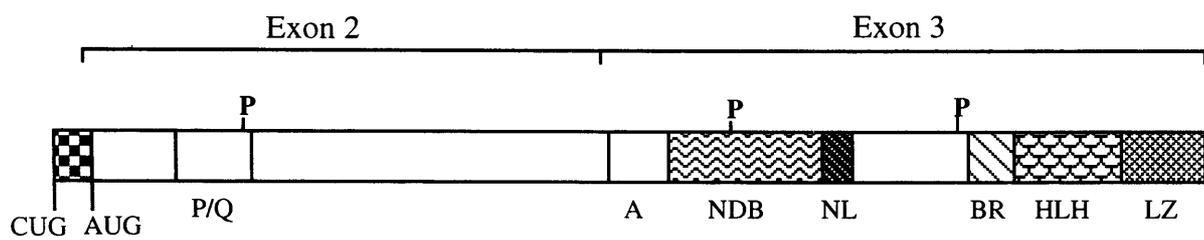


Figure 1.ix. Domain structure of the MYC protein. P/Q: Transactivation domain A: Acidic domain NDB: Nonspecific DNA binding domain NL: Nuclear localization BR: Basic region HLH: Helix-loop-helix domain LZ: Leucine Zipper

C-*myc* is a transcription factor

The observation that Myc is unable to dimerize at physiological concentrations led to the search for an alternative binding partner. The *max* gene was identified, which encodes a protein that dimerizes with Myc via its own helix-loop-helix and leucine zipper domains (Blackwood and Eisenman, 1991). This dimerization is essential for the *myc*-dependent stimulation of cell cycle progression and apoptosis (Amati *et al.*, 1993).

The Myc-Max heterodimer binds DNA at “E-box” sequence elements with the loose consensus CACGTG, and stimulates transcription from them *in vitro* (Blackwell *et al.*, 1993; Blackwell *et al.*, 1990). Positive identification of authentic target genes is difficult, since several other transcription factors recognise E-box motifs *in vivo* (Blackwell *et al.*, 1993). Nonetheless, a number of target genes have been identified. Those with known functions are generally involved in cell growth (Dang, 1999). Examples include ornithine decarboxylase (Bello-Fernandez *et al.*, 1993), involved in DNA synthesis, the translation initiation factors eIF4E (Jones *et al.*, 1996) and eIF2 α (Rosenwald *et al.*, 1993), and the cell cycle regulator p53 (Reisman *et al.*, 1993).

Normal c-*myc* function

C-*myc* is crucially involved in the determination of cell fate, and its degree of expression is functionally related to the selection of a quiescent, proliferating, differentiated or apoptotic state. Being an immediate early response gene, it is rapidly induced by the stimulation of quiescent cultured cells with growth factors (Kelly *et al.*, 1983). Levels of Myc protein remain high throughout the cell cycle (Hann *et al.*, 1985). Conversely, if growing cells are deprived of growth factors, Myc protein levels drop concomitantly with entry into G₀.

There is a causal relationship between Myc expression and cell cycle progression. Artificial reduction of Myc protein levels result in retardation (Shichiri *et al.*, 1993) or outright

inhibition (Heikkila *et al.*, 1987) of entry into S-phase. Furthermore, artificial induction of Myc in the absence of growth factors results in DNA synthesis (Eilers *et al.*, 1991).

Generally, elevated levels of Myc protein are incompatible with differentiation. Artificial elevation of Myc protein levels result in persistent cell cycling, precluding differentiation (Coppola and Cole, 1986; Dmitrovsky *et al.*, 1986). Also, a range of terminally differentiated human tissues have undetectably low levels of Myc (Marcu *et al.*, 1992). Furthermore, in some cases the depletion of Myc protein by an antisense approach will induce differentiation as well as growth arrest (Prochownik *et al.*, 1988).

Finally, a link between Myc expression and entry into apoptosis has emerged in some systems. Serum-starved cells in culture can be stimulated to apoptose by the artificial induction of Myc activity (Evan *et al.*, 1992), and such Myc-induced apoptosis can be inhibited by the addition of cellular survival factors. This and other observations have led to the formulation of a model in which Myc protein is seen as a constitutive inducer of apoptosis, unless the cell is supplied with the correct survival factors, in which case Myc drives proliferation (Prendergast, 1999).

C-myc is a proto-oncogene

The expression of *c-myc* is deregulated in a large number of tumour types. This often results from alterations at the *c-myc* locus such as chromosomal translocations and gene amplifications (Nesbit *et al.*, 1999). In addition to genomic modifications, mechanisms have been described in tumour cell lines that increase Myc levels through enhanced translation or protein stabilization (Paulin *et al.*, 1996; Shindo *et al.*, 1993; West *et al.*, 1995).

Deregulated *c-myc* expression in untransformed cell lines results in partial transformation, reflecting the critical position that *c-myc* occupies in transducing growth inhibitory and stimulatory signals. However, constitutive expression of *c-myc* alone does not

induce a tumorigenic phenotype. Nevertheless, *c-myc* can cooperate with other oncogenes such as activated *ras* and Bcr-Abl in the transformation of many cell types suggesting that secondary genetic events are necessary to induce a malignant phenotype (Marcu *et al.*, 1992). Indeed, transgenic mice constitutively expressing *c-myc* in the B-lymphocyte lineage develop clonal lymphomas after a variable latency period, and some of these tumors carry a mutated *ras* gene (Alexander *et al.*, 1989). These observations indicate that deregulated *c-myc* expression predisposes for but is not sufficient to induce tumourigenesis.

Control of *c-myc* expression

C-myc expression is tightly controlled, and regulation occurs at several levels. This is presumed to reflect the profound implications that over-expression of Myc has for cell proliferation.

Transcription of *c-myc* is regulated both positively and negatively. Promoter structure is complex, accounting for the variation in length of the 5' UTR (Figure 1.viii). Normally, about 75-90% of *c-myc* mRNAs originate from the P2 promoter, resulting in a 394 nt 5' UTR. Most of the other transcripts derive from P1, having a 554 nt 5' UTR (Taub *et al.*, 1984). Two minor cryptic promoters also operate, P0 and P3, which, between them, produce about 5% of *c-myc* messages.

Growth factors upregulate transcription initiation via *cis*-acting sequence elements at the *c-myc* promoter (Marcu *et al.*, 1992). Transcription is also regulated at the level of elongation, with premature termination occurring at the 3' end of exon 1 (Bentley and Groudine, 1986), which can be relieved in cultured cells by the addition of mitogens (Nepveu *et al.*, 1987). The *c-myc* mRNA has an extremely short half-life, of the order of 15 minutes (Brewer, 1998), and modulation of expression also occurs at this level (Yeilding *et al.*, 1998).

Considerable evidence has accumulated for control of Myc expression at the level of translation. The constitutive expression of the entire human *c-myc* gene in murine fibroblasts results in the accumulation of high levels of mRNA in the cytoplasm without a significant increase in the levels of *c-myc* protein (Ray *et al.*, 1989). Furthermore, in cell lines derived from patients with multiple myeloma and Bloom's syndrome *c-myc* expression is de-regulated by translational mechanisms (Paulin *et al.*, 1996; West *et al.*, 1995). In the case of multiple myeloma this is associated with the single point mutation C255U (numbering from the 5' end of the P2 transcript) in the 5' UTR of the mRNA (Paulin *et al.*, 1996).

Chromosomal translocations occurring between the *c-myc* locus and one of the immunoglobulin loci result in de-regulated *c-myc* expression. Breakpoints in exon 1 or intron A occur in approximately 70% of murine plasmacytomas and 50% of human lymphomas (Cory, 1986), suggesting that exon 1 might be contributing to the normal regulation of *c-myc* expression. The high degree of sequence conservation of exon 1 between species supports this idea.

There is abundant evidence that Myc can be synthesized via a cap-dependent translation initiation pathway in various systems, *viz.* the attenuation of *c-myc* translation by secondary structure in its 5' UTR in rabbit reticulocyte lysate and *Xenopus* oocytes (Darveau *et al.*, 1985; Parkin *et al.*, 1988; Stoneley, 1998), the increase of *c-myc* translation in cells overexpressing eIF4E (De Benedetti *et al.*, 1994), and the absence of extraneous AUG codons within known P2 5' UTR sequences.

The intriguing discovery that the *c-myc* 5' UTR contains an IRES active *in vivo* (Nanbru *et al.*, 1997; Stoneley *et al.*, 1998) shows that Myc protein can be translated by two alternative mechanisms. There is also evidence suggesting that scanning through the *c-myc* 5' UTR unwinds the IRES, rendering it inactive (Stoneley *et al.*, 2000b), so that translation can

proceed by one mechanism or the other, but not both. It might be expected, therefore, for the to IRES become active at a time when cap-dependent initiation is impaired.

Apoptosis is one such time, during which translation is inhibited by caspase-mediated cleavage of eIF4G (Marissen and Lloyd, 1998). Indeed, recent work (Stoneley *et al.*, 2000a) shows that the *c-myc* IRES is up-regulated during apoptosis, maintaining Myc levels at a time when cap-dependent initiation is de-activated. Taken together with the observation that *c-myc* mRNA remains polysomally associated after poliovirus-induced eIF4G cleavage (Johannes *et al.*, 1999; Johannes and Sarnow, 1998), these data suggest a mechanism for the *c-myc* IRES reminiscent of the picornaviral paradigm, in which a cleaved product of eIF4G is involved in internally initiated translation.

Thus a great many questions remain regarding the function of the *c-myc* IRES. This work aims to gain a better understanding of IRES structure and structure-function relationships, and of the role of the *c-myc* IRES within the wider context of cellular physiology.

Chapter 2

Materials and Methods

2.1 General Reagents

Unless otherwise stated all chemical reagents were of analytical grade and most were obtained from BDH laboratory supplies (Lutterworth, Leicestershire, UK), Fisons (Loughborough, Leicestershire, UK), ICN Flow Ltd (Thame, Oxfordshire, UK) or Sigma Chemical company Ltd (Poole, Dorset, UK). Products for molecular biological techniques were routinely purchased from Boehringer Mannheim UK Ltd (Lewes, East Sussex, UK), Gibco-BRL (Paisley, Scotland), Stratagene Ltd (Cambridge, UK), New England Biolabs (NEB) (c/o CP Labs, Bishops Stortford, Hertfordshire, UK), MBI Fermentas (c/o Helena Biosciences Ltd, Sunderland, Tyne & Wear, UK) and Pharmacia Biotech (Milton Keynes, Buckinghamshire, UK). Reagents for bacterial cell culture were obtained from Oxoid (Unipath, Basingstoke, Hampshire, UK). Foetal calf serum (FCS) for mammalian tissue culture was obtained from Advanced Protein Products (APP) (Brierly Hill, UK) and Wolf laboratories (York, UK). All tissue culture plastic was supplied by Nunc products (Gibco-BRL) with the exception of six-well plates, which were obtained from Greiner Laboratories (Lake Mary, Florida, US). Radiolabelled chemicals were obtained from Amersham International Plc (Little Chalfont, Buckinghamshire, UK) and NEN Dupont (Hounslow, UK).

2.2 Tissue Culture Techniques

Tissue culture media and supplements

RPMI 1640 medium: Rose Park Memorial Institute 1640 medium, with L-glutamine (Gibco-BRL) was supplemented with 10% foetal calf serum (FCS) (Advanced protein products).

DMEM medium: Dulbecco's modified eagle medium, without sodium pyruvate (Gibco-BRL) was supplemented with 10% FCS.

Cell Lines

HeLa S3: Human cervical epithelioid carcinoma, maintained in DMEM

GM2132: Multiple myeloma patient-derived lymphoblastoid cell line, maintained in RPMI 1640

Maintenance of cell lines

Cell lines were cultured in the appropriate growth medium supplemented with 10% FCS in sterile plasticware (Nunclon, Gibco-BRL). The adherent cells were grown to confluence in 10cm petri dishes and treated with 1X trypsin solution (Gibco-BRL) supplemented with 0.5 mM EDTA, pH 8.0. Approximately 1×10^6 cells were diluted into fresh medium and replated into a new dish. Cells grown in suspension were maintained at concentrations between 5×10^5 - 1×10^6 cells/ ml. All cells were routinely grown at 37°C in a humidified atmosphere containing 5% CO₂.

Calcium phosphate-mediated DNA transfection

Calcium phosphate-mediated DNA transfection of mammalian cells was performed essentially as described in Jordan *et al.* (Jordan, Schallhorn, and Wurm, 1996), with minor

modifications. Approximately 20 hours before transfection, 1×10^6 cells were seeded onto a 10cm plate in 9 ml of complete medium. A solution of 50 μ l of 2.5 M CaCl_2 and 25 μ g of plasmid DNA (20 μ g of luciferase plasmid and 5 μ g of p β -Gal) was diluted with sterile de-ionised water to a final volume of 500 μ l. This 2X Ca/DNA solution was added dropwise to an equal volume of 2X HEPES buffered saline (50 mM HEPES, pH 7.05, 1.5 mM Na_2HPO_4 , 140 mM NaCl) whilst bubbling air through the mixture. The calcium phosphate-DNA coprecipitate was allowed to form for 5 min and was then added slowly to the 9 ml of medium covering the cells. After exposing the cells to the precipitate for 15-20 hours at 37°C, the medium was removed and the cells were washed twice with phosphate buffered saline (4.3 mM Na_2HPO_4 , 1.5 mM KH_2PO_4 , 137 mM NaCl, 2.7 mM KCl, pH 7.4). Subsequently, fresh medium was added and the cells were grown for a further 24 hours before harvesting.

2.3 Bacterial Methods

Culture media and supplements

LB medium: 10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl dissolved in 1 l of de-ionized water.

LB agar plates: 10 g Bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl dissolved in 1 l of de-ionized water and supplemented with 15 g of agar.

SOC medium: 2 g Bacto-tryptone, 0.5 g Bacto-yeast extract, 1 ml of 1 M NaCl, 0.25 ml of 1 M KCl, 1 ml of 2 M MgCl₂, 1 ml of 2 M glucose.

Ampicillin: a stock solution of 50 mg/ml was prepared using sterile de-ionized water. Ampicillin was used at a final concentration of 50 µg/ml.

Bacterial strains

The *E. coli* strain JM109 was used in most bacterial manipulations:

JM109: *e14(mrcA)recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, Δ(lac-proAB), F', traD36, proAB, lacZΔM15*.

Epicurean Coli[®] XL1-blue supercompetent cells: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacZΔM15 Tn10 (Tet)]*

Preparation of competent cells

A single colony from an LB plate was inoculated into 2.5 ml of LB medium and incubated overnight at 37°C with shaking. The entire overnight culture was inoculated into 250 ml of LB medium supplemented with 20 mM MgSO₄ and incubated at 37°C until the A₆₀₀ reached 0.4-0.6. Cells were pelleted by centrifugation at 4,500 g for 5 min at 4°C using a GSA rotor (Sorvall). Following centrifugation, the pellet was gently resuspended in 100 ml of ice-cold filter sterile TFB1 (30 mM KAc, 10 mM CaCl₂, 50 mM MnCl₂, 100 mM RbCl, 15%

glycerol, adjusted to pH 5.8 with 1 M acetic acid). After incubating on ice for 5 min, the cells were centrifuged at 4,500 g for 5 min at 4°C. The pellet was resuspended in 10 ml of ice-cold filter sterile TFB2 (1 mM MOPS, pH 6.5, 75 mM CaCl₂, 10 mM RbCl, 15% glycerol, adjusted to pH 6.5 with 1 M KOH) and the cells were incubated on ice for 1 hour. Finally, the cells were rapidly frozen in an isopropanol/dry ice bath in 200 µl aliquots and stored at -70°C.

Transformation of competent cells

Ligation products or plasmid DNA (10 ng) were added to 50 µl of competent cells and incubated on ice for 20 min. After heating the mixture at 42°C for 2 min, 150 µl of SOC medium was added. Subsequently, the cells were incubated with shaking at 37°C for 45 min. Finally, the sample was spread onto a pre-warmed LB agar plate containing ampicillin and then incubated at 37°C for 16-20 hours.

2.4 Molecular Biology Techniques

CsCl₂-saturated isopropanol (ITC): Mix 10 g of CsCl₂, 10 ml TES, and 40 ml isopropanol.

Shake thoroughly to obtain a saturated solution and use the upper isopropanol phase.

OLB: Mix solutions A, B, and C in the ratio 2:5:3.

A: 1.2 M Tris-HCl, pH 8.0, 0.12 M MgCl₂, 1.75% β-mercaptoethanol, and 0.5 mM of dATP, dGTP and dTTP.

B: 2 M HEPES-NaOH, pH 6.6.

C: 1.6 mg/ml Hexadeoxyribonucleotides in 3 mM Tris-HCl, 0.2 mM EDTA, pH 7.

1X TBE: 89 mM Tris base, 89 mM Boric acid, 2.5 mM EDTA, pH 8.0

TE: 10 mM Tris-HCl pH 8.0, 1 mM EDTA

TAE: 40 mM Tris-Acetate pH 7.6, 1 mM EDTA

TBE: 89 mM Tris-Borate pH 8.3, 2 mM EDTA

Plasmids

pSK+Myc 5' UTR

pGL3R2

pSKLUTR

pRhpF

pβ-Gal

pGL3R2utr

pGL3Rhrv

pSKMΔ1

pSP64RUTR Poly(A)

Plasmid synthesis described in Stoneley, 1998.

Ethanol precipitation of DNA

DNA was precipitated by adding 0.1 volume of 3 M sodium acetate, pH 5.2 and 2 volumes of absolute ethanol. The sample was incubated on ice or at -20°C for 15-30 min following which the DNA was pelleted by centrifugation at 12,000 g for 10 min. Excess salt

was removed from the pellet by washing with 70% ethanol, then the DNA was dried briefly and resuspended in either TE or sterile de-ionized water.

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 12,000 *g* for 5 min. The upper aqueous phase was removed to a separate tube, to which an equal volume of chloroform:isoamyl alcohol was added. Following extraction and separation of the phases, the aqueous layer was transferred to new tube and the nucleic acid was precipitated.

Purification of DNA using glassmilk

Glassmilk was used to purify DNA fragments or to isolate DNA when a change of reaction buffer was required. Up to 5 μg of DNA was incubated with 3 volumes of 6 M NaI and 5 μl of glassmilk for 5 min at room temperature. The glassmilk was pelleted by centrifugation and washed two times in 0.5 ml of wash solution. After the glassmilk was resuspended in 10 μl of sterile de-ionized water, it was incubated at 45-55°C for 5 min to elute the DNA. Following centrifugation, the supernatant was removed to a fresh tube and the elution process was repeated.

Agarose gel electrophoresis

Fragments of DNA were fractionated according to their molecular weight by electrophoresis through agarose gels. Agarose was melted in 1X TBE buffer, cooled and cast into a gel. The gel was submerged in 1X TBE in a horizontal electrophoresis tank. Samples were mixed with 0.2 volumes of 5X TBE loading buffer and separated in the gel at up to

8V/cm. After electrophoresis, the gel was stained with ethidium bromide (1.3 mg/l in 1X TBE) for 15-20 min and the DNA was visualised on a UV transilluminator.

Gel isolation of DNA fragments

Initially, DNA fragments were separated by agarose gel electrophoresis as described, except the gel was prepared and submerged in 0.5X TAE. After staining with ethidium bromide, the fragments were visualised using a low intensity UV lamp. Agarose containing the required fragment was excised from the gel and melted at 55°C in 3 volumes of 6 M NaI. DNA was isolated from this solution using the glassmilk procedure as described.

Synthesis and purification of oligonucleotides

Oligonucleotides were synthesised on an Applied Biosystems model 394 machine (Protein and Nucleic Acid Sequencing Laboratory, Leicester University) at a 0.2 µM scale. Oligonucleotides were purified by ethanol precipitation with 0.1 volume of 3 M sodium acetate, pH 5.2 and 3 volumes of absolute ethanol. Samples were incubated at -20°C for 30 min and the precipitate was pelleted by centrifugation at 12,000 g for 20 min. After washing with 70% ethanol the pellet was dried and then resuspended in 100 µl of TE. The concentration of the oligonucleotide was determined by measuring the absorbance at 260nm.

Oligonucleotides

Details of the oligonucleotides employed are given in Table 2.a.

Restriction enzyme digestion

DNA was digested with restriction enzymes in a total volume of 10-50 µl under the conditions recommended by the suppliers. Reactions were incubated at the appropriate temperature for 1-2 hours.

Mung bean nuclease treatment

Mung bean nuclease was used to remove overhanging ends of endonuclease-cleaved DNA. DNA at 0.1 µg/ µl was treated with 1u/mg mung bean nuclease (NEB) in 1X mung bean buffer (50 mM Na(Ac), 30 mM NaCl, 1 mM ZnSO₄) for 1 hour at 37°C.

Alkaline phosphatase treatment of DNA

Linearised plasmids were treated with calf intestinal alkaline phosphatase (CIAP) to remove phosphate groups from the 5' ends and prevent self-ligation. Following restriction digestion, the restriction enzyme was inactivated by heating the reaction at 65°C for 15 min. Dephosphorylation was performed in a final volume of 50 µl in 1X restriction enzyme buffer. For DNA fragments with overhanging 5' ends the reaction was incubated for 30 min at 37°C using 1 unit of CIAP, after which another unit of enzyme was added and the incubation was repeated. For DNA fragments with blunt ends, the reaction was incubated at 37°C for 15 min followed by 56°C for 15 min using 1 unit of CIAP, and these incubations were repeated after the addition of another unit of CIAP. The reaction was terminated by heating at 75°C for 10 min and the DNA was purified using glassmilk. For those restriction enzymes that are resistant to heat-inactivation, the DNA was first purified using glassmilk and resuspended in 50 µl of 1X CIAP reaction buffer (50 mM Tris-HCl, pH 9.3, 1 mM MgCl₂, 0.1 mM ZnCl₂). The reactions were then performed as described above.

Ligations

Ligations were performed in a total volume of 10 µl. Vector DNA (50 ng) was mixed in a 1:3 molar ratio with insert DNA in a reaction containing 1X T4 DNA ligase buffer (30 mM Tris-HCl pH 7.8, 10 mM MgCl₂, 10 mM DTT, 10 mM ATP) and T4 DNA ligase. For ligations involving fragments with overhanging termini, the reaction was incubated at 16°C for

2-16 hours. Polyethylene glycol 8000 (4%) was included in reactions in which all DNA termini were blunt. The efficiency of these blunt ended ligations was further improved by incubating the reaction for at least 16 hours at 16°C. Alternatively, a cycle ligation was performed in which the temperature was alternated between 10°C and 30°C for 10 second periods over 16 hours. After incubation, 5 µl of the ligation reaction was transformed into competent *E.coli*.

Small scale preparation of plasmid DNA

A single colony of *E.coli* was inoculated into 5 ml of LB media containing ampicillin and incubated overnight at 37°C in a shaking incubator. Approximately 1.5 ml of the culture was decanted into a labelled tube and the bacteria were pelleted by centrifugation. The pellet was resuspended in 100 µl of ice-cold solution I (25 mM Tris-HCl, 10 mM EDTA, 50 mM Glucose, pH 8.0). After a 5 minute incubation at room temperature, 200 µl of solution II (1% SDS, 0.2 M NaOH) was added and the solutions were mixed gently. The sample was incubated on ice for 5 min, following which 150 µl of 7.5 M NH₄Ac, pH 7.6 was added. After briefly mixing the solutions using a vortex, the sample was incubated on ice for a further 5 min. The precipitated matter was pelleted by centrifugation at 12,000 g for 5 min and the supernatant was removed to a fresh tube. Plasmid DNA was ethanol precipitated from this solution as described. Finally, the washed and dried pellet was resuspended in 30 µl of TE. Diagnostic restriction digests were performed using 5 µl of this solution.

Large scale preparation of plasmid DNA

The ammonium acetate method was used to prepare milligram quantities of plasmid DNA. An overnight culture of *E. coli* containing the plasmid was inoculated into 250 ml of LB media supplemented with ampicillin. The culture was grown for 12-16 hours in a 37°C shaking

incubator. Cells were harvested by centrifugation at 5,000 g for 10 min at 4°C. The pellet was resuspended in 3 ml of ice-cold solution I and incubated at room temperature for 5 min. Then, 6 ml of solution II was added and the sample was incubated on ice for 10 min. This solution was neutralised with 4.5 ml of 7.5 M NH₄Ac, pH 7.6 and incubated for a further 10 min on ice. The precipitated matter was pelleted by centrifugation at 10,000 g for 10 min at 4°C and the supernatant was removed to a fresh tube. Isopropanol (0.6 volumes) was added and the solution was incubated at room temperature for 10 min. The insoluble material was pelleted by centrifugation (10,000 g) for 10 min at room temperature. The plasmid DNA in the pellet was resuspended thoroughly in 2 M NH₄Ac, pH 7.4. The insoluble matter was pelleted as before and the supernatant removed to a fresh tube. After the addition of 1 volume of isopropanol, the solution was incubated at room temperature for 10 min and the plasmid DNA was pelleted by centrifugation. Following resuspension of the pellet in 1 ml of sterile de-ionized water, contaminating RNA was removed by the adding 100 µg of RNase A and incubating the solution at 37°C for 15 min. Proteins were then precipitated by the addition of 0.5 volume of 7.5 M NH₄Ac, pH 7.6 and incubating at room temperature for 5 min. The precipitated proteins were pelleted by centrifugation and the supernatant was removed to a fresh tube. Finally, the plasmid DNA was precipitated using 1 volume of isopropanol, pelleted by centrifugation and washed with 70% ethanol. The resulting pellet was resuspended in a volume of 0.5-1 ml of TE.

Caesium chloride gradient purification of plasmid DNA

To achieve efficient formation of a calcium phosphate/DNA co-precipitate, the DNA was further purified on a CsCl₂ gradient. Plasmid DNA was resuspended in 8 ml of TE, into which 10 g of CsCl₂ were subsequently dissolved. The CsCl₂/DNA solution was transferred to an 11.5 ml polyallomer tube and supplemented with 0.5 ml of 10 mg/ml ethidium bromide. If

necessary, additional TE was added to give a final volume of 11.5 ml and the tube was sealed. Plasmid DNA was fractionated on a CsCl₂ gradient by centrifugation of the sample in a Sorvall Ti270 rotor at 50,000 rpm for 20 hours at 4°C. The supercoiled plasmid DNA was removed from the gradient using a syringe and separated from the ethidium bromide by repeated extraction with an equal volume of CsCl₂-saturated isopropanol (ITC). The aqueous solution was diluted with 2 volumes of de-ionized water and the plasmid DNA was precipitated by the addition of an equal volume of isopropanol and 0.1 volume of 3 M NaAc, pH 5.2. After centrifugation at 12,000 g for 10 min the pellet was resuspended in 0.5 ml of de-ionized water and plasmid DNA was ethanol precipitated as described previously. The final pellet was resuspended in 0.25-1 ml of 0.1X TE.

Double stranded DNA sequencing

Plasmid DNA was isolated using the small scale method and contaminating RNA was digested with 1 µg of RNase A at 37°C for 30 min. After RNase treatment, the DNA was ethanol precipitated and resuspended in 10 µl of sterile de-ionized water. The plasmid DNA was denatured by incubating this solution with 0.1 volumes of 2 mM NaOH, 2 mM EDTA, pH 8.0 at 37°C for 15 min. The solution was then neutralised with 0.1 volumes of 3 M KAc, pH 4.8, and 1 volume of isopropanol was added. Following incubation at room temperature for 10 min, the single stranded DNA was pelleted by centrifugation at 12,000 g for 10 min and air-dried. The pellet was resuspended in 10 µl of a 2.5 ng/ µl solution of sequencing primer and 2 µl of annealing buffer (280 mM Tris-HCl, pH 7.5, 100 mM MgCl₂, 350 mM NaCl). The plasmid DNA/primer solution was heated at 75°C for 10 min, and then incubated at 37°C for 10 min, followed by 5 min on ice to achieve primer annealing. Samples were labelled at 20°C for 5 min, in a reaction containing 0.4 µl [α -³⁵S] dATP (12.5 mCi/ ml), 3 µl of labelling mix A (2 µM dGTP, 2 µM dCTP, 2 µM dTTP), and 1 unit of T7 DNA polymerase. Labelling was

terminated by the addition of 2-4 μ l of each termination mix (150 μ M of each dNTP, 10 mM $MgCl_2$, 40 mM Tris-HCl, pH 7.5, 50 mM NaCl, 15 μ M ddNTP G, A, T, or C) and incubated at 42°C for 5 min. Finally, the reaction was stopped by adding 4 μ l of formamide loading dyes (100% de-ionized formamide, 0.1% Xylene cyanol FF, 0.1% Bromophenol blue, 1 mM EDTA). The labelled DNA fragments were fractionated on a 6% polyacrylamide/7 M urea gel following which the gel was dried for 1 hour at 80°C and exposed to either x-ray film (Fuji) or a phosphor screen (Molecular Dynamics) for from 16 hours to several days.

PCR mutagenesis

PCR mutagenesis was adapted from the Stratagene QuikChange™ site-directed mutagenesis instruction manual. Briefly, purified mutagenic primers (125 ng each) and plasmid pskLUTR (250 ng) were combined with 1X *Pfu* polymerase reaction buffer (200 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1% Triton® X-100, 1 mg/ml nuclease-free BSA), 200 μ M dNTPs and 1 μ l (2.5 units) *Pfu* DNA polymerase in a 50 μ l reaction. After heating at 95°C for 30s, the reaction was incubated for 30s at 95°C (denature), 2 min at (calculated T_m -2)°C (anneal) and 11 min at 68°C (extend) sequentially for 18 cycles. Annealing temperatures typically required optimization. The reaction was then cooled on ice for 2 min. 1 μ l of the restriction enzyme Dpn I (10 Units) was added and mixed thoroughly. The mixture was then incubated at 37°C for 1 hour to digest parental DNA. 4 μ l were then transformed into Epicurean Coli XL1-Blue Supercompetent cells as described.

2.5 Biochemical Techniques

Translating cytoplasmic extracts

The method for preparation of translating cytoplasmic extracts was adapted from Brown and Ehrenfeld (Brown and Ehrenfeld, 1979) and Jackson and Pelham (Pelham and Jackson, 1976). 5×10^8 GM 2132 or suspension HeLa cells ($\approx 4 \times 10^5$ / ml) were harvested by centrifugation at 2000 rpm for 10 min in a chilled GS-3 rotor (Sorvall), washed twice in ice-cold Earle's solution, resuspended in 2 packed cell volumes of lysis buffer (20 mM HEPES pH 7.4, 10 mM KCl, 1.5 mM Mg(OAc)₂, 1 mM DTT), and incubated on ice for 10 min. The cells were broken with 25 strokes of a Dounce homogenizer, and 0.1 volume of buffer A (0.2 M HEPES pH 7.4, 1.2 M K(OAc), 40 mM Mg(OAc)₂, 50 mM DTT) was immediately added. Nuclei were pelleted by centrifugation in an SS-34 rotor (Sorvall) for 5 min at 4500 rpm (2,500 g) at 4°C. Mitochondria and other membranes were removed by centrifugation for 15 min at 10000 rpm (12,000 g) at 4°C. Supernatant was transferred to a 50 ml conical tube, and .01 volume each of 0.1 M CaCl₂ and micrococcal nuclease (15,000U/ ml) were added. The mixture was incubated at 20°C for 15 min, .01 volume of 0.2 M EGTA was added, and the mixture was spun in an SS-34 rotor (Sorvall) for 15 min at 10000 rpm. 400 µl aliquots of supernatant were quick-frozen in liquid nitrogen and stored at -70°C.

Ribosomal salt wash

Cells were harvested and homogenized as in the previous method, except that buffer A was omitted. Nuclei and membranes were removed, and ribosomes were pelleted from the supernatant by centrifugation in a TLA 100.3 rotor (Beckman) for 1h at 83000 rpm (370,000 g) at 4°C. The sticky ribosome pellet was resuspended in hypotonic buffer (20 mM HEPES pH 7.4, 10 mM KCl, 1.5 mM Mg(OAc)₂, 1 mM DTT) to an A₂₆₀ value of 240U/ ml. The solution

was made 0.5 M with respect to KCl by addition of 1/8 volume of 4 M KCl, then stirred on ice for 15 min, then re-centrifuged in a Beckman TLA 100.3 rotor for 1h at 83000 rpm (370,000 g) at 4°C. The supernatant was dialysed for 2h at 4°C against 500 volumes of dialysis buffer (5 mM Tris-HCl pH 7.5, 100 mM KCl, 0.05 mM EDTA, 1 mM DTT, 5% glycerol), then quick-frozen in 50 µl aliquots and stored at -70°C.

Nuclear extract and hnRNP A1

HeLa nuclear extract was prepared according to the method of Dignam *et al.* (Dignam *et al.*, 1983) and was a kind gift from Dr. I. C. Eperon, as were wild-type and mutant forms of hnRNP A1.

Nuclear salt wash

The method for nuclear salt wash preparation was adapted from Pasternack *et al.* (Pasternack *et al.*, 1991) and Tata (Tata, 1972). A nuclear pellet was prepared as for translating extract preparation, without the addition of buffer A. The pellet was washed once in lysis buffer and once in 0.25 M sucrose in TKM buffer (5 mM Tris-HCl pH 7.5, 2.5 mM KCl, 5 mM MgCl₂), and resuspended into 2 volumes TKM/2.3 M sucrose, bringing the sucrose concentration to 1.62 M. The mixture was transferred to SW41 tubes (Beckman) which were filled 5/7 full. The mixture was underlaid with 2/5 tube volume TKM/2.3 M sucrose, then centrifuged in a SW41 Ti rotor (Beckman) at 4°C for 1h at 20000 rpm (60,000 g). The supernatant was carefully but completely removed, and the pellet of nuclei resuspended in 2 ml HKM buffer (10 mM HEPES pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT). The mixture was brought to 0.42 M NaCl and stirred for 30 min at 4°C, and the supernatant was recovered by centrifugation in a TLA 100.3 rotor (Beckman) for 1h at 22000 rpm (25,000 g) at 4°C. The supernatant was dialysed for 2h at 4°C against 500 volumes of

dialysis buffer (5 mM Tris-HCl pH 7.5, 100 mM KCl, 0.05 mM EDTA, 1 mM DTT, 5% glycerol), then quick-frozen in 50 μ l aliquots and stored at -70°C .

High-efficiency transcription extract

High-efficiency transcription extract was prepared by a protocol adapted from Shapiro *et al.* (Shapiro *et al.*, 1988). 1×10^9 GM 2132 cells were harvested, pelleted, washed in cold phosphate buffered saline and repelleted. The pellet was resuspended in 5 packed cell volumes of hypotonic buffer (10 mM HEPES pH 7.9, 0.75 mM spermidine, 0.15 mM spermine, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 10 mM KCl), allowed to swell for 10 min on ice and repelleted. Buffer was replaced with 2 original packed cell volumes of hypotonic buffer, and the cells were broken with three strokes of a Dounce homogenizer. 0.1 volume sucrose restore buffer (67.5% sucrose, 50 mM HEPES pH 7.9, 0.75 mM spermidine, 0.15 mM spermine, 10 mM KCl, 0.2 mM EDTA, 1 mM DTT) was quickly added and mixed with two strokes of the pestle. The homogenate was immediately spun in an SW41 Ti rotor (Beckman) at 2°C for 30s at 10000 rpm (16,000 g). The pellet was resuspended in 3 ml nuclear suspension buffer (20 mM HEPES pH 7.9, 0.75 mM spermidine, 0.15 mM spermine, 0.2 mM EDTA, 2 mM EGTA, 2 mM DTT, 25% glycerol), rocked for 30 min at 4°C , then sedimented by centrifugation in an SW41 Ti rotor (Beckman) at 2°C for 90 min at 30000 rpm (150,000 g). The supernatant was carefully removed, and solid ammonium sulphate was gradually added to 0.33 g/ ml. After 20 min rocking at 4°C , the precipitate was collected by centrifugation at 2°C for 20 min at 85,000 g. The pellet was redissolved in 1.0 ml of nuclear dialysis buffer and dialysed twice for 90 min each against 300 volumes of nuclear dialysis buffer (20 mM HEPES pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 2 mM DTT). Small aliquots were quick-frozen and stored in liquid nitrogen.

Preparation of cell lysates from transfected cells

After transfection, the medium was aspirated and the adherent cells were washed twice with phosphate buffered saline (PBS). Cells were lysed by the addition of 800 μ l of either 1X Reporter lysis buffer (Promega) or 1X Passive lysis buffer (Promega) and plates were scraped with a rubber policeman. The lysate was transferred to a tube and the insoluble matter was pelleted by centrifugation. The supernatant was removed to a fresh tube and either enzyme activity was assayed immediately or the lysate was stored at -70°C .

Luciferase assays

The activity of Firefly and *Renilla* luciferases *in vitro* translation reactions or lysates prepared from transfected cells was measured using a Dual-luciferase reporter assay system (Promega). Lysates were prepared using 1X passive lysis buffer as described. 5 μ l of lysate was added to 25 μ l of luciferase assay reagent and light emission was measured over 10 seconds using an OPTOCOMP I luminometer. Assays were performed according to the manufacturers' recommendations except that only 25 μ l of each reagent was used.

β -Galactosidase assays

The activity of β -galactosidase in lysates prepared from cells transfected with p β -Gal was measured using a Galactolight plus assay system (Tropix). 5 μ l of cell lysate was added to 100 μ l of Galactolight reagent (1:100 dilution in Galactolight buffer) and incubated at room temperature for 1 hour. 150 μ l of Accelerator was then added and the reaction was incubated at room temperature for 30s. Enzyme activity was then determined by measuring the light emission from the reaction in a luminometer, as described.

2.6 RNA Methods

In vitro run-off transcription

10 µg 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 10 µl of filter sterile de-ionized water. To synthesise uncapped transcripts, a reaction was set up containing 1X Transcription buffer (80 mM HEPES-KOH, pH 7.5, 6 mM MgCl₂, 2 mM spermidine, 4 mM DTT), 80 mM KOH, 40 units of recombinant RNasin ribonuclease inhibitor, 2 mM of each NTP, 1.5 µg of DNA template, and 40 units of T7, T3, or SP6 RNA polymerase in a final volume of 50 µl. After incubation at 37°C for 2 hours, the DNA template was digested with 10 units of RNase-free DNase I for 15 min at 37°C. Immediately following digestion, the RNA was phenol/chloroform extracted and unincorporated nucleotides were removed by passing the solution through a Sephadex G-50 column. The RNA was precipitated by the addition of 0.5 volume of 7.5 M NH₄Ac and 2.5 volumes of ethanol. After incubation at -70°C for 30min, the RNA was pelleted by centrifugation and washed with 75% ethanol. The pellet was resuspended in 30 µl of filter-sterilised 0.1X TE and the concentration was determined using the absorbance at 260nm. In addition, 0.5 µl of the RNA was subjected to agarose gel electrophoresis to ensure the product was not degraded.

Capped transcripts were synthesised in a reaction containing 1X transcription buffer, 7 mM KOH, 40 units of RNasin, 1 mM ATP, 1 mM UTP, 1 mM CTP, 0.5 mM GTP, 2 mM m⁷G(5')ppp(5')G, 1 µg of DNA template and 20 units of appropriate RNA polymerase in a final volume of 50 µl. After incubation of the reaction for 1 hour at 37°C, the RNA was isolated as described above.

Rabbit reticulocyte translation

Rabbit reticulocyte translation reactions were performed using a standard reticulocyte lysate system (Promega) with minor modifications. Briefly, each reaction contained 8.25 μ l of reticulocyte lysate, 0.25 μ l of RNasin (40 units/ μ l), 1 μ l of 1 mM amino acid mixture and RNA substrate (0.125-20 ng/ μ l) in a final volume of 12.5 μ l.

Coupled transcription/translation

Coupled transcription/translation reactions were assembled in a total volume of 25 μ l, using 50% TNTTM (Promega) coupled lysate, 1X TNTTM reaction buffer, 0.5 μ l 1 mM complete amino acids mix, 0.5 μ l RNasin, 0.5 μ l T3 RNA polymerase and 0.5 μ l linearized plasmid pSKM Δ 1 or pRhpFM DNA, and incubated for 90 min at 30°C.

Cytoplasmic extract translation

Cytoplasmic extract translation reactions were typically performed in a total volume of 20 μ l, containing 10 μ l lysate, 50-100 ng RNA, 4 μ l 5X S10 translation buffer (175 mM HEPES pH7.4, 5 mM ATP, 1 mM GTP, 5 mM DTT, 125 mM creatine phosphate, 2 mg/ ml creatine phosphokinase), 0.25 μ l 1 mM complete amino acid mix, and 0.25 μ l RNase inhibitor. Reactions were incubated at 30°C for 1-3 hours, halted by rapid chilling and either assayed immediately or stored at -70°C.

Chemical structure probing

The chemical probing protocol is adapted from Stern *et al* (Stern *et al.*,). 5 μ g RNA were combined with 5 μ l 10X standard structure probing buffer SSPB (100 mM Tris-HCl pH 7.0, 1 M KCl) or 25 μ l 2X BMK (100 mM potassium pentaborate, 200 mM KCl, pH 8.0) in the case of CMCT treatments. 5 μ l of 100 mM MgAc or 5 mM EDTA were added, and the

mixture was brought to 50 μ l with water. The mixture was heated to 80°C for 3 min and cooled to 4°C over 1h in a PCR machine and then chilled at 0°C for ten min to permit structural equilibration. All chemical treatments were carried out at 0°C for 1 hour. Differing quantities of the various agents were used, typically 1-10 μ l (50 μ l of CMCT stock). Mock-treated samples were prepared in parallel, being treated identically but with the omission of chemical modifying agent.

DMS was diluted 1:12 in ethanol. CMCT was at a concentration of 42 mg/ml in 1X BMK. Kethoxal was diluted with water 1:20.

Chemical treatments were halted by ethanol precipitation after the addition of 50 μ g of carrier tRNA. 100 μ l of 85 mM potassium pentaborate was added to kethoxal-treated samples to stabilize the adduct. The RNA pellet was washed as above, resuspended into 35 μ l sterile filtered water (25 mM potassium pentaborate, pH 7.0 for kethoxal treated RNA) and stored at -70°C.

Primer extension

The procedure for primer extension was adapted from Stern *et al* (Stern *et al.*,). 1 μ l of primer (2 pmol/ μ l) was combined with 1 μ l hybridization buffer (250 mM K-HEPES pH 7.0, 500 mM KCl) and 2.5 μ l RNA (i.e. in molar excess relative to the primer). The mixture was incubated at 85°C for 1 min and allowed to cool at room temperature for 10-15 min. 3 μ l of extension mix was added to the cooled hybrid, consisting of 1 μ l AMV reverse transcriptase (2 units diluted with 50 mM Tris-HCl pH 8.4, 2 mM DTT, 50% glycerol), 0.33 μ l dNTP stock (110 μ M each dGTP, dCTP, dTTP, 6 μ M dATP), 0.66 μ l extension buffer (1.3 M Tris-HCl pH 8.4, 100 mM MgCl₂, 100 mM DTT), 0.5 μ l α ³⁵S-dATP or α ³²P-dATP and 0.5 μ l filtered, sterile water. The reaction mixture was incubated at 37°C for 30 min, at which time 1 μ l of chase mix (1 mM each dGTP, dCTP, dTTP, dATP) was added and incubation continued for a

further 15 min. The reaction was stopped by the addition of 3 μ l 3 M NaAc pH 5.4 and 90 μ l ethanol. The mixture was vortexed, incubated at 0°C for 1 hour and spun in a microcentrifuge at high speed (14,000 g) for 15 min. The supernatant was carefully drawn off, the pellet dried and resuspended into 10 μ l of gel loading buffer (7 M urea, 0.03% xylene cyanol and bromophenol blue dyes).

The products of the reaction were then heated to 100°C for 2.5 min, chilled briefly on ice and 2-5 μ l were quickly loaded onto a 7 M urea 6% polyacrylamide sequencing gel.

2.7 Theoretical Methods

Secondary structure prediction

Secondary structure predictions and dot-plots were generated using the web implementation of the Mfold algorithm (Zuker *et al.*, 1999), incorporating version 3.0 of the Turner rules (Mathews *et al.*, 1999).

Stability calculations

Domain stabilities were calculated using the web implementation of the *efn* algorithm which calculates the free energy of a given sequence and secondary structure. The algorithm does not accommodate pseudoknots, so their stabilities were calculated manually using version 3.0 of the Turner rules (Mathews *et al.*, 1999).

Sequence alignment

Sequence alignments were generated from *c-myc* 5' UTR sequences from human, gibbon, marmoset, woodchuck, pig, cat, rat, mouse and sheep sequences, using the UNIX implementation of the CLUSTALW algorithm, a part of the university of Wisconsin GCG package. Parameters were optimized by trial and error, and were Pairwise 1: 5.00, 2: 1.00 and Multiple 1: 4.00, 2: 1.00, 3: 75%. The alignments were refined by eye.

Covariation analysis

Covariant positions were identified using the Macintosh Hypercard software *Covariation* (Brown, 1991), using a CLUSTALW-derived sequence alignment as input.

Structure searching

The "Y-shaped motif" secondary structure (Le and Maizel, 1997) was codified as input for the C implementation of the RNABOB algorithm, and sought in the available mammalian *c-myc* sequences.

Chapter 3

Structure Determination

3.1 Introduction

All IRESs characterized thus far are, regardless of their origin, dependent for their function upon the attainment of the correct three-dimensional conformation of the mRNA that comprises them. An examination of the *c-myc* IRES reveals that it is long and GC-rich; these features make the presence of extensive structure inevitable. This structure may contribute to the function of the IRES in a number of ways. In order, therefore, to fully understand the mechanism of function of a given IRES it is necessary to determine this structure.

A secondary structure model is an invaluable step toward the ultimate solution of the full tertiary structure of the RNA element. It provides a point from which to design future crystallographic or NMR studies, and its most obvious limitation, that it is a two-dimensional interpretation of a three-dimensional object, is irrelevant to considerations of scanning and translation as one-dimensional processes.

Numerous studies have been made upon the secondary structures of viral IRESs, utilizing both theoretical and practical approaches, and models of cellular IRESs including *c-myc* (Nanbru *et al.*, 1997; Stoneley *et al.*, 1998), PDGF2, and VEGF (Huez *et al.*, 1998) based solely upon phylogenetic and computer prediction have been published, but as yet no experimentally substantiated model of a cellular IRES has been derived. The first proposed secondary structure of the *c-myc* 5' UTR (Nanbru *et al.*, 1997) was derived solely by computer modelling of the human *c-myc* 5' UTR sequence, and cannot be relied upon in the absence of other evidence.

It was clear, therefore, that an investigation into the secondary structure of the *c-myc* IRES was of merit.

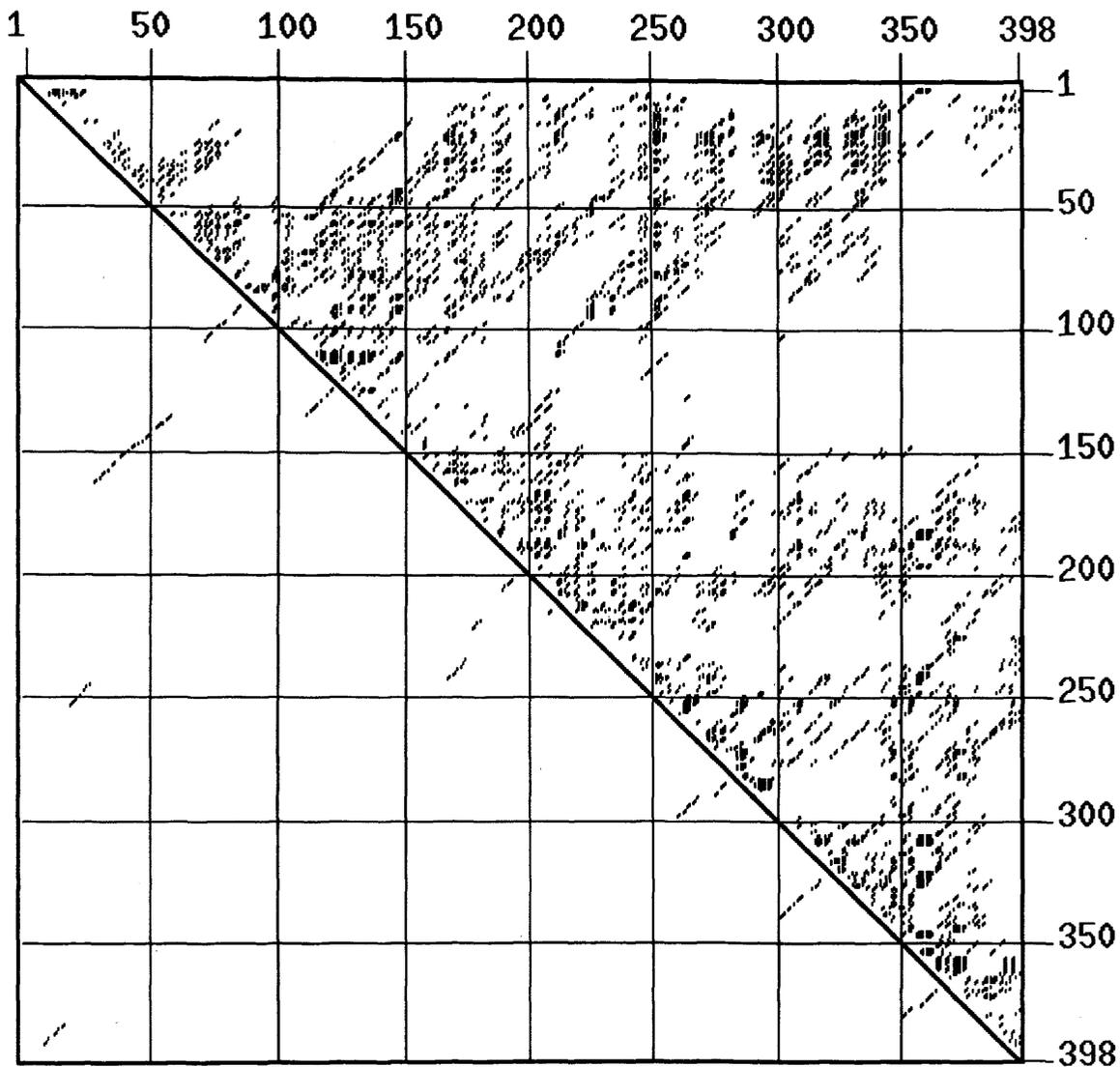
3.2 Theoretical Approaches

It is possible to model the secondary structure of a given RNA molecule solely on the basis of its sequence, by employing one of number of computer algorithms, of which “Mfold” (Mathews *et al.*, 1999; Zuker *et al.*, 1999) is the best known. This approach can yield good results when the molecule to be modelled is very short or very simple in structure. However, for large complex structures, a class that includes most if not all IRESs, the results obtained are generally less satisfactory. As a first step toward modelling the *c-myc* 5' UTR, the RNA sequence was fed to the Mfold algorithm. These data are represented as an energy dotplot (Figure 3.i). The plot shows that a great number of alternative structures are suggested that all lie within 5% of the maximum theoretical stability. Thus this approach in isolation is incapable of supplying a sufficiently definitive model.

