uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Stephen Baird

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Microbiology and Immunology-Specialization: Human and Molecular Genetics)

GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Searching for IRES

TITRE DE LA THÈSE / TITLE OF THESIS

Robert Korneluk

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Martin Holcik and Marcel Turcotte

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

John Bell

Davide Ruggero (absent)

Jocelyn Côté

Ken Dimock

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Searching for IRES

by

Stephen Baird
B.I.S. University of Waterloo, 1984
M.Sc. University of Ottawa, 1988

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in the Human and Molecular Genetics Program

Biochemistry, Microbiology and Immunology Department
Faculty of Medicine
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-25855-2
Our file Notre référence
ISBN: 978-0-494-25855-2

Abstract

The standard method of translation initiation, where the ribosome binding onto mRNA is mediated by initiation factors that congregate at the 5' "cap-nucleotide" of the RNA, is at times, partially disabled. For example, during viral infection, mitosis, and cellular stress, the efficiency of this form of initiation is reduced relative to an alternate mechanism of initiation that utilizes Internal Ribosome Entry Sites (IRESes) contained in the 5' UTR sequence of viral and cellular RNA transcripts. These IRESes often are bound by non-canonical initiation factors that help to recruit the ribosomes to the mRNA upstream of the start codon without the need of a 5' cap structure. In viral IRESes, like HCV and EMCV, RNA structures have been shown to be critical for IRES function. These structures are also similar in other viruses that do not share sequence similarity. HCV for example, is similar to CSFV and BVDV, while the EMCV IRES structure is shared with other cardio-picornaviruses. In contrast, although several cellular IRESes' RNA secondary structures have been determined, no similarities have yet been identified. In order to establish if structure similarities could be found in cellular IRESes, the secondary structure of the X-linked inhibitor of apoptosis (XIAP) IRES was determined and used for an *in silico* search of all human 5' UTRs using the RSEARCH program. Five of the top computational matches were cloned and tested for IRES activity in β-Gal/CAT bicistronic constructs. Of these, the 5' UTRs from Aquaporin 4 and the uncharacterized ELG1 transcripts exhibited IRES activity although at a much lower level than that of the XIAP IRES. The actual RNA secondary structure of the ELG1 IRES did not match the predicted match from RSEARCH but the AQP4 IRES had a similar structure surrounding a polypyrimidine tract. Importantly, deletion of this motif removed all IRES activity similar to the analogous polypyrimidine tract

deletion of the XIAP IRES. As only a slight similarity was found, it seems that RNA structures of IRESes in cellular messages are not shared and do not play the same role as those in viral mRNAs.

Acknowledgements

I would like to thank my three supervisors for their support and guidance. Specifically Martin Holcik, my chief supervisor, for his unlimited patience, optimism, edifying nature and his continued support on all my research endeavours wherever they went, Marcel Turcotte for straightening me out on many aspects of bioinformatics, and Bob Korneluk, my supervisor outside of this thesis for 15 years who has allowed me the opportunity to learn and do research without restrictions, support me in difficult periods, and include me in the excitement of the discoveries. Thanks to the "Lost in Translation" group for discussions and help, specifically Nicoleta for her technical support, summer student, Zachary King, for help on the random UTR cloning and Dr. Stephen Lewis for his critical analysis, interest in good science, and the collaboration of our work. Sylvia Balabania took the first crack at compiling the IRES dataset. Thanks also to the ARC researchers and staff that include longtime work colleagues Jane and Fred. I can't forget Alex who has fed me with food, friendship and support through hard times. Also thanks to my friends, like the Huffmans, and family who have supported my kids and me over these school years.

STEPHEN D. BAIRD

The Apoptosis Research Centre,
December, 2006

To my sons, Nicolas, Nathan, Daniel and David

Table of Contents

List of Abbreviations

Note: for IRES gene names, see Table1; for virus names, see Table 2.