It was thought that a combination of energy minimization and phylogenetic analyses might provide a more robust model. An examination of the genbank database yields a large number of *c-myc* 5' UTR sequences from a range of species, including mammals, birds, fish and amphibia. However, only those sequences from the mammalian species available proved sufficiently similar for the construction of a sequence alignment.

The sequence alignment (Figure 3.ii) reveals a high degree of conservation between species, consistent with conservation of structure. One surprising feature is the presence of a CUA in place of a CUG codon in the sheep sequence, at the otherwise perfectly conserved in-frame alternative translation start codon. This may well represent a sequencing error.

This sequence alignment was used to aid structure prediction in two ways. Firstly, a pool of possible structure models was generated from each sequence. Models from different species were inspected to determine whether particular motifs were held in common. Helical



Optimal energy: -168.0

Basepairs Plotted: 5880

Figure 3.i. Energy dotplot of Mfold-generated suboptimal foldings of the human *c-myc* 5' UTR sequence. The x- and y-axes both represent the sequence of the authentic P2 *c-myc* mRNA 5'UTR. Each point above the diagonal represents a base pair in a predicted secondary structure of a stability within 5% of the calculated most stable structure. Points beneath the line represent base pairs in only the most stable predicted structure (-168.0 kcal/mole).

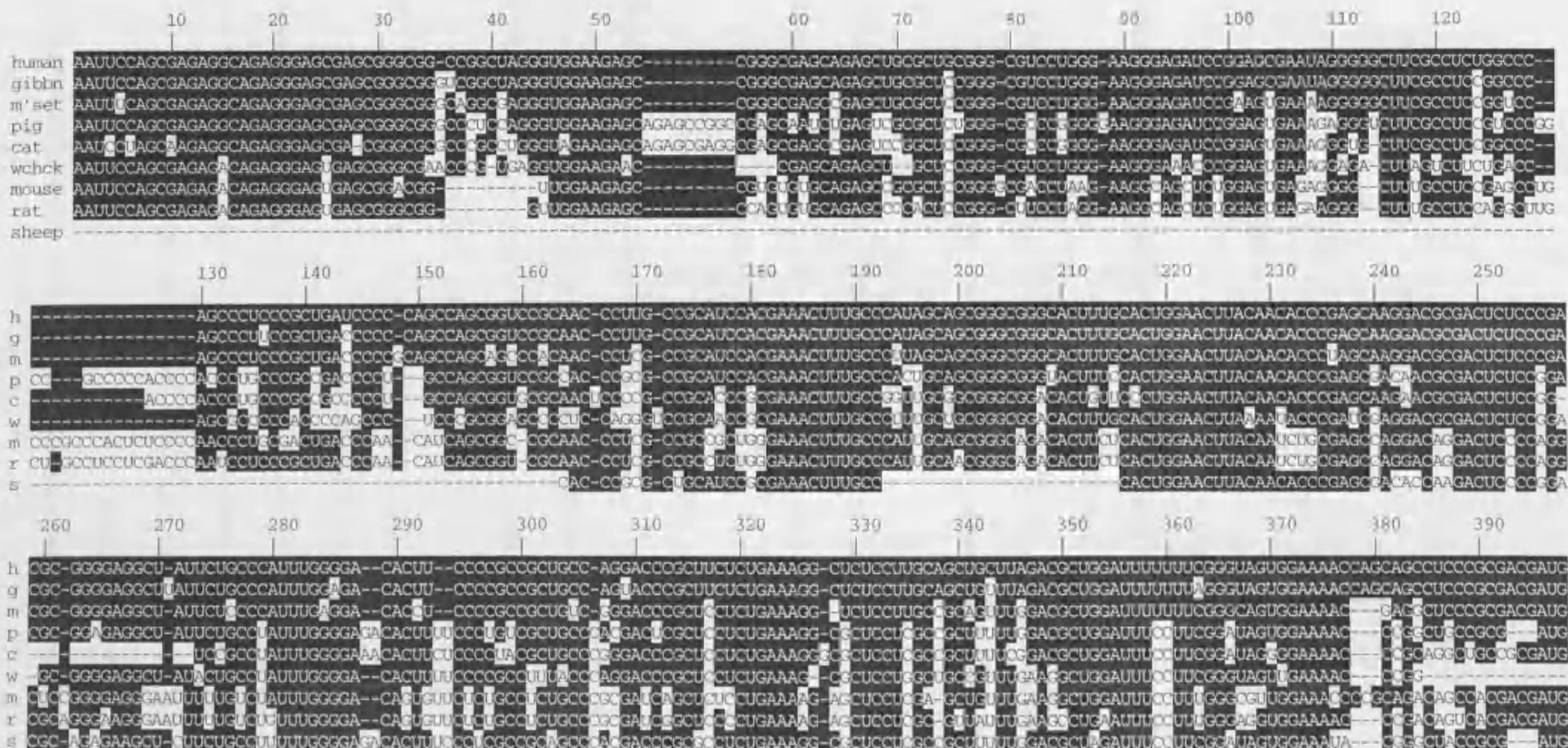


Figure 3.ii. Alignment of all available mammalian *c-myc* 5' UTR sequences showing conservation. Sequences are derived from human, gibbon, marmoset, pig, cat, woodchuck, mouse, rat and sheep tissues. Numbering is from the 5' end of the human P2 transcript. Shaded positions mark bases identical to the human sequence. The sequences derived from sheep and woodchuck are incomplete.

segments that seemed promising were then tested for their ability to form across species by reference to the sequence alignment.

Secondly, the sequence alignment was subjected to analysis by the program “Covariation” (Brown, 1991), which seeks to identify covariant positions within the alignment at which base pairs might be expected to form in all species. Such positions were scrutinized to see if they lay within conserved helical segments.

These approaches enabled a few motifs to be tentatively identified (Figure 3.iii); however, when taken together these motifs do not form a satisfactory completed secondary structure model. Two of the motifs, 1a and 1b, overlap and so cannot co-exist. In the absence of other data one cannot with any great certainty state that either is more likely to form than the other. Furthermore, a large proportion of the sequence cannot be assigned into any conserved structure at all. This suggests that either the sequence alignment is not sufficiently accurate, due to sequencing errors or misalignment, or that much of the secondary structure is in fact not conserved, or that the methods used to identify conserved motifs are overly stringent, excluding functionally equivalent tertiary structures that diverge at the secondary structure level.

This method has been most successfully applied in situations where large numbers of different sequences sharing a very similar structure are easily available, as when sequences have been gathered from viral strains.

In order to obtain direct evidence regarding the structure of the *c-myc* 5'UTR, structure probing experiments were performed.

3.3 Chemical Probing

The plasmid pSK+Myc 5'UTR contains the 5'UTR of the P2 *c-myc* message downstream of a T3 viral RNA polymerase promoter. In order to reduce the likelihood of artifactual secondary contacts between the UTR sequence and flanking vector sequences a deletion was made between the T3 promoter and the AATTCC sequence marking the authentic P2 initiation locus (Figure 3.iv). The presence of flanking sequence beyond the 3' end of the UTR is in one sense advantageous as it permits the extreme 3' end of the UTR sequence to be analyzed by primer extension. Furthermore, *c-myc* IRES function is retained in transient transfection experiments when the UTR is placed in frame upstream of a number of reporter genes, including firefly and sea pansy (*Renilla sp.*) forms of luciferase, CAT, and the authentic *c-myc* coding region (Stoneley, 1998). This suggests that the *c-myc* IRES function (and therefore structure) is unaffected by the nature of the 3' flanking coding sequence, unlike, for example the HCV IRES (Honda *et al.*, 1996).

The resulting plasmid, pSK+ Δ P2, was cleaved with the restriction enzyme *Acc65 I* and used as a substrate for run-off transcription of *c-myc* 5' UTR RNA.

Four complementary DNA oligonucleotides were designed for primer extension analysis of RNA modification (Figure 3.v). All hybridize at regions of high uniqueness, have a strong G/C clamp at the 3' end and are predicted not to self-hybridize into primer-dimers or hairpin structures.

Purified RNA was resuspended into probing buffer containing either 10 mM MgAc or, in its place, 0.5 mM EDTA. The mixture was then rapidly heated and gradually cooled to refold the molecule into its native conformation. This step will also encourage binding or dissociation of magnesium ions.

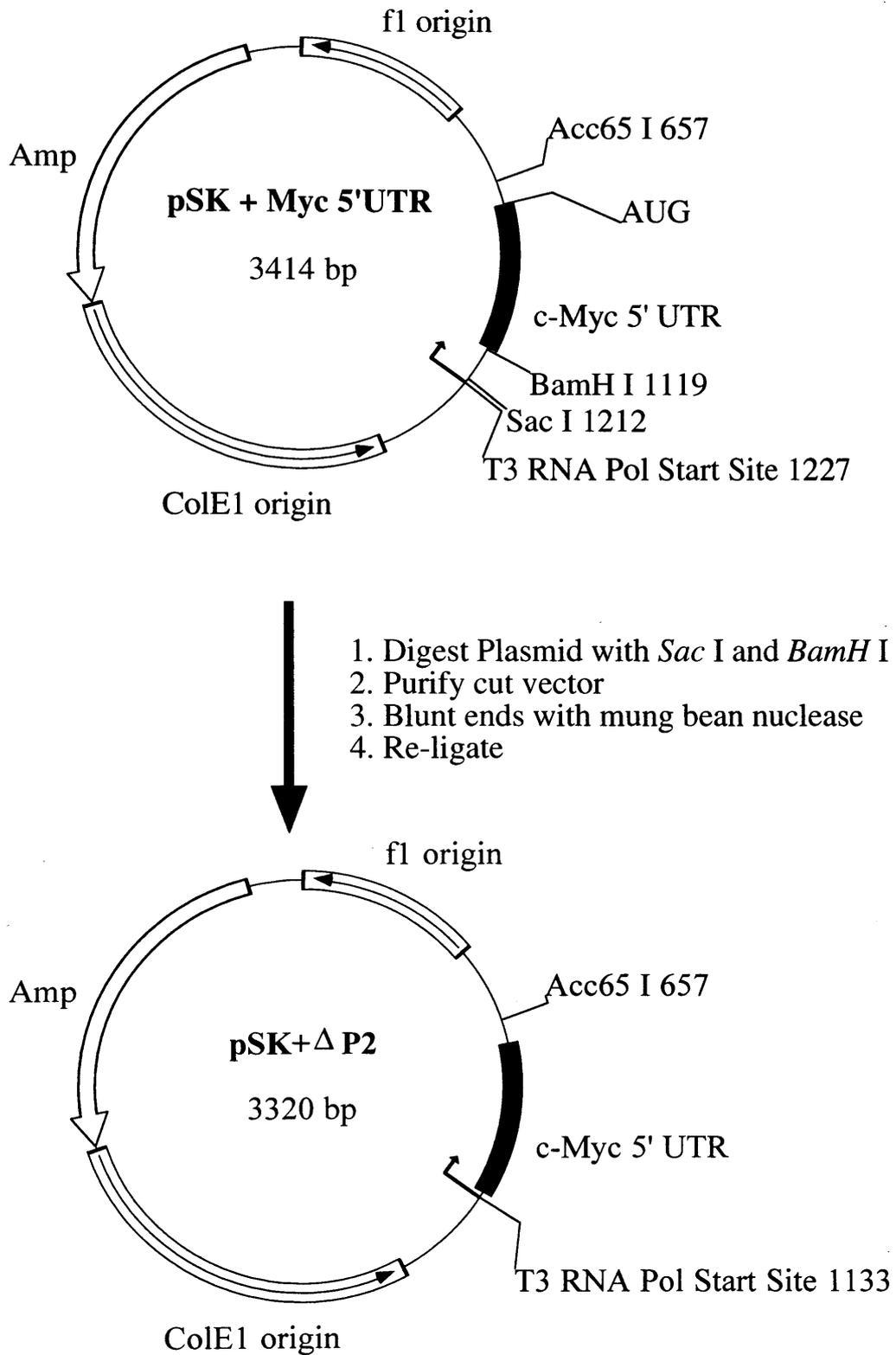


Figure 3.iv. Plasmid pSK+ΔP2 cloning strategy.

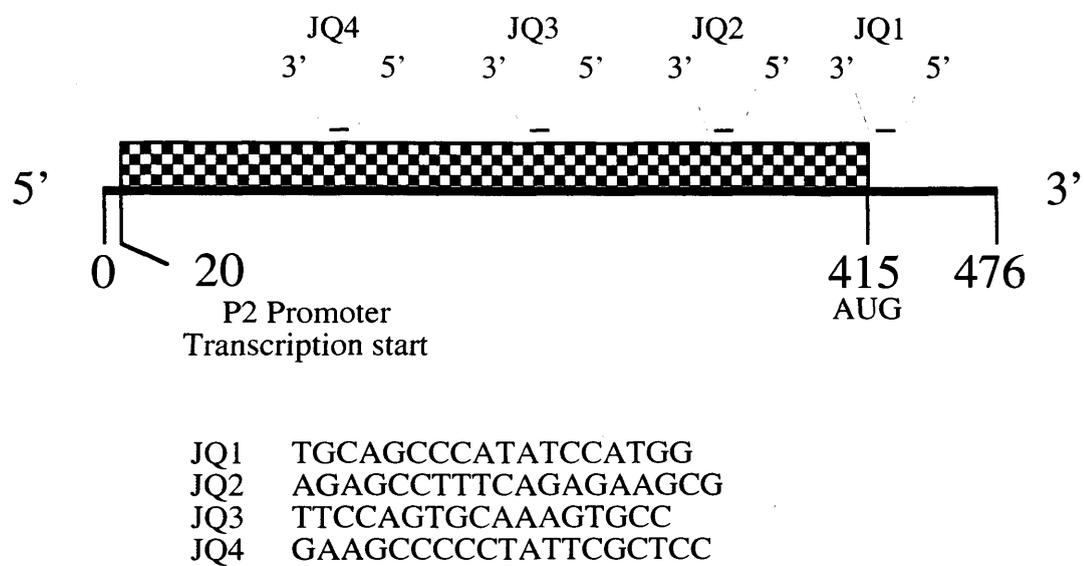


Figure 3.v. Schematic representation of pSK+ Δ P2 RNA transcript used in chemical probing studies showing sites of primer hybridization. Numbering is from the 5' end of the *in vitro* transcribed RNA.

The RNA was then incubated with one of three chemical probing agents, which between them enable the Watson-Crick hydrogen-bonding positions of all four bases to be probed. Each agent specifically adduces itself to one or more bases (Table 3.a). Experiments were performed in duplicate concomitantly with a control treatment, comprising a mock treatment of the RNA renatured in the same buffer and incubated in the absence of chemical probe.

Purified RNAs were then hybridized with one of the oligonucleotides JQ1-4 and the primers were extended in the presence of a radioactively labelled dNTP. The original template vector, pSK+ Δ P2, was sequenced using the same oligonucleotides and label. The products of control and treated RNA-programmed primer extension reactions were then subjected to PAGE alongside the corresponding DNA sequencing ladder to allow identification of modified residues. Since chemically modified residues are unable to base-pair, the progress of the reverse transcriptase is halted one nucleotide before that point. Thus a band that is dark relative to the control lane indicates a chemical modification at the base immediately above the equivalent band on the sequencing ladder (Figures 3.vi-3.ix).

It is notable that background bands are more obvious in autoradiographs of primer extension experiments of DMS-treated RNA. This is unlikely to be a consequence of the reagent used, as it is not observed in repeat experiments (data not shown). The perceived effect might at least partly be due to the fact that of the experiments shown only the DMS experiments were visualised using a shark's-tooth comb, while the others used square-toothed combs, meaning that background bands appear all across the gel, causing an illusion of greater density.

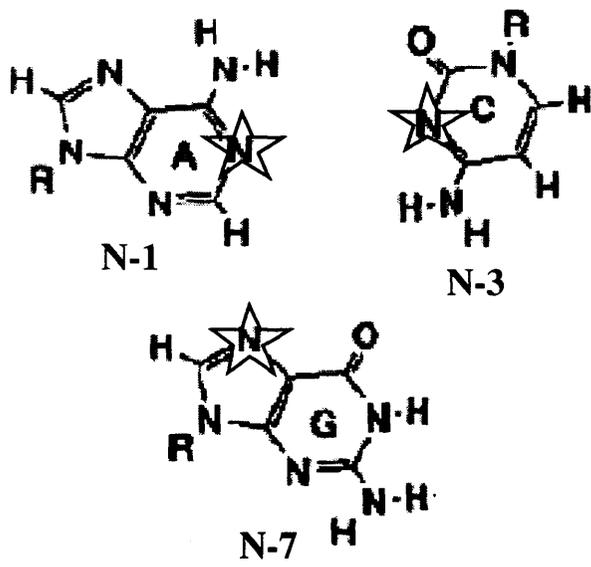
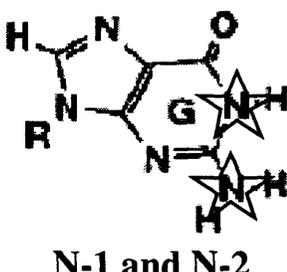
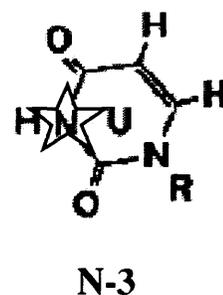
Probe	Site(s) of Modification
Dimethyl Sulphate	 <p>The diagram shows three nucleobases: Adenine (A), Guanine (G), and Uracil (U). In each structure, a star symbol indicates the site of chemical modification. For Adenine, the star is at N-1. For Guanine, the star is at N-7. For Uracil, the star is at N-3. Each structure also shows an 'R' group representing a methyl group from dimethyl sulphate.</p>
Kethoxal	 <p>The diagram shows a Guanine (G) nucleobase with two stars indicating modification sites at N-1 and N-2. An 'R' group is attached to the N-1 position.</p>
CMCT	 <p>The diagram shows a Uracil (U) nucleobase with a star indicating a modification site at N-3. An 'R' group is attached to the N-3 position.</p>

Table 3.a. Table showing sites of action of chemical probes used in this study. Sites of chemical adduction are starred. The methylation of guanine at N-7 is not prevented by Watson-Crick base pairing, does not impede the progress of reverse transcriptase and is therefore undetectable by direct primer extension.

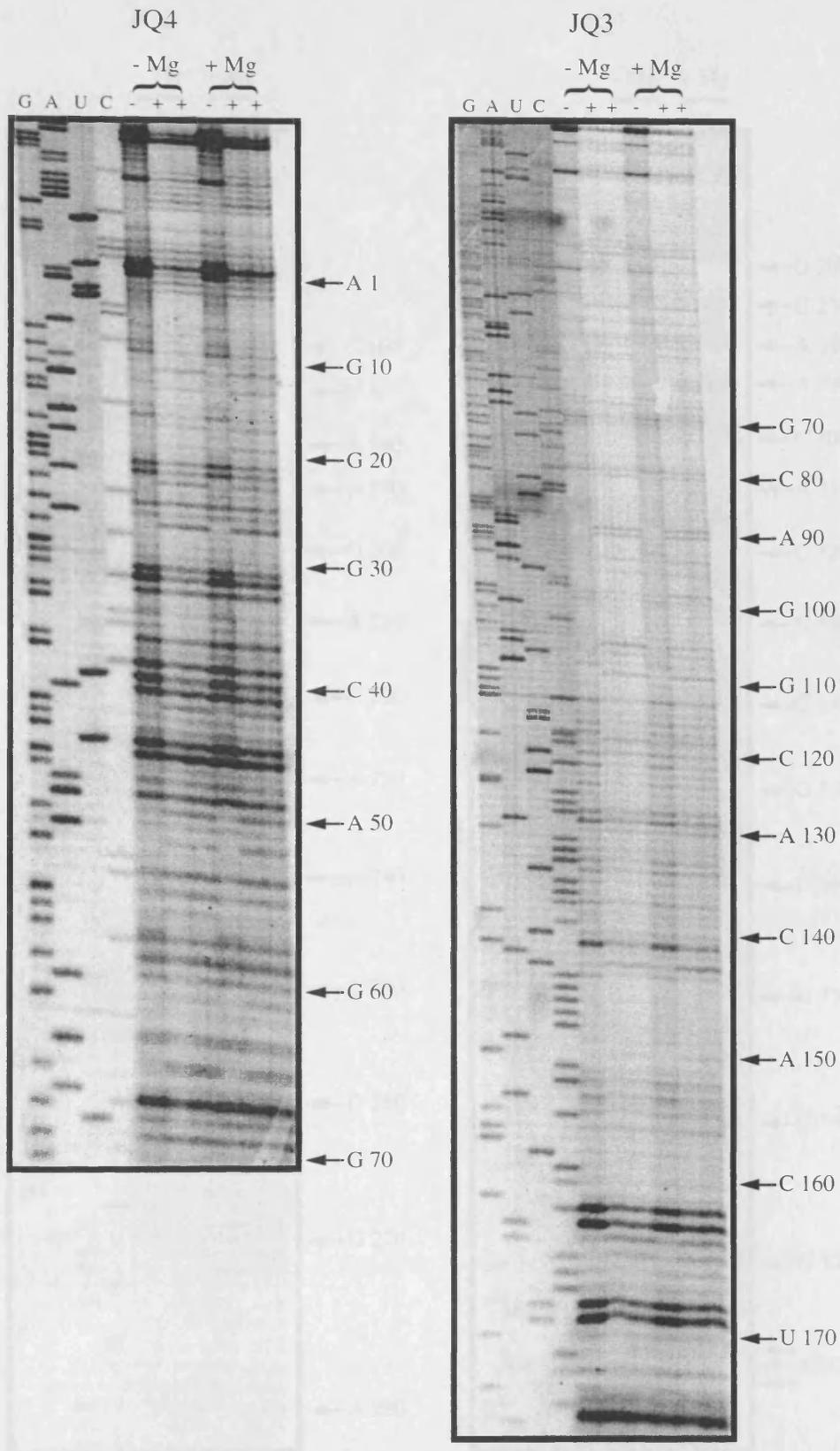


Figure 3.vi. Representative PAGE autoradiographs of ^{35}S -labelled primer extension analysis of *c-myc* 5'UTR RNA treated with DMS *in vitro*. RNA templates were treated in the presence of either 0.5mM EDTA or 10mM Mg^{2+} . Labels mark positions one base below the corresponding base in the sequencing ladder.

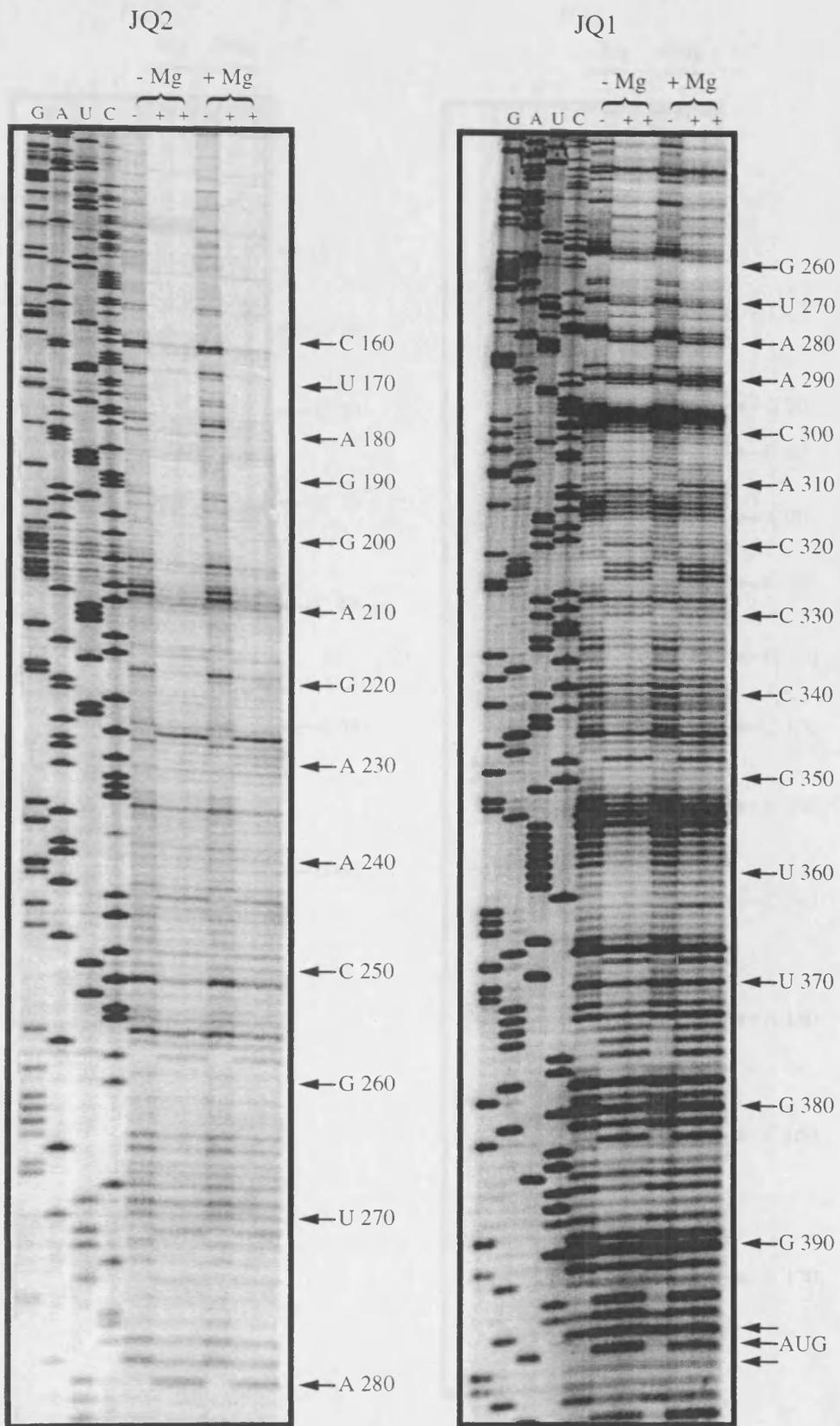


Figure 3.vi. (continued).

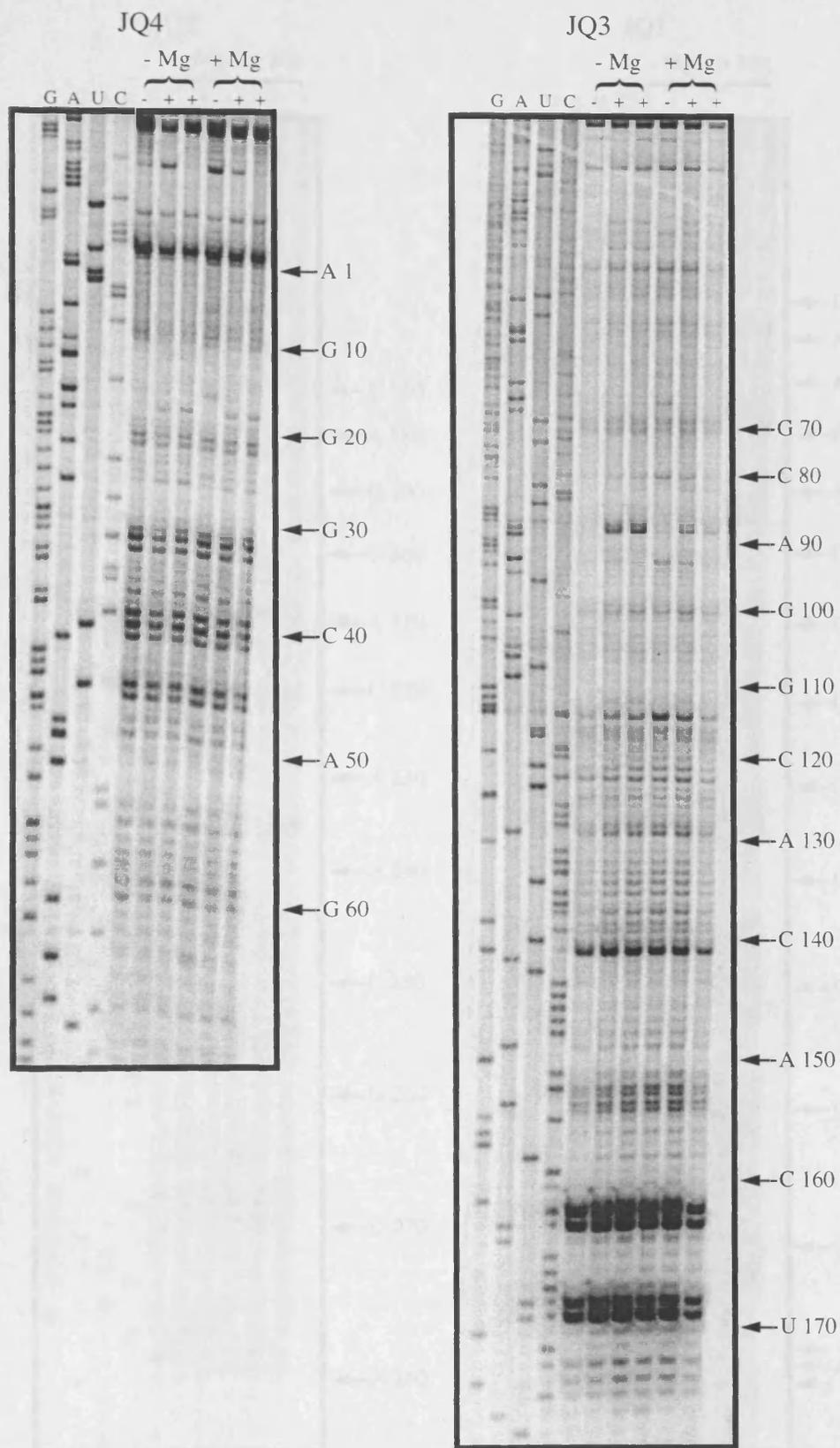


Figure 3.vii. Representative PAGE autoradiographs of ^{35}S -labelled primer extension analysis of *c-myc* 5'UTR RNA treated with kethoxal *in vitro*. RNA templates were treated in the presence of either 0.5mM EDTA or 10mM Mg^{2+} . Labels mark positions one base below the corresponding base in the sequencing ladder.

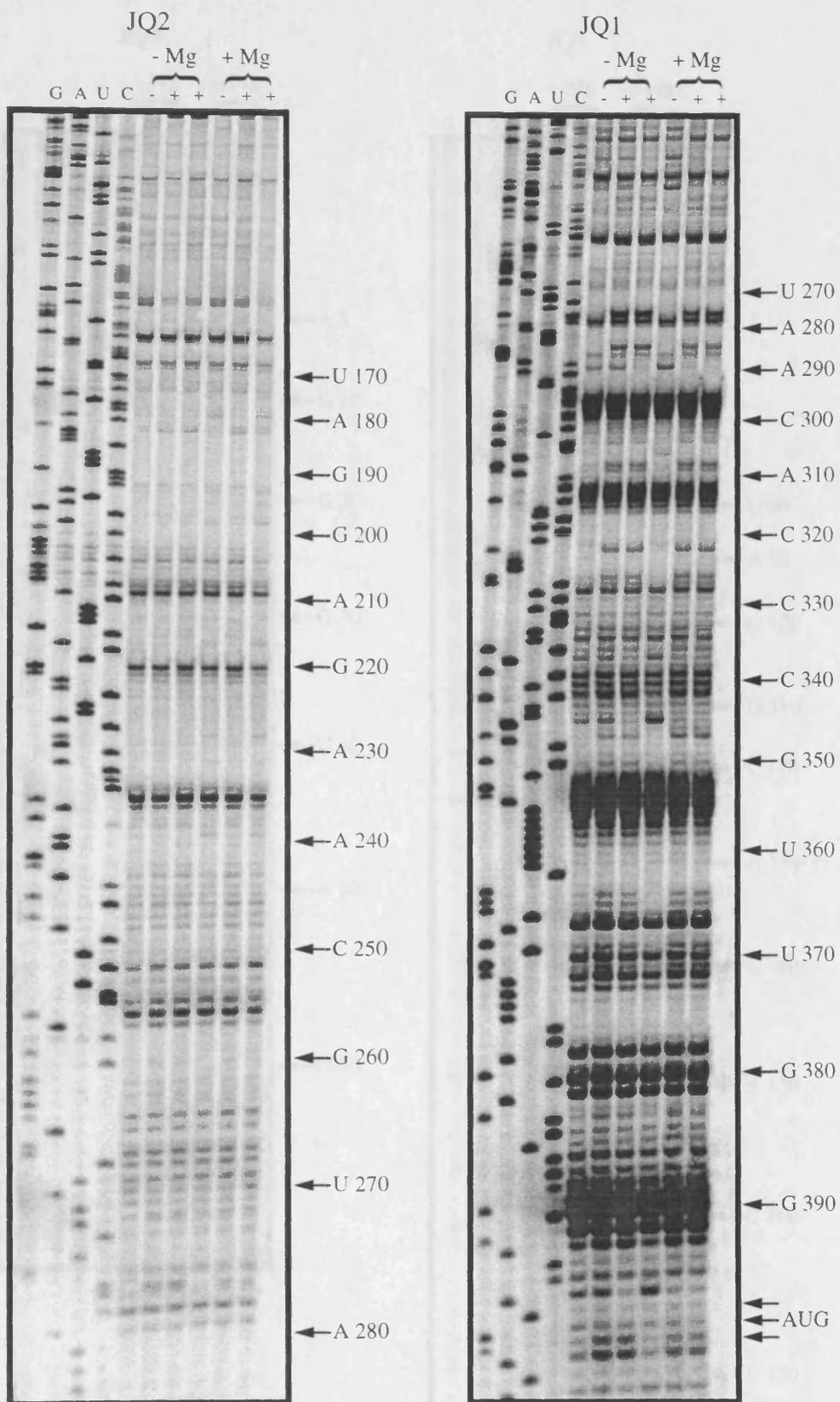


Figure 3.vii. (continued)

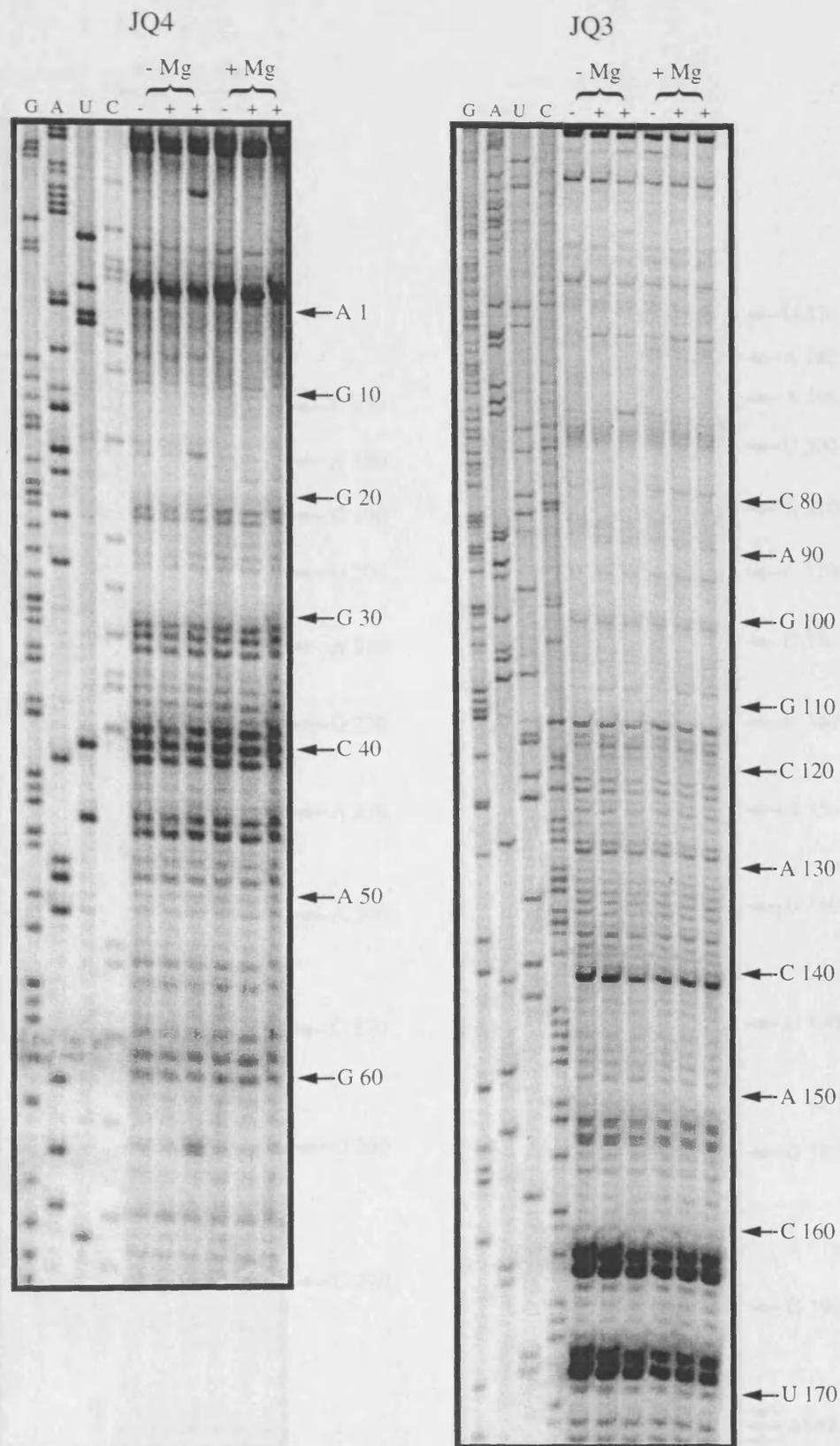


Figure 3.viii. Representative PAGE autoradiographs of ^{35}S -labelled primer extension analysis of *c-myc* 5'UTR RNA treated with CMCT *in vitro*. RNA templates were treated in the presence of either 0.5mM EDTA or 10mM Mg^{2+} . Labels mark positions one base below the corresponding base in the sequencing ladder.

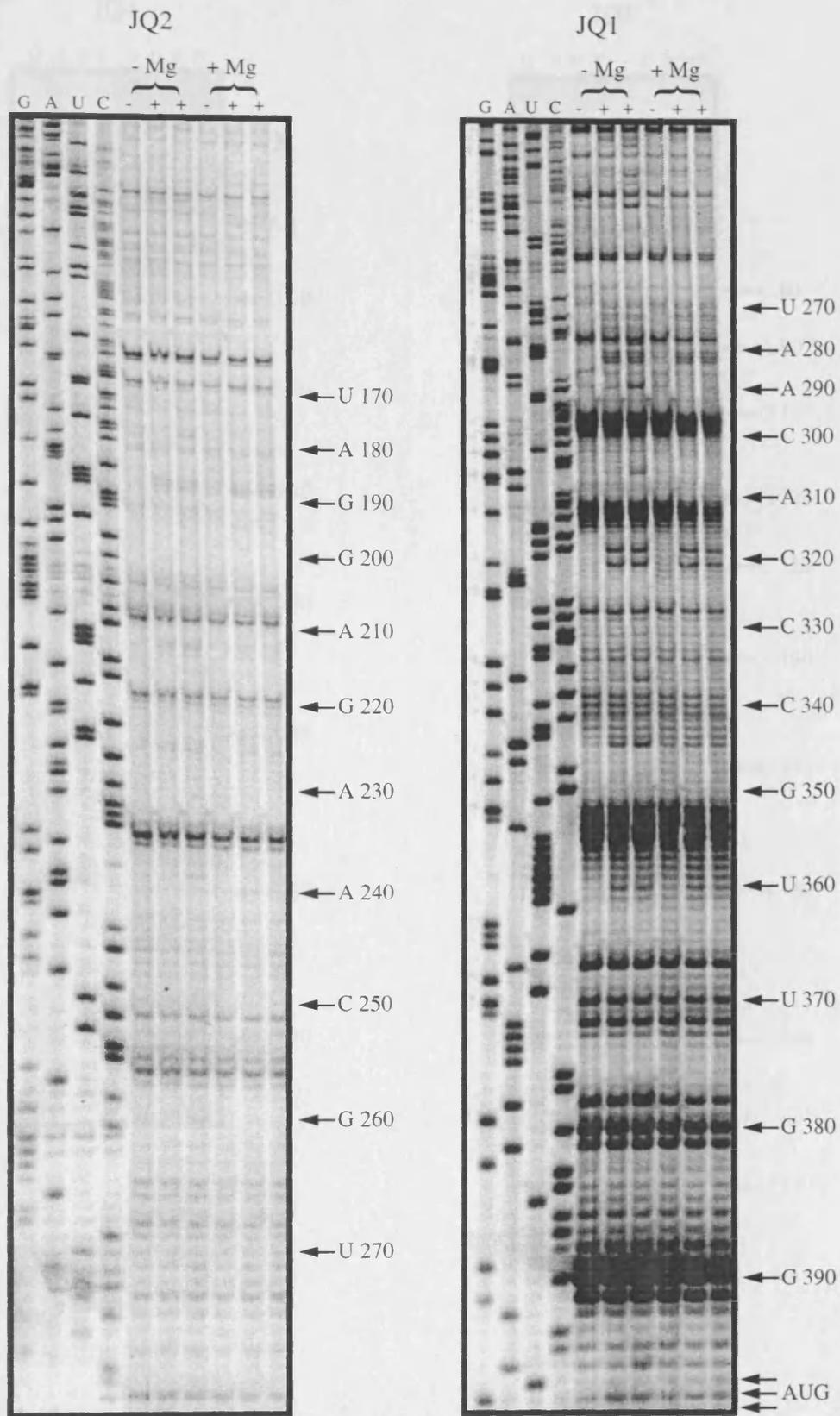


Figure 3.viii. (continued)

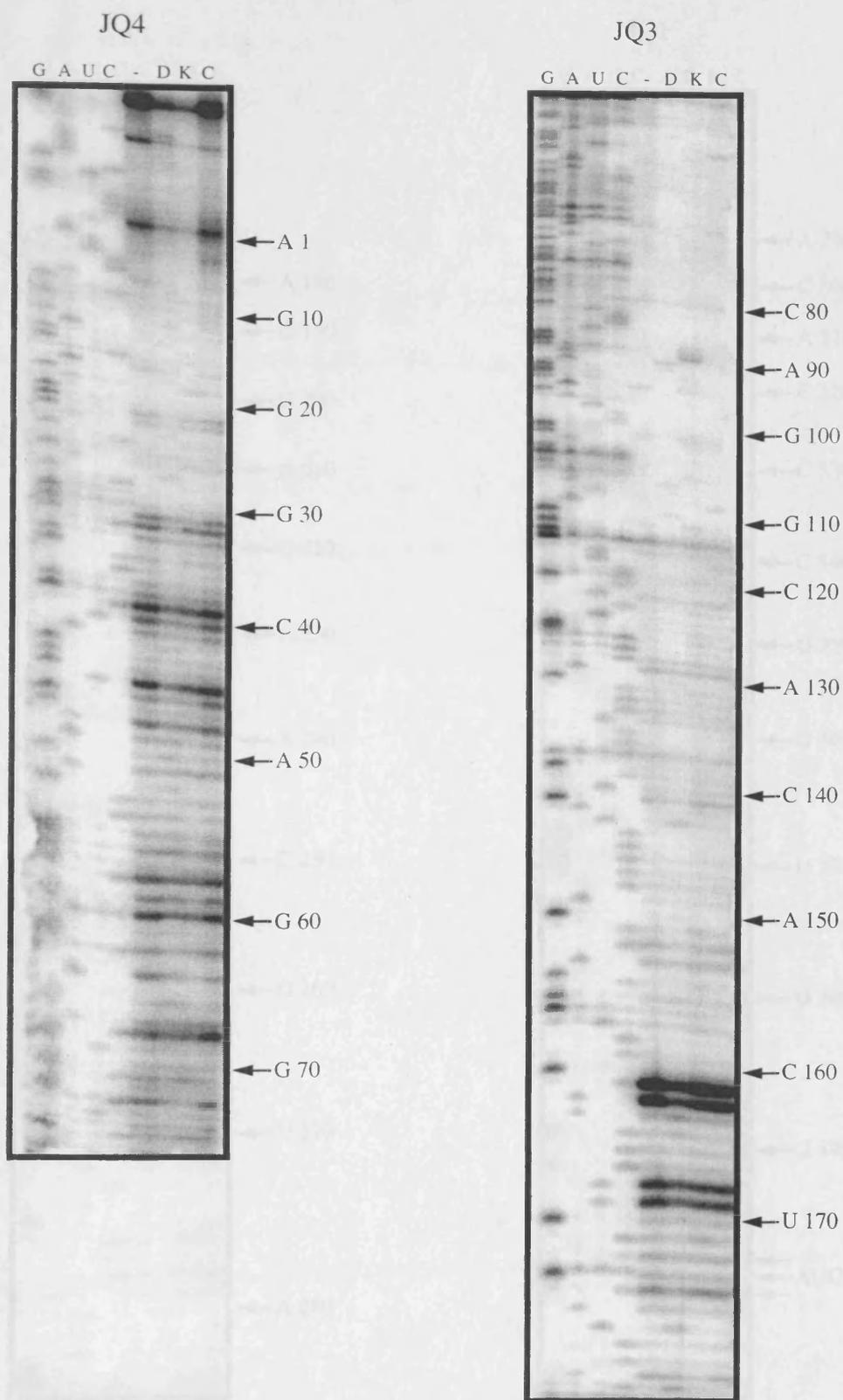


Figure 3.ix. Representative PAGE autoradiographs of ^{32}P -labelled primer extension analysis of *c-myc* 5'UTR RNA treated with DMS, kethoxal and CMCT *in vitro*. RNA templates were treated in the presence of Mg^{2+} . Labels mark positions one base below the corresponding base in the sequencing ladder.

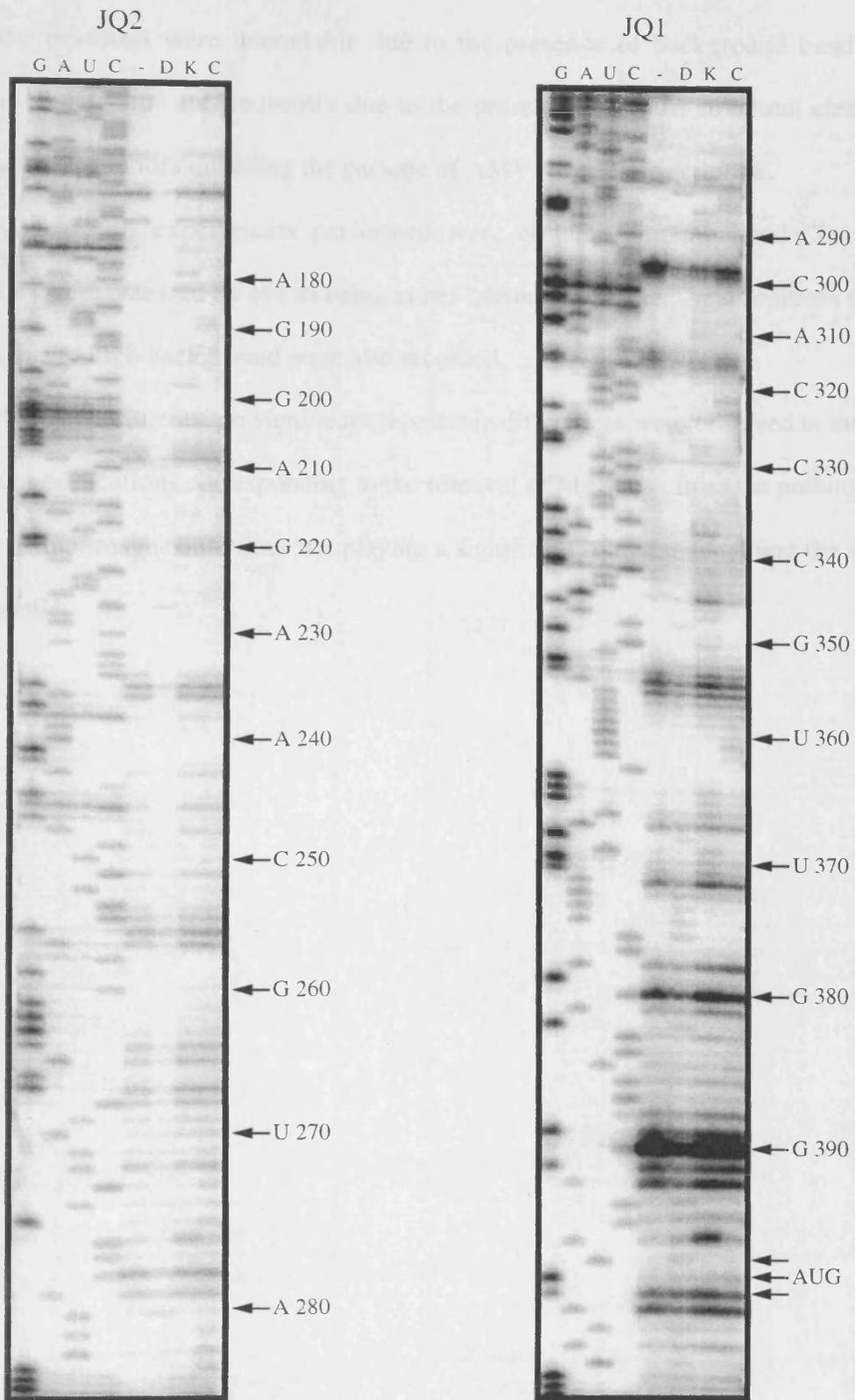


Figure 3.ix. (continued)

Many positions were unreadable due to the presence of background bands in the control lane. Such stops are frequently due to the presence of robust structural elements or particular sequence motifs impeding the passage of AMV reverse transcriptase.

Data from all experiments performed were collated and tabulated (Table 3.b). Modifications were assessed by eye as being either “strong” or “weak”, and positions rendered unreadable due to high background were also recorded.

In these experiments, no significant repeatable differences were observed in the profile of chemical modifications corresponding to the removal of Mg^{2+} ions from the probing buffer. This suggests that magnesium does not playing a significant role in maintaining the structure of this molecule.

Position	Base	Background	Modification
242	G		++
243	G		++
244	A		++
249	A		+
251	T	+	
255	C	+	
258	A		++
260	G		+
262	G		+
263	G		+
264	G	+	
265	G	+	
268	G	+	
269	C	+	
270	T		+
271	A		+
272	T		++
273	T		++
274	C		+
275	T		++
276	G		++
278	C	+	
279	C	+	
280	A		++
281	T		++
282	T		++
283	T		++
287	G		++
288	A		++
289	C	+	
290	A		++
294	C	+	
295	C	+	
296	C	+	
297	C	+	
298	G	+	
299	C	+	
304	G		++
307	A		+
308	G		++
309	G		++
311	C		+
312	C	+	
313	C	+	
314	G	+	
315	C	+	
319	T		++
320	C		++
321	T		++
322	G		++

Position	Base	Background	Modification
323	A		++
324	A		++
325	A		++
326	G		++
327	G		++
330	C		+
331	T	+	
336	G		+
339	G	+	
345	T	+	
346	A		+
347	G		++
348	A		++
350	G		+
352	T	+	
353	G	+	
354	G	+	
355	A	+	
356	T	+	
357	T	+	
358	T	+	
359	T		+
360	T		+
361	T		+
364	G		++
365	G		++
366	G		+
368	A		++
369	G		++
371	G		+
372	G		+
374	A		++
375	A		++
376	A		++
378	C	+	
379	A	+	
380	G	+	
381	C	+	
382	A		++
383	G		++
388	C	+	
389	C	+	
390	G	+	
391	C	+	
393	A		++
395	G	+	
396	A		++

Table 3.b. (Continued).

Position	Base	Background	Modification
242	G		++
243	G		++
244	A		++
249	A		+
251	T	+	
255	C	+	
258	A		++
260	G		+
262	G		+
263	G		+
264	G	+	
265	G	+	
268	G	+	
269	C	+	
270	T		+
271	A		+
272	T		++
273	T		++
274	C		+
275	T		++
276	G		++
278	C	+	
279	C	+	
280	A		++
281	T		++
282	T		++
283	T		++
287	G		++
288	A		++
289	C	+	
290	A		++
294	C	+	
295	C	+	
296	C	+	
297	C	+	
298	G	+	
299	C	+	
304	G		++
307	A		+
308	G		++
309	G		++
311	C		+
312	C	+	
313	C	+	
314	G	+	
315	C	+	
319	T		++
320	C		++
321	T		++
322	G		++

Position	Base	Background	Modification
323	A		++
324	A		++
325	A		++
326	G		++
327	G		++
330	C		+
331	T	+	
336	G		+
339	G	+	
345	T	+	
346	A		+
347	G		++
348	A		++
350	G		+
352	T	+	
353	G	+	
354	G	+	
355	A	+	
356	T	+	
357	T	+	
358	T	+	
359	T		+
360	T		+
361	T		+
364	G		++
365	G		++
366	G		+
368	A		++
369	G		++
371	G		+
372	G		+
374	A		++
375	A		++
376	A		++
378	C	+	
379	A	+	
380	G	+	
381	C	+	
382	A		++
383	G		++
388	C	+	
389	C	+	
390	G	+	
391	C	+	
393	A		++
395	G	+	
396	A		++

Table 3.b. (Continued).

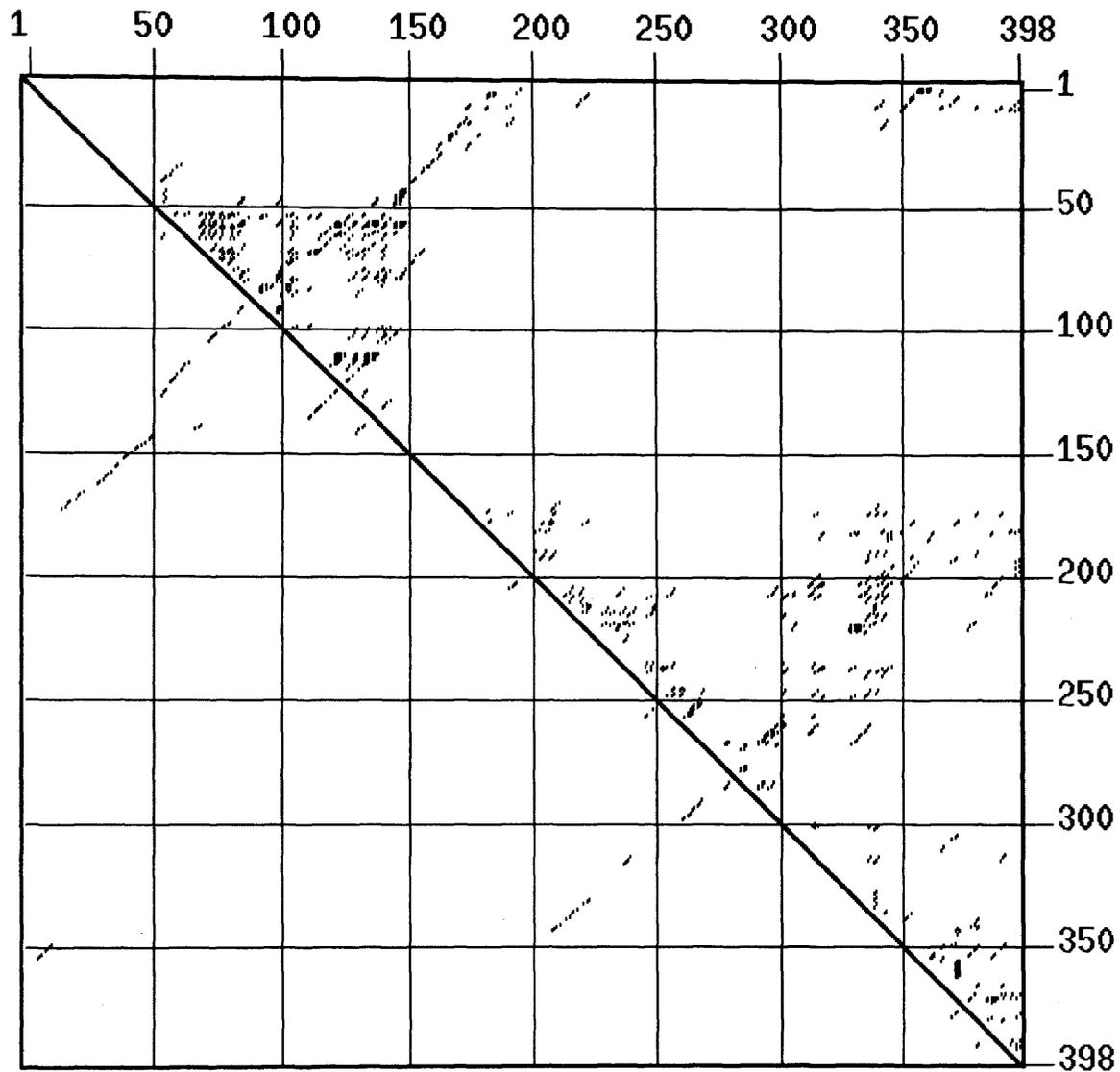
3.4 Structure Modelling

Having obtained a set of data recording the accessibility of certain bases to chemical probes, it was now possible to rule out a great number of helices suggested by preliminary Mfold algorithm results. By constraining the “strongly” modified positions to be single-stranded, a new set of structures were generated, represented in the form of a dotplot in Figure 3.x.

By examining a number of these suboptimal secondary structures with weak modifications superimposed, several well-defined motifs were apparent, including two apical loops and one interior loop (Figure 3.xi). The compound structures terminating in the apical loops were named domain I and domain II. The Mfold algorithm predicts that loop 1 will form as shown in Figure 3.xi.B, a hexaloop with the sequence UGGGAA. By inspection, the loop is more likely to form as GGGAA pentaloop (Figure 3.xi.C), since the pattern of modifications is typical of a GN(n)RA polyloop. If the hexaloop were to form, the closing A/U pair would certainly form a canonical Watson-Crick pairing, and the A would be strongly protected rather than strongly modified. The pattern of modification observed is more consistent with the sheared G/A pair that forms at the base of a GN(n)RA polyloop.

These motifs were then used as constraints upon the Mfold algorithm in addition to strong chemical modifications, yielding the energy dotplot shown in Figure 3.xii.

It is apparent from Figure 3.xii that a number of alternative structures with roughly equivalent stabilities are predicted to form between the well-defined proximal and distal portions of domain 1. Moreover, the superimposition of “weak” modification data does not completely resolve the picture. The sequence alignment, however, reveals sizeable sequence insertions in the species derived from pig, cat, rat, and mouse. There are no equivalent insertions elsewhere in the sequence which might accommodate the formation of species-



Optimal energy: -136.4

Basepairs Plotted: 1179

Figure 3.x. Energy dotplot of Mfold-generated suboptimal foldings of the human *c-myc* 5' UTR sequence with strongly modified bases constrained to be single-stranded. The x- and y-axes both represent the sequence of the authentic P2 *c-myc* mRNA 5'UTR. Each point above the diagonal line represents a base pair in a predicted secondary structure of a stability within 5% of the calculated most stable structure. Points beneath the line represent base pairs in only the most stable predicted structure (-136.4 kcal/mole).

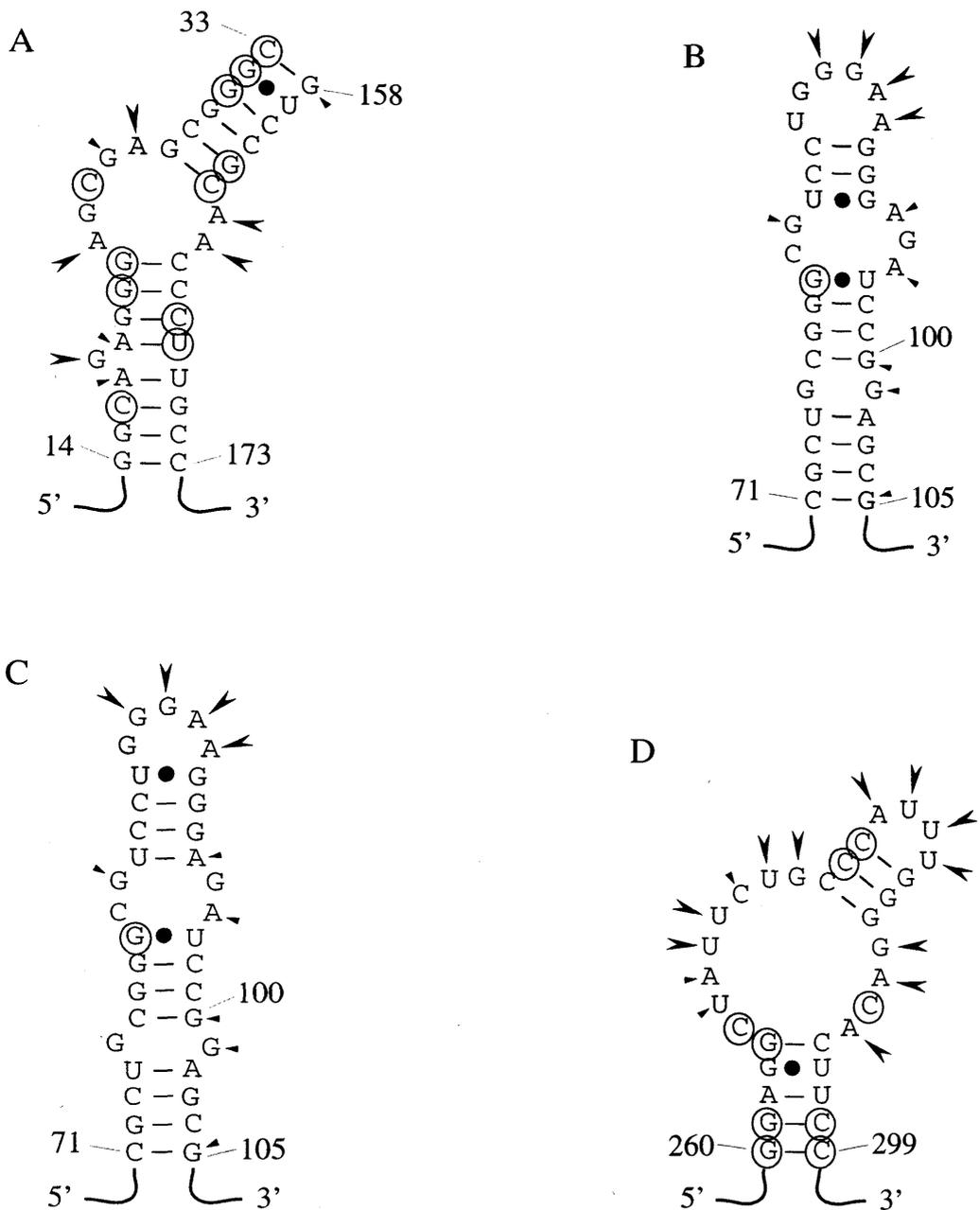
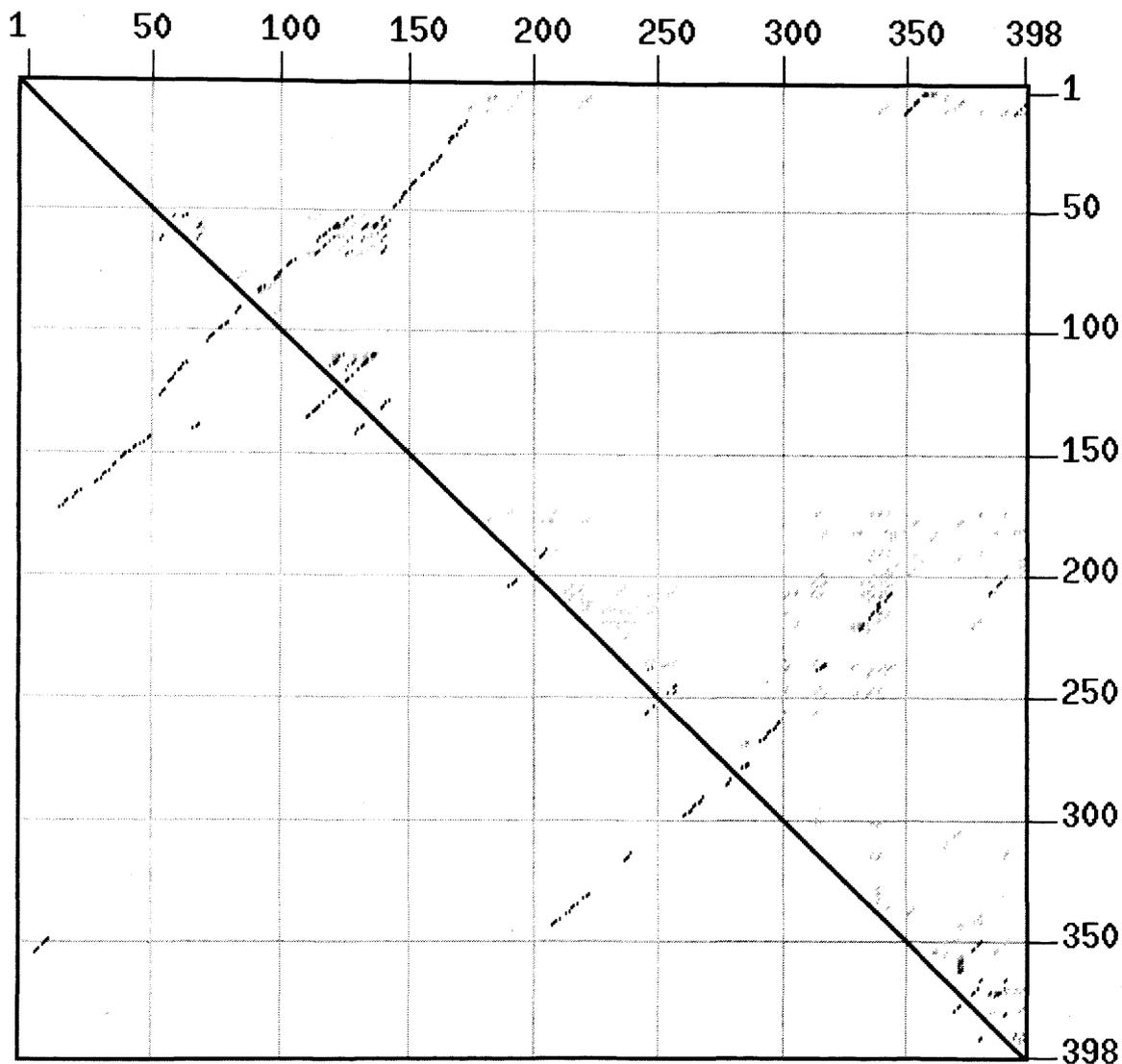


Figure 3.xi. Secondary structure predictions of "well-defined" motifs. A: "Elbow" motif in the 5' proximal region of domain 1 B: Apical stem-loop of domain 1 as predicted to form by the Mfold algorithm C: Revised apical loop 1 incorporating GN(n)RA pentaloop D: Domain II. Large and small arrows mark strong and weak chemical modifications respectively. Sites of reverse transcriptase arrest are circled.



Optimal energy: -136.4

Basepairs Plotted: 820

Figure 3.xii. Energy dotplot of Mfold-generated suboptimal foldings of the human *c-myc* 5' UTR sequence with strongly modified bases constrained to be single-stranded and well-defined motifs constrained to form. The x- and y-axes both represent the sequence of the authentic P2 *c-myc* mRNA 5' UTR. Each point above the diagonal line represents a base pair in a predicted secondary structure of a stability within 5% of the calculated most stable structure. Points beneath the line represent base pairs in only the most stable predicted structure (-136 kcal/mole).

specific helical segments, so the insertions can only be presumed to lie between structured elements, most likely as bulges of slight or insignificant structure. An analogy might be sought in rRNA: the sizeable insertions in eukaryotic rRNA relative to bacterial rRNA are mostly accommodated in this fashion, as is demonstrated by their relative vulnerability to nuclease treatment of intact ribosomes (Holmberg and Nygard, 1997). By eliminating structures with helical elements that bridge this region, which could not form in those species bearing insertions, the structure shown in Figure 3.xiii was obtained. This, however, is at the expense of predicting two sizeable and apparently exposed loops containing a quantity of bases strongly protected from chemical modification, namely nucleotides 135-143 and 117-124.

In its turn, this problem is overcome by the observation that the region 136-142 is complementary to a seven-nucleotide segment further downstream, nucleotides 198-204. Likewise, nucleotides 118-121 and 206-209 are also complementary, accounting for the majority of the protected region. The formation of these helices results in a double pseudoknot structure.

An extensive scrutiny of the region between domains 1 and 2 failed to identify any well conserved structures, or any structures that well accommodated the observed pattern of “weak” chemical modifications. It seems likely that these regions do not take up any single stable structure under the conditions used for the chemical probing, but instead assume a number of alternative conformations. It is not inconceivable that specific IRES-binding factors might act as scaffolds bringing structural order to this region of the UTR. Even if this is the case, there remains little scope for secondary structure held in common between species.

Nucleotides 300-398 are similarly refractory to modelling. It has been previously shown that deletions in this area have little impact upon IRES function (Stoneley *et al.*, 1998), suggesting that the ribosome is scanning through this portion of the UTR. Thus the presence of

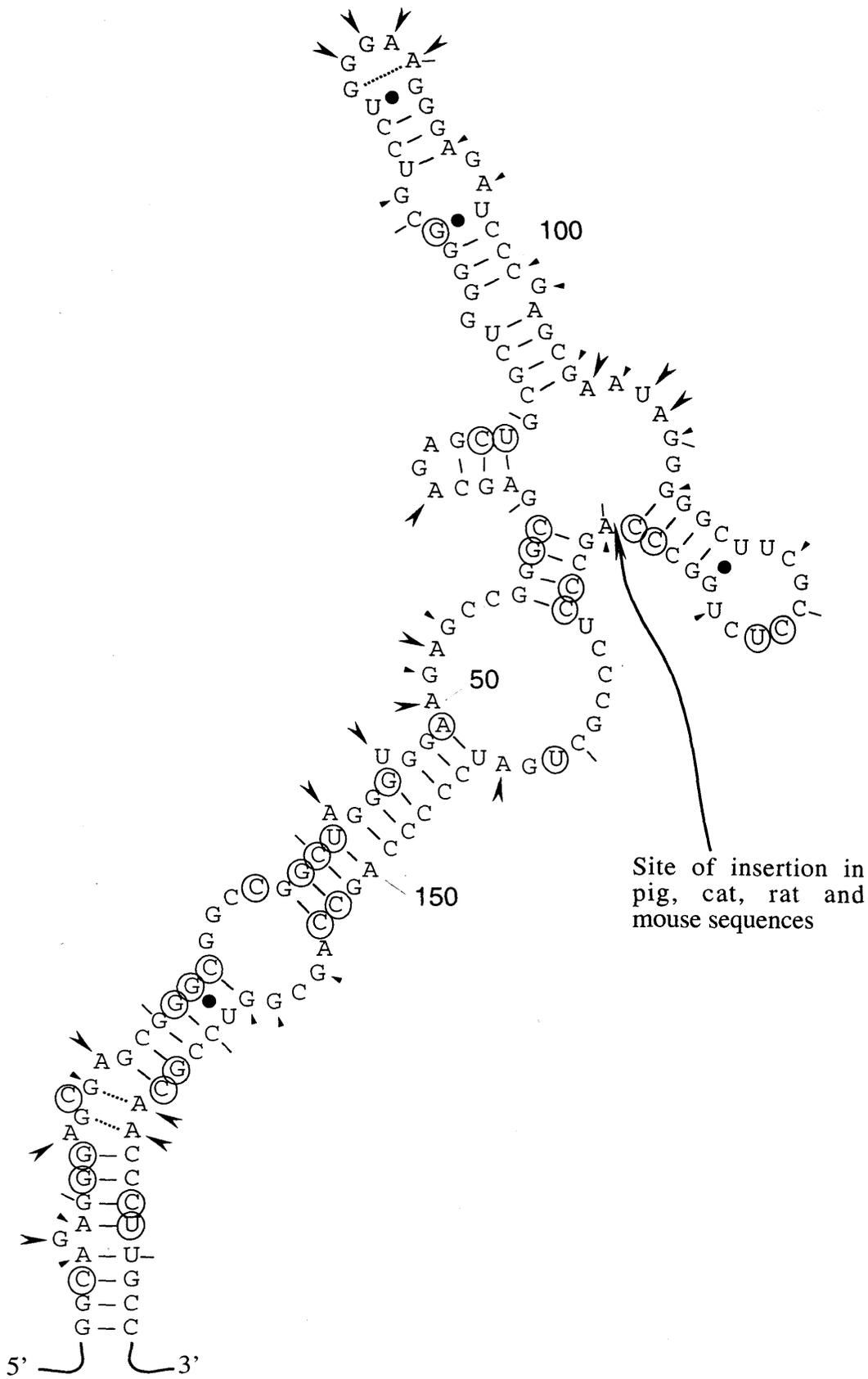


Figure 3.xiii. Secondary structure model of domain 1 constrained to form without bases 129-130 as consecutive paired nucleotides. Note the protected loops 135-143 and 117-124. Large and small arrows mark strong and weak chemical modifications respectively. Sites of reverse transcriptase arrest are circled. Sheared G/A pairings are

stable structures at the 3' end of the sequence would be disadvantageous to Myc protein expression, as it might be expected to block scanning after ribosome entry. These observations also eliminate the idea that long-range secondary interactions between the extreme 5' and 3' ends of the molecule are crucial to IRES function.

The overall secondary structure model is represented as a diagram (Figure 3.xiv) and superimposed upon a sequence alignment (Figure 3.xv).

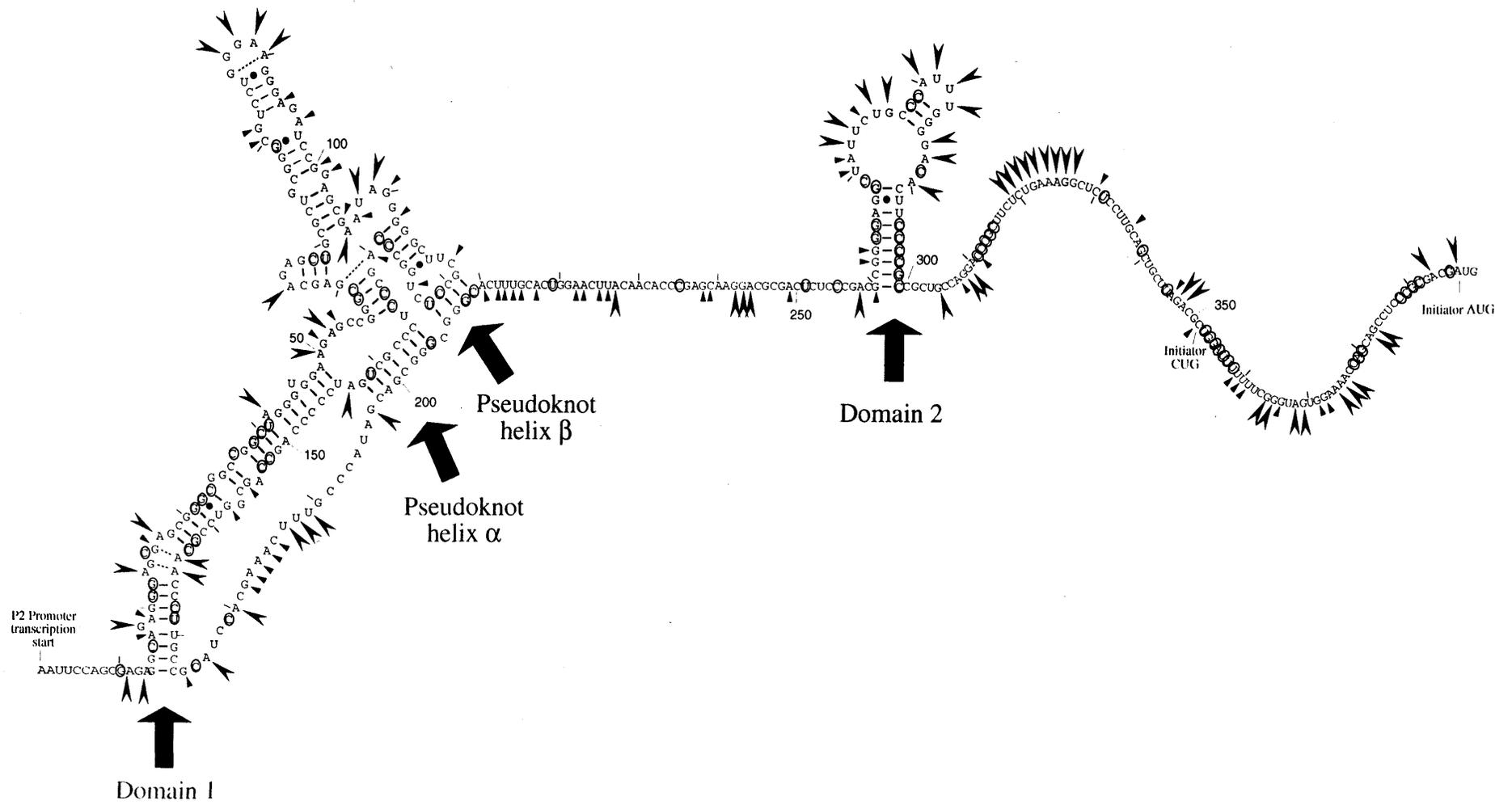


Figure 3.xiv. Secondary structure of the human *c-myc* 5' UTR. >:strong modification. <:weak modification. o:RT block. -:G/A sheared pair.

3.5 Other Strategies

A number of alternative strategies might have been attempted to further test elements of the model. Data supporting or excluding the domain structure might have been obtained by the chemical probing of truncated RNAs corresponding to only domain 1 or domain 2. If the model is correct, the domains would be expected to fold correctly in isolation, and the patterns of protection from chemical attack would be unchanged. If the patterns were altered this would indicate that the truncated RNAs were not folding according to the model, and that the proposed domain structure was erroneous.

Regions of the 5' UTR predicted to be single-stranded could be also be further investigated. Antisense oligonucleotides could be incubated with the folded full-length 5' UTR, and subsequent RNase H treatment followed by primer extension would reveal sites at which DNA/RNA hybrids had been formed. Such cleavages would only be expected to occur between bases which are predicted to be unpaired in the secondary structure model. This approach would be particularly useful in testing the boundaries of the proposed single-stranded regions between domain 1 and the pseudoknots and between the pseudoknots and domain 2.

Another approach that might yield useful structural data is psoralen cross-linking. This reagent, when activated by UV, covalently links closely apposed nucleotides within a folded RNA molecule. Single cross-linked species can then be purified by denaturing PAGE, and the location of the cross-link can be roughly determined either by primer extension or differential electrophoretic mobility studies of the products of targeted RNase H treatment. This approach has been successfully used to identify long-range tertiary interactions within rRNA.

3.6 Features of the Secondary Structure

The human *c-myc* IRES contains two domains of secondary structure. Domain 1 is the larger and more complex, and is predicted to contain two overlapping pseudoknots. Domain 2 contains only two helical segments, separated by a large internal loop.

Domain 1 is capped by pentaloop belonging to the class of G(n)NRA polyloops, stable motifs that are found ubiquitously in structured RNAs and are often involved in RNA-RNA interactions (Abramovitz and Pyle, 1997). Domain 2 is capped by another ubiquitous loop motif, in this case an AUUU tetraloop.

Since the chemical modification agents used only react at specific ring positions on the bases, further scrutiny of the secondary structure and superimposed chemical modification data makes possible the prediction of some non Watson-Crick base pairs (Leontis and Westhof, 1998). In particular sheared G/A pairs may be inferred to form at positions where a kethoxal-unmodified or lightly modified guanine residue lies opposite a strongly DMS modified adenine residue, as shown in Figure 3.xvi. Such pairings are common in internal loops (Gautheret *et al.*, 1994), especially “E loop” motifs as are found in 5s RNA (Correll *et al.*, 1997), and at helix junctions and termini (Heus and Pardi, 1991). A sheared G/A pair distorts the normal A-form helix backbone geometry in such a way that it cannot be connected to a Watson-Crick pair situated 5' of the adenosine (Cheng *et al.*, 1992; Gautheret *et al.*, 1994). Thus sheared pairs are predicted to form in the “elbow motif” internal loop at G24/A63 and G26/A64, in the helix junction at G60/A130 and at the base of the GGGAA pentaloop at G86/A90.

The overlapping double pseudoknot motif in domain 1 is topologically unique among published RNA structures. Multiple pseudoknots have been described in numerous molecules, including rRNA, hepatitis delta virus (HDV) ribozyme (Wadkins *et al.*, 1999), tobacco mosaic

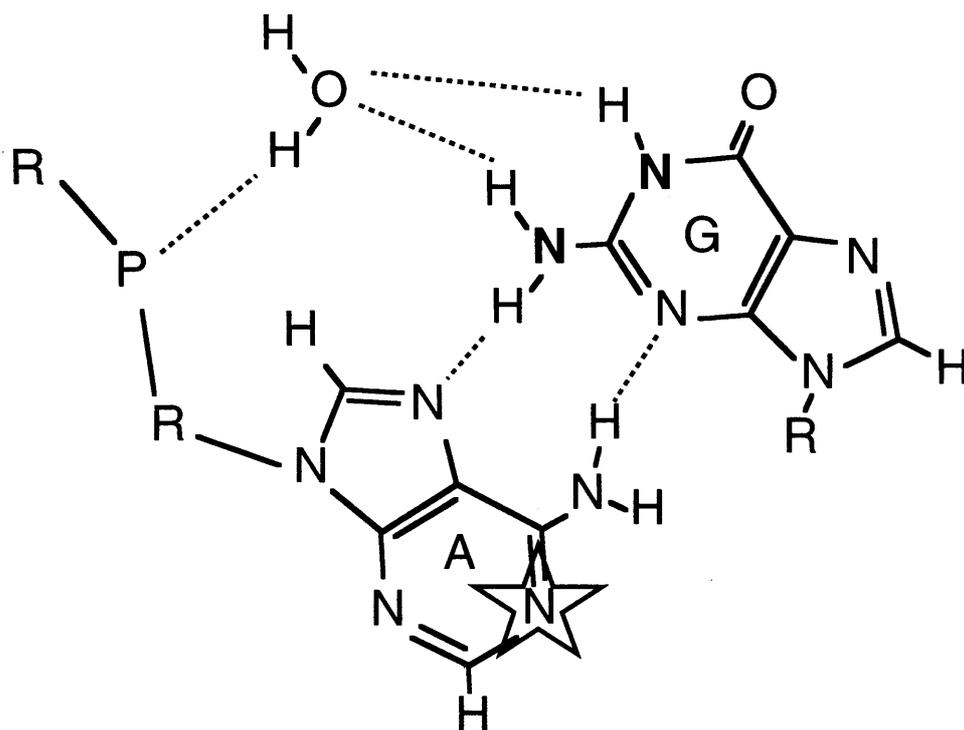


Figure 3.xvi. Sheared G-A base pair showing accessibility to modification by chemical agents. N-1 of adenine (starred) is highly accessible to DMS. N-1 and N-2 (bold) of guanine are shielded from attack by kethoxal by H-bonding with water. In resolved crystal structures, this water molecule also contacts the phosphate group 5' of adenosine.

virus RNA (Felden *et al.*, 1996) and *E. coli* 10Sa RNA (Felden *et al.*, 1997). In these molecules, the pseudoknots are either topologically distinct entities or, in the case of the HDV ribozyme, fully nested (Figure 3.xvii).

Classical pseudoknot interactions are limited in size to a maximum of 7 nucleotides. Pseudoknot-forming helices any longer than this approach a full turn of an A-form helix, and the formation of a true overhand knot. This is most unlikely, as it would require “threading” of one end of the molecule through an RNA loop. At first sight, this rule would seem to rule out the possibility of stacking between helices α and β . However, this is not so; even if helices α and β do stack co-axially, as is suggested in Figure 3.xiv, they escape this difficulty since the upstream portions of the helices are formed by two separate loops, and no threading need occur. Instead, the gap opposite the (presumably flipped) base C205 in the otherwise continuous helix would allow the downstream portion of the molecule to form the helix by entering “sideways”, rather than by threading. Alternatively, the helices might hold a disjointed conformation.

It is appealing to speculate that this double pseudoknot arose by evolutionary pressure in favour of a longer or stronger single pseudoknot-forming helix. As extension of one pseudoknot beyond 7 nucleotides was impossible, the structure might have been augmented by the topologically permissible stratagem of continuing the helix with a distant portion of the molecule, leading to the double pseudoknot structure predicted to form.

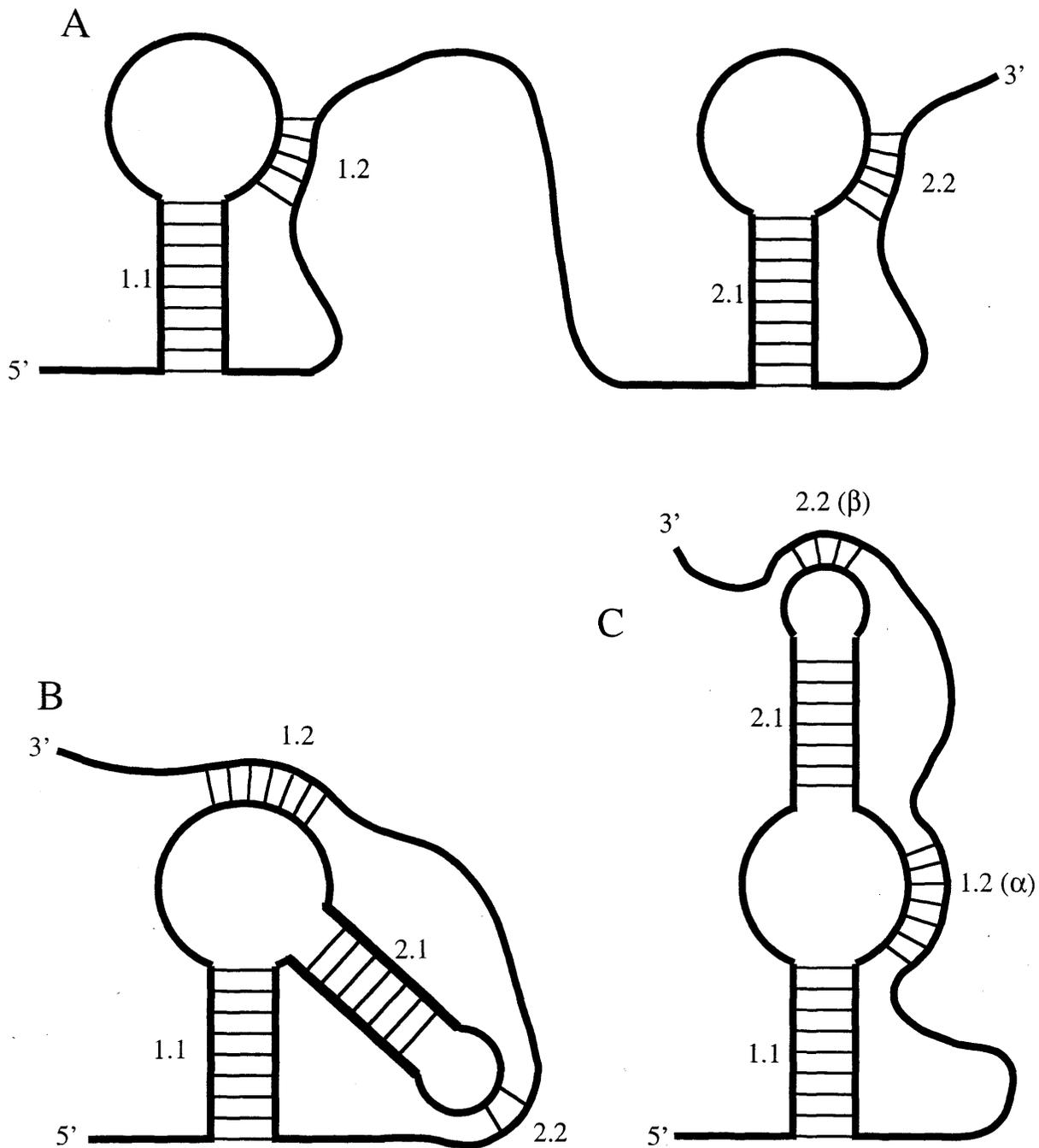


Figure 3.xvii. Three kinds of double pseudoknot. A: Distinct pseudoknots, as described in, for example, *E. Coli* 10Sa RNA. B: The nested double pseudoknot in HDV ribozyme. C: The overlapping double pseudoknot in the *c-myc* IRES. Helices not involved in pseudoknot formation are omitted. Helices are numbered such that n.1 and n.2 form a single pseudoknot structure in the absence of other secondary structure.

3.7 Conservation of Secondary Structure

To what extent is secondary structure of the *c-myc* IRES conserved between species?

The answer to this consequential question determines the extent to which the sequence alignment may be drawn upon as a resource to aid modelling the human IRES. A number of observations assist the formation of a conclusion.

If there are to be grounds for belief in a shared structure between species, there must be a shared function. That is to say that all the sequences aligned must form functional IRESs. The only other sequence to have been tested in such an assay is that derived from the mouse, which does indeed function as an IRES when transiently transfected into a human cell line, and presumably in murine tissues as well (C. Jopling, personal communication). Given that the mouse and rat sequences bear the least degree of sequence identity (excluding the incomplete ovine sequence) with the human 5' UTR, this is circumstantial support for a functional IRES in a range of mammalian species. We still cannot rule out the possibility that, say, only primates and rodents have retained (or developed) *c-myc* IRESs. The balance of probabilities seems to lie in favour of a common IRES, however.

The high degree of sequence conservation between species suggests that there must be some structure held in common. The apical stem-loop structure of domain 1 is an example of a structure predicted to form in all species before any "wet" data was considered, and which was subsequently borne out by chemical probing analysis. Likewise, it is impossible to believe that the almost identical primate sequences bear significantly divergent structures. But to be useful for structural prediction, sequences must be at one and the same time very similar in structure and significantly divergent in sequence- a figure of about 70% sequence identity is often quoted as being optimal (James *et al.*, 1989). This may be reasonably assumed in many situations where phylogenetic analysis has proved to be a powerful tool, such as rRNA and

viral IRES secondary structure prediction. One can be confident of common structure in the former case since the translational machinery is universal and ancient in evolutionary terms, and plays such a vital role in the existence of the cell that it will not bear much structural change. In the latter case, a given viral IRES holds its optimum conformation in the face of constant and massive selective pressures upon the size of its genome, proper interaction with host cell machinery and the ability to out-perform mutant strains.

The case of the *c-myc* IRES lies between these examples. As in a study of rRNAs, the sequences in this alignment come from a range of different species, but in this case the mechanism in question is not even known to be universal among the class of mammals, let alone across kingdoms. As in an alignment of viral IRES sequences, the sequences (probably) bear a common function, but are not exposed to identical evolutionary pressures since they come from different species. Even if we accept as fact the likelihood that they are interacting with homologous protein factors, we have as yet neither identified these factors nor the extent to which they themselves are conserved.

Furthermore, in the case of the viral IRES, the typical requirement is to function as efficiently as possible under all circumstances. The expression of Myc protein, however, is likely to be under much more subtle control. This putative control mechanism would represent another level at which structure variation might creep in, as one species might achieve control in a quite a different fashion to another.

Most tellingly, a thorough scrutiny of the *c-myc* 5' UTR sequence alignment revealed only a few perfectly or nearly perfectly conserved motifs (Figure 3.iii). Of these structures, only one was supported by chemical probing analysis. This forces us to the conclusion that in this case secondary structure is far from perfectly conserved.

This is reflected in the superimposition of the predicted human secondary structure over the aligned sequences. While the alignment has been improved by the insertion of spacers to maintain base pairs where possible, it is clear that some helices will not form in some species.

Chapter 4

Mutagenic analysis

4.1 Introduction

An approach combining structural modelling and mutagenesis permits the establishment of links between structure and function by the creation of mutations with specific effects upon RNA structure, and the assessment of their impact upon IRES function. Numerous studies of this kind have been carried out on viral IRESs, including, for example, aphthovirus and flavivirus IRESs (Martínez-Salas *et al.*, 1996; Wang *et al.*, 1995). Mutagenic analysis can also provide further supporting evidence for specific features of a secondary structure model, or identify the site at which the ribosome first acquires the message into the mRNA binding track, or “lands” (Pilipenko *et al.*, 1994; Reynolds *et al.*, 1996; Rijnbrand *et al.*, 1997).

A number of terminal deletions mutants of the *c-myc* IRES have previously been generated and assayed for their impact upon IRES function (Nanbru *et al.*, 1997; Stoneley *et al.*, 1998), the results of which are summarised in Table 4.a. These data show that 5' deletions of domain 1 gradually ablate IRES function, becoming negligibly low when the first 237 nucleotides are deleted. This corresponds roughly to the 3' boundary of domain 1. Deletions at the 3' end have little impact at first, with the loss of 59 nucleotides having no effect whatsoever upon IRES efficiency. The deletion of 165 nucleotides (1-233), corresponding to the complete removal of domain 2 and most of the spacer region reduces efficiency to about 35%. A deletion that stops just short of domain 2 (1-308) reduces activity to about 60%, suggesting that the spacing between domain 2 and the start codon(s) may be important.

Bases present	% IRES Activity
1-395 (complete)	100
59-395	≈70
101-395	≈40
173-395	≈25
237-395	≈10
255-395	0
1-233	≈35
1-308	≈60
1-339	≈100

Table 4.a. Summary of *c-myc* IRES terminal deletion mutant data. Activity does not drop catastrophically when a minimal element is infringed, but is dispersed along the IRES.

4.2 Mutagenesis

A number of novel mutant forms of the *c-myc* IRES were designed, and synthesized using the Stratagene QuikChange™ protocol (Figure 4.i), using the plasmid pSKLUTR as the template.

The data obtained from such mutants is only of use if the RNA is folding correctly. For example, if a mutant designed to disrupt a particular helix instead folds into a novel structure elsewhere, then any effect upon IRES efficiency cannot be solely ascribed to the disruption of that particular helix. For this reason every care was taken when designing mutations to preserve correct folding as far as possible. The likelihood that a mutant would fold correctly was assessed in every case by inspection of predicted foldings of mutant structures.

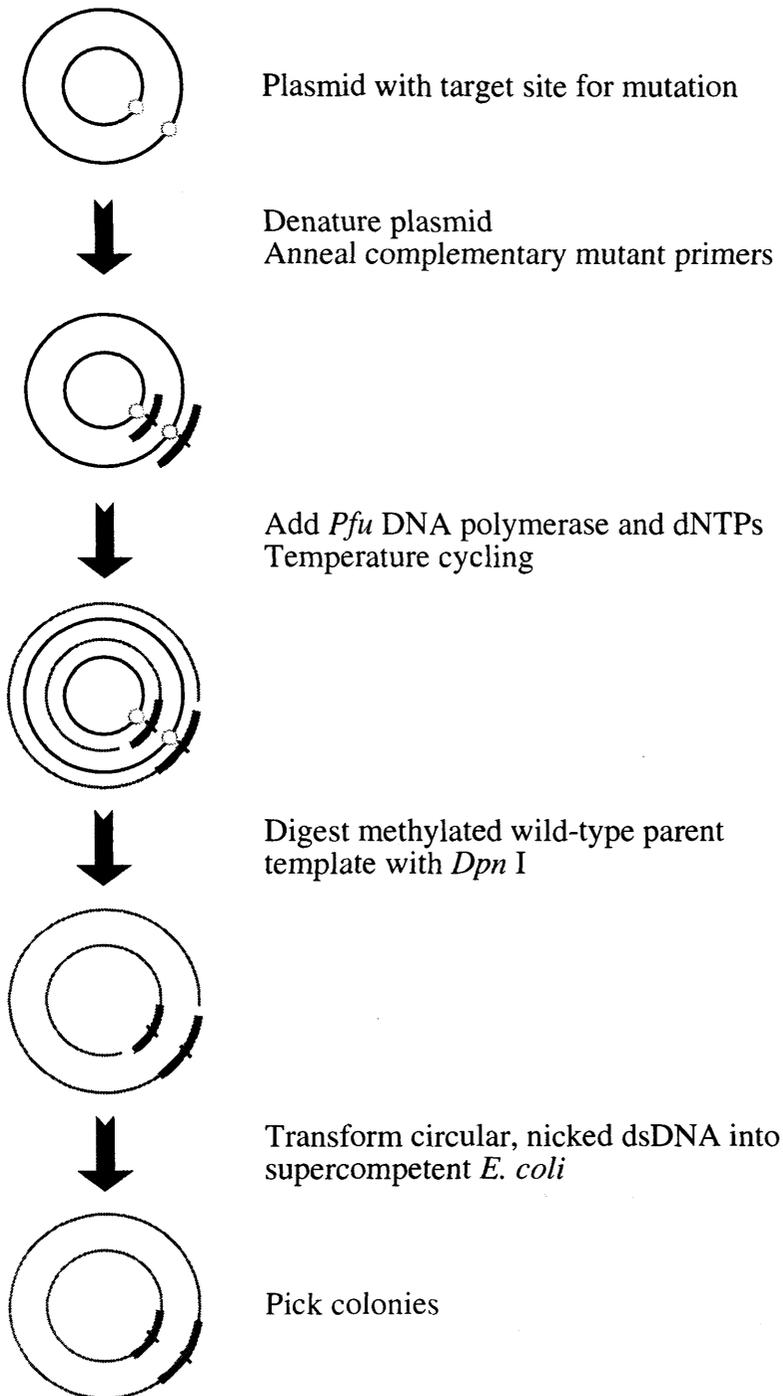


Figure 4.i. Overview of the QuikChange™ site-directed mutagenesis method.

4.3 IRES assay

Mutant UTRs were then excised and ligated into the bicistronic expression vector pRhpF (Figure 4.ii). This plasmid contains a stable hairpin structure with a predicted stability of -58.1 kCal/mol between the two forms of luciferase, upstream of the site at which UTR fragments are inserted. This is to impede ribosomes that re-initiate scanning at the end of the first cistron, and which would otherwise go on to translate the second cistron. This “read-through” effect is undesirable since scanning ribosomes will disrupt IRES structure, affecting perceived levels of efficiency in an unpredictable manner (Stoneley *et al.*, 2000b).

Wild-type and mutant IRES efficiencies are compared by reporter gene assays of transiently transfected HeLa cell lysates. In each experiment, two constructs are mixed and transfected together. One, derived from pRhpF, contains the mutant *c-myc* IRES, and the other, p β -Gal, expresses β -galactosidase, which is assayed and used as a control for transfection efficiency. Transfections using the wild-type *c-myc* IRES were performed in parallel as a further control. Experiments were executed in triplicate. IRES activity was calculated as the average of (IRES-driven firefly luciferase expression/ β -Gal expression), and efficiency expressed as (mutant IRES activity)/(wild-type IRES activity)x100%. Errors were calculated as the standard deviation of the three calculated IRES activities, and expressed as a percentage of the average activity.

4.4 Identification of the Ribosomal Entry Window

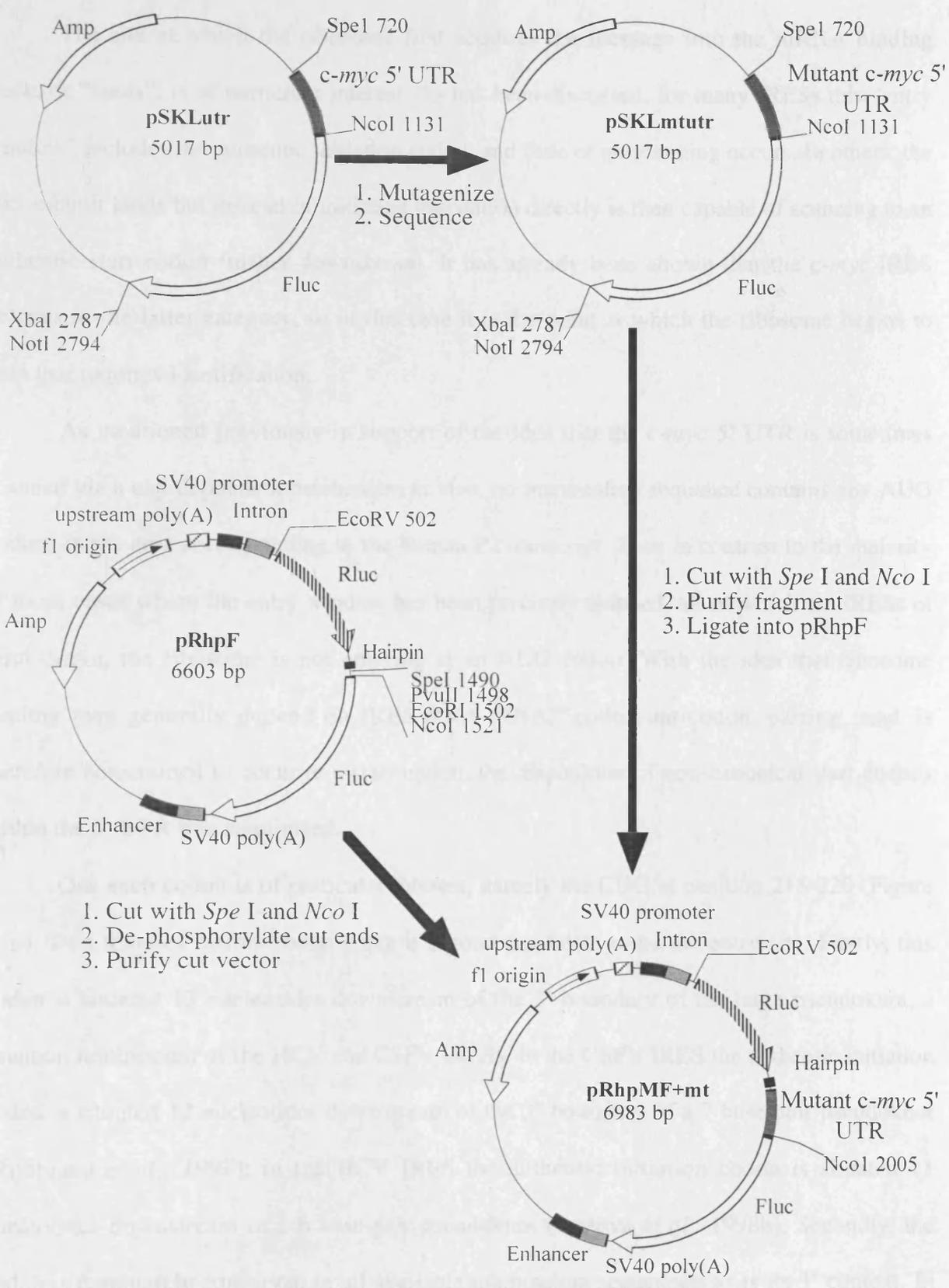


Figure 4.ii. Strategy for cloning mutant forms of the *c-myc* 5' UTR.