4E-BP	eIF4E-binding proteins
AKT	Protein kinase B (originally oncogene from retrovirus AKT8)
CAT	Chloramphenicol-Acetyl Transferase
cDNA	Copy DNA
Cryo-EM	Cryogenic-electron microscopy
dsRNA	Double stranded RNA
eIF	Eukaryotic initiation factor
eRF	Eukaryotic release factor
EST	Expressed Sequence Tags
GCN2	General amino acid Control of gene expression, Non-derepressing protein 2
HEK	Human embryonic kidney
hnRNP	heterogeneous nuclear ribonucleoproteins
HRI	Heme Regulated Inhibitor
IFN	Interferons
ITAFs	IRES-specific cellular trans-acting factors
MAPK	mitogen-activated protein kinase
MCC	Matthews Correlation Coefficient (RNA structure version)
Met-tRNA	Methionine transfer RNA
mRNA	Messenger RNA
NEO PTII	Neomycin phosphotransferase II
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PERK	PKR like ER kinase
PPT	Polypyrimidine tract
P-site	Peptidyl site of ribosome
PTB	Polypyrimidine Tract Binding protein
qRT-PCR	Quantitative RT-PCR
RACE	Rapid Amplification of cDNA Ends
RefSeq	NCBI Reference Sequence Database
RNAi	RNA interference
RT-PCR	Reverse transcriptase - polymerase chain reaction
SCFG	stochastic context-free grammars
SECIS	selenocysteine insertion sequence
snRNA	Small nuclear RNA
TNF	tumor necrosis factor alpha
TRAIL	TNF-related apoptosis-inducing ligand
uORF	Upstream open reading frame
UTR	Untranslated region
β-Gal	β-Galactosidase

List of Figures

List of Tables

Chapter 1

Introduction

The cell has many ways of regulating the production of a protein from a gene. This thesis will focus on one mechanism of initiating the translation of messenger RNA without following the standard pathway used by the majority of mRNAs. Translation initiation via the Internal Ribosome Entry Site (IRES) is a mechanism that allows translation of specific mRNAs because of unique RNA sequence in their untranslated regions (UTRs), which recruit ribosomes. Although some viral IRES share primary sequence or secondary structure similarity, this similarity has not yet been found between known cellular IRES. It is from this premise, from which this thesis is based, that the following hypothesis was generated.

Hypothesis: *RNA structural elements are necessary and conserved between some viral IRESes therefore cellular IRESes also share some necessary structural elements as well.*

This lack of similarity has raised some questions regarding the existence of cellular IRESes (Kozak, 2001; Kozak, 2003). Nevertheless, a sequence containing a cellular IRES can be cloned and tested for function outside of its native gene context but it is not known what commonality exists to allow various IRES to recruit ribosomes without utilizing the standard protein translation initiation mechanism.

The clinical importance of IRES is most evident when considering the live Sabin poliovirus vaccine that is used worldwide is attenuated due to mutations in its IRES (Ehrenfeld, 1996). As some viruses require IRES to replicate, extensive work is done on the Hepatitis C virus IRES due to it being the most common chronic infection in the world. Therapeutics that would specifically effect the action of the HCV IRES would be of world-

wide importance. The clinical importance of cellular IRESes are just coming to light with the example of IRES initiated translation of c-myc and cyclin-D transcripts that are responsible for rapamycin resistant cancer cells in AKT deficient cells (Shi et al., 2005) and lack of IRES mediated translation contributing to the phenotype of X-linked dyskeratosis congenita (Yoon et al., 2006). The understanding of the IRES mechanism of translation initiation will allow new possibilities for therapeutic corrections of pathologies.

There have been many very good reviews on IRES over the years that are helpful in understanding the different facets of this mechanism of translation initiation. Some favorites are: Hellen and Sarnow (Hellen and Sarnow, 2001), Jackson et al. (Jackson et al., 1995), on picornavirus (Belsham and Sonenberg, 1996), FMDV IRES structure/function (Martinez-Salas et al., 2002), structural aspect relevant to medical intervention (Gallego, 2002), with respect to cancer (Holcik, 2004; Stoneley and Willis, 2004), the very critical and controversial (Kozak, 2001; Kozak, 2003) and on stress related IRES (Holcik and Sonenberg, 2005; Holcik et al., 2000a; Lewis and Holcik, 2005).

In this introduction the mechanism of standard "cap-dependent" translation initiation will be explained, the cellular states that inhibit it, and the factors that contribute to IRES "cap-independent" initiation. An overview of what is known about viral and cellular RNA structures of IRES will be given followed by a brief description of some of the available data and software resources.

1.1 Mechanism of cap-dependent translation, inhibition and IRES

The textbook explanation of standard translation initiation has the cap-binding protein, the eukaryotic initiation factor eIF4E, recruited to the 5' end of the mRNA where it binds to the modified "cap" nucleotide, a methyl^{7}GDP (guanadyldiphosphate) on the 5' end of all