4.4 Identification of the Ribosomal Entry Window

The site at which the ribosome first acquires the message into the mRNA binding track, or “lands”, is of particular interest. As has been discussed, for many IRESs this “entry window” includes the authentic initiation codon, and little or no scanning occurs. In others, the 40S subunit lands but instead of initiating translation directly is then capable of scanning to an authentic start codon further downstream. It has already been shown that the *c-myc* IRES belongs to the latter category, so in this case it is the point at which the ribosome begins to scan that requires identification.

As mentioned previously in support of the idea that the *c-myc* 5' UTR is sometimes scanned via a cap-dependent mechanism *in vivo*, no mammalian sequence contains any AUG codons in the area corresponding to the human P2 transcript. Thus in contrast to the majority of those cases where the entry window has been precisely mapped, all of which are IRESs of viral origin, the ribosome is not entering at an AUG codon. With the idea that ribosome landing may generally depend on IRES RNA-tRNA_i^{met} codon-anticodon pairing, and is therefore constrained to occur at a start codon, the disposition of non-canonical start codons within the 5' UTR was scrutinized.

One such codon is of particular interest, namely the CUG at position 218-220 (Figure 4.iii). Two features of this codon make it a good candidate to be the entry site. Firstly, this codon is situated 13 nucleotides downstream of the 3' boundary of the large pseudoknot, a situation reminiscent of the HCV and CSFV IRESs. In the CSFV IRES the authentic initiation codon is situated 12 nucleotides downstream of the 3' boundary of a 7 base-pair pseudoknot (Rijnbrand *et al.*, 1997); in the HCV IRES the authentic initiation codon is situated 11 nucleotides downstream of a 6 base-pair pseudoknot (Pestova *et al.*, 1998b). Secondly, the codon is completely conserved in all available mammalian sequences, as is its 3' context. In

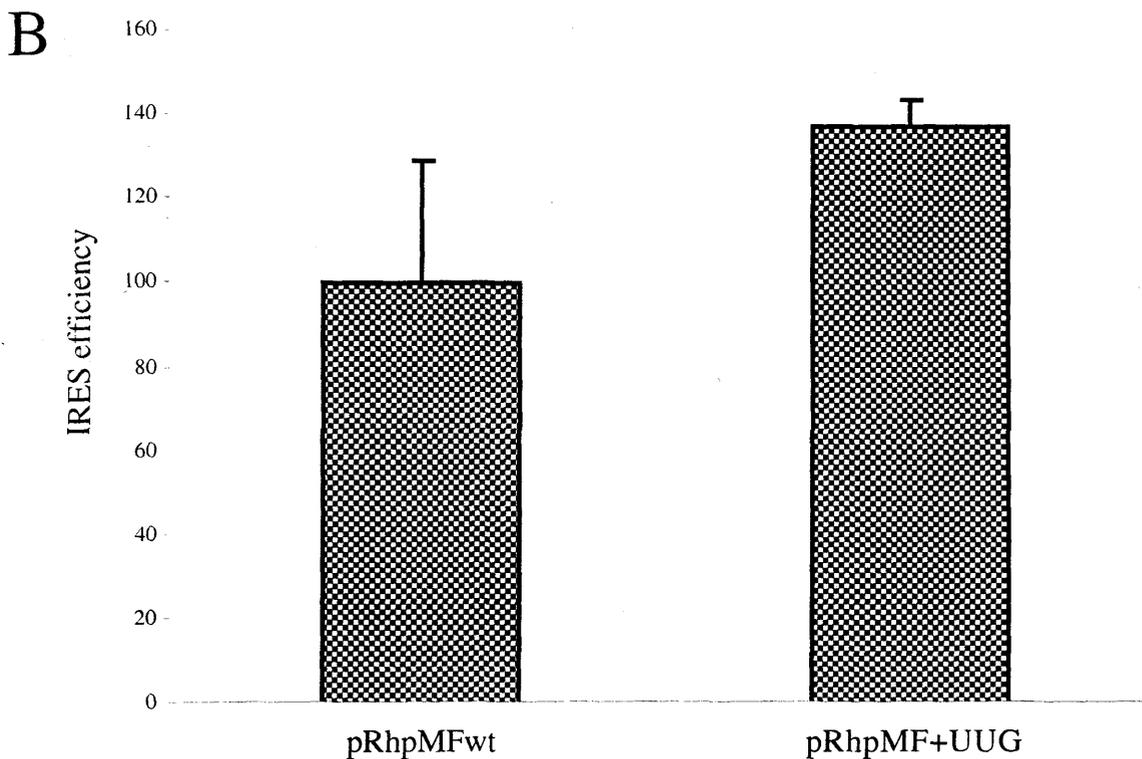
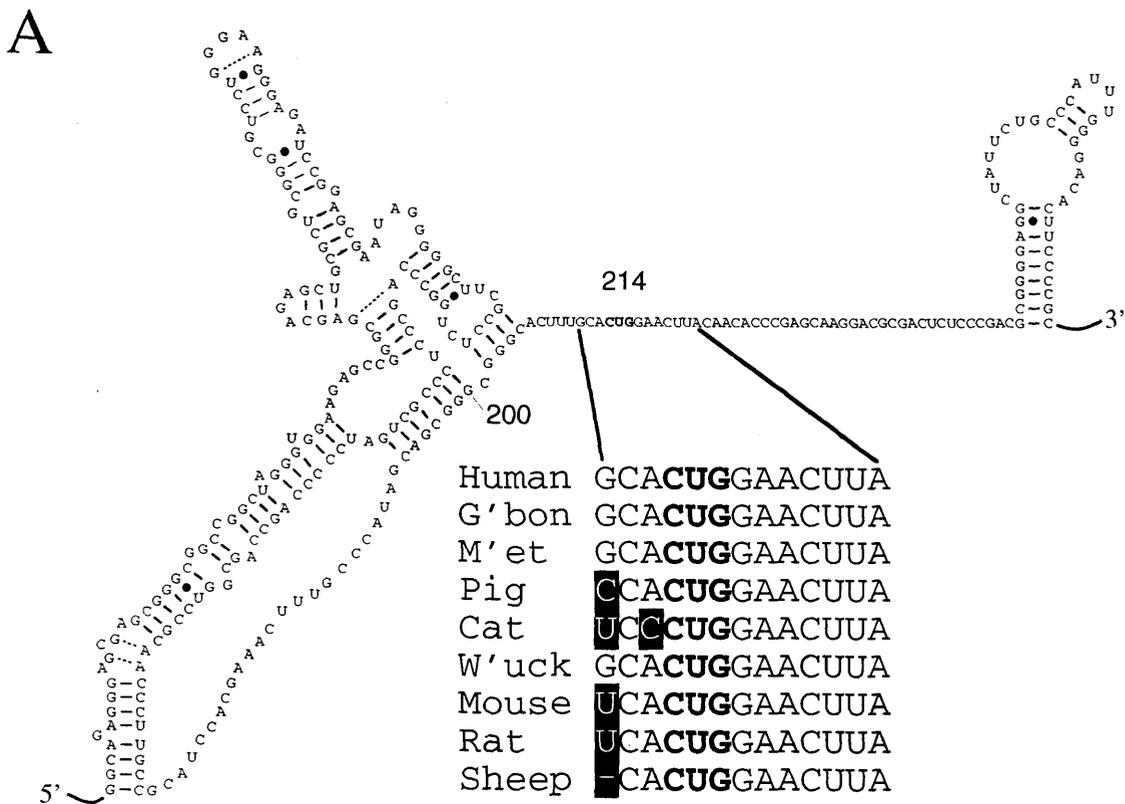


Figure 4.iii. A candidate ribosome landing site. A: Position and context of a conserved CUG codon (bold); non-conserved positions are shaded. B: Effect upon IRES efficiency of codon mutation from CUG to UUG.

order to test the hypothesis that ribosome entry is taking place at this CUG codon, the mutation C218U was produced, changing the codon to a UUG.

If the conserved CUG were indeed involved in ribosome landing, this would presumably be mediated by codon-anticodon pairing between the CUG and the tRNA_i^{met} in the ribosomal P-site. The change from a C to a U at position 1 would be expected to impair this interaction, and consequently IRES efficiency, in agreement with the observation that UUG is apparently never naturally used as a eukaryotic start codon. That no such decrease in IRES efficiency is observed suggests that this codon is not crucial for ribosome entry, and that either ribosome landing is taking place at this point but the acquisition of mRNA into the binding track is unaffected by the change in RNA, or that the entry window is located elsewhere. The apparent rise in IRES efficiency is not easily explained.

As prediction from sequence and secondary structure failed to identify the precise location of the entry window, a more exploratory approach was employed. It is possible to determine if scanning is initiated 5' or 3' of a given point by the introduction by mutagenesis of an AUG start codon that is out-of-frame relative to the downstream reporter. If the AUG lies 5' of the ribosome entry window, and the sequence change does not otherwise affect IRES function, IRES efficiency will not be affected. If, however, the AUG lies downstream of this point, the scanning ribosome will initiate polypeptide synthesis prematurely and expression of the reporter gene will be abolished, or at least very much diminished.

Accordingly, nine AUG codons were individually engineered into the 5' UTR to locate the entry window. In each case the AUG was designed to have a guanosine at position +4, giving the sequence AUGG. This improves the context of the start codon, favouring initiation over readthrough. Likewise, where possible a purine was present at position -3. Mutants and activities are shown in Figure 4.iv.

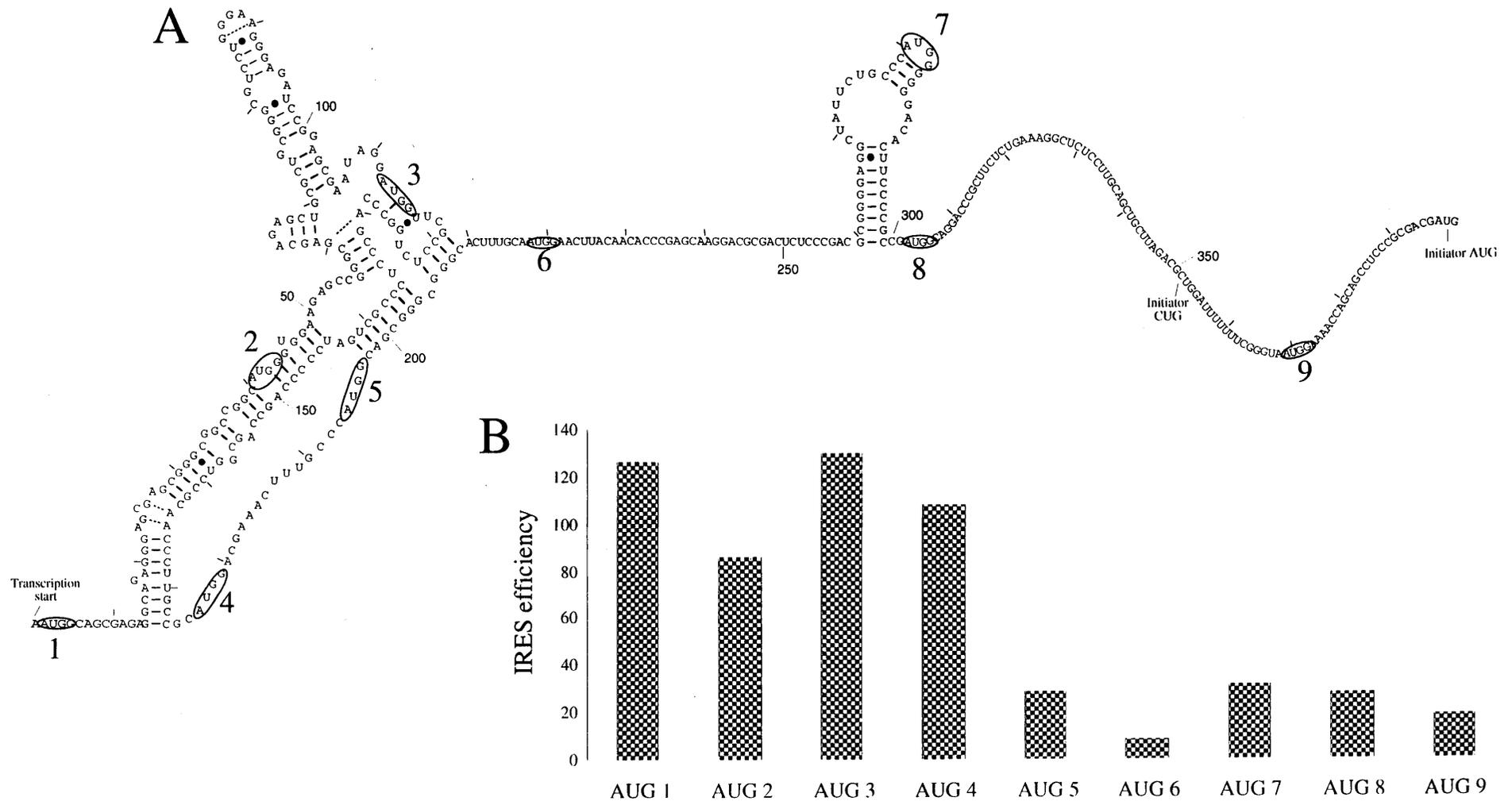


Figure 4.iv. The ribosome entry window. A: Locations of engineered upstream AUG codons. AUGs were present singly in transfected constructs. B: Activities of mutant IRESs. The ribosome is entering at some point between nucleotides 177 and 194.

There is a possibility that any single AUG affects IRES function not via an altered codon-anticodon interaction but rather because the mutant IRES is mis-folded or has abnormal *trans*-acting factor binding properties. This would have the most misleading consequences if it was the root cause of the impairment of AUG mutant 5. In order to more completely rule out this possible source of error it would be necessary to perform control experiments to demonstrate that other base substitutions (such as CUG, UUG and GUG) resulted in IRESs of approximately wild-type activity. However, it remains unlikely that such an artefact has been produced, since many relatively gross mutations have only slight effects upon the activity of the IRES.

AUG mutants 1-4 are all active, while AUGs 5-9 are all significantly impaired, suggesting that the ribosome is starting to scan at some point between AUG 4 and AUG 5, or between nucleotides 177 and 194. Mutants 5, 7 and 8 all lack purines at -3, so their relatively high activity is probably due to a failure of the post-internal entry ribosome to efficiently initiate translation at the inserted AUG. This region contains the longest absolutely conserved sequence segment within the 5' UTR, with the sequence GAAACUUUGCC. It is likely that the ribosome is landing within this segment, and the CUU codon is the strongest candidate for the site of the first codon-anticodon interaction. CUU is the site of translation initiation (and very probably ribosome entry) in the PSIV virus IRES (Sasaki and Nakashima, 1999), and in the *c-myc* IRES is situated 12 nucleotides 3' of the nearest helical element, a spacing held in common with the authentic start codon of the CSFV IRES (Rijnbrand *et al.*, 1997).

4.5 Deletion Mutants

Domain 1 is predicted to have an unusual and elaborate structure, containing two overlapping pseudoknots and a large internal loop. Deletion mutant 1 lacks bases 53-141, ablating the GN(n)RA polyloop and both pseudoknots, and might be expected to have a profound effect upon IRES function.

Deletion mutant 2 excises the entirety of domain 2, which is well conserved in all species except cats. The mutants and their relative activities are shown in Figure 4.v.

Both deletion mutants are significantly reduced in activity, to about 60% of the wild-type IRES. Thus neither deleted region is crucial to IRES function, yet both need to be present for maximal IRES activity. So, neither pseudoknot is required for IRES function, in contrast to the HCV and pestivirus IRESs (Wang *et al.*, 1995). Domain 2, which is situated far downstream of the ribosome entry window, and which might have been expected to exert an inhibitory effect upon scanning (and IRES efficiency) is in fact required for maximal activity. It seems likely, therefore, that domain 2 is involved in the assembly of a competent initiation complex by RNA-RNA or RNA-protein interaction.

The observation that neither of these large internal deletions has catastrophic consequences is perhaps surprising, given the relative ease with which viral IRESs may be disabled by much smaller changes. Clearly a very different mechanism is at work.

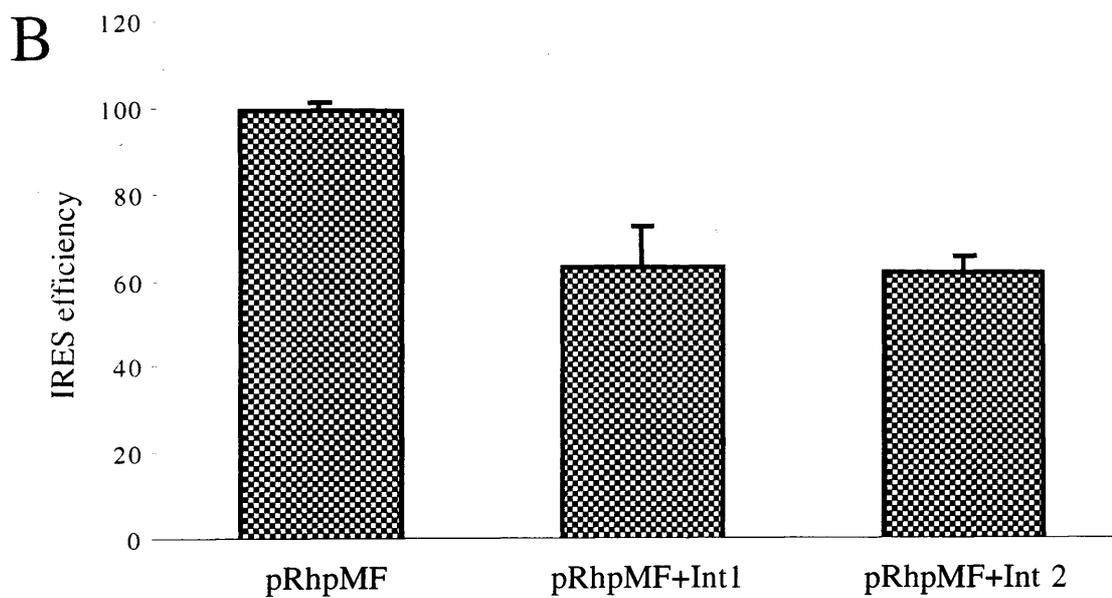
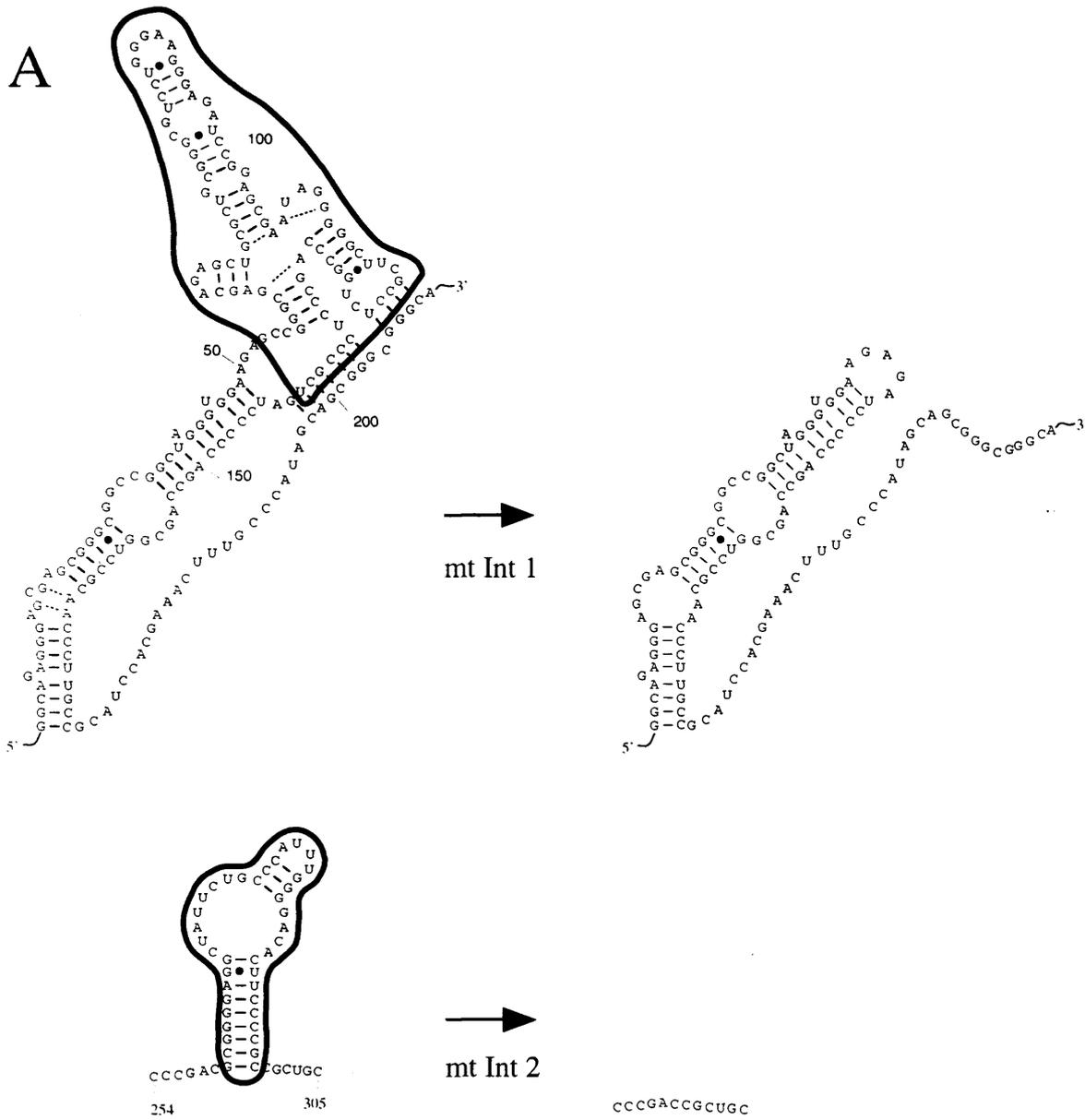


Figure 4.v. Internal deletions. A: Internal deletions of mutants 1 and 2 and their predicted structural consequences. B: Activities of internal deletion mutants 1 and 2.

4.6 Mutations of Domains 1 and 2

In order to dissect more finely the structure-function relationships of domains 1 and 2, a number of base substitution mutations were introduced into the UTR (Figure 4.vi). Mutants A, B, C, D, F and G were all assayed individually, and the double mutants B+C and F+G were also assayed.

Mutant A alters the sequence of the hairpin loop capping domain 1. In the wild-type IRES, this loop has the sequence GGGAA. In mutant A, however, this is altered to GAAUU. This altered loop sequence is no longer a member of the G(n)NRA member of polyloops, which are inherently stable due to sheared G/A pairing, and frequently implicated in RNA-RNA interactions (Abramovitz and Pyle, 1997). The exposure of this loop to the solvent *in vitro* shows that it is not involved in a classical intramolecular interaction, but it may require “scaffolding” proteins for such an interaction to occur, or it may interact with RNA components of the translational machinery. This well conserved and highly exposed loop is also a good candidate for an RNA-protein interaction.

Mutant D disrupts an even more well conserved loop motif. The sequence AUUU has previously been characterised as one capable of forming stable loops, and is believed to exist as a loop in at least one other IRES (Sasaki and Nakashima, 1999), though there is as yet nothing to suggest any more general structural role. In any case, its complete conservation and solvent exposure make it another good candidate for mutagenesis.

Mutants B and C, when present individually, are predicted to disrupt the helical segment supporting the GGGAA loop, resulting in an enlarged stretch of single-stranded RNA. When present simultaneously, the potential for helix formation is restored. The precise topology of this reversed helix will be subtly different from the wild-type form, however, so

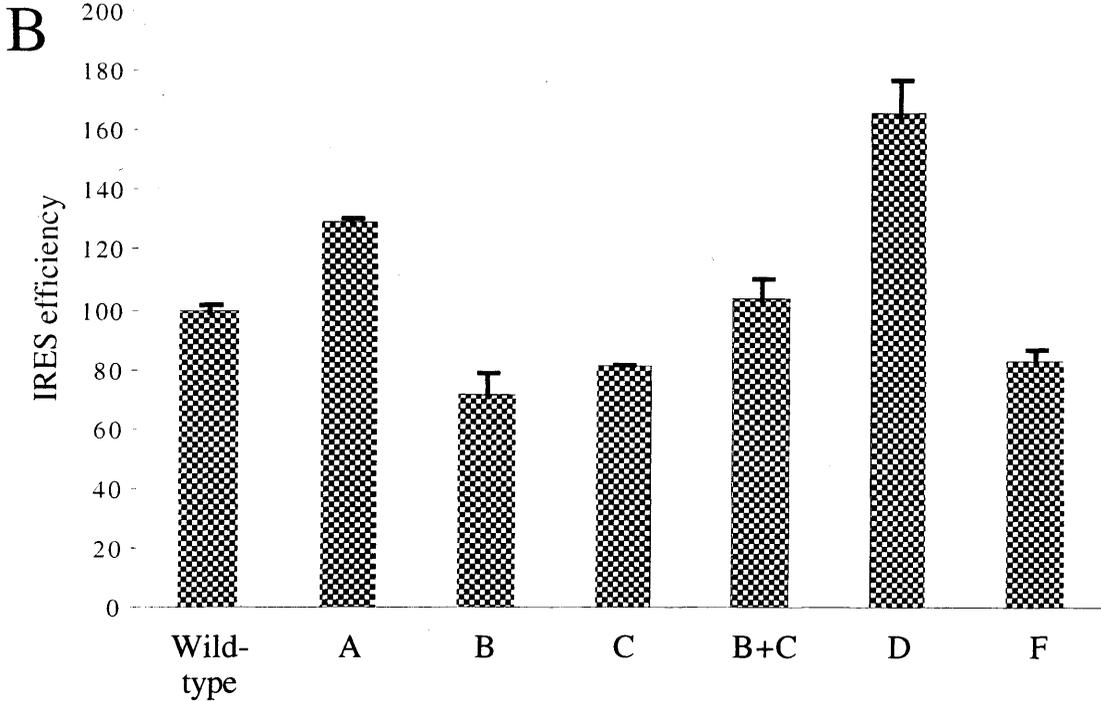
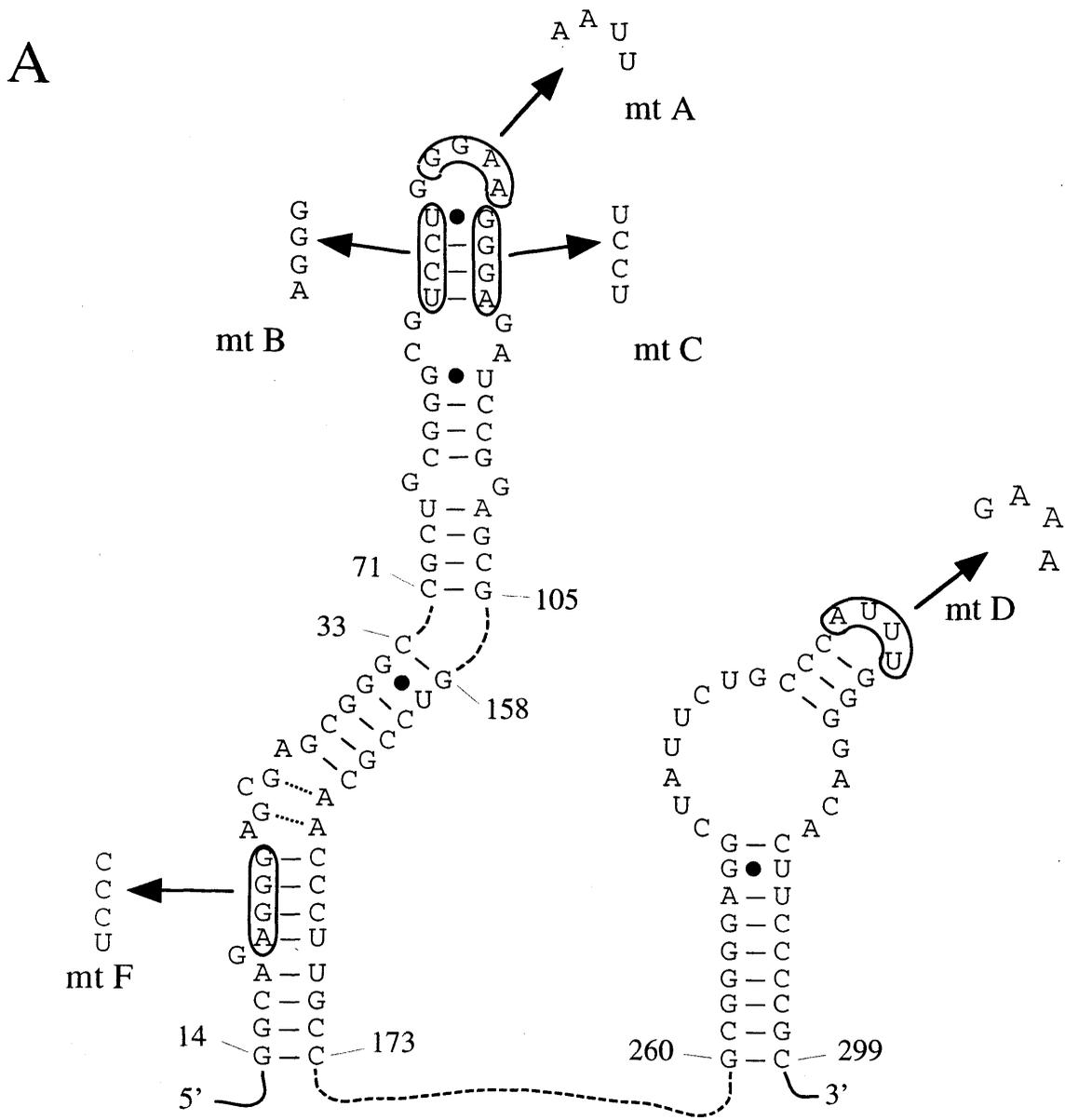


Figure 4.vi. Mutations of domains 1 and 2. A: Base substitutions of domains 1 and 2. B: Activities of mutant IRESs.

whilst there should be no effect upon overall secondary structure within the IRES, the ability to interact with a conformation-specific binding factor might be affected.

Mutant F disrupts a similar helix at the base of domain 1. The presence of a helix at this position is well conserved, but its precise conformation is not, with different species having a range of bulged bases and other discontinuities.

Mutations A and D significantly activate the IRES. These are surprising observations, although not wholly without precedent. The C255U mutation within the *c-myc* IRES was already known to activate (A. Willis, personal communication), and activating mutations have been identified in the IRESs from FMDV (Martínez-Salas *et al.*, 1993) and HCV (Honda *et al.*, 1996). If we accept that a given IRES sequence is functionally optimal, then the existence of mutations that significantly activate the IRES demonstrates that the system is constitutively repressed. This is not easily explained when considering viral IRESs, which are commonly imagined to have evolved under overriding evolutionary pressure in favour of efficiency. It is more understandable when considering a cellular gene, particularly one whose over-expression is associated with as far-reaching effects upon the individual as is *c-myc*.

This repression could be exerted in one of two ways; either the wild-type IRES is failing to make an interaction that favours initiation as efficiently as the mutant, or it is interacting more efficiently with an inhibitory factor than the mutant. It is more likely that the alteration of loop sequences in a more or less undirected fashion reduces the ability to specifically interact than enhances it, leading to the hypothesis that the wild-type loops are making specific, inhibitory contacts.

Mutants B and C, when present individually, both have a negative impact upon IRES efficiency. The double mutation restores activity to above wild-type levels. This observation supports the predicted helix. It is notable that when present singly the mutations, which are

predicted to disrupt the G(n)NRA polyloop, and might be expected to replicate the effect of mutation A instead have the opposite effect. It appears that the helical region in question must be intact for the IRES to function most efficiently, but that the sequence of the terminal loop exerts an inhibitory effect. It is unlikely that the loop sequence could maintain its inhibitory influence when the supporting helix is disrupted, as the sheared G/A pair and subsequent loop structure is unlikely to form unless adjacent to another base pair. Thus the repressive effects of mutants B and C are probably partially masked by the activating effect of terminal loop disruption.

Mutants F reduces IRES efficiency, albeit to a lesser extent than B or C, suggesting that this helical element is present *in vivo*, and is required for the proper function of the IRES.

4.7 Pseudoknot Mutations

The most striking features of domain 1 are the two pseudoknot motifs. Both are completely conserved among primates. Pseudoknot α is also conserved in the porcine sequence, pseudoknot β in cats. Neither is conserved in rodent species. Pseudoknot α is the larger of the two, being seven nucleotides in length, the upper size limit for a pseudoknot-forming helix. Pseudoknot β is four nucleotides in length. Single and double mutations of the pseudoknot helices were generated as shown in Figure 4.vii.

As with mutants B and C, and F and G, the single mutations are predicted to disrupt helices, the double mutants to restore them. Mutants H and I respectively alter 5' and 3' components of pseudoknot α , while J and K alter pseudoknot β .

Mutants H and I in isolation both significantly enhance IRES function, and the double mutant has an activity that is close to the wild-type. These observations support the prediction that helix α will form, and show that the presence of pseudoknot α is inhibitory to IRES function. This is in contrast to the pseudoknots present in the IRESs of HCV and pestiviruses, which are necessary for IRES function.

Mutant J also enhances IRES function, albeit to a much lesser extent than either mutant H or I, whilst neither mutant K nor the double mutant are discernably altered in activity. Thus the existence of pseudoknot helix β is not supported by these data. However, given that the formation of helix α might be expected to bring the components of helix β into apposition, and that the presence of helix α (the larger and more 5' of the two pseudoknot helices) might be expected to mask the effects of helix β mutation upon IRES activity according to the model outlined below, it was decided to retain helix β as a part of the secondary structure model.

The only other inhibitory helical segment that has been characterised in IRESs is stem-loop IV of HCV (Honda *et al.*, 1996). This small stem-loop element has a calculated stability

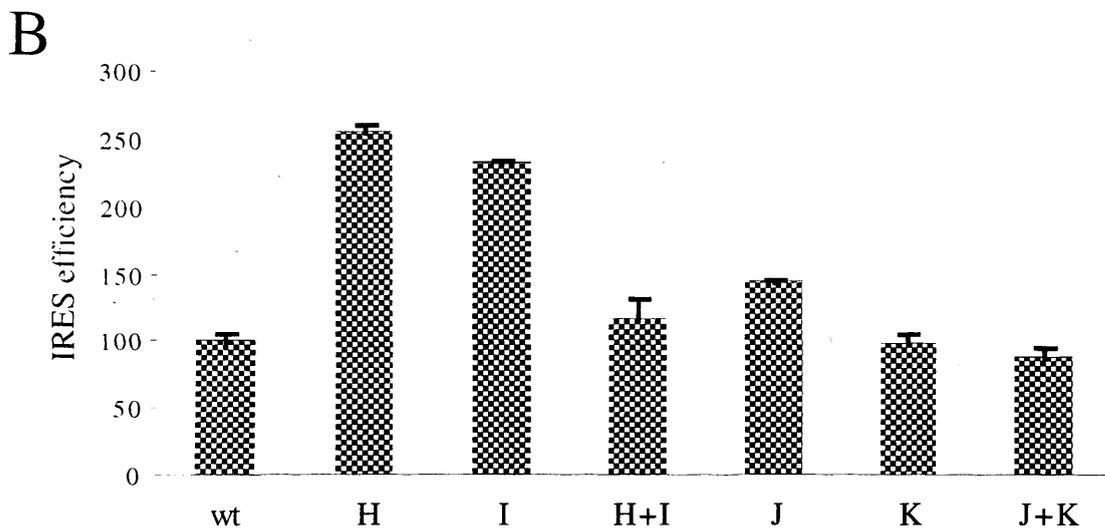
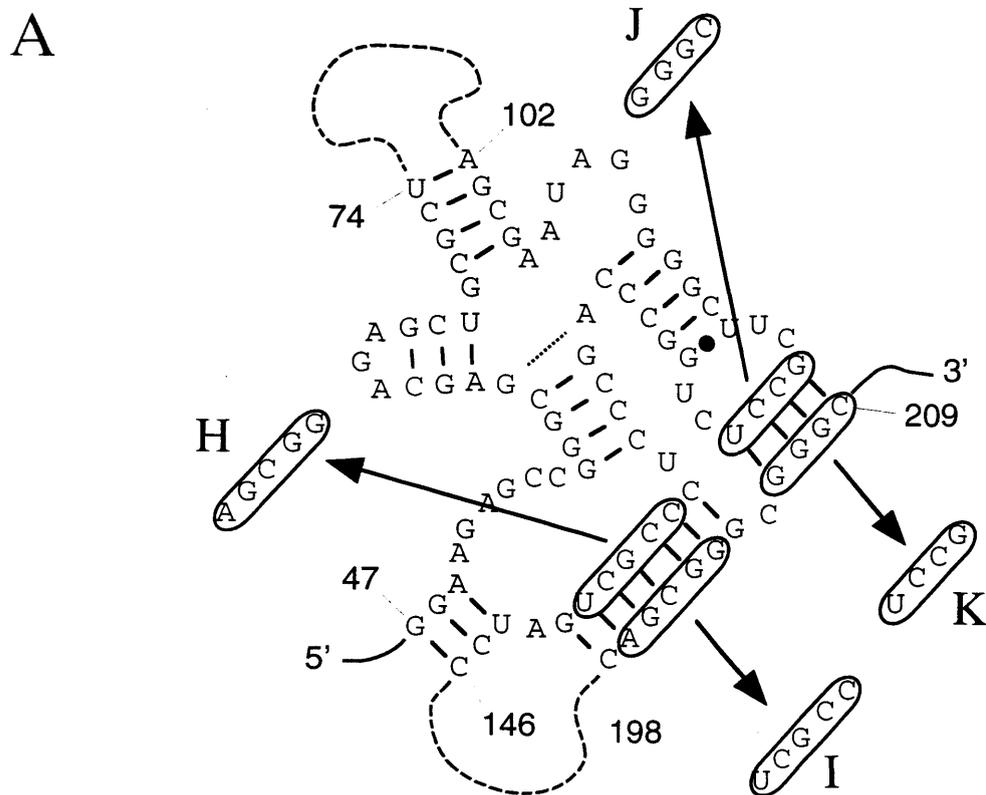


Figure 4.vii. Pseudoknot mutations. A: Base substitution mutations of pseudoknots α and β . Double mutants were designed to restore pairing potential. B: Effects of pseudoknot mutations upon IRES efficiency.

of -6.0 kCal/mol and contains the authentic initiation codon (Figure 1.v). Mutant studies show that its stability is inversely proportional to IRES efficiency. As monocistronic constructs bearing defective HCV IRESs with 5' deletions show no relationship between stability of this element and efficiency of cap-dependent initiation, stem-loop IV only inhibits internally initiated translation. This is explained by suggesting that the element specifically interferes with the acquisition of the RNA by the ribosome ("landing"), rather than stalling the 40S subunit on the message.

The situation is at least analogous to the effect of a hairpin positioned very close to the 5' cap structure of a monocistronic message. When positioned 12 nt from the cap, a stem-loop with a calculated stability of -30 kCal/mol was seen to inhibit 40S ribosomal subunit landing (Kozak, 1989). When placed 52 nt from the 5' end, it was readily traversed by the scanning ribosome. The cut-off points for stability and spatial separation at which such structure no longer presents a barrier to ribosome binding have not been empirically determined. However, given that ribosomes are capable of initiating translation at start codons positioned directly adjacent to the 5' cap, the spatial limit is likely to coincide with the distance between the P-site and the 3' edge of the mRNA binding track, between 12 and 17 nt (Kozak, 1997; Pestova *et al.*, 1998a).

A similar situation exists in the *c-myc* IRES; the 3' component of pseudoknot α is positioned at most 21 nt downstream of the ribosomal landing site as determined by the AUG insertions. If the conserved CUU codon is the actual insertion site, this distance is reduced to 9 nt. Thus the pseudoknot helices, having a calculated combined stability of -21 kCal/mol, might be expected to present a significant barrier to 40S subunit binding.

Furthermore, in the presence of an intact pseudoknot α , there is an inadequate length of single-stranded RNA for a 40S subunit to bind. The length of single-stranded RNA between

the base of domain 1 and pseudoknot α is 21 nucleotides. The mRNA binding track can be divided into three portions: 5' of the P-site (11 nt (Kozak, 1977; Rijnbrand *et al.*, 1997)), the P-site itself (3 nt) and 3' of the P-site (12-17 nt (Kozak, 1997; Pestova *et al.*, 1998a)), making a total *minimum* of 26 nt. Thus, the single-stranded region is at least 5 nt too short to accommodate the 40S subunit. According to this model, the ablation of pseudoknot α is likely to enhance IRES activity to a greater extent than the ablation of pseudoknot β , as it will not only reduce the stability of the proposed helices, it will also increase the length of ssRNA available to be bound by the 40S subunit.

4.8 Mutations of the Spacer Region

The region separating domains 1 and 2 of the IRES is enigmatic. It is generally solvent-exposed, seemingly supporting very little structure. It is not especially well conserved at the sequence level, and seems unlikely to be a region of crucial importance to IRES function. Yet it is in this region that the multiple myeloma-correlated C255U mutation lies. This point mutation activates the IRES, alters the profile of proteins bound by the IRES (Paulin *et al.*, 1998), and is predicted to favour the formation of a small stem-loop with a stability of -0.7 kCal/mol two nucleotides upstream of domain 2.

Three mutations were designed in an effort to determine which role(s) this mutation was playing in IRES activation (Figure 4.viii). Mutant E is predicted to favour the formation of small stem-loop in the same position as that favoured by C255U mutation. The only difference between the predicted structures of mutant IRES E and the C255U mutant IRES is the identity of the base pair closing the loop, resulting in a stem-loop with a stability of -2.1 kCal/mol. Mutant L was designed to slightly reduce the spacing between domains 1 and 2. Mutant M was designed to shorten the spacer region to the same extent as the C255U-associated stem-loop stacking onto domain 2.

The activating effect of mutant E supports the predicted structural consequences of the C255U mutation. The greater enhancement seen in the presence of mutant E than C255U might reflect the slightly greater stability of the mt E hairpin.

The activating effect of mutation L shows that a slight reduction of the spacing between domains 1 and 2 (less than that caused by the formation of the C255U-associated stem-loop) can also replicate the activating effect. The negative impact of mutation M rules out the idea that the up-regulating effect of C255U is due to a reduction in spacing caused by stacking of helices.

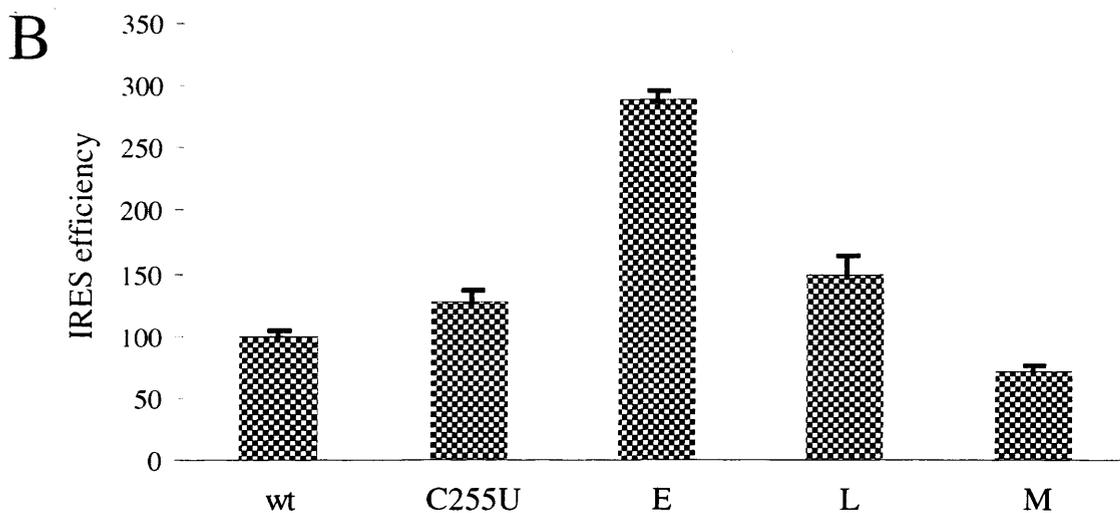
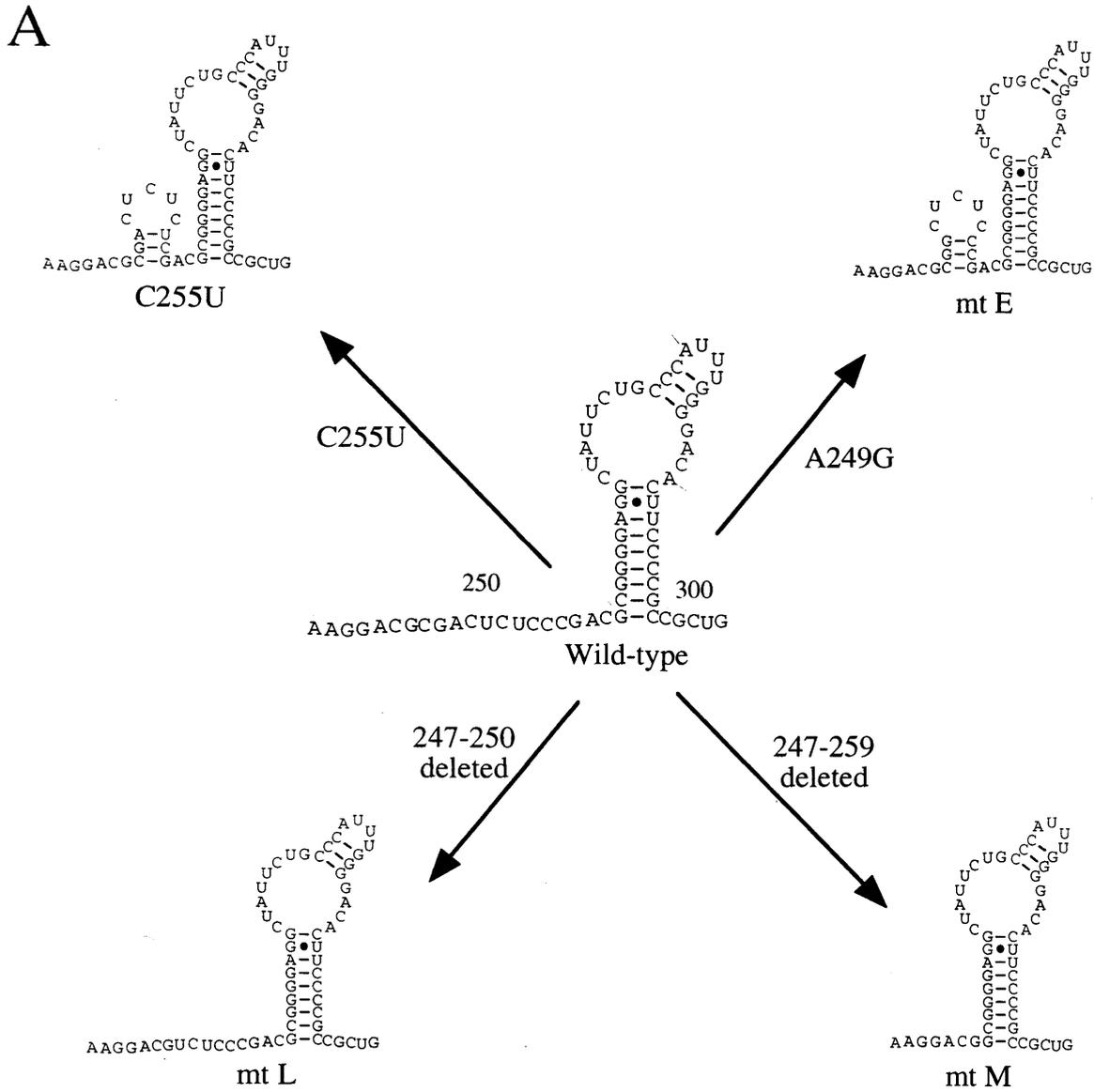


Figure 4.viii Mutations of the spacer region. A: Mutants synthesized: C255U is associated with the human neoplasm multiple myeloma, mt E replicates the predicted

Chapter 5

In vitro Analysis

5.1 Introduction

Although *c-myc* has only recently been shown to contain an IRES, other effects of its 5' UTR upon translation have previously been studied. In rabbit reticulocyte lysate, both murine and human *c-myc* transcripts bearing exon 1 sequences are translated significantly less efficiently than those lacking the majority of the 5' UTR (Darveau *et al.*, 1985; Parkin *et al.*, 1988). Deletion analysis revealed that this 5' UTR-mediated translational repression is a property of the entire region rather than a specific motif (Parkin *et al.*, 1988). This repression is presumably due to the structural elements within the UTR impeding the progress of scanning ribosomes.

A comparison of the distribution of endogenous *c-myc* polysomes in Burkitt's lymphoma cell lines demonstrated that exon 1 sequences do not affect the *in vivo* translational efficiency of *c-myc* mRNAs (Nilsen and Maroney, 1984) in that system. Furthermore, exon 1 does not inhibit the translation of exogenously expressed *c-myc* mRNAs in cultured cell lines or in HeLa cell extracts (Butnick *et al.*, 1985; Parkin *et al.*, 1988). Nevertheless, sequences from both the murine and human exon 1 repress the translation of heterologous reporter mRNAs micro-injected into *Xenopus* oocytes (Fraser and Browder, 1995; Parkin *et al.*, 1988). Thus, the effect of the *c-myc* 5' UTR on translation initiation depends on the assay system. This may reflect the competence of these systems to translate mRNAs with structured leader sequences. Alternatively it has been suggested that the translation of *c-myc* mRNAs may require a non-canonical *trans*-acting factor present in cultured cells but lacking in reticulocyte lysate and *Xenopus* oocytes (Parkin *et al.*, 1988).

All assays of the *c-myc* IRES described thus far have involved transfection of DNA constructs into the nucleus of the cell, where transcription, RNA processing and export to the cytoplasm follow. It has been shown that if circular or bicistronic RNAs bearing the *c-myc* 5' UTR are transfected into the cytoplasm, the IRES is inactive (Carter *et al.*, 1999; Stoneley, 1998). Moreover, transfection of a T7 promoter-bearing DNA constructs into an engineered vaccinia virus-infected cell line that constitutively expresses T7 RNA polymerase in the cytoplasm results in no IRES activity (Stoneley *et al.*, 2000b). These observations have led to the formulation of the “nuclear event hypothesis”, which states that, for the IRES to function, the RNA must undergo some special experience that occurs solely in the nucleus. This event might be the binding of a specific protein factor that is required for IRES function and is restricted to the nucleus. Alternatively, it might be an RNA-processing event with an impact upon structure, such as a transient interaction with an RNA chaperone to favour the formation of a particular structure, or even the chemical modification of one or more bases.

The cricket paralysis virus IRES shows a similar pattern of activity in DNA/RNA transfection systems (P. Sarnow, personal communication) and yet it is active in a translation system derived from HeLa cytoplasm. This might possibly be due to contamination of the lysate used with nuclear factors. As with *c-myc*, the precise cause of this pattern of activity remains unclear.

It has been demonstrated that the *c-myc* IRES does not function in rabbit reticulocyte lysate. This is not an uncommon observation for viral IRESs. For example, the HRV IRES fails to function in RRL, but is active when RRL is supplemented with ribosomal salt wash (Brown and Ehrenfeld, 1979).

An *in vitro* assay is a pre-requisite for full mechanistic characterisation of the IRES. In particular, a cell-free system makes possible the identification of *trans*-acting factors by the

screening of subcellular fractions. *In vitro* experiments are also generally quicker to perform and conditions are much more easily manipulated. Thus attempts were made to develop a system in which the *c-myc* IRES could be assayed *in vitro*.

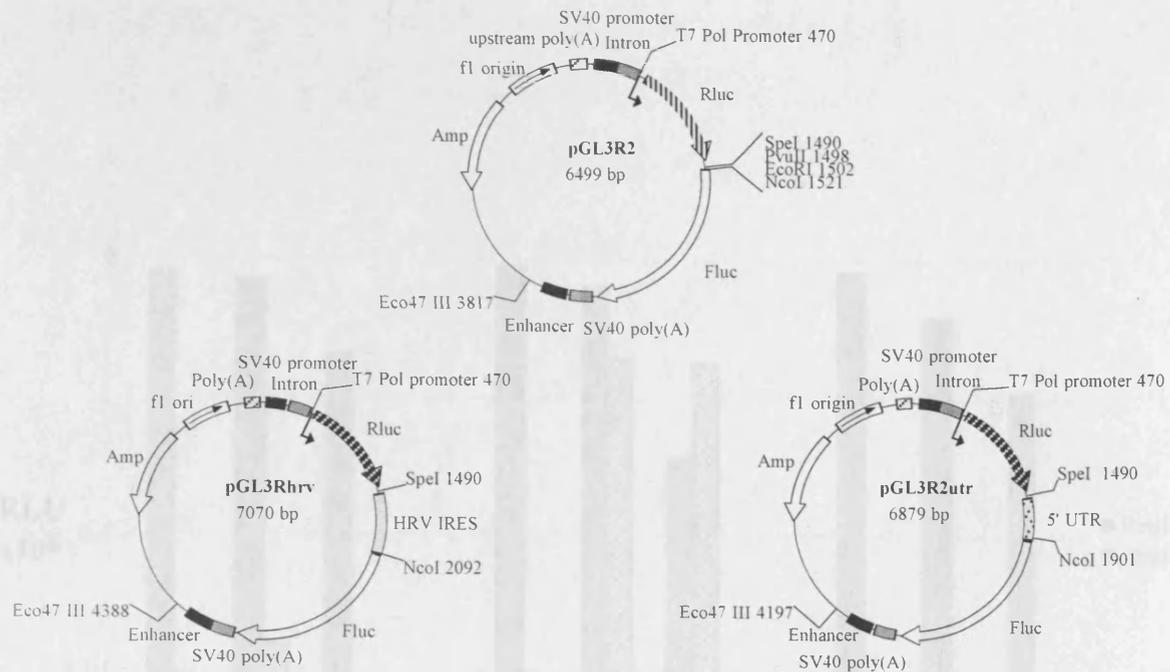
5.2 Translation in Rabbit Reticulocyte Lysate

It has been previously shown that the *c-myc* IRES fails to function in unsupplemented rabbit reticulocyte lysate (RRL) (Stoneley, 1998). Many IRESs of viral origin have been shown to be similarly inactive, but are activated by the addition of *trans*-acting factors. Ribosomal salt wash (RSW) is a rich source of *trans*-acting translation factors, containing both canonical translation initiation factors and other proteins that are associated with polysomes, which might reasonably be expected to include factors required for the function of cellular IRESs (Brown and Ehrenfeld, 1979). Accordingly, a ribosomal salt wash fraction was prepared from HeLa cells and added to RRL-based *in vitro* translation reactions (Figure 5.ii), programmed with *in vitro* transcribed RNAs (Figure 5.i).

As expected, the addition of 10% RSW has no significant effect upon the expression of either cistron from constructs that have no insert. The low level of firefly luciferase expression observed is due to re-initiation by ribosomes that fail to dissociate from the mRNA after completing translation of the first cistron, or “readthrough”. Following the addition of 37.5% RSW, *Renilla* luciferase expression is slightly diminished, perhaps due to the presence of salt in the RSW, or perhaps high levels of RNA-binding protein interfering with translation.

In the absence of RSW, there is practically no HRV IRES function. 10% RSW has a massively stimulating effect as *trans*-acting factors are supplied, and again 37.5% RSW has an inhibitory effect upon *Renilla* luciferase expression, but in this case the effect is greater than that observed when no IRES is present. This might be due to the sequestration of ribosomal subunits and/or canonical initiation factors by IRES-specific *trans*-acting factors, rendering them unavailable specifically for cap-dependent initiation.

In the absence of RSW, there is even less expression from the firefly cistron downstream of the *c-myc* IRES relative to the insert-free construct. This presumably reflects a



1. Linearize with *Eco47III*
2. Transcribe with T7 RNA pol
3. Purify RNA

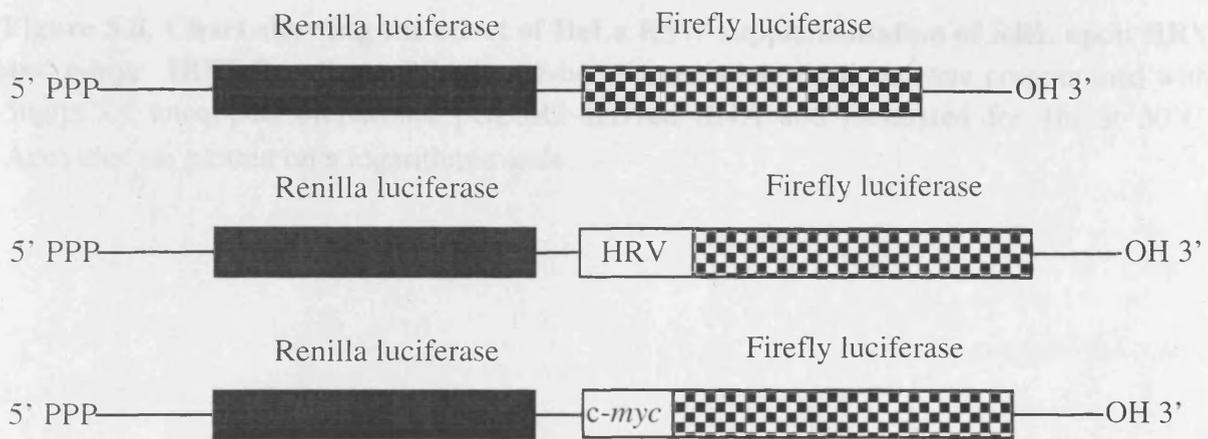


Figure 5.i. Scheme of synthesis of bicistronic RNAs from the pGL3R2 series of vectors. Uncapped bicistronic RNAs containing either no insert, the HRV IRES or the *c-myc* 5' UTR were generated.

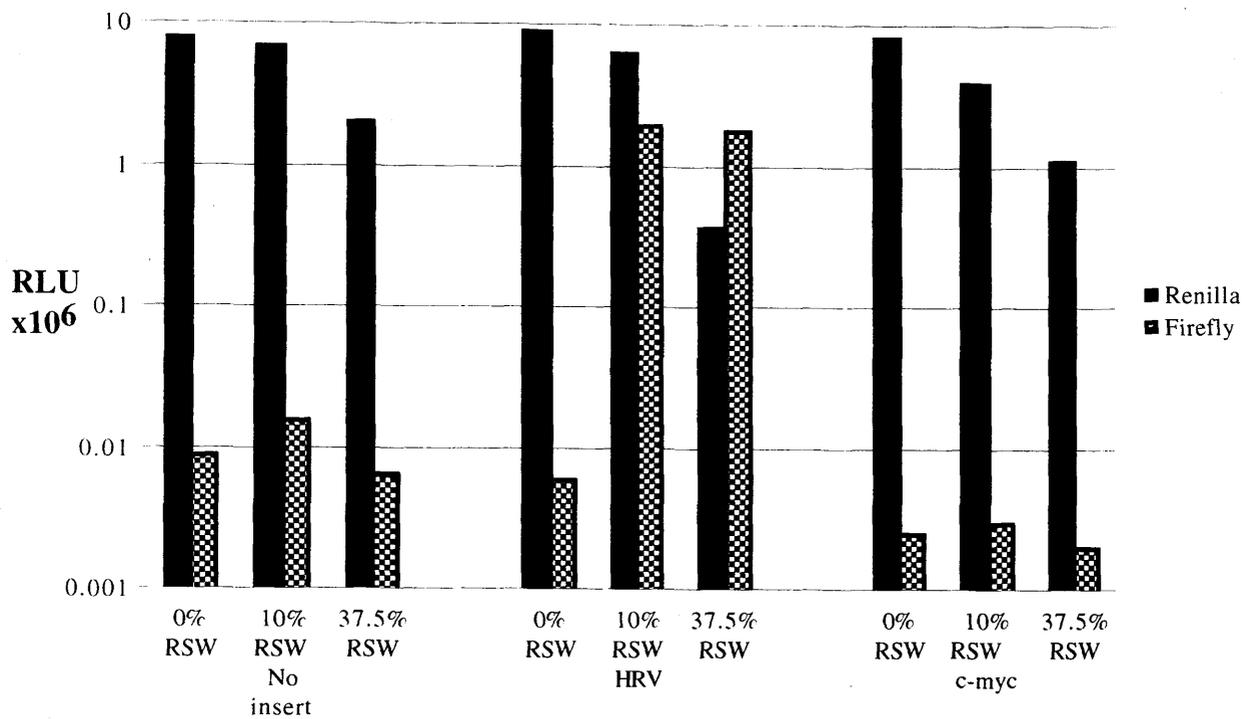


Figure 5.ii. Chart showing the effect of HeLa RSW supplementation of RRL upon HRV and c-myc IRES function. RRL lysate-based translation reactions were programmed with 5ng/ μ l of uncapped bicistronic pGL3R2-derived RNA and incubated for 1hr at 30°C. Activities are plotted on a logarithmic scale.

barrier effect of greater length and secondary structure upon readthrough ribosomes. Thus, addition of RSW fails to activate the *c-myc* IRES. This might be for a number of reasons: an inhibitory factor in RSW might be present in excess, the IRES might require not only human *trans*-acting factors but also human translation machinery, nuclear factors might be required, or the some aspect of the RNA structure might be defective in this *in vitro* system.

5.3 Translation in HeLa Cell Cytoplasmic Lysate

Despite the very high degree of conservation of ribosomal proteins and RNA, it is possible that a human cellular IRES would fail to interact correctly with ribosomes of, say, lapine origin. To test this, a translating extract of HeLa cells was prepared and programmed with bicistronic messages (Figure 5.iii).

Again, the activity of the HRV IRES is seen to rise upon the addition of RSW. The increase is less dramatic than that observed in the previous experiment since even unsupplemented HeLa extract contains significant quantities of the *trans*-acting factors enriched in the RSW fraction.

In unsupplemented extract, the firefly expression from the *c-myc* construct is 4-fold greater than that from the empty construct, although this drops to under 3-fold as RSW is added. In the previous experiment the firefly expression from the *c-myc* construct was 3.5-fold *less* than readthrough, an overall differential of 14-fold. If calculated as (firefly/*Renilla*), *c-myc* IRES activity in unsupplemented HeLa lysate is 600-fold greater than in unsupplemented RRL.

The observations are consistent with an IRES dependent upon human translation machinery for function, but which is inhibited by a factor enriched in HeLa-cell derived RSW. This is difficult to reconcile with the very high IRES activity seen *in vivo*, however. It seems equally likely that this increase in firefly expression from the *c-myc* construct might be due to a non-specific internal initiation activity inherent to the HeLa lysate. It is important to remember that all initiation described so far has proceeded in the absence of a 5' cap structure, demonstrating that even the upstream cistron is able to be translated in a cap-independent (but not necessarily end-independent) fashion.

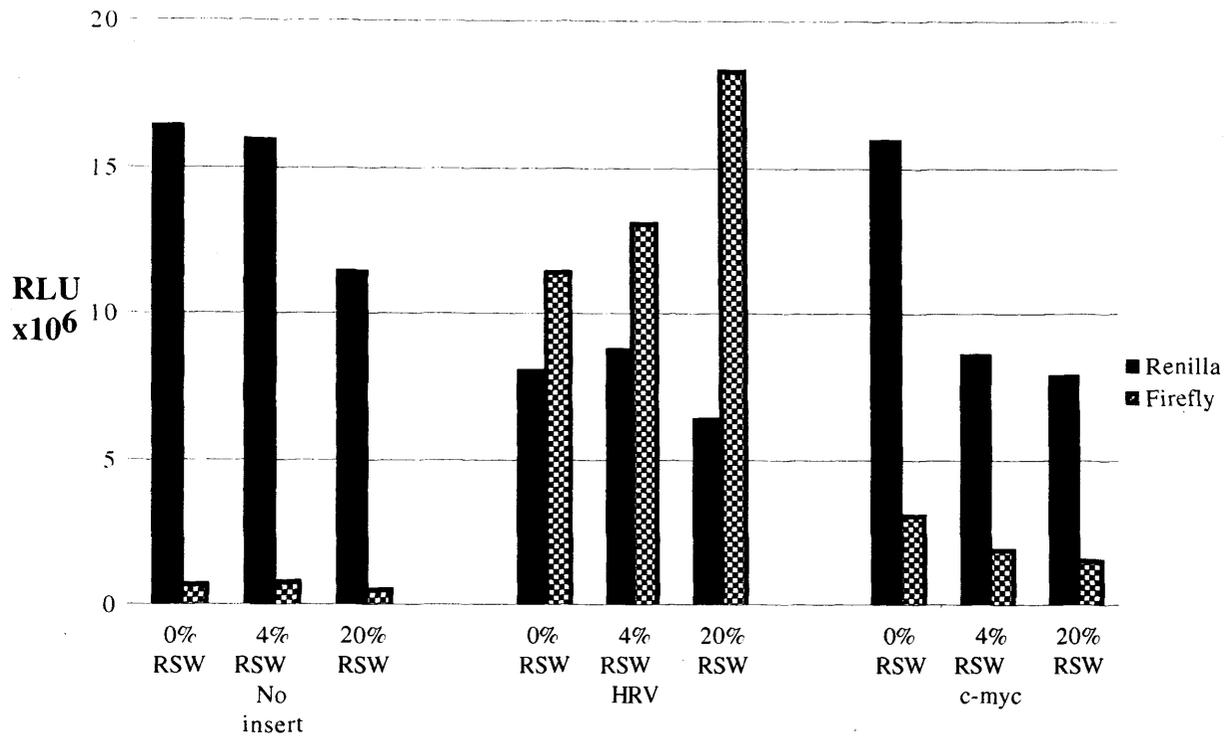


Figure 5.iii. Chart showing IRES activity in HeLa cytoplasmic extract with and without RSW supplementation. HeLa S10 lysate-based translation reactions were programmed with 2.5ng/ μ l of uncapped bicistronic pGL3R2-derived RNA and incubated for 2hr at 30°C.

5.4 Wild-type vs. mutant IRES activity *in vitro*

If the firefly luciferase expression described in the previous experiment is really due to the presence of the *c-myc* IRES activity, then a difference in activity would be expected between wild-type and mutant forms of the IRES that had been shown to have different activities *in vivo*. Therefore bicistronic RNAs containing the wild-type and C255U *c-myc* inserts were prepared and assayed in HeLa cytoplasmic extract (Figure 5.iv). Also, RNAs bearing the AUG mutants AUG5 and AUG6 were compared (Figure 5.v). These RNAs were all derived from the pRhpfM series of plasmids, and consequently bore hairpin structures between the *Renilla* cistron and the IRES insert (Figure 5.vi).

The chart reveals that the ratio between firefly and *Renilla* luciferase expression from both wild-type and C255U mutant constructs is not significantly influenced by the concentration of RNA present in the reaction, so there is no evidence for saturation of any IRES activity. Nor is there any significant difference in expression levels from either cistron between the wild-type and mutant RNAs, as is seen when the IRES is active *in vivo* (Figure 4.viii). Firefly luciferase expression is generally repressed with regard to prior experiments since the presence of the inter-cistronic hairpin is reducing ribosomal read-through.

As described in chapter 4, the AUG5 mutant lies upstream of the ribosomal entry window, and does not have a negative impact upon IRES efficiency *in vivo*, while the AUG6 mutant lies downstream, and abrogates IRES efficiency almost completely. If the IRES was active in this *in vitro* system, then the firefly expression from the AUG5 construct would be expected to exceed that from the AUG6 construct to a degree commensurate with the activity of the IRES. In fact, it is clear that firefly expression levels are negligible in both constructs. This shows that the low levels of firefly expression seen in experiments so far are dependent upon the presence of an inter-cistronic spacer that contains no start codons whatsoever,

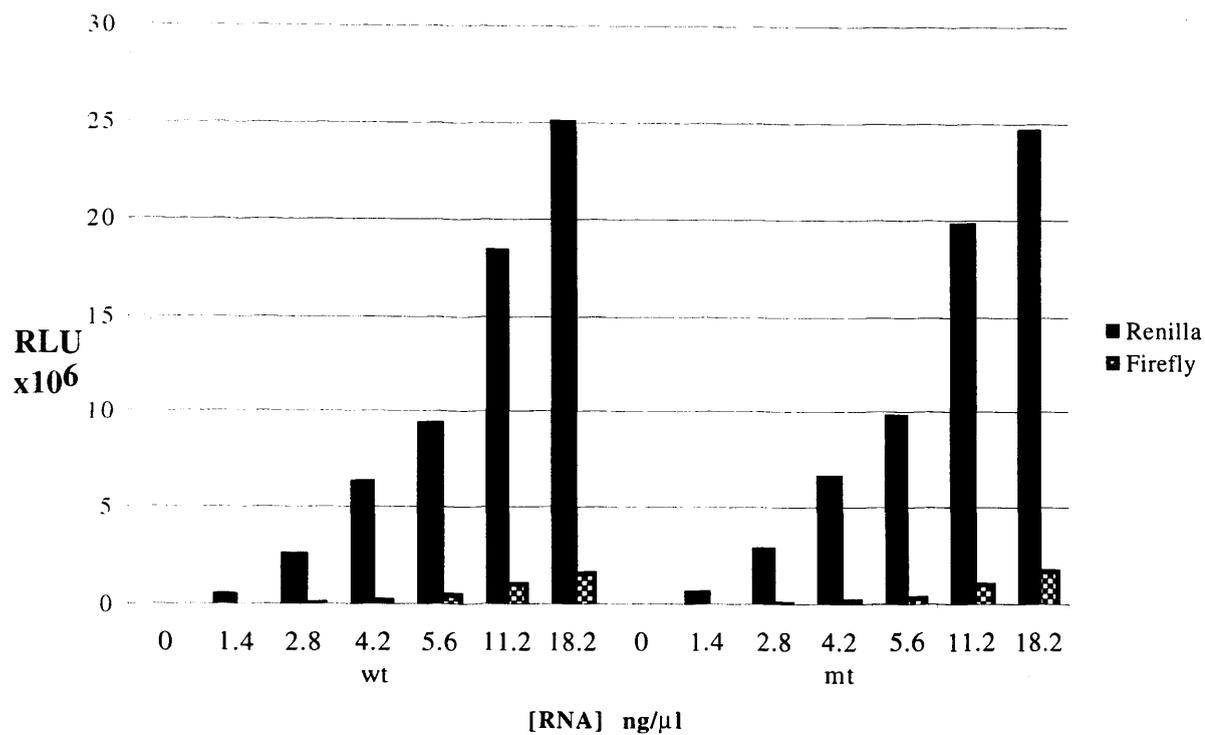


Figure 5.iv. Activities of the wild-type and C255U mutant IRES-bearing constructs in HeLa cytoplasmic extract. HeLa S10 lysate-based translation reactions were programmed with varying quantities of uncapped hairpin-containing bicistronic RNA and incubated for 1hr at 30°C.

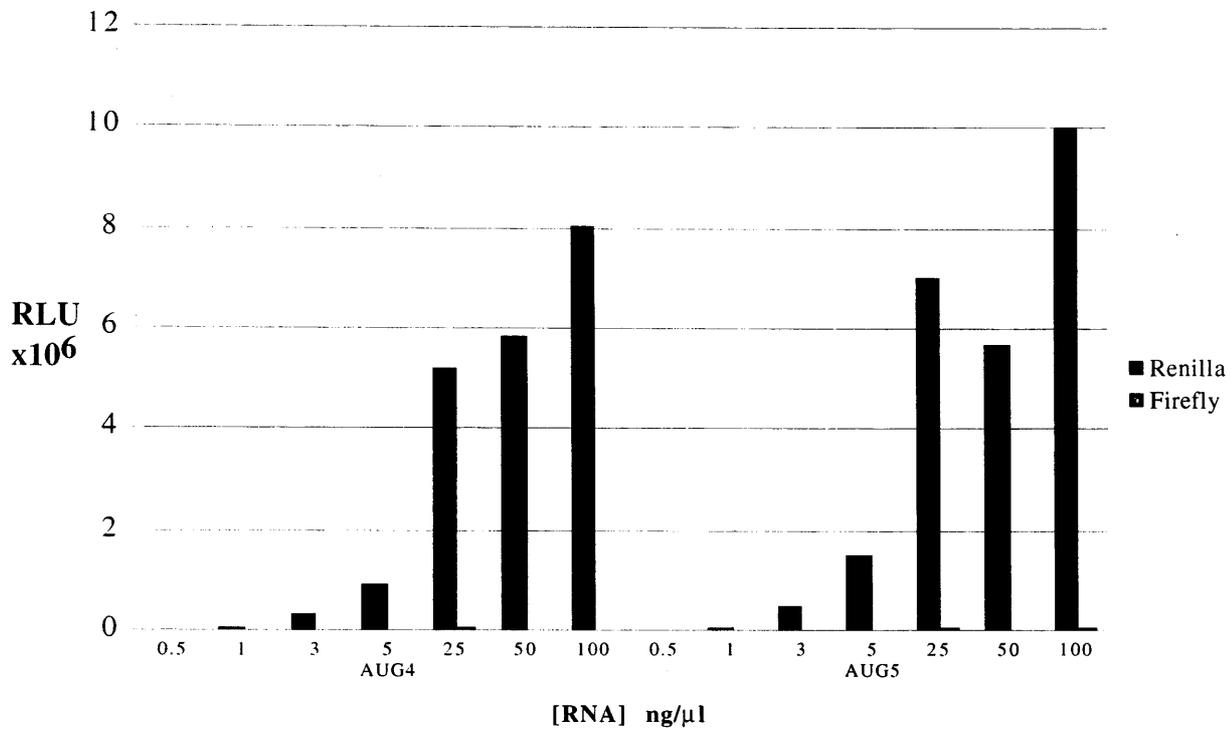
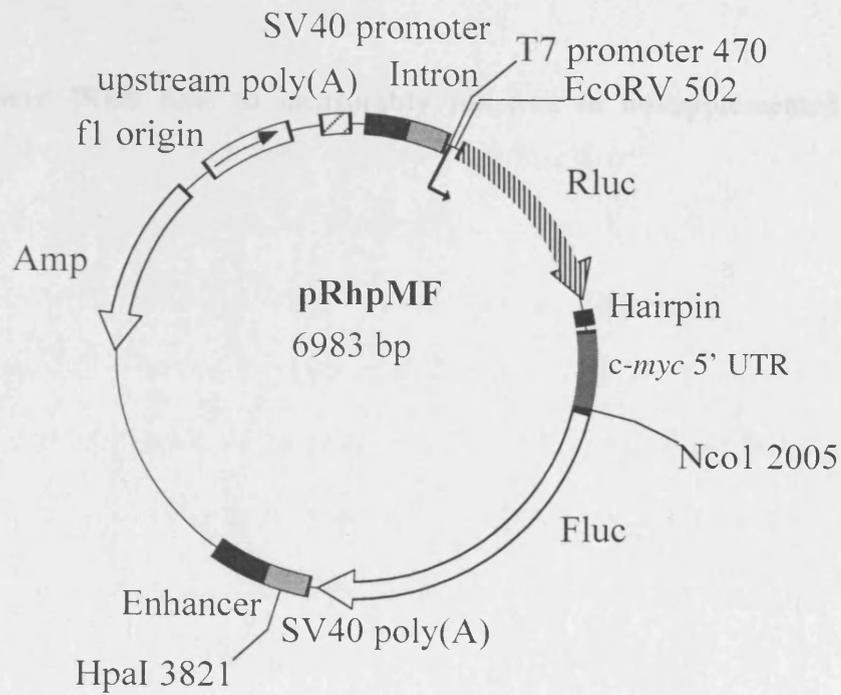


Figure 5.v. Activities of the AUG4 and AUG5 mutant IRES-bearing constructs in HeLa cytoplasmic extract. HeLa S10 lysate-based translation reactions were programmed with varying concentrations of uncapped RNA derived from plasmids pRhpFM+AUG4 and pRhpFM+AUG5, and incubated for 1hr at 30°C.



1. Linearize with *Hpa* I
2. Transcribe with T7 RNA pol
3. Purify RNA

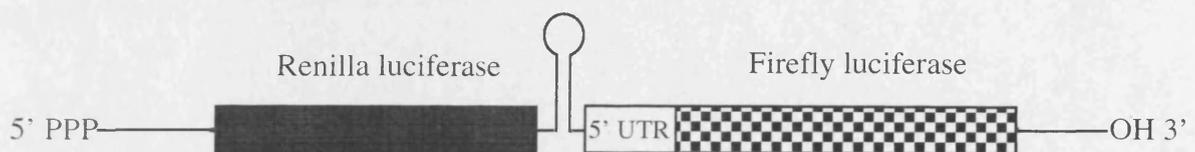


Figure 5.vi. Scheme of synthesis of hairpin-containing bicistronic RNAs from the pRhpMF series of plasmids. Mutant forms of the plasmid were created by the strategy shown in Figure 4.ii.

regardless of their position relative to the ribosome entry window. Thus this expression is due to ribosomes reading through from the first cistron, and the enhancement of “IRES activity” seen in Figure 5.iii is due solely to a readthrough-enhancing effect of the *c-myc* IRES insertion.

Thus the *c-myc* IRES fails to measurably function in unsupplemented HeLa cytoplasmic lysate.

5.5 Translation in GM 2132 Cell Cytoplasmic Lysate

The translational control of Myc protein levels was first noted in cell lines derived from individuals suffering from multiple myeloma (Paulin *et al.*, 1996). In the light of more recent studies, it seems likely that the over-expression of *c-myc* in these cell lines is due to relative activation of the IRES. Experiments in which bicistronic constructs are transiently transfected into myelomatous cell lines show that the IRES is more strongly activated in these lines than in HeLa cells (A. Willis, personal communication). The highest levels of Myc protein were observed in the myelomatous plasma-cell line GM 2132.

This might reflect a difference in the *trans*-acting factors present in the cells, such as a greater concentration of an activating factor, or a reduction in the concentration of a repressor. Thus a translating extract was prepared from GM 2132 cells, and tested for the ability to drive IRES-dependent translation *in vitro* (Figure 5.vii).

Expression from both cistrons is reduced in GM 2132 lysate relative to HeLa cell lysate. It is interesting that minimum levels of expression are seen with a 25:75 mixture of HeLa and GM 2132 lysates, and there is no explanation for this phenomenon at this time. The replacement of HeLa with GM 2132 cytoplasm does not enhance the expression of firefly relative to *Renilla* luciferase from either wild-type or C255U IRES-bearing constructs; it is reduced to about a quarter in both cases. This shows that the GM 2132 extract is less capable of read-through initiated translation than the HeLa extract. The presence of a 5' cap has no significant activating effect upon the IRES.

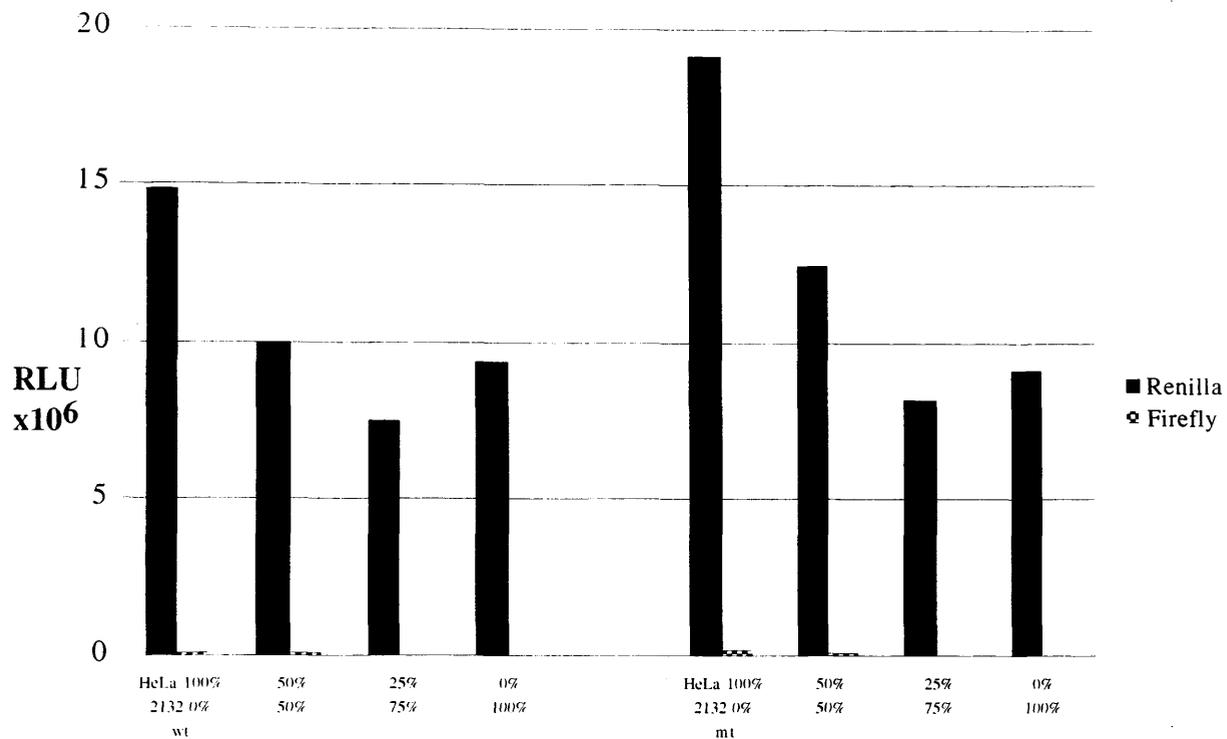


Figure 5.vii. Effect of transition from HeLa to 2132 cytoplasmic extract upon expression from wild-type and C255U IRES-bearing constructs. Translation reactions were programmed with 5ng/ μ l m7GpppG-capped RNA derived from plasmids pRhpfm and pRhpfmmt, and incubated at 30°C for 1hr.

5.6 Addition of Specific Protein Factors

Myc protein is known to regulate its own expression at the level of transcription, and has also been shown to have at least the potential to bind RNA. Thus it was hypothesized that the *c-myc* IRES might interact with the protein in a way that stimulated its own translation. In order to test this hypothesis, TNT™ coupled transcription/translation reactions were programmed with the Myc-encoding plasmid pSKMΔ1 (Figure 5.viii) and, as a control, pRhpFM, which lacks a suitable viral RNA polymerase promoter. A HeLa cell derived translating extract was programmed with a polyadenylated bicistronic RNA bearing the wild-type *c-myc* IRES, and supplemented with varying proportions of control and Myc protein-containing coupled transcription/translation reactions (Figure 5.ix). Polyadenylated RNA was transcribed from the plasmid pSP64RUTRL Poly (A) (Figure 5.x).

The addition of Myc protein is not sufficient to activate the *c-myc* IRES. Nor does the presence of a polyA tail have any significant effect upon IRES function.

Several researchers have identified functional interactions between IRESs and heterogeneous nuclear ribonucleoproteins, specifically hnRNP I or PTB (Hunt and Jackson, 1999), hnRNP L (Hahn *et al.*, 1998) and hnRNP C (Sella *et al.*, 1999). As hnRNP A1 was available and shows interesting RNA-binding properties, it was assayed for its effect upon an *in vitro* translation reaction. A 10-fold molar excess of a mixture of varying proportions of wt hnRNP A1 and an RNA-binding defective mutant were added to a HeLa cell cytoplasmic lysate programmed with polyadenylated bicistronic mRNA bearing the *c-myc* IRES (Figure 5.xi).

The wild-type form of hnRNP A1 appears to stimulate translations of both cistrons, relative to the mutant form. However the ratio of firefly:*Renilla* luciferase expression is not significantly altered, and the IRES remains inactive.

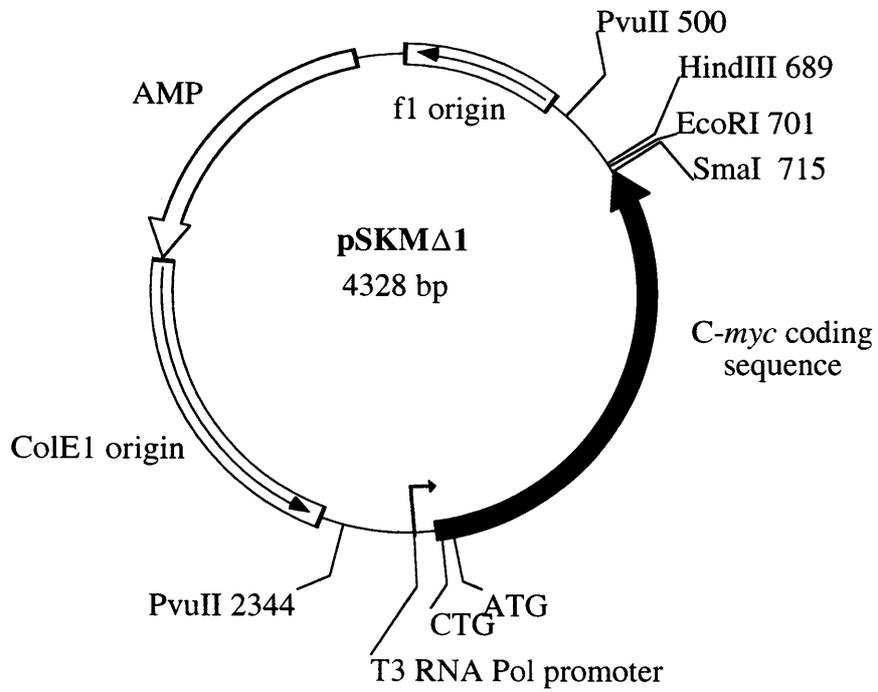


Figure 5.viii. Plasmid pSKMΔ1. The plasmid was linearized with *SmaI* before addition to the coupled transcription/translation reaction.

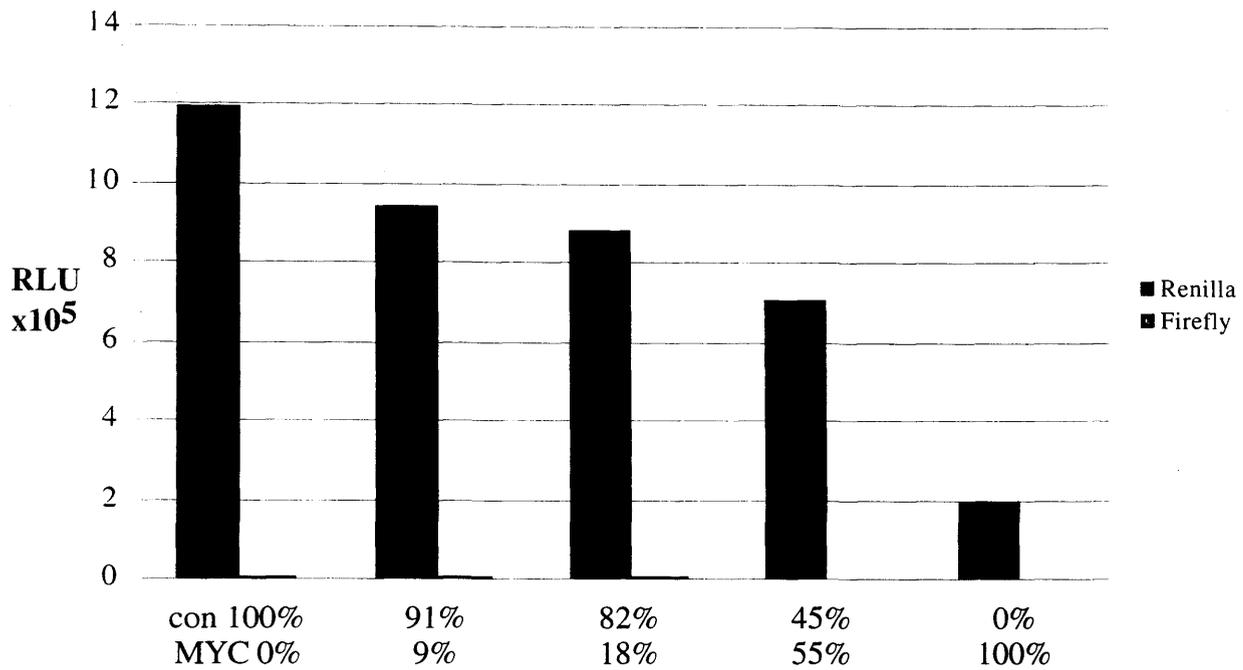
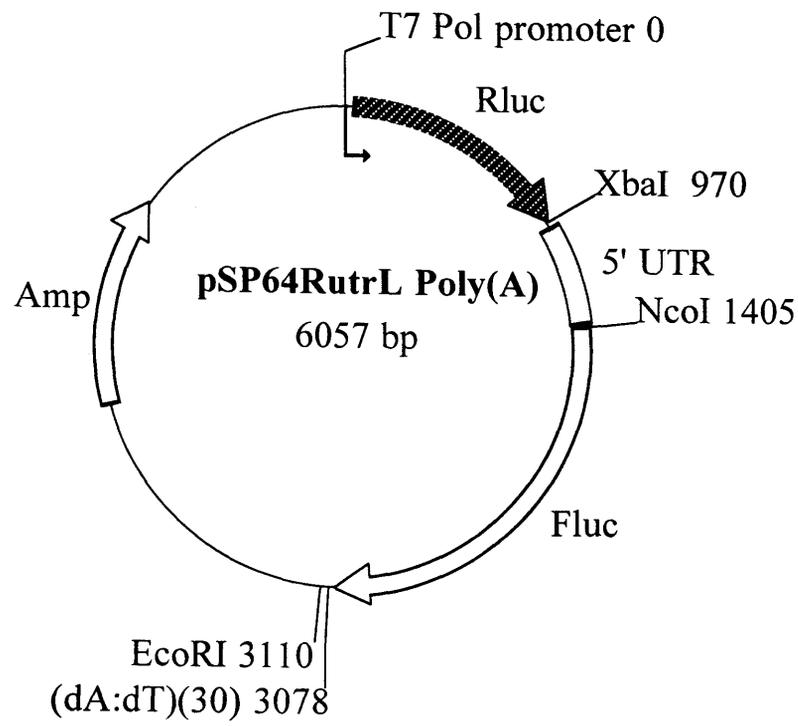


Figure 5.ix. Effect of supplementation of HeLa cytoplasmic translation extract with MYC protein upon *c-myc* IRES function. HeLa translation reactions were programmed with 4ng/ μ l uncapped polyadenylated RNA derived from the plasmid pSP64RUTRL Poly(A) and incubated at 30°C for 90 minutes. TNT™ mix containing variable proportions of Myc- and control plasmid-programmed reactions made up 22% of the final reaction volume.



1. Linearize with *EcoRI*
2. Transcribe with T7 RNA pol
3. Purify RNA

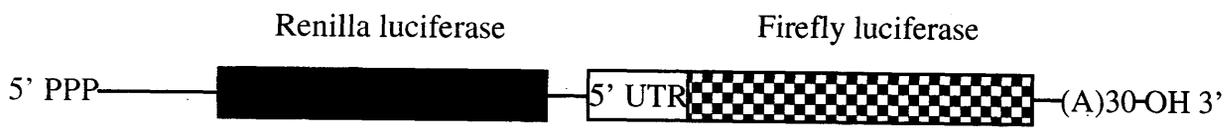


Figure 5.x. Scheme of synthesis of polyadenylated bicistronic RNA from the plasmid pSP64RUTRL Poly(A).

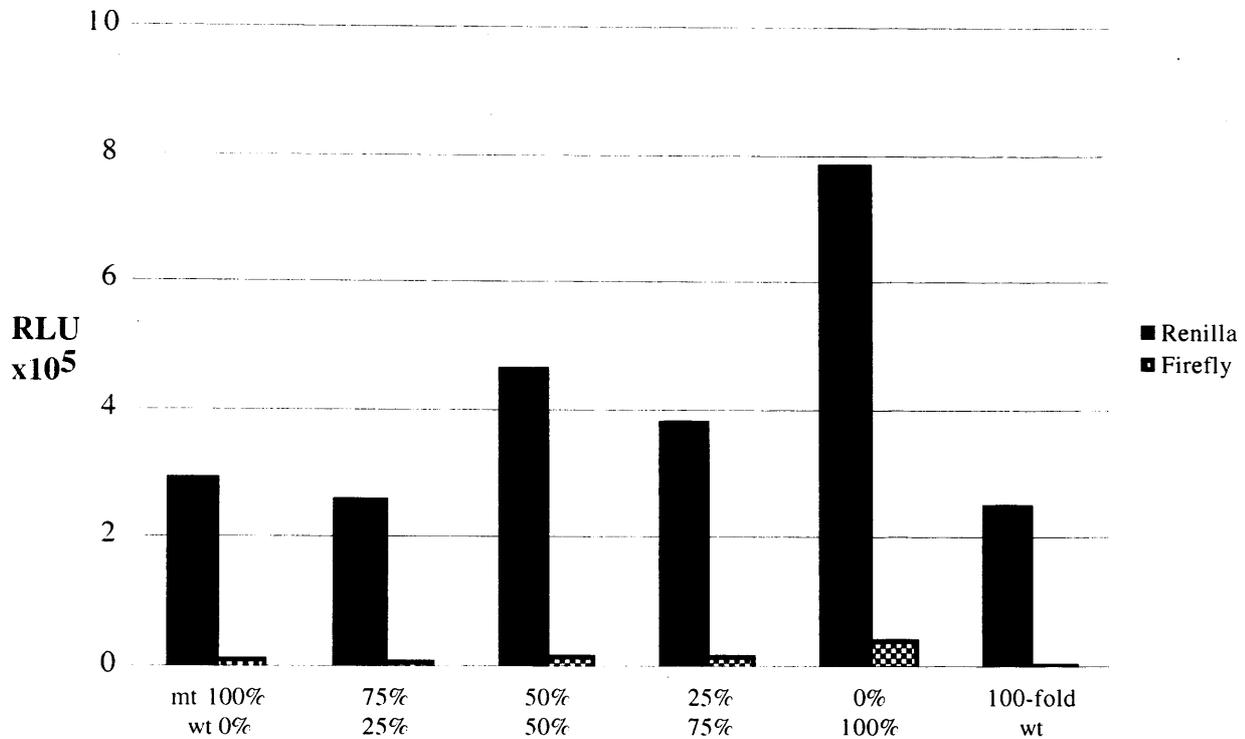


Figure 5.xi. Effect of supplementation of HeLa cytoplasmic translation extract with wild-type and mutant forms of hnRNP A1 upon *c-myc* IRES function. HeLa translation reactions were programmed with 4ng/ μ l uncapped polyadenylated RNA derived from the plasmid pSP64RUTRL Poly(A) and incubated at 30°C for 90 minutes. The last column shows the effect of the addition of a 100-fold molar excess of the wild-type protein.

5.7 Re-folding of IRES RNA

In some systems, translation from *in vitro* transcribed viral IRESs is enhanced by heating the RNA immediately before adding to the translation reaction. This has the effect of unfolding all secondary structure, and upon cooling the RNA refolds into a more favourable conformation. The effect may be due to the IRES RNA assuming incorrect conformations during its handling between transcription and translation, or it may reflect a requirement for the *trans*-acting factors to be assembled into the IRES during the folding process. In order to determine whether the *c-myc* IRES could be activated in this way, uncapped bicistronic RNA derived from the pRhpMF plasmid was heat denatured at 85°C for 90 min before addition to a translating HeLa cytoplasmic extract (Figure 5.xii).

In this case, RNA re-folding has no activating effect upon the IRES. The diminished translation from re-folded RNA may be due to partial degradation of the template during heat-treatment.

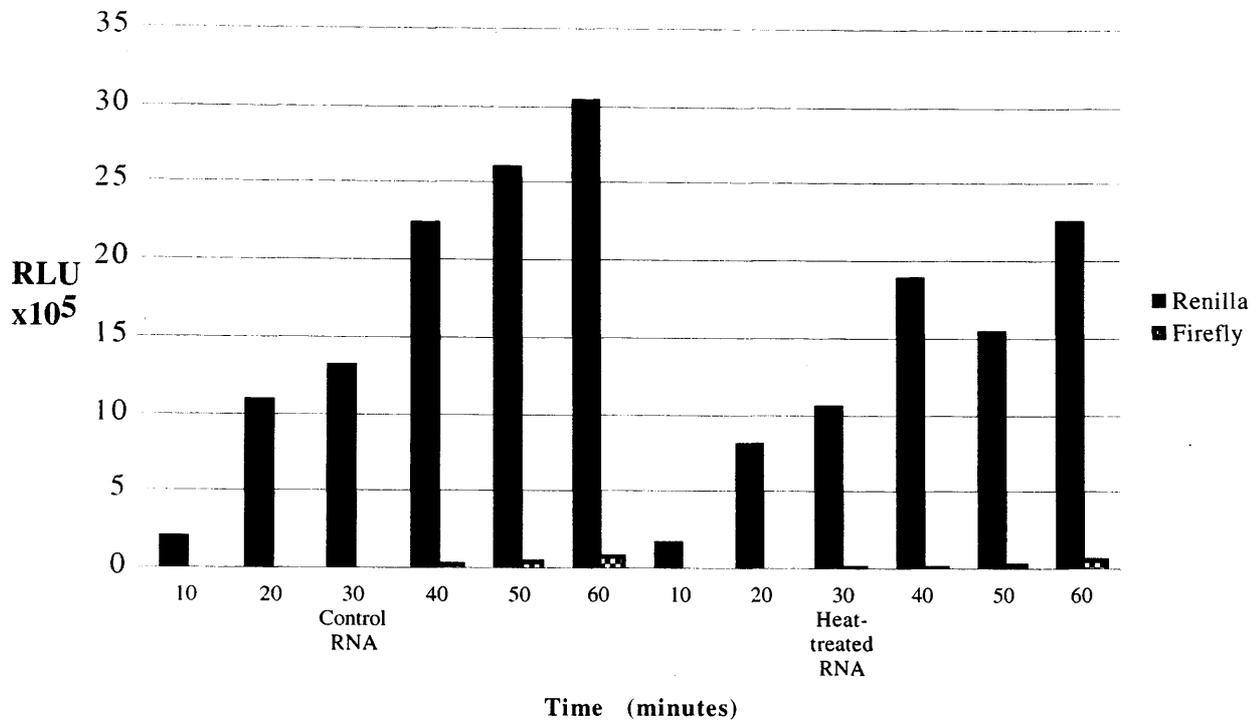


Figure 5.xii. Effect of heat-denaturation of template RNA upon *c-myc* IRES function. cytoplasmic extract. HeLa S10 lysate-based translation reactions were programmed with 5 ng/ μ l of uncapped RNA derived from plasmid pRhpfM.

5.8 Addition of Nuclear Extracts

One possible explanation for the inactivity of the *c-myc* IRES in the systems described so far is the provided by the hypothesis that the IRES RNA undergoes a nuclear experience that primes it for function after passage thorough the nuclear pore. If this nuclear event was simply the binding of a factor abundant in the nuclear compartment before export, then the supplementation of an *in vitro* translation reaction with nuclear extract might be expected to reconstitute IRES function. Accordingly, a commercially available HeLa nuclear extract with *in vitro* splicing activity (HNE) was added to the HeLa cytoplasmic extract (Figure 5.xiii).

The addition of HNE has a generally inhibitory effect upon translation, and no significant enhancement of IRES efficiency is observed. This suggests that either the nuclear event is not simply the binding of a nuclear factor, or that this binding step is not reconstituted in the system used, or that for some reason this extract lacks the appropriate factor (s).