cellular mRNAs. With this, the initiation factor eIF4G binds to both the cap-binding protein and the mRNA. eIF4G is called the "scaffold" protein as it also binds eIF4A, an ATP dependent helicase, which is responsible for unwinding the secondary and tertiary structure of the RNA during translation. The kinase MnkI also binds eIF4G and regulates eIF4E binding activity through phosphorylation. The initiation factors eIF4A, eIF4G and eIF4E are also known collectively as the protein complex eIF4F. These factors are key to the recruitment of the ribosome to the 5' cap structure of mRNA in the standard mechanism of initiation called "cap-dependent" translation. The 40S small subunit of the ribosome, the initial part of the ribosome to bind the mRNA, binds the charged start codon tRNA, Met-$tRNA_i$ with eIF2 and GTP, to its P site. This binding is promoted by eIF1A, the eIF4 factors and eIF3. Together these proteins and tRNAs, now called the 43S complex, are believed to scan along the mRNA in an ATP dependent process looking for the proper place to start translation. At the point where the proper start codon is found by the 43S scanning complex, the GTP with eIF2 is hydrolyzed to GDP in the presence of eIF5. Several factors dissociate leaving the 40S subunit with the Met-tRNA anticodon basepaired to the start codon. The large 60S ribosomal subunit then joins the small subunit and protein synthesis begins. For a review on translation initiation factors see Dever (Dever, 2002) and on structural aspects of initiation factors see Sonenberg and Dever (Sonenberg and Dever, 2003).

At some points of a cell's life this standard mechanism of translation initiation is compromised, but at these times, some mRNAs use an alternate form of initiation that does not require the cap nucleotide at the 5' end of the mRNA as a congregation site for initiation factors and is called "cap-independent" translation. This was first observed in picornavirus infections where the uncapped RNA viral genomes of the polio and encephalomyocarditis

virus were efficiently translated in eukaryotic cells through the binding of the ribosome to an internal portion of the 5' UTR of the viral RNA (Jang et al., 1988; Pyronnet et al., 2001). A viral protease cleaves the two forms of eIF4G inhibiting host protein synthesis (Gradi et al., 1998a; Gradi et al., 1998b; Svitkin et al., 1999). Even though the cleaved form of eIF4GI does not bind to eIF4E, it has been shown to participate in the translation of capped mRNA, but much less efficiently than viral RNA (Ali et al., 2001).

There are several other mechanisms that lower the efficiency of cap-dependant translation initiation inside a cell besides viral infection. During mitosis, the eIF4E-binding proteins (4E-BPs) are hypophosphorylated and competitively bind onto the cap binding protein, eIF4E, preventing eIF4E from forming the eIF4F initiation complex (Pyronnet et al., 2001). The phosphorylation states of eIF4E and the 4E-BPs in different cellular conditions have been well reviewed by Gingras et al. (Gingras et al., 1999).

At times of cellular perturbation, changes in protein levels and mRNA levels do not always correlate (Ideker et al., 2001; Nishizuka et al., 2003). Cellular stresses and the induction of apoptosis cause inhibition of standard translation initiation through the phosphorylation of the eIF2α subunit by one of the four known eIF2α kinases in mammalian cells: HRI, PKR, PERK and GCN2 (Proud, 2005). The initiation factor eIF2 is the adapter protein which binds the initiator Met-tRNA and GTP as part of the 43S preinitiation complex. Phosphorylation of the α subunit of eIF2 creates tighter binding to eIF2B preventing the GDP-GTP exchange activity of eIF2B needed for the recycling of eIF2 for the successive rounds of initiation of protein synthesis. During apoptosis, the cell still requires the *de novo* synthesis of proteins required for the orderly breakdown of the cell, but the standard protein translation initiation machinery is also slowed down with changes to the phosphorylation of eIF4G (Ling et al., 2005), eIF4E and 4E-BPs (Clemens, 2001) as well

as caspase cleavage of several canonical initiation factors, eIF4B, eIF3, eIF2α and proteins of the eIF4G family (Clemens et al., 2000). Other molecular events, like the hyperphosphorylation of eIF4GII (Pyronnet et al., 2001) or Hsp27 binding to eIF4G during heat shock, hinder the formation of eIF4F (Cuesta et al., 2000).