It is possible that the nuclear fraction of GM 2132 cells is responsible for their relatively high *c-myc* IRES activity. In order to test this hypothesis, and also to obtain nuclear extracts of alternative and precisely controlled provenance, two extracts of GM 2132 cell nuclei were prepared. One was prepared by salt disruption of sucrose gradient-purified nuclei, and is referred to as nuclear salt wash (NSW). The other was prepared by a protocol used to make high-activity transcription extracts (Shapiro *et al.*, 1988) and is referred to as HTE. Their respective effects upon reporter gene expression are shown in Figures 5.xiv-xv. In these experiments, polyA-tail bearing RNAs derived from pSP64UTR were used.

None of the nuclear extracts added stimulate translation of the firefly cistron. Instead, in all cases some component of the nuclear extracts is seen to repress translation in general at higher concentrations. This might be due to ribonuclease activity destroying the templates, or perhaps competition between translation initiation factors and nuclear RNA-binding proteins.

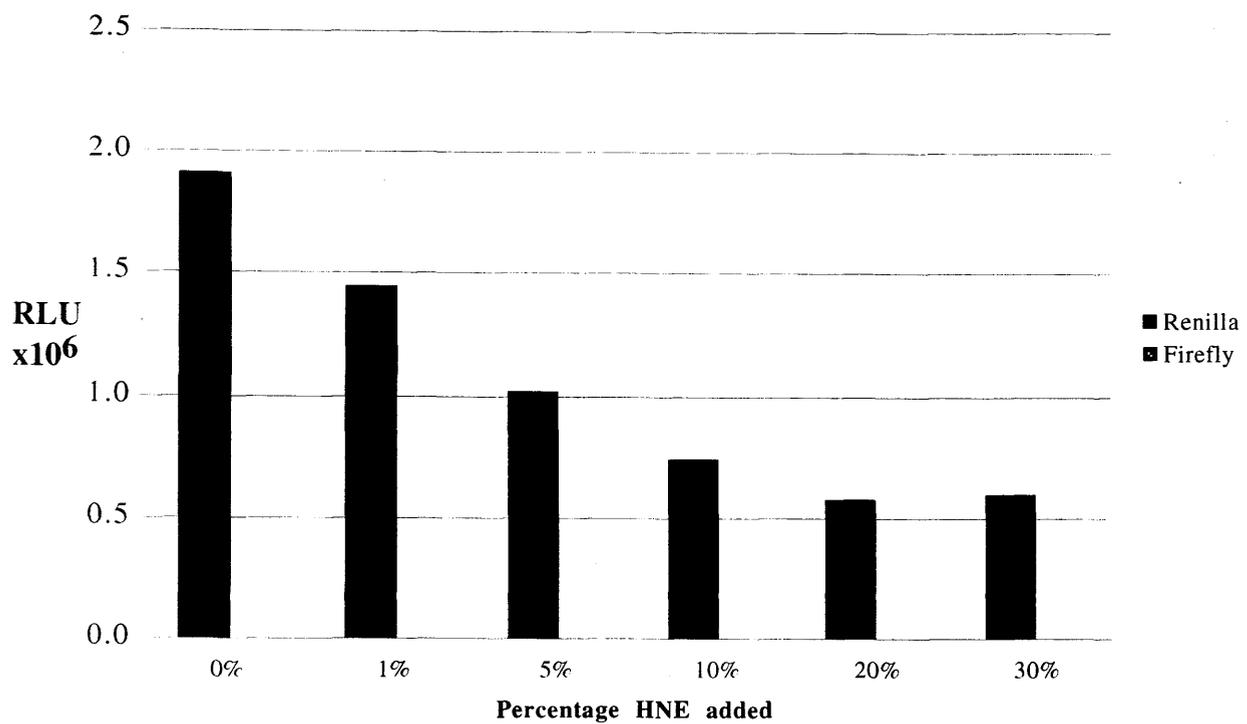


Figure 5.xiii. Effect of supplementation of HeLa cytoplasmic translation extract with HeLa nuclear extract upon *c-myc* IRES function. HeLa S10 lysate-based translation reactions were programmed with 5ng/ μ l of uncapped hairpin-containing bicistronic RNA derived from plasmid pRhpFM and incubated for 90 minutes at 30°C.

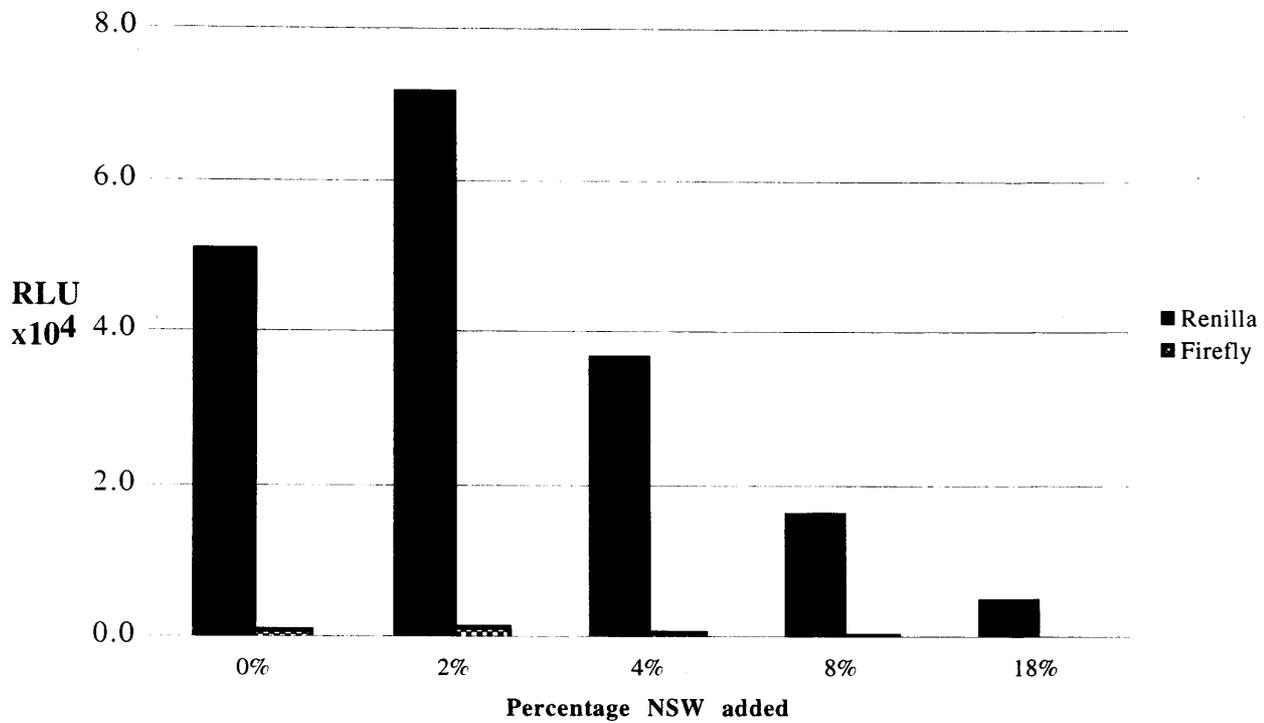


Figure 5.xiv. Effect of supplementation of HeLa cytoplasmic translation extract with 2132 cell nuclear salt wash extract upon *c-myc* IRES function. HeLa translation reactions were programmed with 4ng/ μ l uncapped polyadenylated RNA derived from the plasmid pSP64RUTRL Poly(A) and incubated at 30°C for 1 hour.

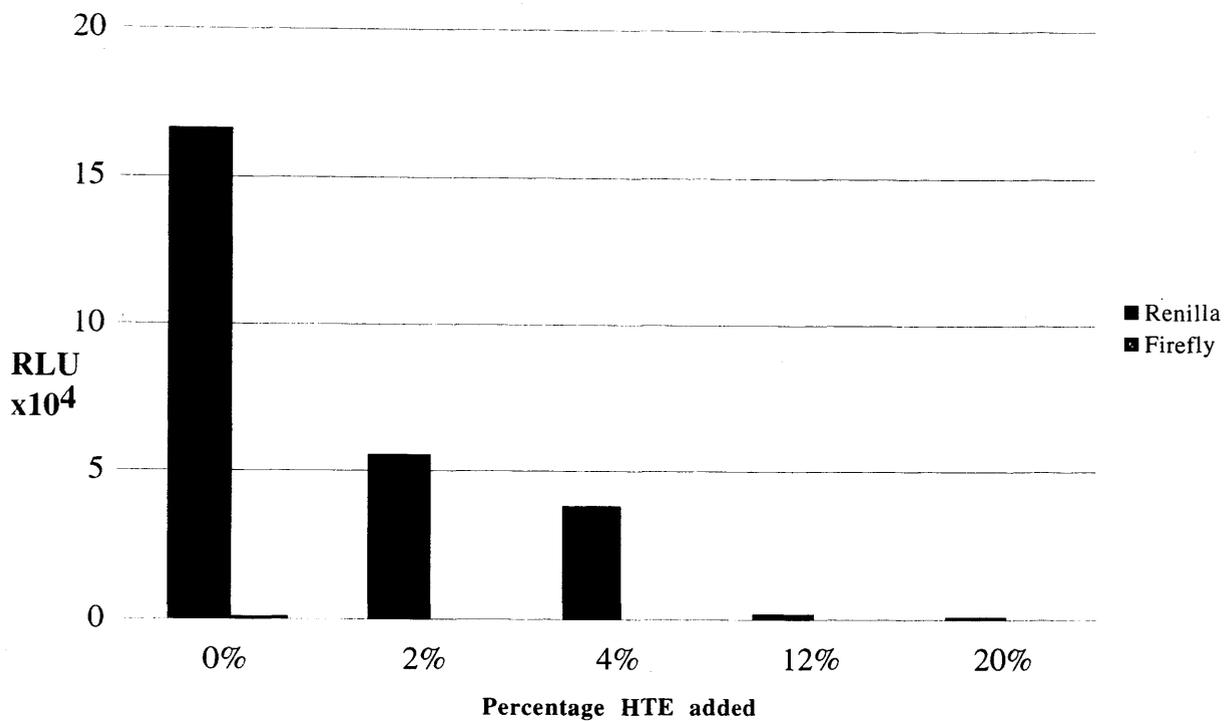


Figure 5.xv. Effect of supplementation of HeLa cytoplasmic translation extract with 2132 high-activity transcription extract upon *c-myc* IRES function. HeLa translation reactions were programmed with 5ng/ μ l uncapped polyadenylated RNA derived from the plasmid pSP64RUTRL Poly(A) and incubated at 30°C for 2 hours. Activities are plotted on a linear scale.

The stimulating effect of the NSW lysate upon translation seen when present as 2% of the reaction volume is notable. This is probably due to a RSW-like activity from salt disruption of polysomes associated with the nuclear envelope.

If an IRES-specific factor or factors are present in any of these nuclear extracts, then either they are failing to interact, or their interaction is rendered futile by some other activity inherent to the unfractionated lysate.

The addition of nuclear extracts to translating systems is qualitatively vastly different to the addition of the traditional source of translation-specific factors, RSW. RSW is by definition derived from active or potentially active translation machinery components, and even factors that are present in RSW in tiny amounts are likely to retain the potential for proper function. Nuclear extracts are derived from a compartment that might ordinarily be considered to be inimical to translation, as it favours a completely different set of processes. So, if the nuclear event is merely the persistent binding of a nuclear factor, it is perhaps not too surprising that this step cannot be reconstituted by supplying a crude cocktail of nuclear proteins, many of which will interact with RNA in a decidedly “nuclear” fashion. Still, the hypothesis most directly supported by the data presented in this chapter is that the *c-myc* IRES is not potentiated by the mere presence of a specific nuclear factor.

Transcription, RNA maturation and nuclear export are the major pre-translational phases in the history of any RNA message. It is not hard to imagine consequences of any one of these processes that might have an effect upon IRES function: co-transcriptional factor binding, base modification, or factor exchange concomitant with nuclear export. Having imagined such mechanisms, however, it is difficult to come up with any precedents that fit the observed data. Specific base modifications are known to occur, and known to be vital to the function of rRNA, but there has been no suggestion that such modifications take place on

mRNAs. The co-transcriptional binding of hnRNPs and splicing factors is a step in the processing of most if not all messages, and, for example, the splicing factor PTB has been shown to activate rhinovirus (Borman *et al.*, 1993) (Hunt and Jackson, 1999) and FMDV (Niepmann *et al.*, 1997) IRESs. However, there is no evidence to suggest that the binding of PTB to IRES RNA specifically occurs in the nuclear compartment during RNA processing *in vivo*, and PTB is abundant and active in the RSW fraction. There have been hints that nuclear export may be accompanied by factor binding, but there is less evidence that such factors remain irreversibly associated with their messages. Furthermore, translation is such a dynamic process, and so generally characterised by the dissociation and re-association of its component parts, that the postulation of an early and irreversible RNA-protein binding step is hard to accept. Even if this were the case, one would still expect the protein in question to be present in the RSW fraction.

Tightly and lasting binding of proteins to specific mRNAs prior to nuclear export is exemplified by the phenomenon translationally masked mRNPs. It has long been known that mRNAs can exist in an inactive cytoplasmic form, packaged as mRNP particles that are inaccessible to the translational machinery (Spirin, 1966). These mRNPs are in stable storage, ready to be activated by an external cue; this is typical of free mRNPs in germ cells and other dormant states.

Analysis of masked mRNPs from *Xenopus* oocytes reveals that one of the major proteins involved is the oocyte-specific DNA-binding transcription factor FRGY2 (Deschamps *et al.*, 1992), and a similar protein was subsequently identified in the masked mRNPs of murine spermatocytes (Kwon *et al.*, 1993).

Intriguingly, masking (presumably mediated by FRGY2) is far more evident upon transcripts innate to the oocyte than synthetic mRNAs that are microinjected into either the

cytoplasm or the nucleus(Bouvet and Wolffe, 1994). Thus, effective masking in the cytoplasm seems to be coupled to transcription. It is suggested that nascent RNA is either better loaded with masking protein, or more capable of conserving a repressive secondary structure upon loading with protein than is the pre-synthesised and microinjected RNA.

The situation is reminiscent of the nuclear experience necessary for the activity of the *c-myc* IRES, except that in this case translation is prevented rather than potentiated by the nuclear event. Conceivably the nuclear event that activates the *c-myc* IRES is closely analogous, and is due to the co-transcriptional binding of a nuclear factor.

How, then, to determine the nature of the nuclear event, and establish an *in vitro* assay for the *c-myc* IRES? A number of strategies might yield results. Microinjection of IRES-bearing bicistronic RNA constructs into cell or *Xenopus* oocyte nuclei is possible, and if IRES activity resulted would demonstrate that no co-transcriptional event is necessary, and prompt further experiments with fractionated nuclear extracts. Alternatively, cytoplasmic mRNPs might be purified from a quantity of cultured cells transfected with IRES-bearing bicistronic DNA constructs, and used to program *in vitro* translation reactions. If pure RNA was seen to show IRES activity, this would suggest that the nuclear event is a process that modifies the mRNA itself. If crude mRNPs showed IRES function, this would support the co-transcriptional factor binding hypothesis, and provide a starting point for the purification and identification of the trans-acting factor(s) involved.

Chapter 6

Discussion

6.1 A Mechanistic Model

IRES architecture

The *c-myc* IRES is peculiar. In contrast to viral IRESs, there is no sharp boundary that when crossed by a terminal deletion causes a catastrophic failure in function. Rather, IRES function is divided between two structural domains, flanking the ribosome landing site. Furthermore, a high proportion of *c-myc* IRES mutants assayed are as active or more active than the wild-type, a rare occurrence in viral IRESs. The differences in function, and consequent different requirements for control of expression, between Myc and viral proteins are likely to account for the divergence in IRES mechanism.

Diffuse function

The *c-myc* gene has retained the ability to be translated via a cap-dependent mechanism. No AUG codons have crept into the 5' UTR by genetic drift. The 5' UTR inhibits translation by about 50% in rabbit reticulocyte lysate (Stoneley, 1998), as would be expected given the predicted stability of domain 1, but it does not inhibit translation of monocistronic messages in HeLa cell extract (Butnick *et al.*, 1985), a system that does *not* support *c-myc* IRES function. The factor(s) present in HeLa cell lysate that facilitate scanning through the 5' UTR have not been identified, and it is not known whether this activity affects all 5' UTRs, or only acts upon certain messages. Thus, in a human system, the P2 *c-myc* 5' UTR does not provide a significant barrier to scanning ribosomes, and has specifically retained this property. This may also be partially achieved by a "modular" IRES, in which each module provides a

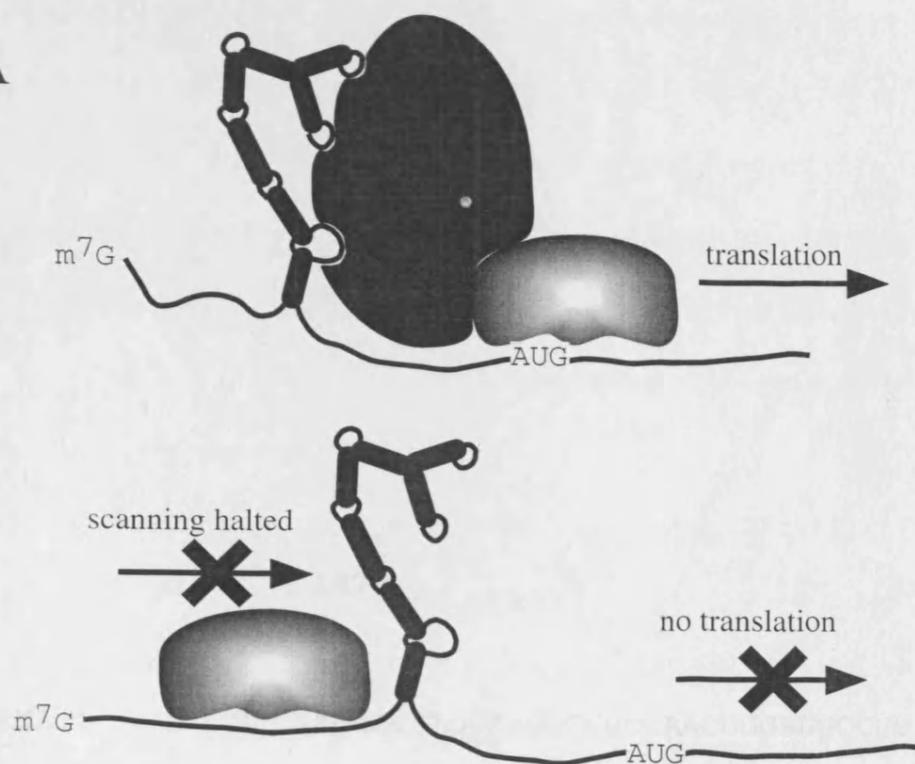
different IRES component, but avoids any single folded domain of a stability which would prevent the progress of a scanning ribosome (Figure 6.i).

Ribosome capture

In all forms of translation initiation, a necessary early step is the recruitment of a ribosome to the initiation complex. As has been discussed, in cap-dependent initiation this first contact appears to be mediated by a number of eIFs. In picornaviruses it is likely to depend on some of the same contacts, with the unproved participation of IRES/18S RNA complementarity (Pilipenko *et al.*, 1992), whilst in HCV, the IRES RNA alone is capable of binding 40S subunits in the correct orientation (Pestova *et al.*, 1998b). In a “synthetic IRES” system (De Gregorio *et al.*, 1999) it is shown that internal initiation can be performed by a laboratory-designed leader sequence that binds a chimeric protein containing the conserved central domain of eIF4G, suggesting that an eIF4G/eIF4A complex has an innate ability to capture and “launch” scanning ribosomes that does not depend upon precise orientation with RNA. Thus, hypothetically, a cellular IRES need do no more than bind eIF4G.

There is little direct evidence to show what this first contact might be in the case of the *c-myc* IRES. Certainly, it does not require *intact* eIF4G for function (Johannes and Sarnow, 1998; Stoneley *et al.*, 2000a), but it may at least require some C-terminal cleavage product, as do the picornaviridae. Identification of the *trans*-acting factors required for IRES function depends upon the development of a functional IRES system *in vitro*. There is also some scope for pairing between 3'-terminal 18S RNA and *c-myc* IRES entry window RNA, as shown in Figure 6.ii.

A



B

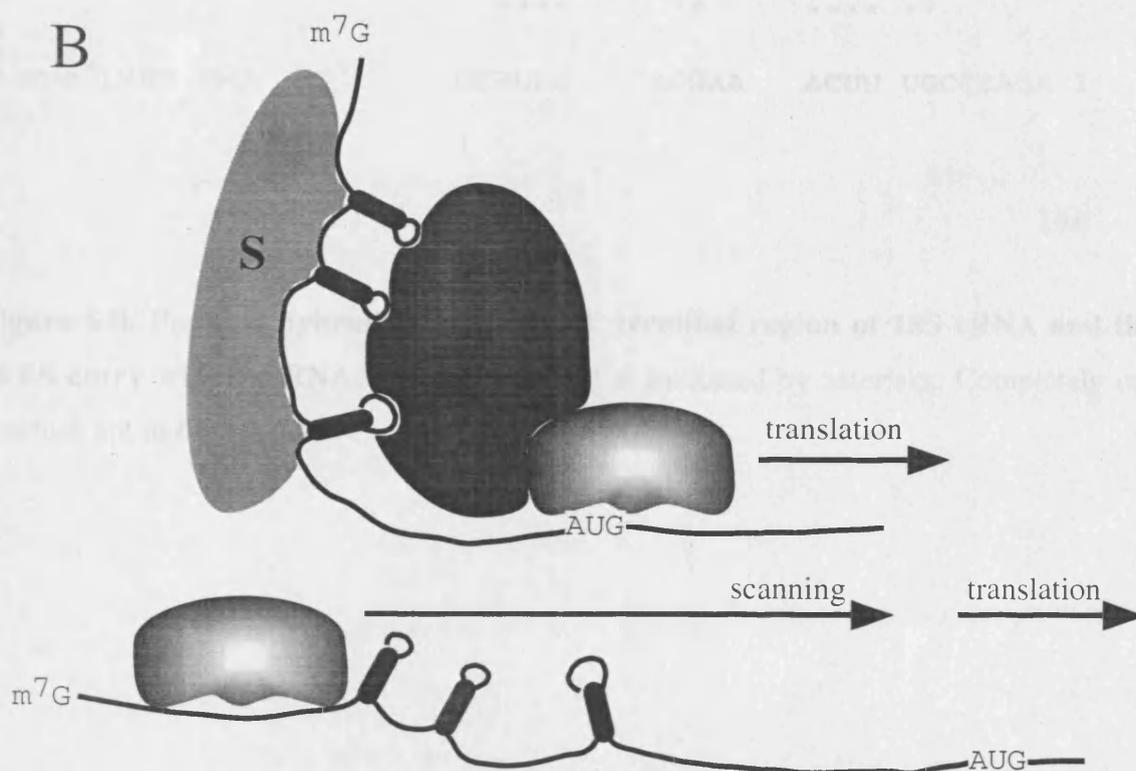


Figure 6.i. Modular IRESs and scanning. A: mRNA with a large stable IRES structure incompatible with scanning, and consequently incapable of translation via a cap-dependent mechanism. B: A modular IRES making the same contacts with trans-acting factors as IRES A with the assistance of a scaffold protein (S). No single folded domain is of a stability that inhibits scanning, allowing both cap-dependent and cap-independent translation initiation of the message.

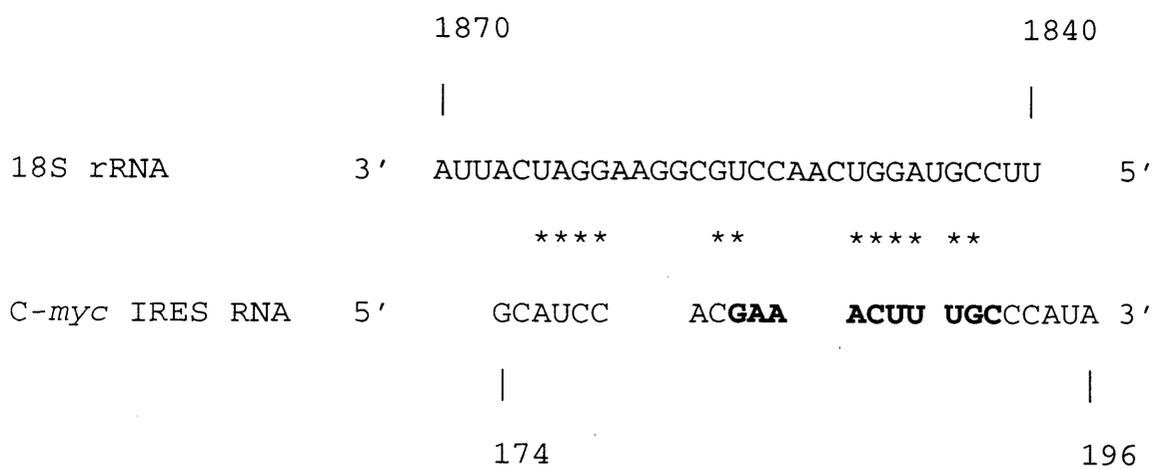


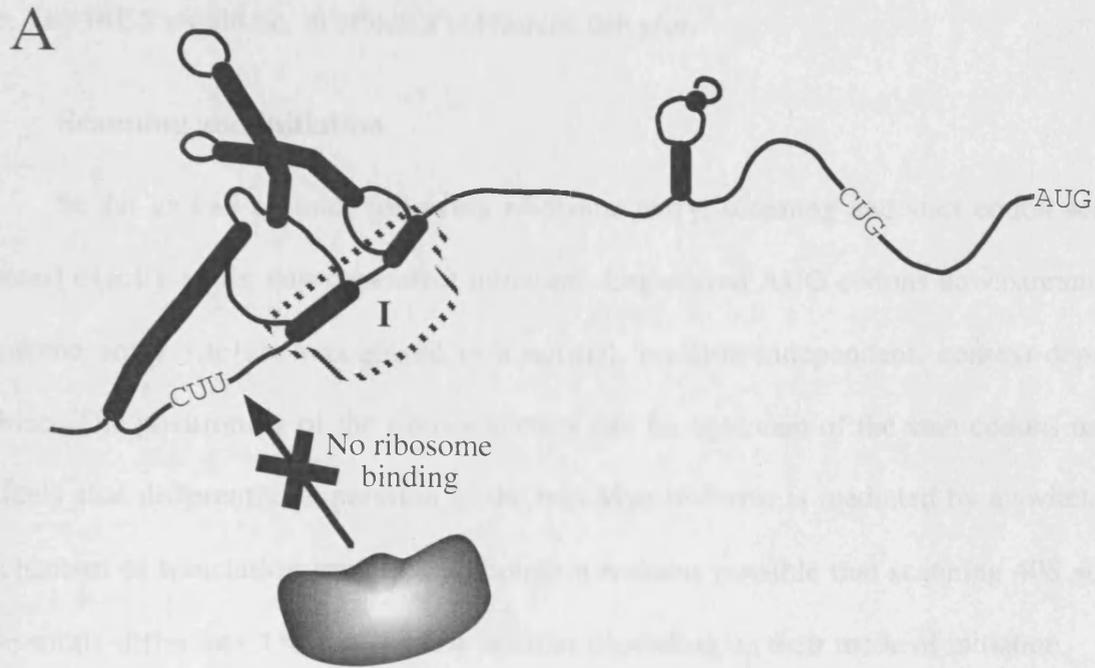
Figure 6.ii. Possible hybridization of the 3' terminal region of 18S rRNA and the c-*myc* IRES entry window RNA. Pairing potential is indicated by asterisks. Completely conserved residues are in bold type.

Ribosome entry

Whatever the nature of the initial contact, the 40S subunit must overcome the double pseudoknot structure at the 3' edge of the entry window before landing. This might be achieved in a number of ways. If the initial contact is with the entry window, then pseudoknot unwinding must be the first event. If the IRES binds eIF4G, then pseudoknot unwinding could possibly be performed by the helicase activity of eIF4A. Alternatively, the 40S subunit may be recruited by a different mechanism, and ribosome landing depends upon some other pseudoknot-disrupting procedure.

It is tempting to regard the pseudoknots as a switch, which must be specifically disrupted in order for IRES-mediated Myc expression to occur. A switch-like role has been proposed for the inhibitory domain IV helix in the HCV IRES (Honda *et al.*, 1996), which it is suggested may be stabilized by interaction with a protein expressed during viral infection, forming a negative feedback loop helping to maintain a steady, sub-lytic state of viral infection. One can speculate that an unknown protein factor might stabilize the pseudoknot(s) present in the *c-myc* IRES, thus blocking ribosome entry, or that a stimulating factor might cause their destabilization (Figure 6.iii).

Alternatively, some aspect of IRES structure might more directly transduce a cellular physiological state to the translational machinery. It has very recently been proposed that low intracellular levels of polyamines de-repress *c-myc* translation (Frostedesjo and Heby, 2000). Classically, polyamine-mediated de-repression of translation is mediated via a uORF encoding the peptide MAGDIS (Ruan *et al.*, 1996). No *c-myc* transcript contains such a uORF. As polyamines are believed to be a major determinant of RNA structure (Yoshida *et al.*, 1999) and have been shown to stabilize A-form helices (Antony *et al.*, 1999), one might speculate that *Myc* translation is modulated by their direct influence upon IRES structure; for example,



Inhibitor release?
 Activator binding?
 Polyamine depletion?

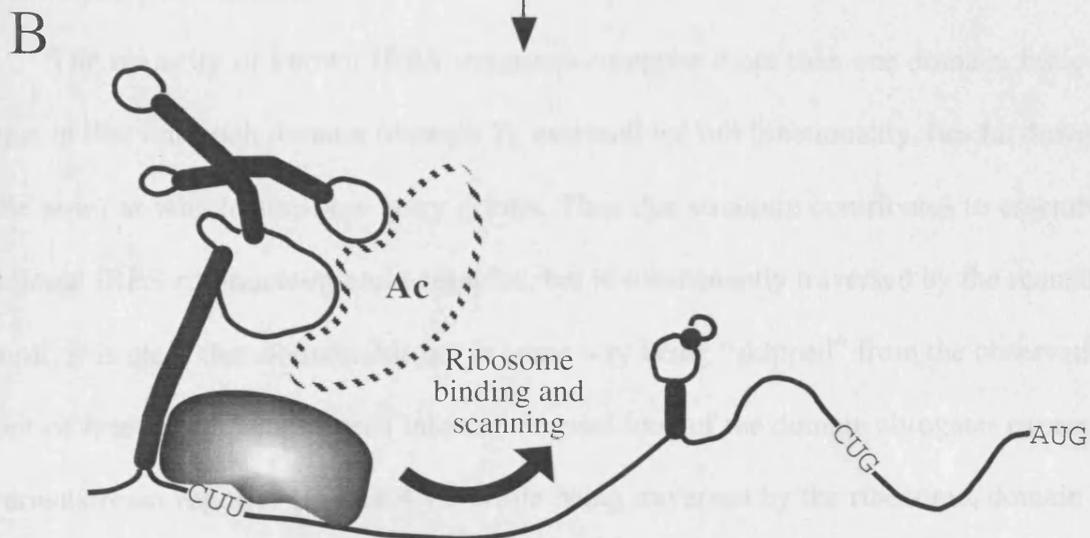


Figure 6.iii. Model of IRES regulation by pseudoknot destabilization. A: 40S subunits are unable to land when pseudoknot a is intact; possibly it is stabilized by an inhibitor protein "I". B: Pseudoknot helices disrupted by inhibitor release/ activator binding/ polyamine depletion, allowing ribosome entry and scanning.

if high polyamine levels stabilized pseudoknots α and β , then low polyamine levels would destabilize these helices, allowing IRES-driven translation initiation to proceed at a greater rate. The IRES would be, in effect, a polyamine detector.

Scanning and initiation

So far as can be told, following ribosome entry, scanning and start codon selection proceed exactly as for cap-dependent initiation. Engineered AUG codons downstream of the ribosome entry site are recognised in a normal, position-independent, context-dependent fashion. The positioning of the ribosome entry site far upstream of the start codons makes it unlikely that differential expression of the two Myc isoforms is mediated by a switch in the mechanism of translation initiation, although it remains possible that scanning 40S subunits have subtly different CUG-recognising abilities depending on their mode of initiation.

It is perhaps significant that the spacing between the *c-myc* IRES ribosome landing site and the first start codon is somewhere between 157 and 174 nucleotides; the equivalent spacing in type I picornavirus is generally 155-160 nucleotides. The significance (if any) of this conformity is unclear.

The majority of known IRES structures comprise more than one domain, but *c-myc* is unique in that one such domain (domain 2), essential for full functionality, lies far downstream of the point at which ribosome entry occurs. Thus this structure contributes to assembly of a functional IRES ribonucleoprotein complex, but is subsequently traversed by the scanning 40S subunit. It is clear that domain 2 is not in some way being “skipped” from the observation that an out-of frame AUG engineered into the terminal loop of the domain abrogates expression of the downstream reporter (Figure 4.v). While being traversed by the ribosome, domain 2 must be unwound, breaking any inter- or intra-molecular interactions in which it is engaged, thus paradoxically casting it simultaneously in the roles of IRES activator and impeder of scanning

(Figure 6.iv). In order to allow the 40S subunit to pass, any tertiary contacts made by the domain would have to be transitory (i.e. have a high on/off rate), or be very weak (i.e. have a low affinity), or both. Conceivably, the activating mutation D (Figure 4.vii) operates by disturbing the kinetics of this interaction in such a way that the ability of domain 2 to participate in ribosome landing is disturbed less than its ability to impede the scanning ribosome.

If domain 2 is involved in every IRES-driven initiation event, then it will undergo a cycle of folding - IRES RNP complex binding - IRES RNP complex release - unwinding - 40S subunit transit for as long as the IRES is active. As every 40S subunit that enters the message in this way goes on to disrupt the fabric of the IRES, such a cycle might play a role in limiting the rate of initiation.

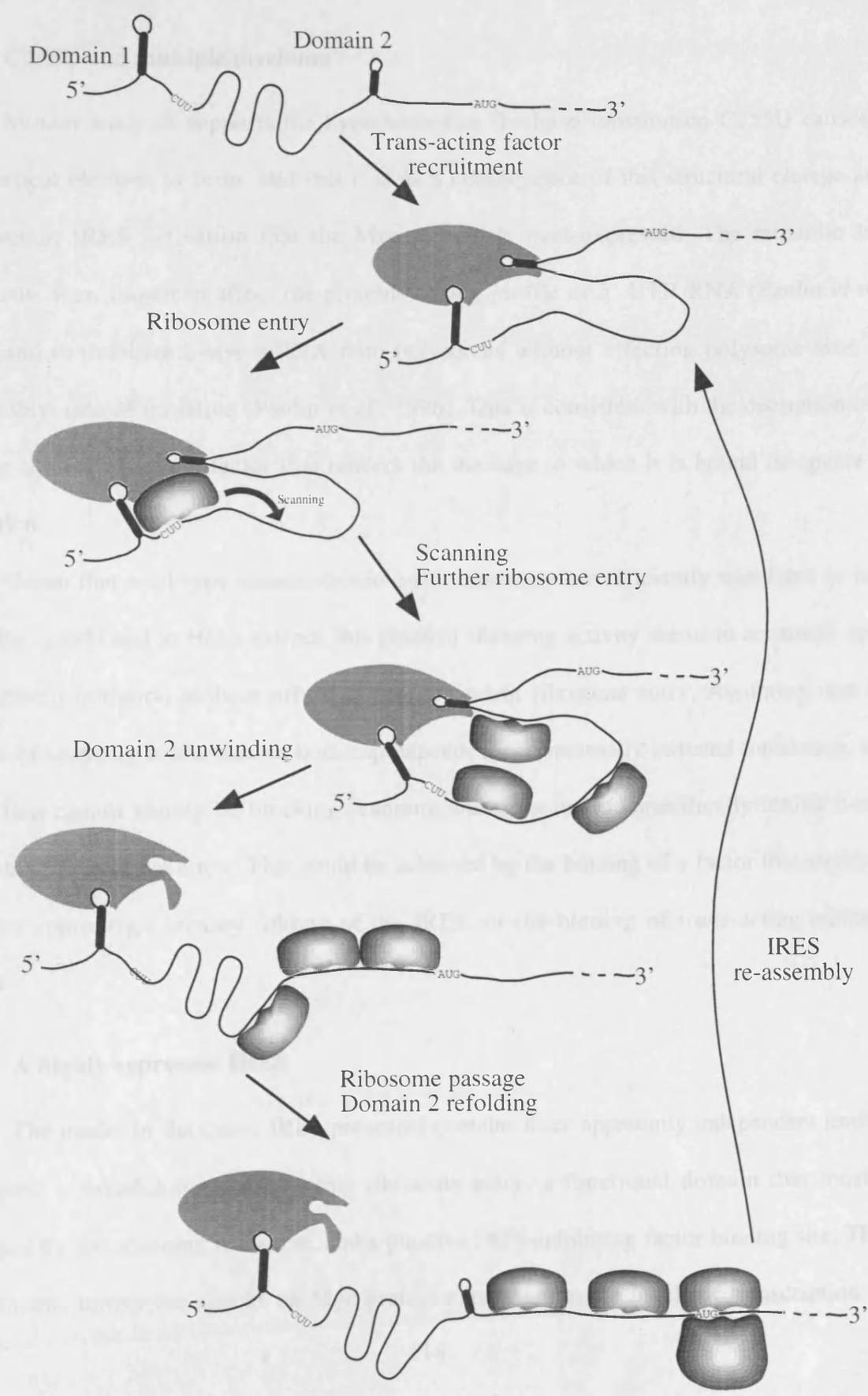


Figure 6.iv. Partial disruption of IRES structure as a consequence of scanning.

6.2 The *c-myc* IRES and Disease

C255U and multiple myeloma

Mutant analysis supports the hypothesis that the base substitution C255U causes a short helical element to form, and that it is as a consequence of this structural change and concomitant IRES activation that the Myc protein is over-expressed. The mutation has previously been shown to affect the protein-binding profile of 5' UTR RNA (Paulin *et al.*, 1998), and to mobilize *c-myc* mRNA onto polysomes without affecting polysome size or, presumably, rate of initiation (Paulin *et al.*, 1996). This is consistent with the disruption of a binding site for a protein factor that renders the message to which it is bound incapable of translation.

Given that wild-type monocistronic *c-myc* messages are efficiently translated *in vivo* (Stoneley, 1998) and in HeLa extract, this putative silencing activity seems to act solely upon IRES-driven initiation without affecting cap-dependent ribosome entry. Assuming that the process of scanning is identical in both cap-dependent and internally initiated translation, this interaction cannot simply be blocking scanning, but must instead specifically inhibit *c-myc* IRES-driven ribosome entry. This could be achieved by the binding of a factor that sterically prevents appropriate tertiary folding of the IRES, or the binding of *trans*-acting initiation factors.

A highly repressed IRES

The model of the *c-myc* IRES presented contains three apparently independent limiting strategies: a pseudoknot that prevents ribosome entry, a functional domain that must be traversed by the scanning ribosome, and a putative IRES-inhibiting factor binding site. These mechanisms mirror the checks on Myc protein expression at the levels of transcription and

mRNA and protein stability. They may provide an explanation for the paradoxical failure of the IRES to function *in vitro*.

Translational deregulation of Myc expression has been described in numerous human neoplasia and cancer-prone syndromes (Cory, 1986; Macpherson *et al.*, 1992; Taub *et al.*, 1984; West *et al.*, 1995), and is likely to be behind a significant proportion of those tumour types where Myc levels are elevated by unknown mechanisms, which form the majority (Ryan and Birnie, 1996). The additional limiting mechanisms here suggested may either be interpreted as cumulative precautions against such deregulation, or exploitable fallibilities imposed by the needs of normal *c-myc* expression. Answers to such questions await the complete description of the physiological roles of the *c-myc* IRES.

Bibliography

- Abramovitz, D.L. and Pyle, A.M. (1997) Remarkable morphological variability of a common RNA folding motif: The GNRA tetraloop-receptor interaction. *J. Mol. Biol.*, **266**, 493-506.
- Alexander, W.S., Bernard, O., Cory, S. and Adams, J.M. (1989) Lymphomagenesis in E-*myc* transgenic mice can involve Ras mutations. *Oncogene*, **4**, 575-581.
- Amati, B., Littlewood, T.D., Evan, G.I. and Land, H. (1993) The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO J.*, **12**, 5083-5087.
- Antony, T., Thomas, T., Shirahata, A. and Thomas, T.J. (1999) Selectivity of polyamines on the stability of RNA-DNA hybrids containing phosphodiester and phosphorothioate oligodeoxyribonucleotides. *Biochemistry*, **38**, 10775-10784.
- Bello-Fernandez, C., Packham, G. and Cleveland, J.L. (1993) The ornithine decarboxylase gene is a transcriptional target of *c-myc*. *Proc. Natl. Acad. Sci. USA*, **90**, 7804-7808.
- 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.
- Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N. and Weintraub, H. (1993) Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell Biol.*, **13**, 5216-5224.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) Sequence-specific DNA binding by the c-Myc protein. *Science*, **250**, 1149-1151.
- Blackwood, E.M. and Eisenman, R.N. (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science*, **51**.
- Blyn, B.L., Chen, R., Semler, B.L. and Ehrenfeld, E. (1995) Host cell proteins binding to domain IV of the 5' noncoding region of poliovirus RNA. *J. Virol.*, **69**, 4381-4389.

- Blyn, L.B., Towner, J.S., Semler, B.L. and Ehrenfeld, E. (1997) Poly(rC) binding protein 2 is required for translation of poliovirus RNA. *J. Virol.*, **71**, 6243-6246.
- Boeck, R. and Kolakofsky, D. (1994) Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.*, **13**, 3608-3617.
- 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., 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. *Nucl. Acids Res.*, **23**, 3656-3663.
- Bouvet, P. and Wolffe, A.P. (1994) A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. *Cell*, **77**, 931-941.
- Brewer, G. (1998) Characterization of c-myc 3' to 5' mRNA decay Activities in an in vitro system. *J. Biol. Chem.*, **273**, 34770-34774.
- Brown, B.A. 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.
- Brown, J.W. (1991) Phylogenetic analysis of RNA structure on the Macintosh computer. *CABIOS*, **7**, 391-393.
- Bukh, J., Purcell, R.H. and Miller, R.H. (1992) Sequence analysis of the 5' noncoding region of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA*, **89**, 4942-4946.
- Butnick, N.Z., Miyamoto, C., Chizzonite, R., Cullen, B.R., Ju, G. and Skalka, A.M. (1985) Regulation of the human c-myc gene: 5' noncoding sequences do not affect translation. *Mol. Cell. Biol.*, **5**, 3009-3016.

- Carter, P.S., Jarquin-Pardo, M. and DeBenedetti, A. (1999) Differential expression of Myc1 and Myc2 isoforms in cells transformed by eIF4E: evidence for internal ribosome repositioning in the human c-myc 5'UTR. *Oncogene*, **18**, 4326-4335.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K.H., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
- Chaudhuri, J., Das, K. and Maitra, U. (1994) Purification and characterisation of bacterially expressed mammalian translation initiation-factor-5 (eIF-5)-demonstration that eIF-5 forms a specific complex with eIF-2. *Biochem.*, **33**, 4794-4799.
- Chaudhuri, J., Kausik, S. and Umadas, M. (1997) Function of eukaryotic translation initiation factor 1A (eIF 1A) (formerly call eIF 4C) in initiation of protein synthesis. *J. Biol. Chem.*, **272**, 7883-7891.
- Cheng, J.-W., Chou, S.-H. and Reid, B.R. (1992) Base-pairing geometry in GA mismatches depends entirely on the neighbouring sequences. *J. Mol. Biol.*, **228**, 1037-1041.
- Cigan, A.M., Pabich, E.K. and Donahue, T.F. (1988) tRNA-iMet functions in directing the scanning ribosome to the start site of translation. *Science*, **242**, 93-97.
- Coppola, J.A. and Cole, M.D. (1986) Constitutive c-myc oncogene expression blocks mouse erythroleukemia cell differentiation but not commitment. *Nature*, **320**, 760-763.
- Correll, C.C., Freeborn, B., Moore, P.B. and Steitz, T.A. (1997) Metals, motifs and recognition in the crystal structure of a 5S rRNA domain. *Cell*, **91**.
- Cory, S. (1986) Activation of cellular oncogenes in hemopoietic cells by chromosome translocation. *Adv. Cancer Res.*, **47**, 189-234.
- Dang, C.V. (1999) c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell Biol.*, **19**, 1-11.

- Darveau, A., Pelletier, J. and Sonenberg, N. (1985) Differential efficiencies of *in vitro* translation of mouse *c-myc* transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. USA*, **82**, 2315-2319.
- De Benedetti, A., Joshi, B., Graff, J.R. and Zimmer, S.G. (1994) CHO cells that are transformed by the translation initiation factor eIF-4E display increased levels of *c-myc* expression, but require concomitant overexpression of Max for tumorigenicity. *Mol. Cell. Diff.*
- De Gregorio, E., Preiss, T. and Hentze, M.W. (1999) Translation driven by an eIF4G core domain *in vivo*. *EMBO J.*, **18**, 4865-4874.
- deQuinto, S. and MartinezSalas, E. (1997) Conserved structural motifs located in distal loops of aphthovirus internal ribosome entry site domain 3 are required for internal initiation of translation. *J. Virol.*, **71**, 4171-4175.
- Deschamps, S., Viel, A., Garrigos, M., Denis, H. and le Maire, M. (1992) mRNP4, a major mRNA-binding protein from *Xenopus* oocytes is identical to transcription factor FRGY2. *J. Biol. Chem.*, **267**, 13799-13802.
- Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucl. Acids Res.*, **11**, 1475-1489.
- Dmitrovsky, E., Kuehl, W.M., Hollis, G.F., Kirsh, I.R., Bender, T.P. and Segal, S. (1986) Expression of a transfected human *c-myc* oncogene inhibits differentiation of a mouse erythroleukemia cell line. *Nature*, **322**, 748-750.
- Dorner, A.J., Semler, B.L., Jackson, R.J., Hanecak, R., Duprey, E. and Wimmer, E. (1984) In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate. *J. Virol.*, **50**, 507-514.

- Eilers, M., Schirm, S. and Bishop, J.M. (1991) The MYC protein activates transcription of the α -prothymosin gene. *EMBO J.*, **10**, 133-141.
- 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.
- Felden, B., Florentz, C., Giegé, R. and Westhof, E. (1996) A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5' upstream pseudoknots in the 3' terminus of tobacco mosaic virus RNA. *RNA*, **2**.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J.P., Atkins, J.F. and Gesteland, R.F. (1997) Probing the structure of the *Escherichia coli* 10Sa RNA (tm RNA). *RNA*, **3**, 89-103.
- Fraser, S.D. and Browder, L.W. (1995) Translational regulation in *Xenopus* by the c-myc I 5'UTR. *Developmental Biology*, **170**, 757.
- Frostesjo, L. and Heby, O. (2000) Polyamine depletion up-regulates c-Myc expression, yet induces G(1) arrest and terminal differentiation of F9 teratocarcinoma stem cells. *J. Cell. Biochem.*, **76**, 143-152.
- Futterer, J. and Hohn, T. (1996) Translation in plants - rules and exceptions. *Plant Mol. Biol.*, **32**, 159-189.
- Gautheret, D., Konings, D. and Gutell, R.R. (1994) A major family of motifs involving G•A mismatches in ribosomal RNA. *J. Mol. Biol.*, **242**, 1-8.
- Geballe, A.P. and Morris, D.R. (1994) Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.*, **19**, 159-164.
- Gmyl, A.P., Pilipenko, E.V., Maslova, E.V., Belov, G.A. and Agol, V.I. (1993) Functional and genetic plasticities of the poliovirus genome: quasi-infectious RNAs modified in the 5'-untranslated region yield a variety of pseudo-revertants. *J. Virol.*, **67**, 6309-6316.

- Goumans, H., Thomas, A., Verhoeven, A., Voorma, H.O. and Benne, R. (1980) The role of eIF-4C in protein synthesis initiation complex formation. *Biochem. Biophys. Acta*, **608**, 39-46.
- Hahn, B., Kim, Y.K., Kim, J.H., Kim, T.Y. and Jang, S.K. (1998) Heterogeneous nuclear ribonucleoprotein L interacts with the 3' border of the internal ribosomal entry site of hepatitis C virus. *J. Virol.*, **72**, 8782-8788.
- Haller, A.A. and Semler, B.L. (1992) Linker scanning mutagenesis of the internal ribosome entry site of poliovirus RNA. *J. Virol.*, **65**, 5886-5894.
- Hann, S.R., Dixit, M., Sears, R.C. and Sealy, L. (1994) The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. *Genes Dev.*, **8**, 2441-2452.
- Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W. and Eisenman, R.N. (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell*, **52**, 185-195.
- Hann, S.R., Sloan-Brown, K. and Spotts, G.D. (1992) Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities in response to methionine deprivation. *Genes Dev.*, **6**, 1229-1240.
- Hann, S.R., Thompson, C.B. and Eisenman, R.N. (1985) c-myc oncogene protein synthesis is independent of the cell cycle in human and avian cells. *Nature*, **314**, 366-369.
- Hannig, E.M. (1995) Protein synthesis in eukaryotic organisms: new insights into the function of translation initiation factor eIF-3. *BioEssays*, **17**, 915-919.
- Heikkila, R., Schwab, G., Wickstrom, E., Loke, S.L., Pluznik, D.H., Watt, R. and Neckers, L.M. (1987) A c-myc antisense oligodeoxynucleotide inhibits entry into S phase but not progress from G0 to G1. *Nature*, **328**, 445-449.

- Hellen, C.U.T. and Pestova, T.V. (1999) Translation of hepatitis C virus RNA. *Journal of Viral Hepatitis*, **6**, 79-87.
- Hellen, C.U.T., Pestova, T.V. and Wimmer, E. (1994) Effect of mutations downstream of the internal ribosomal entry site on initiation of poliovirus protein-synthesis. *J. Virol.*, **68**, 6312-6322.
- Hellen, C.U.T., 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 ribosomal entry is identical to the nuclear pyrimidine tract-binding protein. *Proc. Natl. Acad. Sci. USA*, **90**, 7642-7646.
- Hershey, J. (1999) Structures of eIF-3 and initiation complexes. *C.N.R.S. - Conference Jacques Monod. New insights into the mechanism of mRNA translation: the significance of RNA structure*, Aussois, France, p. 5.
- Heus, H.A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, **253**, 191-194.
- Hinnebusch, A.G. (1996) Translational control of GCN4: gene specific regulation by phosphorylation of eIF2. In Hershey, J.W.B., Mathews, M.B. and Sonenberg, N. (eds.), *Translational control*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 199-244.
- Hinnebusch, A.G. (1997) Translational regulation of yeast GCN4. *J. Biol. Chem.*, **272**, 21661-21664.
- Holmberg, L. and Nygard, O. (1997) Mapping of nuclease-sensitive sites in native reticulocyte ribosomes. *Eur. J. Biochem.*, **247**, 160-168.