Despite these cellular conditions that change the initiation machinery for normal "cap-dependent" translation, some cellular mRNAs and viral RNAs still retain the ability to recruit ribosomes to a region of their 5' UTR to initiate translation in a "cap-independent" manner with Internal Ribosome Entry Sites. There have been at least 85 cellular IRES and 39 viral IRES described in the literature so far as given in Table 1 and Table 2 found in Results section 3.1. The standard method of defining IRES activity has been the ability of the sequence to initiate translation of the second open reading frame or cistron from a transcript of a synthetically engineered gene with two separate protein coding sequences. Normally in eukaryotic cells, a ribosome falls off of a transcript after reaching the stop codon of the protein coding sequence where the eukaryotic release factors eRF1 and eRF3 are bound. In a bicistronic message the second protein coding sequence would be synthesized if the sequence between the two coding sequences could recruit ribosomes. There has been some criticism and caveats attached to the use of bicistronic constructs for measuring IRES activity (Hellen and Sarnow, 2001; Kozak, 2001; Sherrill et al., 2004). Checks need to be made for promoter activity in the UTR, reinitiation of ribosome on the 2nd ORF, aberrant splicing (Holcik et al., 2005) and inconsistent values (Hennecke et al., 2001) of the dual luciferase reporter gene construct. Re-evaluation of the 5' UTRs of PDGF (Han et al., 2003), PIM-1 (Wang et al., 2005) and the human cyclin-dependent kinase inhibitor, p27kip1(Liu et al., 2005) have shown that they do not have IRESes as was initially thought (Bernstein et al., 1997; Johannes et al., 1999; Miskimins et al., 2001) but the sequences are

able to function as promoters. A statistically rigorous methodology has also been published to evaluate the output values of the bicistronic constructs (Jacobs and Dinman, 2004). The only drawback to this method becoming universally used is that the minimal sample size might require 25 to 50 measurements, which is more than the 3 to 9 sample measurements that are usually done.

The problems in assessing IRES activity must be kept in mind when using the published IRES sequences as a dataset for bioinformatic tools. Like all biological databases, the data are not perfect. In this case, some sequence activity may have been inadvertently misinterpreted.

1.2 mRNA binding proteins

When the function of some canonical translation initiation factors have been disabled or limited in the cell to make cap-dependent translation less efficient, other RNA binding proteins have been found to be required for or to enhance IRES mediated translation initiation. RNA binding proteins have many functions which effect translation, from localization of mRNA in the cytoplasm (zipcode) (Bohl et al., 2000), stabilization of message (AU-rich elements) (Barreau et al., 2005), metabolite riboswitch (Sudarsan et al., 2003), and translation silencing (YB-1) (Evdokimova et al., 2006) . Our interest, in this case, is in IRES-specific cellular trans-acting factors (ITAFs). Some of the mRNA binding proteins implicated in IRES mediated translation are: Polypyrimidine Tract Binding protein PTB/hnRNP I (Cho et al., 2005; Giraud et al., 2001; Mitchell et al., 2001; Mitchell et al., 2003; Mitchell et al., 2005; Pickering et al., 2003), La Autoantigen (Bhattacharyya and Das, 2005; Holcik and Korneluk, 2000; Marash and Kimchi, 2005), hnRNP A1 (Bonnal et al., 2005), hnRNP C1/C2 (Holcik et al., 2003; Millard et al., 2000; Sella et al., 1999), hnRNP K

(Evans et al., 2003), DAP5/p86 (Henis-Korenblit et al., 2000; Marash and Kimchi, 2005; Nevins et al., 2003; Warnakulasuriyarachchi et al., 2004), Unr (Mitchell et al., 2001; Mitchell et al., 2003; Tinton et al., 2005), p60 (Vagner et al., 1996), HuR (Millard et al., 2000), and PCBP1 (Pickering et al., 2003). For a more comprehensive list, see the review by Stonely and Willis (Stoneley and Willis, 2004).

One group of ITAFs is the many heterogeneous nuclear ribonucleoproteins (hnRNPs) that bind onto transcripts and form ribonucleoprotein complexes. They play a key role in pre-mRNA processing as well as mRNA export, localization, stability and translation (Dreyfuss et al., 2002). As an example, PTB has been connected with several functions, such as splicing repression (Spellman and Smith, 2006), pre-mRNA 3'end processing (Le Sommer et al., 2005), and mRNA localization (Cote et al., 1999). It has also been shown to be involved in activation of both viral (Bieleski et al., 2004; Sanderbrand et al., 2000; Wollerton et al., 2001) and cellular (Giraud et al., 2001; Mitchell et al., 2003; Mitchell et al., 2005; Pickering et al., 2003) IRES, but inhibitory to IRES activity in Unr (Cornelis et al., 2005) or Bip (Kim et al., 2000). Several IRES have a polypyrimidine tract at the 3'end that have been shown to be important for activity (Kaminski et al., 1994) and a recognition site for PTB (Kolupaeva et al., 1996). The PTB consensus recognition sequence shown to be important in the HCV IRES sequence is CYYYYCYYYY(G/Y)G, where Y is a pyrimidine (Anwar et al., 2000). It is not known if the location of this sequence within the HCV IRES structure is important as only the first four bases are consistently in single stranded regions, but some believe the binding site needs to be within double stranded regions (Mitchell et al., 2005). PTB has at least three splicing isoforms (Wollerton et al., 2001) and caspase (Back et al., 2002a) or viral protease cleavage products (Back et al., 2002b), which effect IRES activity to differing degrees, as well as tissue specific paralogs (Gooding et al., 2003; Pilipenko et al.,

2001). The protein also contains 4 RNA binding motifs so it is not surprising that other recognition sites for PTB/RNA interactions have also been found which do not match this consensus sequence (Wollerton et al., 2001). It is understandable from the variety of PTB forms available that there does not seem to be an unequivocal consensus sequence for PTB.