- Honda, M., Brown, E.A. and Lemon, S.M. (1996) Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA*, **2**, 955-968.
- Huang, H., Yoon, H., Hannig, E.M. and Donahue, T.F. (1997) GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.*, **2396-2413**.
- Huang, J. and Schneider, R.J. (1991) Adenovirus inhibition of cellular protein synthesis involves inactivation of the cap binding protein. *Cell*, **65**, 271-280.
- Huez, I., Creancier, L., Audigier, S., Gensac, M.C., Prats, A.C. and Prats, H. (1998) Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell Biol.*, **18**, 6178-6190.
- 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 translation of human rhinovirus RNA. *Genes Dev.*, **13**, 437-448.
- 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.
- Jackson, R.J. (1991) The ATP requirement for initiation of eukaryotic translation varies according to the messenger-rna species. *Eur. J. Biochem.*, **200**, 285-294.
- Jackson, R.J. (1999) Cellular RNA-binding proteins involved in internal initiation of picornaviral RNA translation. *C.N.R.S. - Conference Jacques Monod. New insights into the mechanism of mRNA translation: the significance of RNA structure*, Aussois, p. 5.
- Jackson, R.J. and Kaminski, A. (1995) Internal initiation of translation in eukaryotes: The picornavirus paradigm and beyond. *RNA*, **1**, 985-1000.

- James, B.D., Olsen, G.J. and Pace, N.R. (1989) Phylogenetic comparative analysis of RNA secondary structure. *Methods Enzymol.*, **180**, 227-239.
- Jang, S.K., Krausslich, H.G., Nicklin, M.J.H., Duke, G.M., Palmenberg, A.C. and Wimmer, E. (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J. Virol.*, **62**.
- Jang, S.K., Pestova, T.V., Hellen, C.U.T., Witherell, G.W. and Wimmer, E. (1990) Cap-independent translation of picornavirus RNAs- structure and function of the internal ribosomal entry site. *Enzyme*, **44**, 292-309.
- Johannes, G., Carter, M.S., Eisen, M.B., Brown, P.O. and Sarnow, P. (1999) Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc. Natl. Acad. Sci. USA*, **96**, 13118-13123.
- 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.
- Jones, R.M., Branda, J., Johnston, K.A., Polymenis, M., Gadd, M., Rustgi, A., Callanan, L. and Schmidt, E.V. (1996) An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. *Mol. Cell Biol.*, **16**, 4754-4764.
- Kaminski, A., Belsham, G.J. and Jackson, R.J. (1994) Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. *EMBO J.*, **13**, 1673-1681.
- Kaminski, A. and Jackson, R.J. (1998) The polypyrimidine tract binding protein (PTB) requirement for internal initiation of translation of cardiovirus RNAs is conditional rather than absolute. *RNA*, **4**, 626-638.

- Karayiannis, P., Pickering, J., Zampino, R. and Thomas, H.C. (1998) Natural history and molecular biology of hepatitis G virus GB virus C. *Clinical and Diagnostic Virology*, **10**, 103-111.
- Kato, G.J. and Dang, C.V. (1992) Function of the c-Myc oncoprotein. *FASEB J.*, **6**, 3065-3072.
- Kelly, K., Cochran, B.H., Stiles, C.D. and Leder, P. (1983) Cell-specific regulation of the *c-myc* gene by lymphocyte mitogens and platelet-derived growth factor. *Cell*, **35**, 603-610.
- Kozak, M. (1977) Nucleotide sequences of 5'-terminal ribosome-protected initiation regions from two reovirus messages. *Nature*, **269**.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.*, **15**, 8125-8137.
- Kozak, M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structures in eukaryotic mRNAs. *Mol. Cell. Biol.*, **9**, 5134-5142.
- Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA*, **87**, 8301-8305.
- Kozak, M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl. Acad. Sci. USA*, **92**, 2662-2666.
- Kozak, M. (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.*, **16**, 2482-2492.
- Kozak, M. (1998) Primer extension analysis of eukaryotic ribosome-mRNA complexes. *Nucl. Acids Res.*, **26**, 4853-4859.
- Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187-208.

- Kuge, S. and Nomoto, A. (1987) Construction of viable deletion and insertion mutants of the Sabin strain of type 1 poliovirus: function of the 5' noncoding sequence in viral replication. *J. Virol.*, **61**, 1478-1487.
- Kuhn, R., Luz, N. and Beck, E. (1990) Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J. Virol.*, **64**, 4625-4631.
- Kwon, Y.K., Murray, M.T. and Hecht, N.B. (1993) Proteins homologous to the *Xenopus* germ cell-specific RNA-binding proteins p54/p56 are temporally expressed in mouse male germ cells. *Dev. Biol.*, **158**, 90-100.
- Lamphear, B.J., Kirchweger, R., Skern, T. and Rhoads, R.E. (1995) Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. *J. Biol. Chem.*, **270**, 21975-21983.
- Lamphear, B.J., Yan, R., Yang, F., Waters, D., Liebig, H.-D., Klump, H., Kuechler, E., Skern, T. and Rhoads, R.E. (1993) Mapping the cleavage site in protein synthesis initiation factor eIF-4 γ of the 2A proteases from human coxsackievirus and rhinovirus. *J. Biol. Chem.*, **268**, 19200-19203.
- Lawson, T.G., Lee, K.A., Maimone, M.M., Abramson, R.D., Dever, T.E., Merrick, W.C. and Thach, R.E. (1989) Dissociation of double-stranded polynucleotide helical structures by eukaryotic initiation factors, as revealed by a novel assay. *Biochemistry*, **28**, 4729-4734.
- Lawson, T.G., Ray, B.K., Dodds, J.T., Grifo, J.J., Abramson, R.D., Merrick, W.C., Betsch, D.F., Wise, H.L. and Thach, R.E. (1986) Influence of proximal secondary structure on the translational efficiency of eukaryotic messenger-RNAs and on their interaction with initiation factors. *J. Biol. Chem.*, **261**, 3979-3989.
- Le, S.Y. and Maizel, J.V. (1997) A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucl. Acids Res.*, **25**, 362-369.

- Le, S.Y. and Maizel, J.V. (1998) Evolution of a common structural core in the internal ribosome entry sites of picornavirus. *Virus Genes*, **16**, 25-38.
- Le, S.Y., Siddiqui, A. and Maizel, J.V. (1996) A common structural core in the internal ribosome entry sites of picornavirus, hepatitis C virus, and pestivirus. *Virus Genes*, **12**, 135-147.
- Lemon, S.M. and Honda, M. (1997) Internal ribosome entry sites within the RNA genomes of hepatitis C virus and other flaviviruses. *Semin. Virol.*, **8**, 274-288.
- Leontis, N.B. and Westhof, E. (1998) The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA*, **4**, 1134-1153.
- Lorsch, J.R. and Herschlag, D. (1998a) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry*, **37**, 2180-2193.
- Lorsch, J.R. and Herschlag, D. (1998b) The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry*, **37**, 2194-2206.
- Luukkonen, B.G.M., Tan, W. and Schwartz, S. (1995) Efficiency of reinitiation of translation on HIV type 1 mRNAs is determined by the length of the upstream ORF and by intercistronic distance. *J. Virol.*, **69**, 4086-4094.
- Macpherson, A.J.S., Chester, K.A., Robson, L., Bjarnason, I., Malcolm, A.D.B. and Peters, T.J. (1992) Increased expression of c-myc proto-oncogene in biopsies of ulcerative-colitis and Crohn's colitis. *Gut*, **33**, 651-656.
- Mader, S., Lee, H., Pause, A. and Sonenberg, N. (1995) The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 γ and the translational repressors 4E-binding proteins. *Mol. Cell Biol.*, **15**, 4990-4997.

- Marcotrigiano, J., Gingras, A.-C., Sonenberg, N. and Burley, S.K. (1997) Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell*, **89**, 951-961.
- Marcu, K.B., Bossone, S.A. and Patel, A.J. (1992) *myc* function and regulation. *Annu. Rev. Biochem.*, **61**, 809-860.
- Marissen, W. and Lloyd, R.E. (1998) Eukaryotic translation initiation factor 4G is targeted for proteolytic cleavage by caspase 3 during inhibition of translation in apoptotic cells. *Mol. Cell Biol.*, **18**, 7565-7574.
- Martínez-Salas, E., Regalado, M.P. and Domingo, E. (1996) Identification of an essential region for internal initiation of translation in the aphthovirus internal ribosome entry site and implications for viral evolution. *J. Virol.*, **70**, 992-998.
- Martínez-Salas, E., Sáiz, J.-C., Dávila, M., Belsham, G.J. and Domingo, E. (1993) A single nucleotide substitution in the internal ribosome entry site of foot-and-mouth disease virus leads to enhanced cap-independent translation *in vivo*. *J. Virol.*, **67**, 3748-3755.
- Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911-940.
- Meerovitch, K., Svitkin, Y.V., Lee, H.S., Lejbkowitz, 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.
- Methot, N., Pickett, G., Keene, J.D. and Sonenberg, N. (1996) In vitro RNA selection identifies RNA ligands that specifically bind to eukaryotic translation initiation factor 4B: the role of the RNA recognition motif. *RNA*, **2**, 38-50.

- Meyer, K., Petersen, A., Niepmann, M. and Beck, E. (1995) Interaction of eukaryotic initiation factor eIF-4B with a picornavirus internal translation initiation site. *J. Virol.*, **69**, 2819-2824.
- Moustakas, A., Sonstegard, T.S. and Hackett, P.B. (1993) Effects of the open reading frames in the *Rous Sarcoma* virus leader RNA on translation. *J. Virol.*, **67**, 4350-4357.
- Nanbru, C., Lafon, I., Audiger, S., Gensac, G., S., V., Huez, G. and Prats, A.-C. (1997) Alternative translation of the proto-oncogene *c-myc* by an internal ribosome entry site. *J. Biol. Chem.*, **272**, 32061-32066.
- Nau, M.M., Brooks, B.J., Battey, J., Sausville, E., Gazdar, A.F., Kirsch, I.R., McBride, O.W., Bertness, V., Hollis, G.F. and Minna, J.D. (1985) *L-myc*, a new *myc*-related gene amplified and expressed in human small cell lung cancer. *Nature*, **318**, 69-73.
- Nepveu, A., Levine, R.A., Campisi, J., Greenberg, M.E., Ziff, E.B. and Marcu, K.B. (1987) Alternative modes of *c-myc* regulation in growth factor-stimulated and differentiating cells. *Oncogene*, **1**, 243-250.
- Nesbit, C.E., Tersak, J.M. and Prochownik, E.V. (1999) MYC oncogenes and human neoplastic disease. *Oncogene*, **18**, 3004-3016.
- Niepmann, M., Petersen, A., Meyer, K. and Beck, E. (1997) Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus. *J. Virol.*, **71**, 8330-8339.
- Nilsen, T.W. and Maroney, P.A. (1984) Translational efficiency of *c-myc* mRNA in Burkitt lymphoma cells. *Mol. Cell Biol.*, **4**, 2235-2238.
- Ohlmann, T. and Jackson, R.J. (1999) The properties of chimeric picornavirus IRESes show that discrimination between internal translation initiation sites is influenced by the identity of the IRES and not just the context of the AUG codon. *RNA*, **5**, 764-778.

- Ohlmann, T., Rau, M., Pain, V.M. and Morley, S.J. (1996) The C-terminal domain of eukaryotic protein synthesis initiation factor (eIF) 4G is sufficient to support cap-independent translation in the absence of eIF4E. *EMBO J.*, **15**, 1371-1382.
- Parkin, N., Darveau, A., Nicholson, R. and Sonenberg, N. (1988) *cis*-Acting translational effects of the 5' noncoding region of c-myc mRNA. *Mol. & Cell. Biol.*, **8**, 2875-2883.
- Pasternack, M.S., Bleier, K.J. and McInerney, T.N. (1991) Granzyme A binding to target cell proteins. *J. Biol. Chem.*, **266**, 14703-14708.
- 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. *Nucl. Acids Res.*, **26**, 3097-3103.
- Paulin, F.E.M., West, M.J., Sullivan, N.F., Whitney, R.L., Lyne, L. and Willis, A.E. (1996) Aberrant translational control of the c-myc gene in multiple myeloma. *Oncogene*, **13**, 505-513.
- Peabody, D.S. and Berg, P. (1986) Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Mol. Cell Biol.*, **6**.
- Pelham, H.R.B. and Jackson, R.J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.*, **67**, 247-256.
- Pelletier, J., Flynn, M.E., Kaplan, G., Racaniello, V. and Sonenberg, N. (1988) Mutational analysis of upstream AUG codons of poliovirus RNA. *J. Virol.*, **62**, 4486-4492.
- 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, I., Lin, C.H., McAfee, J.G. and Patton, J.G. (1997) Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA*, **3**, 764-778.

- Pestova, T. (1999) The mechanism of ribosomal subunit joining during initiation of eukaryotic translation. *C.N.R.S. - Conference Jacques Monod. New insights into the mechanism of mRNA translation: the significance of RNA structure*, Aussois, France, p. 5.
- Pestova, T.V., Borukhov, S.I. and Hellen, C.U.T. (1998a) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, **394**, 854-859.
- Pestova, T.V., Hellen, C.U. and Wimmer, E. (1991) Translation of poliovirus RNA: role of an essential cis-acting oligopyrimidine element within the 5' nontranslated region and involvement of a cellular 75-kilodalton protein. *J. Virol.*, **65**, 6194-6204.
- Pestova, T.V., Shatsky, I.N., Fletcher, S.P., Jackson, R.T. and Hellen, C.U.T. (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 Dev.*, **12**, 67-83.
- 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.
- Pilipenko, E.V., Gmyl, A.P., Maslova, S.V., Belov, G.A., Sinyakov, A.N., Huang, M., Brown, T.D.K. and Agol, V.I. (1994) Starting window, a distinct element in the cap-independent internal initiation of translation on picornaviral RNA. *J. Mol. Biol.*, **241**, 398-414.
- Pilipenko, E.V., Gmyl, A.P., Maslova, S.V., Svitkin, Y.V., Sinyakov, A.N. and Agol, V.I. (1992) Prokaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. *Cell*, **68**, 119-131.
- Pooggin, M.M., Futterer, J., Skryabin, K.G. and Hohn, T. (1999) A short open reading frame terminating in front of a stable hairpin is the conserved feature in pregenomic RNA leaders of plant pararetroviruses. *J. Gen. Virol.*, **80**, 2217-2228.

- Prendergast, G.C. (1999) Mechanisms of apoptosis by c-Myc. *Oncogene*, **18**, 2967-2987.
- Prochownik, E.V., Kukowska, J. and Rogers, C. (1988) c-myc antisense transcripts accelerate differentiation and inhibit G1-progression in murine erythroleukemia cells. *Mol. Cell Biol.*, **8**, 3683-3695.
- Ray, B.K., Lawson, T.G., Abramson, R.D., Merrick, W.C. and Thach, R.E. (1986) Recycling of messenger RNA cap-binding proteins mediated by eukaryotic initiation factor 4B. *J. Biol. Chem.*, **261**, 1466-1470.
- Ray, R., Thomas, S. and Miller, D.M. (1989) Mouse fibroblasts transformed with the human c-myc gene express a high level of mRNA but a low level of c-Myc protein and are non-tumorigenic in nude mice. *Oncogene*, **4**, 593-600.
- Reisman, D., Elkind, N.B., Roy, B., Beamon, J. and Rotter, V. (1993) c-Myc trans-activates the p53 promoter through a required downstream CACGTG motif. *Cell Growth and Differentiation*, **4**, 57-65.
- Reynolds, J.E., Kaminski, A., Carroll, A.R., Clarke, B.E., Rowlands, D.J. and Jackson, R.J. (1996) Internal initiation of translation of hepatitis C virus RNA: The ribosome entry site is at the authentic initiation codon. *RNA*, **2**, 867-878.
- Rijnbrand, R., vanderStraaten, T., vanRijn, P.A., Spaan, W.J.M. and Bredenbeek, P.J. (1997) Internal entry of ribosomes is directed by the 5' noncoding 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.
- Rivera, V.M., Welsh, J.D. and Maizel, J.V. (1988) Comparative sequence analysis of the 5' noncoding region of the enteroviruses and rhinoviruses. *Virology*, **165**, 42-50.

- Robertson, M.E.M., Seamons, R.A. and Belsham, G.J. (1999) A selection system for internal ribosome entry site (IRES) elements; analysis of the requirement for a conserved GNRA tetraloop in the encephalomyocarditis virus. *Translation UK*, University of Dundee, p. 31.
- Rogers, G.W., Richter, N.J. and Merrick, W.C. (1999) Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.*, **274**, 12236-12244.
- Rosenwald, I.B., Rhoads, D.B., Callanan, L.D., Isselbacher, K.J. and Schmidt, E.V. (1993) Increased expression of eukaryotic translation initiation factors eIF-4E and eIF-2 α in response to growth induction by *c-myc*. *Proc. Natl. Acad. Sci. USA*, **90**, 6175-6178.
- Ruan, H., Shantz, L.M., Pegg, A.E. and Morris, D.R. (1996) The upstream open reading frame of the mRNA encoding S-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. *J. Biol. Chem.*, **271**, 29576-295782.
- Ryan, K.M. and Birnie, G.D. (1996) *Myc* oncogenes: the enigmatic family. *Biochemical Journal*, **314**, 713-721.
- 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**, 1219-1226.
- Schneider, R.J. (1999) Translational control during stress and adenovirus translation. *C.N.R.S. - Conference Jacques Monod. New insights into the mechanism of mRNA translation: the significance of RNA structure*, Aussois, p. 5.
- Schwab, M., Alitalo, K., Klempnauer, K.H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M. and Trent, J. (1983) Amplified DNA with limited homology to MYC cellular oncogene is shared by human neuro-blastoma cell-lines and a neuro-blastoma tumor. *Nature*, **305**, 245-248.

- Sella, O., Gerlitz, G., Le, S.Y. and ElroyStein, O. (1999) Differentiation-induced internal translation of c-sis mRNA: Analysis of the cis elements and their differentiation-linked binding to the hnRNP C protein. *Mol. Cell Biol.*, **19**, 5429-5440.
- Shapiro, D.J., Sharp, P.A., Wahli, W.W. and Keller, M.K. (1988) A high-efficiency HeLa cell nuclear transcription extract. *DNA*, **7**, 47-55.
- Shichiri, M., Hanson, K.D. and Sedivy, J.M. (1993) The effects of c-myc expression on proliferation, quiescence and the G0 to G1 transition in nontransformed cells. *Cell Growth and Differentiation*, **4**, 93-104.
- Shindo, H., Tani, E., Matsumoto, T., Hashimoto, T. and Furuyama, J. (1993) Stabilization of c-myc protein in human glioma-cells. *Acta Neuropathologica*, **86**, 345-352.
- Spirin, A.S. (1966) On 'masked' forms of messenger RNA in early embryogenesis and in other differentiating systems. *Curr. Top. Dev. Biol.*, **1**, 1-38.
- Stern, S., Moazed, D. and Noller, H.F. (1988) Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.*, **164**, 481-489.
- Stewart, S.R. and Semler, B.L. (1997) RNA determinants of picornavirus cap-independent translation initiation. *Semin. Virol.*, **8**, 242-255.
- Stoneley, M. (1998) Functional analysis of the 5' untranslated region of the c-myc proto-oncogene. *Department of Biochemistry*. University of Leicester, Leicester.
- Stoneley, M., Chappell, S.A., Jopling, C.L., Dickens, M., MacFarlane M. and Willis, A.E. (2000a) c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis. *Mol. Cell Biol.*, **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 ribosomal entry segment. *Oncogene*, **16**, 423-428.

- Stoneley, M., Subkhankulova, T., Le Quesne, J.P.C., Coldwell, M.J., Jopling, C.L., Belsham, G.J. and Willis, A.E. (2000b) Analysis of the *c-myc* IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. *Nucl. Acids Res.*, **28**, 687-694.
- Tata, J.R. (1972) Isolation of nuclei from liver and other tissues. *Methods Enzymol.*, **XIIB**, 253-262.
- Taub, R., Moulding, C., Battey, J., Latt, S., Lenoir, G.M. and Leder, P. (1984) Activation and somatic mutation of the translocated *c-myc* gene in Burkitt-lymphoma cells. *Cell*, **36**, 339-348.
- Trachsel, H. and Staehelin, T. (1979) Initiation of mammalian protein synthesis. The multiple functions of the initiation factor eIF-3. *Biochem. Biophys. Acta*, **565**, 305-314.
- Valcarcel, J. and Gebauer, F. (1997) Post-transcriptional regulation: The dawn of PTB. *Curr. Biol.*, **7**, R705-R708.
- 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.
- Wadkins, T.S., Perrotta, A.T., Ferré-D'Amaré, A.R., Doudna, J.A. and Been, M.D. (1999) A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA*, **5**, 720-727.
- Wang, C.Y., Le, S.Y., Ali, N. and Siddiqui, A. (1995) An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis-C virus 5'-noncoding region. *RNA*, **1**, 526-537.
- West, M.J., Sullivan, N.F. and Willis, A.E. (1995) Translational upregulation of the *c-myc* oncogene in Bloom's syndrome cell lines. *Oncogene*, **13**, 2515-2524.

- Williams, M.A. and Lamb, R.A. (1989) Effects of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins. *J. Virol.*, **63**, 28-35.
- Willis, A.E. (1999) Translational control of growth factor and proto-oncogene expression. *Int. Journal of Biochem. & Cell Biol.*, **31**, 73-86.
- Yang, Q. 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. *Nucl. Acids Res.*, **25**, 2800-2807.
- Yeilding, N.M., Procopio, W.N., Rehman, M.T. and Lee, W.M.F. (1998) c-myc mRNA is down-regulated during myogenic differentiation by accelerated decay that depends on translation of regulatory coding elements. *J. Biol. Chem.*, **273**, 15749-15757.
- Yoon, H. and Donahue, T.F. (1992) The *Sui1* suppressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during tRNA^{Met} recognition of the start codon. *Mol. Cell Biol.*, **12**, 248-260.
- Yoshida, M., Meksuriyen, D., Kashiwagi, K., Kawai, G. and Igarashi, K. (1999) Polyamine stimulation of the synthesis of oligopeptide-binding protein (OppA) - Involvement of a structural change of the Shine-Dalgarno sequence and the initiation codon AUG in OppA mRNA. *J. Biol. Chem.*, **274**, 22723-22728.
- Zhang, L., Liu, W.Q. and Grabowski, P.J. (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA*, **5**, 117-130.
- Zuker, M., Mathews, D.H. and Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski, J. and Clark, B.F.C. (eds.), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers.

Publications

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 ribosomal entry segment. *Oncogene* **16**(3), 423-428.

Stoneley, M., Subkhankulova, T., Le Quesne, J. P. C., Coldwell, M. J., Jopling, C. L., Belsham, G. J., and Willis, A. E. (2000b). Analysis of the *c-myc* IRES; a potential role fo cell-type specific trans-acting factors and the nuclear compartment. *Nucleic Acids Research* **28**(3), 687-694.



SHORT REPORT

C-Myc 5' untranslated region contains an internal ribosome entry segment

Mark Stoneley, Fiona EM Paulin, John PC Le Quesne, Stephen A Chappell and Anne E Willis

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

Translation in eukaryotic cells is generally initiated by ribosome scanning from the 5' end of the capped mRNA. However, initiation of translation may also occur by a mechanism which is independent of the cap structure and in this case ribosomes are directed to the start codon by an internal ribosome entry segment (IRES). Picornaviruses represent the paradigm for this mechanism, but only a few examples exist which show that this mechanism is used by eukaryotic cells. In this report we show data which demonstrate that the 5' UTR of the proto-oncogene *c-myc* contains an IRES. When a dicistronic reporter vector, with *c-myc* 5' UTR inserted between the two cistrons, was transfected into both HepG2 and HeLa cells, the translation of the downstream cistron was increased by 50-fold, demonstrating that translational regulation of *c-myc* is mediated through cap-independent mechanisms. This is the first example of a proto-oncogene regulated in this manner and suggests that aberrant translational regulation of *c-myc* is likely to play a role in tumorigenesis.

Keywords: internal ribosome entry segment (IRES); *c-myc* 5' UTR; internal initiation; translation

The 5' untranslated region (UTR) of *c-myc* (which is well conserved amongst species) plays a significant role in modulating the steady state levels of the *c-myc* protein. A translational control mechanism residing in the first exon was originally postulated by Saito *et al.* arising from differential hypothetical secondary structures as a result of chromosomal translocations (Saito *et al.*, 1983). In addition, *c-myc* mRNAs lacking exon 1 were found to be translated more efficiently *in vitro* when compared to full length transcripts (Darveau *et al.*, 1985). Furthermore, a 240 nt restrictive element within exon 1 of murine *c-myc* was isolated and shown to inhibit translation of heterologous mRNAs in rabbit reticulocyte lysate and wheat germ extract (Parkin *et al.*, 1988) demonstrating that the 5' UTR is highly structured and inhibitory to the scanning mechanism of translation. However, early studies *in vivo* examining translational efficiencies of *c-myc* mRNA in Burkitt's lymphoma cell lines suggested that both truncated and full length transcripts were translated with equal efficiencies (Nilsen and Maroney, 1984). Moreover, when expressed in numerous cell lines or translated in HeLa cell extracts the 5' UTR does not inhibit translation of either *c-myc* or reporter genes (Butnick *et al.*, 1985; Parkin *et al.*, 1988). This disparity suggests that non-canonical factors, which are lacking in rabbit

reticulolysate and wheat germ extract, facilitate *c-myc* translation through the 5' UTR *in vivo* (Parkin *et al.*, 1988).

Translational regulation mediated through the 5' UTR is not unique to *c-myc*. Many cellular mRNAs encoding proto-oncogenes, growth factors, receptors and transcription factors possess long, highly structured 5' UTRs which affect their regulation (Gray and Hentze, 1994). For the overwhelming majority of eukaryotic mRNAs, where initiation of protein synthesis occurs via a cap-dependent mechanism (involving binding of the eukaryotic initiation factor (eIF) 4E, to the 7-methyl G cap of the mRNA, for review see Hershey, 1991), such elements influence translation of the mRNA by repressing this cap-dependent mechanism. Alternatively, structured 5' UTRs may contain an internal ribosome entry segment (IRES) which allows cap-independent translation. These IRESes are capable of directing ribosomes to an internal start codon which may be some considerable distance (600–1000 nts) from the 5' end of the message (for reviews see Jackson *et al.*, 1994, 1995; Jackson and Kaminski, 1995). The eukaryotic mRNAs which have so far been demonstrated to contain IRESes include the human immunoglobulin heavy chain binding protein (Macejak and Sarnow, 1991), basic fibroblast growth factor (Vagner *et al.*, 1995) and eukaryotic initiation factor 4G (Gan and Rhoads, 1996). These may exemplify a group of mRNAs whose translation is required even when cap-dependent activity is compromised. However, to date no mechanisms have been elucidated for these eukaryotic IRESes and the cellular circumstances under which internal ribosome entry is required have yet to be fully defined (Sarnow, 1989; Vagner *et al.*, 1995).

In cell lines derived from patients with Bloom's syndrome and Multiple Myeloma we have shown that de-regulated *c-myc* expression occurred by a translational mechanism (West *et al.*, 1995; Paulin *et al.*, 1996) and in the latter case a specific mutation was found in the 5' UTR of *c-myc*. In this paper we demonstrate that the 5' UTR of *c-myc* contains an IRES. This is the first example of a proto-oncogene which can utilise such a method to initiate protein synthesis and deregulation of *c-myc* via such a mechanism would have profound implications for tumorigenesis.

c-myc 5' UTR does not inhibit translation in cultured cells

Four promoters have been identified in the *c-myc* proto-oncogene; P1, P2, P3 and P0 which give rise to transcripts of approximately 2.4 kb, 2.25 kb, 2.0 kb

and 3.1 kb respectively (Battey *et al.*, 1983; Bentley and Groudine, 1986; Yang *et al.*, 1985). P2 is the major promoter from which 75–90% of cellular transcripts originate, with P1 producing only 10–25% (Stewart *et al.*, 1984). Transcripts initiated at P0, P1 and P2 give rise to 5' untranslated regions of approximately 1000, 600 and 400 nucleotides respectively. *In vitro* studies on the 5' UTR of the P2 transcript have demonstrated that this region is highly structured and inhibitory to ribosome scanning (Darveau *et al.*, 1985; Parkin *et al.*, 1988). To

determine the effect of 5' UTR *in vivo*, we inserted a 396 bp segment into the plasmid pGL3 (Promega) directly upstream of the coding region for firefly luciferase to create the plasmid construct pGL3utr (Figure 1a). The pGL3utr construct and the control vector pGL3 were transfected into HeLa and HepG2 cells. The 5' UTR was not found to inhibit the downstream luciferase expression (Figure 1b) hence the activity of luciferase produced from pGL3utr was 1.2–1.6-fold higher than that produced from the control vector pGL3 (Figure 1b).

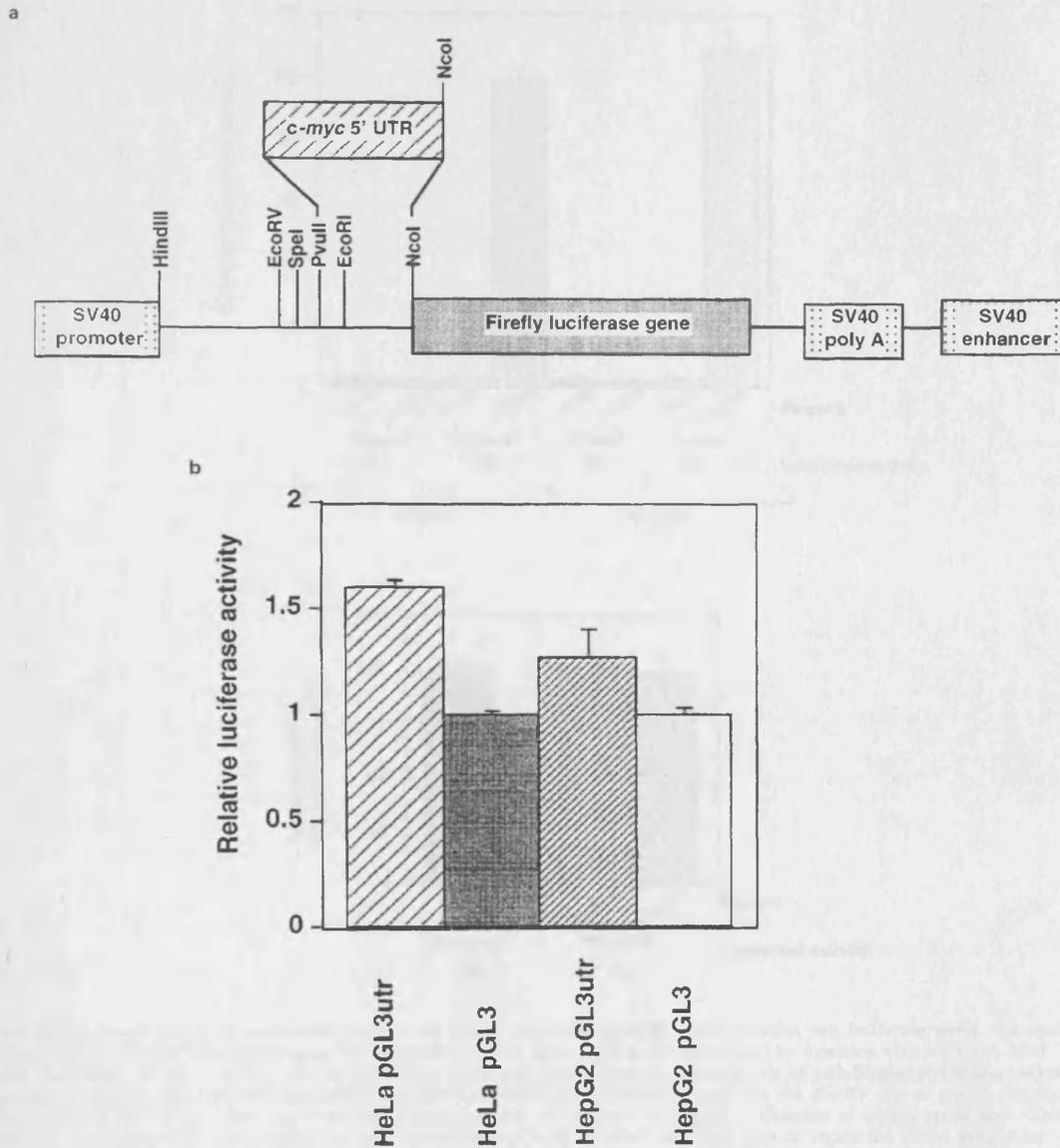


Figure 1 (a) Construction of monocistronic vectors. The *c-myc* 5' UTR was amplified using the primers FP2501, 5'-TAATTCAGCGAGAGGCAGA-3' and MS4519 5'-ATACCATGGTTCGCGGGAGGCTGCT-3'. Amplification resulted in a fragment of 396 bp which is contained within the region from 2501–4519 in the genomic sequence (Watt *et al.*, 1983). This sequence was inserted into the control vector pGL3 (Promega) proximal to the firefly luciferase (FL) gene, using the *PvuII* and *NcoI* sites creating the vector pGL3utr. (b) The effect of the *c-myc* 5' UTR on a downstream cistron. HeLa and HepG2 cells were transfected with 20 µg of the luciferase constructs (pGL3 or pGL3utr) and 5 µg of the β-galactosidase construct pcDNA3.1/HisB/LacZ (Invitrogen) by the calcium phosphate method (Ausubel *et al.*, 1987). Cells were harvested after 48 h and luciferase expression was determined using a luciferase assay system (Promega) and β-galactosidase expression was determined using a Galactolight plus system (Tropix). Both activities were measured in a 1253 Luminometer (BioOrbit). Variations in transfection efficiency were corrected by normalising luciferase activity to β-galactosidase activity. The results presented are an average of three independent experiments

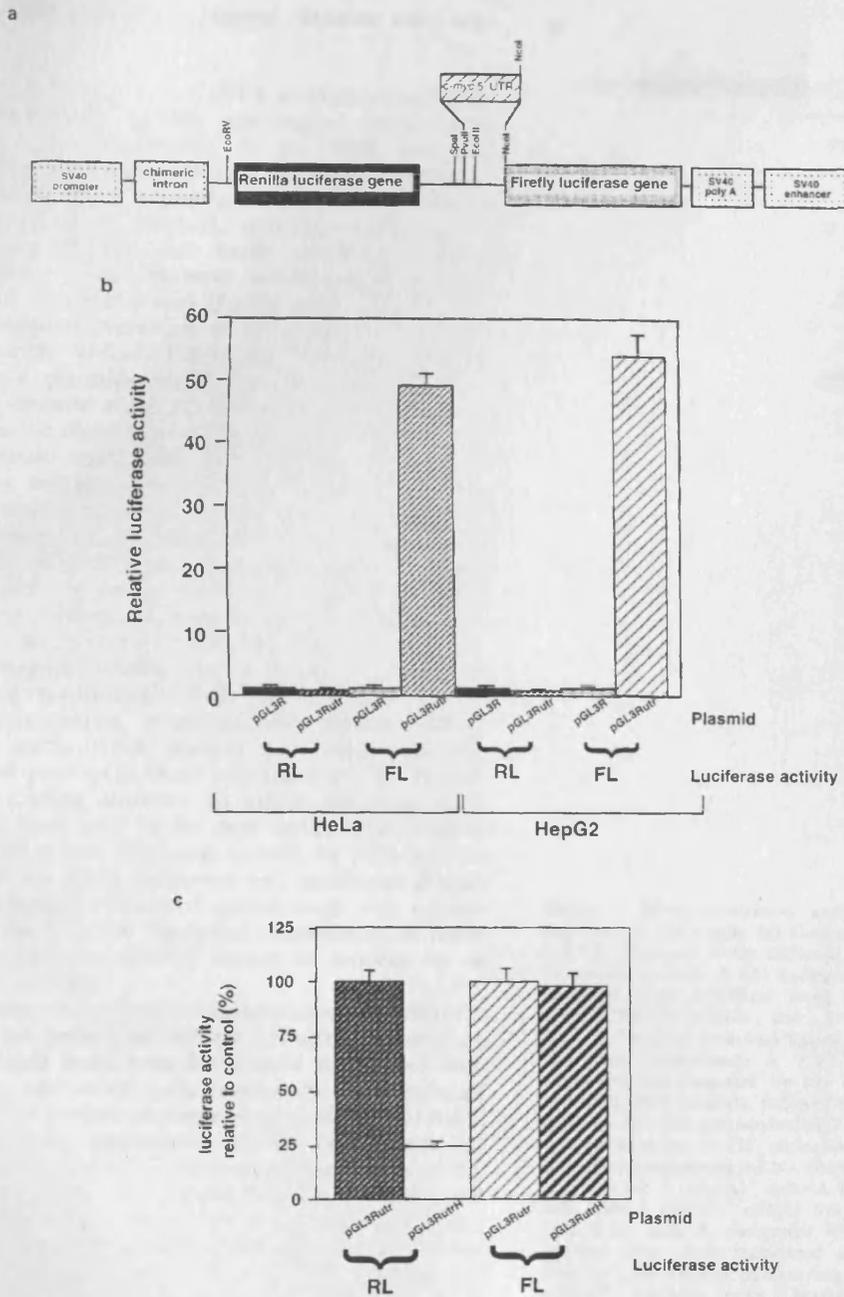


Figure 2 (a) Construction of dicistronic vectors. To create the vector pGL3R which contains two luciferase genes, the coding region of the *Renilla* luciferase (RL) gene was obtained from the vector pRL-CMV (Promega) by digestion with *NheI* and *XbaI*. To extend the length of the 3' UTR, this fragment was blunt end ligated into the *HindIII* site of pSKBluescript (Stratagene) and subsequently excised using *EcoRV* and *XhoI*. This DNA segment was blunt-end ligated into the *EcoRV* site of pGL3. Finally, a chimeric intron from pRL-CMV was blunt-end ligated into the *HindIII* site to minimise utilisation of cryptic splice sites. The 5' UTR of *c-myc* (generated as in Figure 1a) was inserted into pGL3R at *PvuII* and *NcoI* sites to create the vector pGL3Rutr. (b) Expression of Renilla and firefly luciferase from dicistronic mRNAs in HeLa and HepG2 cells. The dicistronic constructs were transfected into HeLa and HepG2 cells as before. Both luciferase activities were measured using the Dual-Luciferase reporter assay system (Promega). The values were normalised to β -galactosidase activity as in Figure 1 and luciferase activities obtained are expressed relative to those obtained for pGL3R. The results presented are an average from three independent experiments. (c) An oligonucleotide cassette 5'-AGATCTGGTACCGAGCTCCCCGGGCTGCAGGAT-3' and 5'-ATCCTGCAGCCCCGGGACCTCGGTACCAGATCT-3' containing an internal *PstI* site was inserted into the *EcoRV* site of pGL3Rutr. This vector was digested with *PstI* and *EcoRV* and the same oligonucleotide cassette was excised with *PstI* and ligated into these sites creating a 60 bp palindromic sequence upstream of the Renilla coding sequence. This results in the production of a hairpin structure with an energy of -55 Kcal/mol. The vector was transfected into HeLa cells and luciferase activity measured as before. The luciferase activities were normalised to β -galactosidase and expressed relative to those obtained from vector pGL3Rutr. The results presented are an average of three independent experiments.

c-myc 5' UTR contains an internal ribosome entry segment

The inability of the c-myc 5' UTR to inhibit translation *in vivo* compared to the previously demonstrated inhibition *in vitro* (Parkin *et al.*, 1988 and our unpublished data) led us to test the hypothesis that it contains an IRES. We inserted the c-myc 5' UTR into a dicistronic reporter plasmid, pGL3R, in the spacer between the Renilla and firefly luciferase cistrons (Figure 2a). These plasmid constructs were then transfected into HeLa and HepG2 cells. The c-myc 5' UTR stimulated expression of the downstream cistron approximately 50-fold (Figure 2b) when compared to the control plasmids which lack this sequence. The apparent increase in the translation of the downstream cistron on the dicistronic message containing the c-myc 5' UTR could result from various mechanisms. The 5' UTR may contain an element which stimulates readthrough past the Renilla luciferase cistron and thus causes reinitiation of translation at the downstream cistron. Alternatively, monocistronic firefly luciferase mRNAs may be produced by transcriptional, RNA cleavage or splicing mechanisms. Finally, the c-myc 5' UTR may direct internal ribosome entry.

To determine whether the 5' UTR is capable of stimulating readthrough from the upstream to the downstream cistron, a palindromic sequence which forms a stable RNA hairpin (-55 Kcal/mol) was introduced into pGL3Rutr upstream of the Renilla luciferase coding sequence to inhibit ribosome scanning. The stem loop in the new vector, pGL3RutrH, reduced the renilla luciferase activity by 75% but the activity of the firefly luciferase was unaffected (Figure 2c). If enhanced ribosomal readthrough was responsible for the 5' UTR dependent stimulation of firefly luciferase then this activity should be reduced by an equivalent amount.

To address the potential fragmentation of dicistronic mRNAs we performed RNase protection assays on RNA isolated from both HeLa cells transfected with pGL3Rutr and mock transfected cells. In transfected cells a 725 nt probe complementary to 624 nts of the 5' UTR containing dicistronic mRNA (see Figure 3a) protected a fragment of the expected size (Figure 3b, lane 3). In addition, protected fragments of 395 and 382 nts were detected in both transfected and mock transfected samples resulting from hybridisation of the probe to endogenous c-myc transcripts (Figure 3b, lanes 2 and 3). The presence of functional monocistronic transcripts would result in smaller protected fragments of at least 101 nts in length, and since no products of this size were detected the increased expression of firefly luciferase must occur on intact dicistronic mRNAs.

Thus we conclude that c-myc 5' UTR contains an IRES. The small, but reproducible, reduction in the expression of luciferase from the upstream cistron of between 15-20% in the cells which contain the plasmid pGL3Rutr, is also consistent with this hypothesis (Figure 2b). This probably reflects a competition between cap-dependent and IRES-dependent translation on the dicistronic mRNA and this phenomenon has also been observed for Bip, picornavirus and eIF4G IRESes (Macejak and Sarnow, 1991; Borman and Jackson, 1992; Gan and Rhodes, 1996).

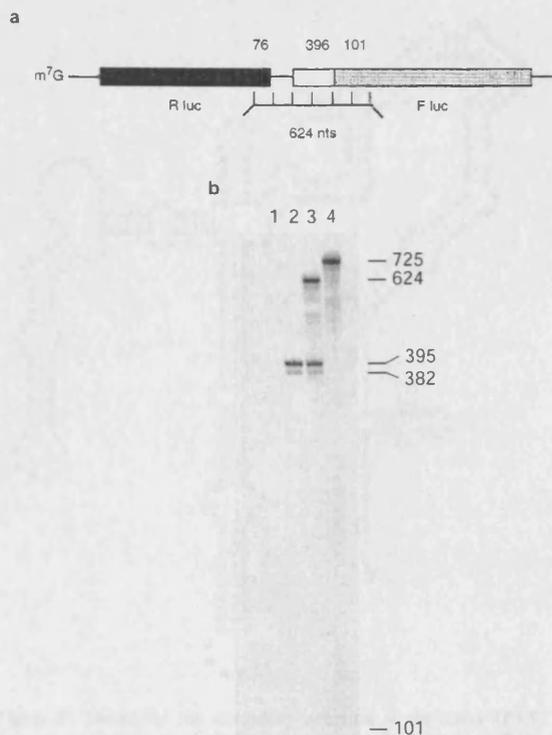


Figure 3 RNase protection analysis of dicistronic mRNAs expressed in HeLa cells. (a) Diagram of the 5' UTR containing mRNA hybridised to the antisense RNA probe used for RNase protection analysis. A 624 nucleotide DNA fragment was PCR amplified from pGL3Rutr using the primers 5'-GCAAGAA-GATGCACCTGATG-3' and 5'-GCGTATCTCTTCAGAGC-CTT-3'. This was blunt-end ligated into the *Sma*I site of pSK⁺ Bluescript (Stratagene). A [³²P]CTP (800 Ci/mmol) labelled riboprobe was generated by run off transcription on a *Xho*I restricted DNA template, followed by DNase I digestion and gel isolation on a 4% polyacrylamide/7 M urea gel. (b) Ribonuclease protection assay. A 725 nucleotide RNA probe was used to protect complementary mRNA fragments. Lane 1, 10 µg of yeast tRNA. Lane 2, poly(A)⁺ mRNA from mock transfected HeLa cells. Lane 3, poly(A)⁺ mRNA from HeLa cells transfected with pGL3Rutr. Lane 4, undigested RNA probe. Total RNA was isolated from mock transfected and transfected HeLa cells. Poly(A)⁺ mRNA was purified from 10 µg of total RNA using oligo(dT) magnetic beads (Dynatec Inc). RNA samples were hybridised with 5 × 10⁵ c.p.m. of riboprobe at 45°C for 16 h in hybridisation buffer (40 mM PIPES pH 6.4, 400 mM NaCl, 1 mM EDTA, 80% deionised formamide). Single stranded RNA was digested using RNase ONE (Promega). The products were size fractionated on a 4% polyacrylamide/7 M urea gel and visualised by phosphorimage analysis (Molecular Dynamics). Product sizes were determined using ³²P-dCTP labelled pBR322 *Hpa*II restriction fragments

Mapping the c-myc IRES

To define the boundaries of the c-myc IRES a series of plasmid constructs was generated containing decreasing lengths of the sequence coding for the 5' UTR. The ability of these truncated sequences to promote internal ribosome entry on a dicistronic mRNA was compared

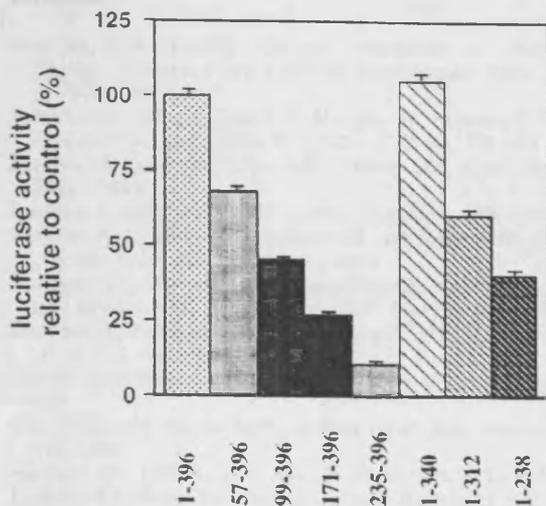


Figure 4 Downstream cistron activity from dicistronic mRNAs containing fragments of the *c-myc* 5' UTR in HeLa cells. The 5' deletion series was generated by restricting the *c-myc* 5' UTR with *NciI*, *AccIII* and *AvaI* giving fragments of 340, 298 and 162 respectively. A fragment of 226 bp was amplified using the oligonucleotides FP 2670 5'-TGCCATCCACGAAACTTT-3' and MS4519 (see Figure 1). These sequences were inserted into pGL3R at the *PvuII* and *NcoI* sites. Deletions from the 3' end were produced by digesting the *c-myc* 5' UTR with *PvuII*, *EcoR0901I* and *AvaI*, generating fragments of 340, 312 and 238 bp. Fragments were inserted by blunt-end ligation into the *PvuII* site of pGL3R. The resulting constructs were then transfected into HeLa cells and luciferase activity measured and calculated as before. The reduction in luciferase activity from the downstream cistron is expressed as percentage of the values obtained with pGL3Rutr

to the full length 5' UTR. Removal of 56 nts from the 5' end decreased the activity of the downstream cistron by 33% and larger deletions of 98, 170 and 234 nts resulted in a corresponding reduction in internal ribosome entry of 56, 74 and 91% respectively (Figure 4). Thus the 5' border of this translational element lies within 56 nts of the 5' end of exon 1.

At the 3' end, deletion of 56 nts had no effect on the activity of the downstream cistron. However, deletions further upstream removing 84 and 158 nts reduced the efficiency of the internal ribosome entry by 40 and 60% respectively (Figure 4). Hence the 3' end of the optimally effective IRES lies between 312 and 340 nts from the 5' end. Furthermore, this analysis suggests a mechanistic distinction between viral IRESes and the *c-myc* IRES. In various viral IRESes 3' end deletions that lie within the IRES completely ablate internal ribosome entry (Pelletier and Sonenberg, 1988; Borman and Jackson, 1992; Borman et al., 1995; Reynolds et al., 1995), whereas *c-myc* IRES 3' end deletions result in a gradual loss of activity. This may reflect a structural difference between cellular and viral IRESes. We have performed phylogenetic and energy minimization analyses on the *c-myc* 5' UTR and obtained a model for the secondary structure (Figure 5). Noteworthy features include the high degree of foldback in the structure when compared to those predicted for viral IRESes, and the absence of cryptic AUGs in all sequences examined.

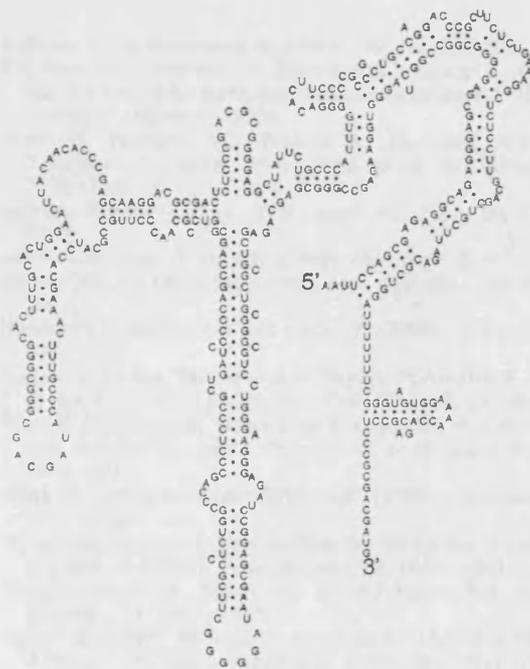


Figure 5 Model for the secondary structure of the *c-myc* IRES. Energy minimization analysis was performed using M Zuker's 'mfold' package (Zuker, 1989). Phylogenetic analysis was performed on aligned sequences derived from human, gibbon, marmoset, woodchuck, mouse, rat, cat, sheep and pig tissues. The model was drawn using the 'CARD' program (Winnenpenninckx et al., 1995)

The preceding data demonstrate that the *c-myc* 5' UTR contains a translational element capable of directing internal ribosome entry and we propose that *c-myc* protein synthesis may therefore be initiated by such a mechanism. This suggests that the *c-myc* protein can be translated under situations where initiation from the 5' cap structure and ribosome scanning is reduced. There are a number of situations where modulation in the levels of *c-myc* protein via internal ribosome entry may be required including the onset of proliferation, during mitosis where cap-dependent translation is reduced (Jackson et al., 1995) and following DNA damage (Sullivan and Willis, 1989).

Deregulation of the *c-myc* proto-oncogene through enhanced internal ribosome entry could play a pivotal role in tumour development. We have described previously two cases where deregulation of *c-myc* by translational mechanisms occurs in cell lines derived from patients with Bloom's syndrome (West et al., 1995) and in multiple myeloma (Paulin et al., 1996).

Further work is merited to investigate the mechanism of action of this IRES, the pathophysiological circumstances under which it is used and the effects that the mutation in this region has on the aberrant translational regulation of *c-myc* in multiple myeloma.

Acknowledgements

This work was supported by grants from the MRC (FEMP) and LRF (SAC). MS and JPCLeQ are funded by PhD studentships from the MRC.

References

- Ausubel RM. (1987). *Current Protocols in Molecular Biology*. Greene J (ed.). Wiley-Interscience: New York, pp. 9.11–9.17.
- Batty J, Moulding C, Taub R, Murphy W, Stewart T, Potter H, Lenoir G and Leder P. (1983). *Cell*, **34**, 779–797.
- Bentley DL and Groudine M. (1986). *Mol. Cell. Biol.*, **6**, 3481–3489.
- Borman A and Jackson RJ. (1992). *Virology*, **188**, 685–696.
- Borman AM, Bailly J-L, Girard M and Kean KM. (1995). *Nucleic Acids Res.*, **23**, 3656–3663.
- Butnick NZ, Miyamoto C, Chizzonite R, Cullen BR, Ju G and Skalka AM. (1985). *Mol. Cell. Biol.*, **5**, 3009–3016.
- Darveau A, Pelletier J and Sonenberg N. (1985). *Proc. Natl. Acad. Sci. USA*, **82**, 2315–2319.
- Gan W and Rhoads RE. (1996). *J. Biol. Chem.*, **271**, 623–626.
- Gray NK and Hentze MW. (1994). *Mol. Biol. Reports*, **19**, 195–200.
- Hershey JW. (1991). *Ann. Rev. Biochem.*, **60**, 717–755.
- Jackson RJ, Hunt SL, Gibbs CL and Kaminski A. (1994). *Mol. Biol. Reports*, **9**, 147–159.
- Jackson RJ, Hunt SL, Reynolds JE and Kaminski A. (1995). *Curr. Top. Microbiol. Immunol.*, **203**, 1–29.
- Jackson RJ and Kaminski A. (1995). *RNA*, **1**, 985–1000.
- Macejak DG and Sarnow P. (1991). *Nature*, **353**, 90–94.
- Nilsen TW and Maroney PA. (1984). *Mol. Cell. Biol.*, **4**, 2235–2238.
- Parkin N, Darveau A, Nicholson R and Sonenberg N. (1988). *Mol. Cell. Biol.*, **8**, 2875–2883.
- Paulin FEMP, West MJ, Sullivan NF and Willis AE. (1996). *Oncogene*, **13**, 505–513.
- Pelletier J and Sonenberg N. (1988). *Nature*, **334**, 320–325.
- Reynolds JE, Kaminski A, Kettinen KJ, Grace K, Clarke BE, Carroll AR, Rowlands DJ and Jackson RJ. (1995). *EMBO J.*, **14**, 6010–6020.
- Saito H, Hayday AC, Wiman K, Hayward WS and Tonegawa S. (1983). *Proc. Natl. Acad. Sci. USA*, **80**, 7476–7480.
- Sarnow P. (1989). *Proc. Natl. Acad. Sci. USA*, **86**, 5795–5799.
- Sullivan NF and Willis AE. (1989). *Oncogene*, **4**, 497–502.
- Spotts GD and Hann SR. (1990). *Mol. Cell. Biol.*, **10**, 3952–3964.
- Stewart TA, Bellve AR and Leder P. (1984). *Science*, **226**, 707–710.
- Vagner S, Gensac MC, Maret A, Bayard F, Amalric F, Prats H and Prats AC. (1995). *Mol. Cell. Biol.*, **15**, 35–44.
- Watt R, Nishikura K, Sorrentino J, ar-Rushdi A, Croce CM and Rovera G. (1983). *Proc. Natl. Acad. Sci. USA*, **80**, 6307–6311.
- West M, Sullivan NF and Willis AE. (1995). *Oncogene*, **11**, 2515–2524.
- Winnenpenninckx B, Van de Peer Y, Backeljau T and De Wachter R. (1995). *Biotechniques*, **18**, 1060–1063.
- Yang J, Bauer SR, Mushinski JF and Marcu KB. (1985). *EMBO J.*, **4**, 1441–1447.
- Zuker M. (1989). *Methods in Enzymology*, Dahlberg JE and Abelson JN (eds). Academic Press Inc: New York, pp. 262–288.

Analysis of the *c-myc* IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment

Mark Stoneley, Tatyana Subkhankulova, John P. C. Le Quesne, Mark J. Coldwell, Catherine L. Jopling, Graham J. Belsham and Anne E. Willis*

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

Received October 21, 1999; Revised and Accepted December 8, 1999

ABSTRACT

The 5' UTR of *c-myc* mRNA contains an internal ribosome entry segment (IRES) and consequently, *c-myc* mRNAs can be translated by the alternative mechanism of internal ribosome entry. However, there is also some evidence suggesting that *c-myc* mRNA translation can occur via the conventional cap-dependent scanning mechanism. Using both bicistronic and monocistronic mRNAs containing the *c-myc* 5' UTR, we demonstrate that both mechanisms can contribute to *c-myc* protein synthesis. A wide range of cell types are capable of initiating translation of *c-myc* by internal ribosome entry, albeit with different efficiencies. Moreover, our data suggest that the spectrum of efficiencies observed in these cell types is likely to be due to variation in the cellular concentration of non-canonical translation factors. Interestingly, the *c-myc* IRES is 7-fold more active than the human rhinovirus 2 (HRV2) IRES and 5-fold more active than the encephalomyocarditis virus (EMCV) IRES. However, the protein requirements for the *c-myc* IRES must differ significantly from these viral IRESs, since an unidentified nuclear event appears to be a pre-requisite for efficient *c-myc* IRES-driven initiation.

INTRODUCTION

The proto-oncogene *c-myc* is required for both cell proliferation and programmed cell death (apoptosis), and de-regulated *c-myc* expression is associated with a wide range of cancers (1,2). It is therefore not surprising that *c-myc* gene expression is tightly controlled at multiple levels (3). The post-transcriptional regulation of *c-myc* involves alterations in the stability of both the mRNA and the protein (4-7), and the control of *c-myc* translation (8-12)

In common with many other genes involved in the regulation of cell growth, the *c-myc* mRNA has a long and potentially highly structured 5' untranslated region (UTR, located in exon 1). Multiple transcription start sites exist within the gene, giving rise to four transcripts (P0, P1, P2 and P3, with sizes of ~3.1, 2.4, 2.25 and 2.0 kb respectively; 13-15), with the

predominant mRNA (P2) having a 5' UTR of ~400 nt. It has been suggested that mRNAs with structured 5' UTRs, such as *c-myc*, are poorly translated due to their reduced ability to associate with the cap-binding complex, the eukaryotic initiation factor 4F (eIF4F). Indeed, over-expression of the cap-binding protein eIF4E, which is believed to be a limiting component of this complex, causes an increase in the translation of mRNAs with structured 5' UTRs such as *c-myc* (16-18). Furthermore, in certain circumstances the translational regulation of *c-myc* is mediated by phosphorylation and inactivation of the eIF4E inhibitor protein 4EBP1 (19).

It has also been shown that the 5' UTR of *c-myc* contains an internal ribosome entry segment (IRES) (11,12). IRESs were originally identified in the 5' UTRs of picornaviral RNAs and these complex structural elements allow ribosomes to enter at a considerable distance (often >1000 nt) from the 5' end of the mRNA (20-22). Several eukaryotic mRNAs have the potential to initiate translation by an internal ribosome entry mechanism and interestingly many of the mammalian IRESs identified to date have been found in genes whose protein products are associated with the control of cell growth, e.g. *c-myc*, fibroblast growth factor -2 (FGF-2), platelet derived growth factor (PDGF) and vascular endothelial growth factor (VEGF) (11,12,23-26).

The region of *c-myc* mRNA that contains the IRES is located downstream of the P2 promoter (12). Approximately 75-90% of *c-myc* transcripts are synthesised from this promoter (3). Therefore, the majority of *c-myc* mRNAs have the potential to initiate translation via internal ribosome entry. The *c-myc* IRES appears to function under conditions where cap-dependent translation is compromised. Indeed, we have recently shown that the *c-myc* IRES is utilised during apoptosis when cap-dependent translation is reduced due to cleavage of eIF4G (27). Furthermore, in poliovirus-infected HeLa cells, in which there is a substantial reduction in cap-dependent protein synthesis due to the proteolysis of eIF4G and sequestration of eIF4E, *c-myc* mRNAs remain associated with heavy polysomes (28). However, since there is some evidence that *c-myc* mRNA can also be translated by a cap-dependent mechanism, to date it has not been possible to assess the contribution that either mechanism makes to the synthesis of c-Myc polypeptides (12,19).

In this study we present further evidence for the existence of an IRES in the *c-myc* 5' UTR. In addition our data confirm that

*To whom correspondence should be addressed. Tel: +44 116 2523363; Fax: +44 116 2523369; Email: aew5@le.ac.uk

Present addresses:

Mark Stoneley, Department of Biochemistry and Molecular Biology, The University of Leeds, Leeds LS2 9JT, UK
Graham J. Belsham, BBSRC, Institute for Animal Health, Pirbright, Woking GU3 0NF, UK

c-myc mRNAs can also be translated by a cap-dependent mechanism. This has led us to propose that both mechanisms operate *in vivo*. We demonstrate that the *c-myc* IRES is active (with one exception) in all cell lines of human origin tested, although there is a wide variation in its efficiency, whereas the IRES is not active in cell lines of murine origin. When compared to IRESs of picornaviral origin, the *c-myc* IRES is 7- and 5-fold more active than the IRESs derived from HRV and EMCV, respectively. Finally we provide evidence that the *c-myc* IRES depends on a prior nuclear event for efficient initiation of translation.

MATERIALS AND METHODS

Cell culture

All cell lines were grown at 37°C in Dulbecco's modified Eagle's medium supplemented with 10% foetal calf serum, in a humidified atmosphere containing 5% CO₂. The cell lines HeLa (Human cervical epitheloid carcinoma), HepG2 (Human hepatocyte carcinoma), HK293 (Human embryonic kidney cell line immortalised with adenovirus DNA), Balb/c-3T3 (Murine embryonic fibroblast cell line), MCF7 (Human breast carcinoma), Cos-7 [Monkey epithelial cell line (CV-1) immortalised with SV40 DNA] and MEL cells (murine erythroleukaemic cells) were purchased from the American type culture collection. The cell line MRC5 (human lung fibroblast) was a kind gift from Dr M. MacFarlane (MRC-Human Toxicology Unit, Leicester, UK). The human SV40 immortalised fibroblast cell line GM637 was obtained from NIGMS.

Plasmid constructs

The plasmids pGL3, pGML (formerly pGL3utr), pRF and pRMF (formerly pGL3R and pGL3Rutr) have been described previously (12). cDNA encoding the HRV2 IRES was obtained from the plasmid pXLJ(10-605) (a gift from Dr R. Jackson, University of Cambridge) and inserted into pRF between the *PvuII* and *NcoI* sites, thus creating pRhvF. To obtain the sequence encoding the EMCV IRES, a polymerase chain reaction (PCR) was performed using the oligonucleotides 5'-GATGACTAGTCCGCCCTCTCCCTCCCCC-3' and 5'-GATGCCATGGC-CATATTATCATCGTGTT-3', with pCAGSIP (an expression vector that contains the EMCV-IRES; a gift from Dr S. Monkley, University of Leicester, UK) as a template. Subsequently, the PCR product was inserted into pRF between the *SpeI* and *NcoI* sites to generate pRemcvF.

A DNA fragment containing a 60 bp palindromic sequence was amplified from pGL3RutrH (12) in a PCR using the oligonucleotides 5'-ACCTCGAGAGATATCTGGTACCGAGCTC-3' and 5'-ACAAGCTTAGATCTGGTACCGAGCTC-3'. This fragment was inserted into pGL3 and pGML at the *SpeI* site, thus creating pHpL and pHpML, respectively.

The *c-myc* P2 cDNA was obtained by reverse transcription and PCR amplification of HeLa cell total RNA, using Super-script reverse transcriptase and *Taq* DNA polymerase (Life Technologies Inc). The fragments encoding the P2 *c-myc* cDNA from -396 to +6 and +7 to +1320 were amplified using the primer sets 5'-TAATTCAGCGAGAGGCAGA-3' with 5'-GGGCATCGTCGCGGGAGGCTG-3', and 5'-CTCAAC-GTTAGCTTACCAAC-3' with 5'-CGGAATTCTTACGCA-CAAGAGTTGCCGAT-3', respectively. These sequences

were inserted sequentially into pSK+-bluescript (Stratagene) using the *SmaI* and *EcoRI* sites thus recreating the entire P2 cDNA in the plasmid pSKMyc. The construct pSKMycΔ1 containing the P2 sequence from -56 to +1320 bp, was obtained by inserting a 1381 bp *PvuII*-*EcoRI* fragment derived from pSKMyc into pSK+ bluescript between the *SmaI* and *EcoRI* sites. Both constructs were linearised with *HindIII* prior to performing *in vitro* transcription reactions.

To create the bicistronic plasmids, pCRF and pCRMF, DNA fragments containing the Firefly luciferase (*luc*) coding region or a 5' UTR-*luc* fusion were excised from pSKL and pSKutrL, respectively. These sequences were inserted into pRL-CMV (Promega) downstream of the *Renilla* luciferase coding region at the *XbaI* site.

The constructs in the pSP64R(x)L Poly A series were generated in two stages. Initially, the *Renilla* luciferase coding region was obtained from pRL-CMV and inserted into pSP64 Poly A (Promega) at the *XbaI* site. Subsequently, DNA fragments containing the luciferase coding region, a *c-myc* 5' UTR-*luc* fusion and a HRV2 IRES-*luc* fusion were excised from pGL3, pGML and pRhvF, respectively, and blunt-end ligated into the *SmaI* site of pSP64RPoly A downstream of the *Renilla* luciferase sequence. Constructs in this series were digested with *EcoRI* prior to inclusion in an *in vitro* transcription reaction. The resulting transcripts have a 3' terminal polyadenylate tail of 30 residues.

DNA transfections

Calcium phosphate-mediated DNA transfection of mammalian cells, with the exception of MRC5, MEL and GM637 cells, was performed essentially as described by Jordan *et al.* (29). The remaining cell lines were transfected with FuGene6 (Roche) according to the manufacturer's protocols.

In vitro run-off transcription and *in vitro* translation reactions

Plasmid constructs were linearised and *in vitro* transcriptions were performed using either SP6 (pSP6R(x)L series) or T3 (pSKMyc and pSKMycΔ1) polymerase as previously described. Capped transcripts were synthesised in a reaction containing 2 mM m⁷(5')ppp(5')G, 0.5 mM GTP and 1 mM of the remaining nucleotides. All RNAs were purified using size exclusion chromatography and quantified using the absorbance at 260 nm. In addition, the integrity of each transcript was verified using agarose gel electrophoresis and ethidium bromide staining.

In vitro translation reactions were performed using rabbit reticulocyte lysate (Promega) according to the manufacturer's recommendations. The translation products were fractionated by SDS-polyacrylamide gel electrophoresis and visualised using phosphorimager analysis (Molecular Dynamics).

Cationic liposome-mediated RNA transfection

Cationic liposome-mediated RNA transfection of mammalian cells was performed as described previously (30). Capped and polyadenylated transcripts were synthesised using *in vitro* run-off transcription on an *EcoRI* linearised pSP64R(x)L poly(A) template. Approximately 2 × 10⁵ HeLa cells were transfected with 5 μg of RNA previously incubated with 12.5 μg of Lipofectin (Life Technologies Inc.). After 8 h of transfection, cells were harvested and processed for reporter gene analysis.

Reporter gene analysis

The activity of Firefly luciferase in lysates prepared from cells transfected with pGL3, pGML, pHpML and pHpL was measured using a luciferase reporter assay system (Promega). Light emission was measured either over 1 s using a 1253 luminometer (Bio-Orbit) or over 10 s using an Optocomp-1 Luminometer (MGM instruments). The activity of both Firefly and *Renilla* luciferase in cell lysates with bicistronic luciferase plasmids was measured using the Dual-luciferase reporter assay system (Promega). Assays were performed according to the manufacturer's recommendations. The activity of β -galactosidase in lysates prepared from cells transfected with pcDNA3.1/HisB/*lacZ* was measured using a Galactolight plus assay system (Tropix).

RESULTS

c-myc translation initiation can occur by internal ribosome entry and the conventional cap-dependent mechanism

We and others have shown that *c-Myc* protein synthesis can occur in a cap-dependent manner and by internal ribosome entry (11,12,17,19). To assess the contribution that these two disparate mechanisms make to *c-myc* expression, a palindromic sequence capable of forming a stable RNA hairpin (~ 55 kcal/mol) was introduced into the control luciferase reporter construct, (pGL3) and the 5' UTR containing construct (pGML, previously known as pGL3utr) at the *SpeI* site (Fig. 1A). As a consequence, ribosome scanning from the cap structure of the transcripts produced by the new constructs (pHpL and pHpML) should be severely impeded, whereas ribosomes entering at a site distal to the hairpin will be unaffected. HeLa cells were transfected with pGL3, pGML, pHpL or pHpML and in agreement with our previously published data, the *c-myc* IRES does not inhibit translation of the downstream Firefly luciferase reporter gene. Moreover, we consistently observe that there is a slight elevation in expression of this enzyme in the presence of the IRES (Fig. 1B). In cells transfected with the construct pHpL there is a 200-fold reduction in the amount of luciferase produced when compared to the control vector pGL3 (Fig. 1B). Hence, as expected the RNA hairpin structure inhibits cap-dependent translation initiation. However, in cells transfected with pHpML, in which the *c-myc* IRES lies downstream of the RNA hairpin, luciferase expression is stimulated by ~ 67 -fold when compared to pHpL. These data demonstrate that the 5' UTR can promote efficient translation initiation despite the presence of an RNA structure which blocks ribosome scanning from the 5' end and thus provide further support for the presence of an IRES within this leader sequence. Nevertheless, it is notable that the RNA hairpin does reduce luciferase expression from a transcript containing the *c-myc* 5' UTR by 3-fold. This observation would indicate that mRNAs originating from the P2 promoter must also support a cap-dependent scanning mechanism in addition to internal initiation.

Comparison of *c-myc* IRES-mediated internal initiation in a range of cell types

We have shown previously that the *c-myc* IRES is capable of promoting translation of the downstream cistron on a bicistronic mRNA in both HeLa and HepG2 cells. To investigate how

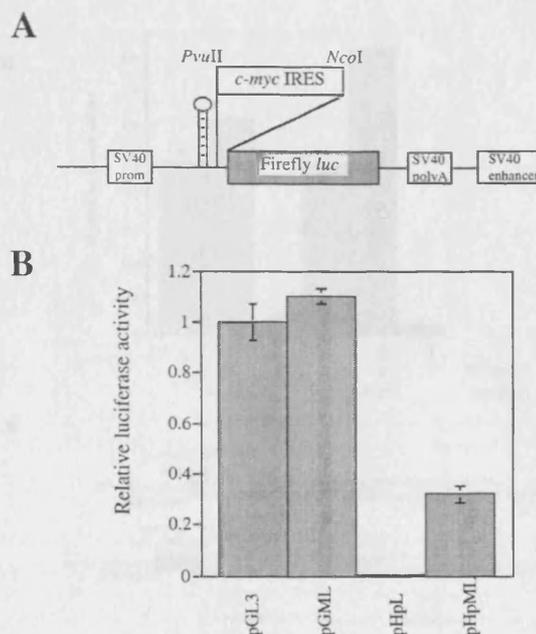


Figure 1. A comparison between the efficiency of IRES-mediated translation and scanning. (A) A diagrammatic representation of the monocistronic hairpin containing plasmids pHpL and pHpML. The hairpin was inserted into the *SpeI* site upstream of the *PvuII* site. (B) HeLa cells were transfected (in triplicate) with the plasmids shown and Firefly luciferase activity is expressed relative to the transfection control β -galactosidase. All experiments were performed on three independent occasions.

widely the IRES is utilised, a range of cell types derived from different tissues, including Cos-7, MCF7, Balb/c-3T3, MEL, MRC5, HK293, GM637, HeLa and HepG2 were co-transfected with either pRF or pRMF and pcDNA3.1/HisB/*lacZ* (Fig. 2A). The expression from both *Renilla* and Firefly luciferase cistrons was assayed and normalised to the transfection control β -galactosidase. Between cell types, significant variation in the level of readthrough re-initiation was observed on the control bicistronic plasmid (data not shown). Accordingly, the efficiency of the IRES is represented as a ratio of FL to RL expressed from pRMF. In each cell line, the presence of the *c-myc* IRES in the mRNA did not significantly alter *Renilla* luciferase expression and indeed, the largest difference was observed in HeLa cells, in which the *c-myc* IRES reduced *Renilla* luciferase activity by $\sim 11\%$ (data not shown; 12). However, it is clear that the efficiency of *c-myc* IRES-driven translation varies widely between cell lines (Fig. 2B). Hence the IRES is most active in HeLa cells, followed by MRC5, HepG2, GM637, HK293 and Cos-7. Interestingly, the IRES is almost inactive in the MCF7 cells suggesting that these cells may lack a factor which is essential for IRES-mediated translation. Alternatively, these cells could express a higher level of a specific inhibitor of internal initiation. One possible explanation for the inactivity of the human *c-myc* IRES in cell lines of murine origin, Balb/c-3T3

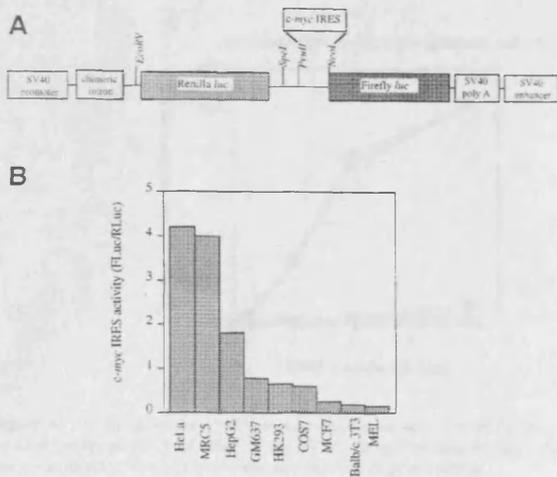


Figure 2. A comparison of the efficiency of *c-myc* IRES initiated translation in cell lines of different origin. (A) A schematic representation of the bicistronic reporter plasmids pRF and pRMF. (B) IRES activity is expressed using the ratio of downstream cistron expression to upstream cistron expression (Fluc/RLuc) with any differences in transfection efficiencies corrected for using the β -galactosidase transfection control. All experiments were performed in triplicate on three independent occasions.

and MEL cells, is that the function of the IRES displays species specificity. However, we have recently shown that this is not the case, since the *c-myc* IRES isolated from murine cells is active in HeLa cells and yet also relatively inactive in Balb/c-3T3 cells (data not shown).

***c-myc* P2 transcripts can be translated by a cap-dependent mechanism in Balb/c 3T3 cells, MCF-7 cells and in reticulocyte lysates**

The relative inactivity of the *c-myc* IRES in Balb/c 3T3 and MCF7 cells enabled us to analyse the effect of the P2 5' UTR on cap-dependent translation initiation. To this end, these cell lines were transfected with the monocistronic control construct, the 5' UTR-containing constructs, pGL3 and pGML, and the *c-myc* 5' UTR construct containing the hairpin pHpML respectively. The P2 5' UTR does not inhibit cap-dependent translation initiation, at least in these cell lines (Fig. 3A). However, the additional presence of the hairpin structure was sufficient to prevent scanning demonstrating that the *c-myc* IRES is relatively inactive in these cell types and consequently *c-myc* is translated by a cap-dependent mechanism (Fig. 3A).

To further investigate the impact of the P2 5' UTR on cap-dependent translation initiation we turned to reticulocyte lysate. This system cannot support internal ribosome entry on the *c-myc* leader sequence (our unpublished data; 31), therefore the contribution of the 5' cap structure can be assessed directly. Thus, rabbit reticulocyte lysate was primed with capped or uncapped *c-myc* transcripts, either bearing the P2 5' UTR sequence (*myc*) or lacking this element (*myc Δ 1) (Fig. 3B). Two species of *c-myc* protein can arise from the P2 transcripts by use of alternate translation initiation codon (CUG or AUG), which give rise to protein products with apparent molecular*

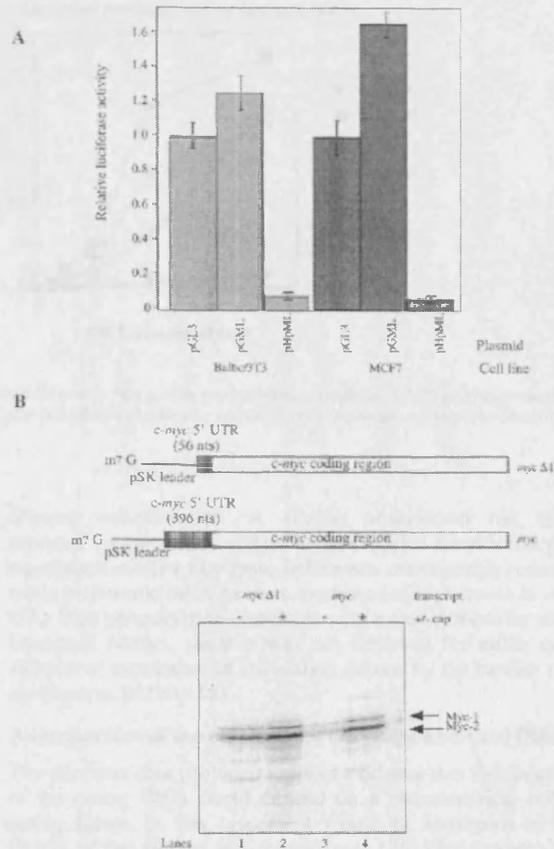


Figure 3. Cap-dependent translation of *c-myc* P2 transcripts in Balb/c 3T3, MCF7 cells and in rabbit reticulocyte lysates. (A) MCF7 and Balb/c cells were transfected with the plasmids pGL3, pGML or pHpML and the Firefly luciferase activity measured as described previously. (B) *c-myc* transcripts bearing 56 nt (*myc Δ 1) or 396 nt (*myc*) of the *c-myc* 5' UTR were synthesised *in vitro* using linearised plasmids pSKM Δ 1 or pSKM respectively. Rabbit reticulocyte lysate was programmed with 5 ng/ μ l of either capped (+) or uncapped (–) *myc* or *myc Δ 1 transcripts. Radiolabelled polypeptides synthesised in the reaction were then fractionated by SDS-PAGE and detected using phosphorimager analysis.**

weights of 67 and 64 kDa respectively (32). As expected, capping the *myc Δ 1 RNA stimulated the synthesis of both Myc-1 and 2 polypeptides (Fig. 3B, lanes 1 and 2). This modest effect of 2–2.5-fold is consistent with the previously reported values for relatively unstructured RNAs using this system (33). In the absence of a cap structure, the *c-myc* 5' UTR reduced the synthesis of both the AUG and CUG-initiated polypeptides by ~90% (Fig. 3B, lanes 1 and 3). It is likely that structural elements within the 5' UTR are responsible for this effect since this element is GC-rich. However, the synthesis of both proteins was enhanced by 14–16-fold on capping of the *myc* transcript (Fig. 3B, lanes 3 and 4), with the result that the*

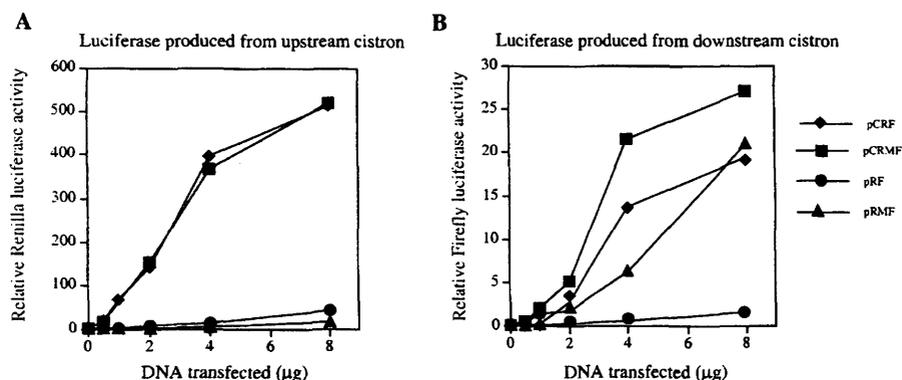


Figure 4. The effect of the CMV promoter/enhancer on *c-myc* IRES directed internal initiation. HeLa cells were transfected with the CMV promoter/enhancer based plasmids pCRF or pCRMf or the SV40 promoter/enhanced based plasmids, pRF and pRMF. (A) *Renilla* and (B) Firefly luciferase activity was determined and normalised to that of the transfection control, β -galactosidase.

5' UTR inhibits translation initiation by only 50%. Hence, the P2 5' UTR strongly attenuates the translational efficiency of uncapped *c-myc* transcripts. Nevertheless, much of this repression is relieved by the presence of a 5' cap. Therefore, translation initiation on the P2 transcript is strongly cap-dependent in the reticulocyte lysate system.

Overexpression of bicistronic mRNAs inhibits the function of the *c-myc* IRES

Thus far, we have demonstrated that in many cell lines *c-myc* translation can occur by the alternative mechanism of internal ribosome entry. However, *c-myc* can also be translated by the conventional cap-dependent mechanism in certain backgrounds. One model that would explain the cell-type specific variation in the efficiency of *c-myc* IRES-driven translation posits that non-canonical *trans*-acting factors are required for the recruitment of the 40S ribosome to this element. In this scenario, the activity of one or more of these factors is considerably reduced in the Balb/c-3T3 and MCF7 cell lines. Further evidence in support of this model was provided by experiments in which the bicistronic mRNAs were overexpressed using the powerful cytomegalovirus (CMV) promoter/enhancer region; this transcriptional element has been shown to result in significantly higher levels of expression than the SV40 promoter/enhancer (34). The *Renilla* luciferase activity measured in cells transfected with a CMV-based control bicistronic plasmid pCRF was significantly greater than that achieved with the analogous plasmid, pRF (~27-fold, Fig. 4A, compare pCRF *Renilla* luciferase to pRF *Renilla* luciferase). However, in cells transfected with the 5' UTR-containing construct, pCRMf, there was not a corresponding increase in Firefly luciferase activity when compared to pRMF. Transfection with 4 or 8 μ g of pCRMf produced only 4- or 1.25-fold more Firefly luciferase than pRMF, respectively (Fig. 4B). Consequently, using the CMV promoter/enhancer, the apparent activity of the *c-myc* IRES when calculated relative to readthrough is only 1.5–2 fold compared to 50-fold for the SV40 based constructs (Fig. 4B). These data suggest that a *trans*-acting factor, which is required for initiation of translation via the *c-myc* IRES, is present at a

limiting concentration. A similar observation has been reported for the entero- and rhinovirus IRESs; the efficiency of translation mediated by these IRESs was considerably reduced when bicistronic mRNAs were expressed at high levels *in vivo* (35). This phenomenon correlates with a requirement for non-canonical factors, since it was not observed for either cap-dependent translation or translation driven by the cardio- and aphthovirus IRESs (35).

A comparison of the efficiency of the *c-myc* and viral IRES

The previous data provided indirect evidence that the function of the *c-myc* IRES could depend on a non-canonical *trans*-acting factor. In this respect, it would be analogous to the IRESs of the entero- and rhinoviruses (36). To compare the efficiency of the *c-myc*, HRV and EMCV IRESs, HeLa cells were transfected with the plasmids pRF, pRMF, pRhVf, pRmCVf. The activities of *Renilla* and Firefly luciferase were determined and normalised to that of the transfection control, β -galactosidase (Fig. 5). Expression of the upstream cistron, *Renilla* luciferase, was not greatly affected by the presence of the EMCV, HRV or the *c-myc* IRES in the intercistronic region (data not shown). A comparison of the downstream cistron activities revealed that all of these elements stimulated Firefly luciferase expression (Fig. 5). However, the extent to which expression from the downstream cistron was enhanced differed widely between these IRESs. In fact, the *c-myc* IRES elevated Firefly luciferase activity by 70.8-fold, whilst the HRV and EMCV IRESs caused a lesser stimulation of 9.6- and 14-fold, respectively. Thus, these data suggest that both of these IRESs are less efficient in this system at promoting internal ribosome entry than the *c-myc* IRES.

c-myc IRES-driven translation requires a nuclear event

It has been suggested previously that efficient translation driven by the IRES located in the 5' UTR of the immunoglobulin heavy chain binding protein (Bip) requires a nuclear event (37). Moreover, two specific nuclear protein factors have been identified which interact with the Bip IRES (38). To test whether the *c-myc* IRES also has such a requirement for

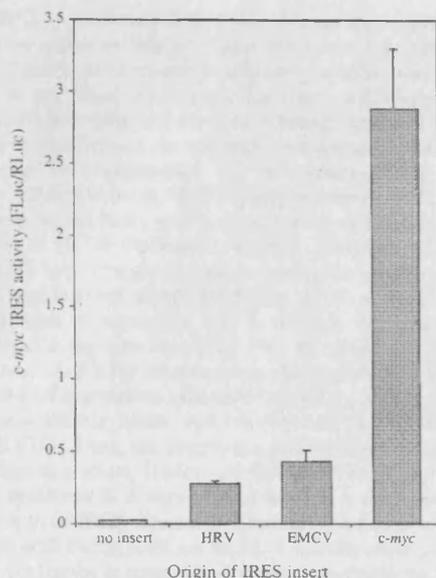


Figure 5. A comparison of the efficiency of HRV, EMCV and *c-myc* IRES-initiated internal ribosome entry on bicistronic mRNAs transcribed in the nucleus. HeLa cells were transfected in triplicate with either the control plasmid pRF, the *c-myc* IRES containing plasmid pRMF, the HRV IRES containing plasmid pRhrvF or the EMCV IRES containing plasmid pRemvF. Upstream cistron (*Renilla* luciferase) and downstream cistron (Firefly luciferase) activities were determined and normalised to that of the transfection control.

nuclear factors, the plasmid constructs pSP64RL poly(A), pSP64R(*c-myc*)L poly(A) and pSP64R(hrv)L poly(A) were generated. Bicistronic transcripts containing an m⁷GpppG cap structure and a polyadenylated tail at the 5' and 3' termini, respectively, were synthesised from each of the plasmids in the pSP64RL(x)Lpoly(A) series by *in vitro* run-off transcription (Fig. 6A). Cationic liposomes were used to encapsulate equimolar quantities of each transcript and introduce them into the cytoplasm of HeLa cells. After a period of 8 h, the expression from the upstream and downstream cistrons was monitored (Fig. 6B and C). In cells transfected with the control bicistronic transcript, RLuc, the *Renilla* luciferase cistron was translated efficiently, whilst little expression of the downstream cistron was observed (Fig. 6B and C). Insertion of the HRV IRES between the two cistrons resulted in a 52-fold stimulation of Firefly luciferase activity when compared to the expression due to readthrough-re-initiation (Fig. 6C). In contrast, the expression of the downstream cistron was only enhanced by 1.4-fold on the *Re-mycL* transcript (Fig. 6B). Thus, the *c-myc* IRES cannot stimulate the translation of the downstream cistron on a bicistronic mRNA introduced directly into the cytoplasm. To confirm these data, the plasmids pRF and pRMF were transfected into human TK143 cells previously infected with a recombinant vaccinia virus that expresses the T7 RNA polymerase (vTF7-3) (39). The presence of a T7 RNA polymerase promoter upstream of the *Renilla* luciferase cistron in pRF and pRMF results in the transcription of bicistronic mRNAs in the cytoplasmic compartment. However, the *c-myc* 5' UTR did not promote

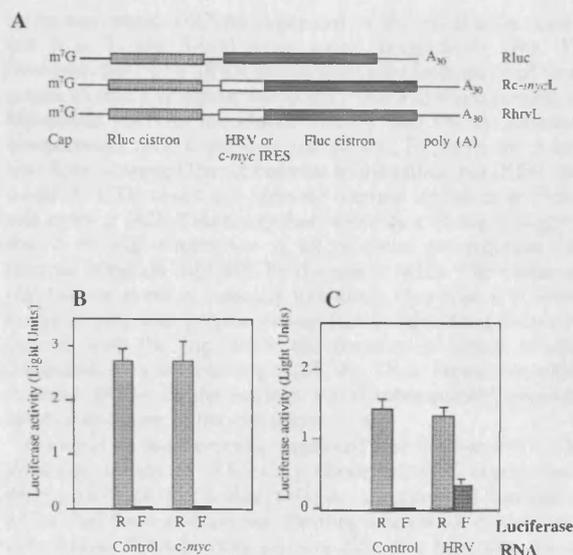


Figure 6. A comparison of the efficiency of the *c-myc* IRES and HRV-IRES directed internal initiation on mRNAs introduced directly into the cytoplasm. (A) A diagrammatic representation of the control (RLuc), *c-myc* 5' UTR containing (*Re-mycL*) and HRV IRES containing (RhrvL) bicistronic RNAs. Transcripts were synthesised *in vitro* and possess both a 5' cap structure and a 3' terminal polyadenylated tail of 30 residues. (B) HeLa cells were transfected with RLuc (control) or *Re-mycL* (*c-myc*) by lipofection. After 8 h *Renilla* (R) and Firefly (F) luciferase expression was determined. (C) Similarly, HeLa cells were transfected with RLuc (control) or RhrvL (HRV IRES) and *Renilla* and Firefly luciferase activities determined.

internal initiation on mRNAs transcribed in the cytoplasm using the T7/vaccinia system (data not shown). In contrast, the IRESs of the entero- and rhinoviruses have been shown to function efficiently using bicistronic mRNAs expressed in this manner (35,40). These data appear to suggest a fundamental difference between the function of the entero- and rhinovirus IRESs and that of *c-myc*. The *c-myc* IRES is only able to promote internal initiation on transcripts expressed in the nucleus, however the HRV element is capable of performing this task on mRNAs that do not originate in this compartment. Therefore, we propose that a nuclear event is a pre-requisite for efficient *c-myc* internal initiation.

DISCUSSION

We and others have shown previously that the 5' UTR of *c-myc* contains an IRES (11,12). We have investigated several features of the *c-myc* IRES and compared its activity in a range of cell lines and to IRESs of viral origin.

First, using a stable RNA structure to substantially impede ribosome scanning from the 5' cap, we have demonstrated that efficient translation initiation can be restored by positioning the *c-myc* 5' UTR downstream of this inhibitory element (Fig. 1). This observation provides further evidence that the P2 leader sequence can support internal entry of ribosomes via an IRES. In these experiments, internal initiation directed by the

c-myc IRES is apparently 3-fold less efficient than cap-dependent translation initiation (but see later). However, reporter mRNAs are translated with comparable efficiency whether the 5' UTR is present or not. Thus, we suggest that *c-myc* mRNAs originating from the P2 promoter are capable of being translated via a cap-dependent mechanism in addition to internal initiation. This hypothesis is strengthened by two observations. First, a reporter mRNA bearing the P2 leader sequence was translated efficiently in cell lines with a significantly reduced capacity to promote 5' UTR-mediated internal initiation (Fig. 3A). Second, in reticulocyte lysate, a system in which the *c-myc* IRES is inactive (our unpublished data; 31), *c-myc* P2 transcripts are translated in a manner that is strongly dependent on the presence of a cap structure (Fig. 3B). In agreement with these data, Carter *et al.* (31) have recently shown that the considerable repression of translation initiation caused by the P1 5' UTR in rabbit reticulocyte lysate can be relieved by the addition of eIF4F/E (31). Thus, we propose a dual mechanism for *c-myc* translation initiation. Under conditions where cap-dependent protein synthesis is compromised there is a shift from a cap-dependent to an IRES-directed mechanism of translation initiation. In accord with this hypothesis, we have recently shown that *c-myc* protein synthesis is maintained during apoptosis by virtue of the IRES, whereas overall cap-dependent translation is significantly inhibited (27).

We have also identified several factors that influence the efficacy of the *c-myc* IRES. Expression of bicistronic mRNAs containing the *c-myc* IRES in a panel of cell lines demonstrated that the activity of this element is critically dependent on cellular origin (Fig. 2). Although the IRES stimulated protein synthesis from the downstream cistron in all the cell lines tested, there was a 20-fold disparity between HeLa and MCF7 cells, the lines in which the IRES is most and least active, respectively. This cell-type specific variation in IRES activity implies that the function of this element could be modulated by non-canonical *trans*-acting factors. In this regard, we have recently demonstrated that ribonuclear protein complexes assembled on the *c-myc* 5' UTR *in vitro* using cell extracts from different cell lines vary distinctly in composition (41). Furthermore, overexpression of bicistronic mRNAs using the powerful CMV promoter/enhancer drastically reduced the apparent efficiency of the *c-myc* IRES (Fig. 4). We speculate that the concentration of a *trans*-acting factor essential for *c-myc* IRES-driven translation initiation is limiting under these conditions. The low concentration of this factor could also explain why *c-myc* internal initiation appears to be 3-fold less efficient than cap-dependent translation (Fig. 1) since transcripts expressed from the monocistronic constructs (pGL3, pGML, pHpL and pHpML) accumulate to a level approximately an order of magnitude higher than those produced from the bicistronic constructs (pRF and pRMF) (our unpublished observations). Significantly, the characteristics described above are not unique to the *c-myc* IRES. Both cell-type specific variations in IRES activity and saturation of IRES function have also been described for the better defined IRESs of the enterovirus and rhinoviruses (35,40). The activity of these elements is known to be dependent on host-specific *trans*-acting factors suggesting that the *c-myc* IRES has similar requirements.

A comparison of the *c-myc* IRES to those of the human rhinovirus (HRV) and encephalomyocarditis virus (EMCV),

using bicistronic mRNAs expressed in the nucleus, revealed that it is 7- and 5-fold more active, respectively (Fig. 5). However, the *c-myc* IRES differs markedly from those of viral origin, in that it is almost completely inactive when present in bicistronic mRNAs introduced directly into the cytoplasmic compartment (Fig. 6 and data not shown). Furthermore, it has also been observed that in contrast to the poliovirus IRES, the *c-myc* 5' UTR could not promote internal initiation in HeLa cell extracts (42). Taken together, these data strongly suggest that a nuclear experience is an essential pre-requisite for internal initiation mediated by the *c-myc* IRES. The nature of this nuclear event is currently unknown. However, it is interesting to note that several nuclear factors have been shown to interact with the Bip IRES, the function of which is also dependent on a nuclear origin (37,38). Thus, factors recruited to these IRESs in the nucleus could subsequently promote internal initiation in the cytoplasm (37).

Carter *et al.* have recently suggested that the *c-myc* 5' UTR does not contain an IRES (31). However, these experiments were performed in reticulocyte lysate, a specialised translation extract known to contain very limiting amounts of nuclear and cytoplasmic RNA binding proteins (33). We have also found that the *c-myc* IRES cannot function in reticulocyte lysate (data not shown). In this respect it is similar to the IRESs of the enterovirus and rhinoviruses, which function inefficiently or not at all in this system. Indeed, reticulocyte lysate must be supplemented with cytoplasmic extracts to support efficient enterovirus internal initiation (36). Most importantly, to our knowledge no eukaryotic cellular IRES has been shown to promote internal initiation in this system. Using bicistronic mRNAs expressed in the nucleus of cell lines, we and others identified an IRES in the *c-myc* 5' UTR (11,12). This finding has been supported by the observation that *c-myc* mRNAs are efficiently translated in poliovirus-infected HeLa cells and in cells undergoing apoptosis (27,28). Here we present further evidence that *c-myc* mRNAs can be translated by internal initiation and we provide additional mechanistic insights. Our data support a model in which both non-canonical *trans*-acting factors and a nuclear experience participate in *c-myc* internal ribosome entry. In the light of these results, it is hardly surprising that the *c-myc* IRES does not function in the reticulocyte lysate system. Finally, we are currently attempting to identify the cytoplasmic and nuclear factors involved in the formation of ribonuclear protein complexes with the *c-myc* 5' UTR. The effect of these factors on *c-myc* internal initiation can then be rigorously tested in cell-free extracts.

ACKNOWLEDGEMENTS

This work was funded by a grant from the Cancer Research Campaign (M.S. and T.S.). J.P.C.L.Q., M.J.C. and C.L.J. hold MRC studentships.

REFERENCES

- Henriksson, M. and Lüscher, B. (1996) *Adv. Cancer Res.*, **68**, 109–182.
- Prengergast, G.C. (1999) *Oncogene*, **18**, 2967–2987.
- Marcu, K.B., Bossone, S.A. and Patel, A.J. (1992) *Ann. Rev. Biochem.*, **61**, 809–860.
- Ross, J. (1995) *Microbiol. Rev.*, **59**, 16–95.
- Lee, C.H., Leeds, P. and Ross, J. (1998) *J. Biol. Chem.*, **273**, 25261–25271.
- Lüscher, B. and Eisenmann, R.N. (1988) *Mol. Cell. Biol.*, **8**, 2504–2512.

7. Shindo.H., Tani.E., Matsumoto.T., Hashimoto.T. and Furuyama.J. (1993) *Acta Neuropathol.* **86**, 345–352.
8. Saito.H., Hayday.A.C., Wiman.K., Hayward.W.S. and Tonegawa.S. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 7476–7460.
9. Butnick.N.Z., Miyamoto.C., Chizzonite.R., Cullen.B.R., Ju.G. and Skalka.A.M. (1985) *Mol. Cell. Biol.*, **5**, 3009–3016.
10. Parkin.N.T., Darveau.A., Nicholson.R. and Sonenberg.N. (1988) *Mol. Cell. Biol.*, **8**, 2875–2883.
11. Nanbru.C., Lafon.L., Audigier.S., Gensac.M.-G., Vagner.S., Huez.G. and Prats.A.-C. (1997) *J. Biol. Chem.*, **272**, 32061–32066.
12. Stoneley.M., Paulin.F.E.M., Le Quesne.J.P.C., Chappell.S.A. and Willis.A.E. (1998) *Oncogene*, **16**, 423–428.
13. Battey.J., Moulding.C., Taub.R., Murphy.W., Stewart.T., Potter.H., Lenoir.G. and Leder.P. (1983) *Cell*, **34**, 779–787.
14. Bentley.D.L. and Groudine.M. (1986) *Mol. Cell. Biol.*, **6**, 3481–3489.
15. Yang.J.-Q., Bauer.S.R., Mushinski.J.F. and Marcu.K.B. (1985) *EMBO J.*, **4**, 1441–1447.
16. Koromilas.A.E., Lazaris-Karatzas.A. and Sonenberg.N. (1992) *EMBO J.*, **11**, 4153–4158.
17. De Benedetti.A., Joshi.B., Graff.J.R. and Zimmer.S.G. (1994) *Mol. Cell. Differ.*, **2**, 347–371.
18. De Benedetti.A. and Rhoads.R.E. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8212–8216.
19. West.M.J., Stoneley.M. and Willis.A.E. (1998) *Oncogene*, **17**, 769–780.
20. Jackson.R.J., Hunt.S.L., Reynolds.J.E. and Kaminski.A. (1995) *Curr. Top. Microbiol. Immunol.*, **203**, 1–29.
21. Jackson.R.J. and Kaminski.A. (1995) *RNA*, **1**, 985–1000.
22. Jackson.R.J., Hunt.S.L., Gibbs.C.L. and Kaminski.A. (1994) *Mol. Biol. Reports*, **19**, 147–159.
23. Vagner.S., Gensac.M.-C., Maret.A., Bayard.F., Amalric.F., Prats.H. and Prats.A.-C. (1995) *Mol. Cell. Biol.*, **15**, 35–44.
24. Bernstein.J., Sella.O., Le.S. and Elroy-Stein.O. (1997) *J. Biol. Chem.*, **272**, 9356–9362.
25. Stein.L., Itin.A., Einat.P., Skaliter.R., Grossman.Z. and Keshet.E. (1998) *Mol. Cell. Biol.*, **18**, 3112–3119.
26. Miller.D., Dibbens.J., Damert.A., Risau.W., Vadas.M. and Goodall.G. (1998) *FEBS Lett.*, **434**, 417–420.
27. Stoneley.M., Chappell.S.A., Jopling.C.L., Dickens.M., MacFarlane.M. and Willis.A.E. (2000) *Mol. Cell. Biol.*, **20**, in press.
28. Johannes.G. and Sarnow.P. (1998) *RNA*, **4**, 1500–1513.
29. Jordan.M., Schallhorn.A. and Wurm.F.M. (1996) *Nucleic Acids Res.*, **24**, 596–601.
30. Dwarki.V.J., Malone.R.W. and Verma.I.M. (1993) *Methods Enzymol.*, **217**, 644–654.
31. Carter.P.S., Pardo-Jarquin.M. and DeBenedetti.A. (1999) *Oncogene*, **18**, 4326–4335.
32. Hann.S.R., King.M.W., Bentley.D.L., Anderson.C.W. and Eisenman.R.N. (1988) *Cell*, **52**, 185–195.
33. Svitkin.Y.V., Ovchinnikov.L.P., Dreyfuss.G. and Sonenberg.N. (1996) *EMBO J.*, **15**, 7147–7155.
34. Sutherland.L.C. and Williams.G.T. (1997) *J. Immunol. Meth.*, **207**, 179–183.
35. Borman.A.M., LeMercier.P., Girard.M. and Kean.K.M. (1997) *Nucleic Acids Res.*, **25**, 925–932.
36. Belsham.G.J. and Sonenberg.N. (1996) *Microbiol. Rev.*, **60**, 499–513.
37. Izuka.N., Chen.C., Yang.Q., Johannes.G. and Sarnow.P. (1995) *Curr. Top. Microbiol. Immunol.*, **203**, 156–177.
38. Yang.Q. and Sarnow.P. (1997) *Nucleic Acids Res.*, **25**, 2800–2807.
39. Fuerst.T.R., Niles.E.G., Studier.F.W. and Moss.B. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 8122–8126.
40. Roberts.L.O., Seamons.R.A. and Belsham.G.J. (1998) *RNA*, **4**, 520–529.
41. Paulin.F.E.M., Chappell.S.A. and Willis.A.E. (1998) *Nucleic Acids Res.*, **26**, 3097–3103.
42. Pelletier.J. and Sonenberg.N. (1988) *Nature*, **334**, 320–325.