The binding site recognition of the ITAF La is very promiscuous as well, and its requirement for IRES function is not always clear. For example, an early *in vitro* study had shown that the HCV IRES could function without any non-canonical factors, but recent *in vivo* studies show that La is required for HCV IRES translation (Costa-Mattioli et al., 2004; Shimazaki et al., 2002). The RNA binding protein Unr is required for HRV IRES activity, and although all 5 cold shock domains of Unr are necessary for RNA binding, the binding seems to be sequence non-specific (Brown and Jackson, 2004).

Although there is no overall consistency as to which RNA binding proteins are required for IRES activity they may be consistently involved in ribonucleoprotein complexes that exist during specific cellular contexts of stress, cell cycle or particular mechanism of viral control over cellular functions. Using microarrays, some RNA binding proteins have been shown to interact with a specific group of transcripts during times of cellular perturbation (Tenenbaum et al., 2002; Tenenbaum et al., 2003). It has been postulated that RNA binding proteins play a role in co-regulating the translation of groups of proteins analogous to bacterial operons (Keene and Tenenbaum, 2002). Therefore, knowing an RNA binding protein that binds an IRES in a specific cellular context would suggest that other IRES in that context may also be bound to the same protein.

The binding of the canonical initiation factors also effect IRES activity. When eIF4E, the cap binding protein, has been removed (Hernandez et al., 2004), the Drosophila *reaper* IRES initiates translation more efficiently than capped messages. The eIF4G family

member, DAP5/p97, is cleaved to DAP5/p86 and enhances IRES-mediated translation (Henis-Korenblit et al., 2002; Hundsdoerfer et al., 2005; Marash and Kimchi, 2005; Nevins et al., 2003; Warnakulasuriyarachchi et al., 2004). The importance of ribosomal proteins in IRES translation initiation was shown using a genome-wide RNAi screen of Drosophila genes. 112 cellular genes were found that were required for infection by the IRES dependent Drosophila C virus (Cherry et al., 2005). Over 50% of these were genes of ribosomal proteins, two of which, when deleted, effected IRES but not cap-dependent translation. This suggests that some ITAFs will not be unique to only IRES translation machinery but will include components of the ribosome as well.

The ITAFs are abundant in the cell and seem quite ubiquitous but are not required in all examples of known IRES activity. They can also exist in several forms of post-translational modification and bind a wide range of sequence motifs. Their own regulation and regulated cellular localization would control the IRES function. Specific canonical initiation factors and ribosomal proteins also play a role. It has not been shown that the mRNA binding proteins which are used for induction of IRES activity in the subsets of the published IRESes have shared binding recognition sites or structural motifs, with the exception of some IRESes which bind PTB (Mitchell et al., 2005).

1.3 Functional classes

There appears to be differences among IRESes as to which proteins are necessary to bind to the UTR in order to recruit the ribosome for translation. When this is coupled with the results of different IRESes initiating translation with varying efficiency dependent upon which cell type or cellular context they are measured in (Nevins et al., 2003), the results suggest that there exists several IRES classes. Many groups have pointed out this

observation already. Several groups have compared IRES actvity in several cell lines and found that specific IRES will have more activity in one specific cell line relative to another IRES (Jopling and Willis, 2001; Jopling et al., 2004; Nevins et al., 2003; Stoneley et al., 2000). This may be due to the different available protein factors or activated pathways in each cell line and therefore different subsets of mRNA maybe selected for translation in different cells or tissues at any one time. This may also explain the lack of primary sequence similarity between the cellular IRES. For example, activition of AKT through p38 MAPK and ERK pathway initiates translation for mRNAs containing c-myc and cyclin D IRESes but not transcripts with the 5' UTR of $P27^{kip1}$ (Shi et al., 2005). The link may be the ITAFs PCBP1, PCBP2, and hnRNP K which are known to be required by the c-myc IRES (Evans et al., 2003), and are regulated by phosphorylation (Shi et al., 2005).

Cell specificity of IRES activity is apparent within the 4 different 5' UTRs of FGF1, each exhibiting some IRES activity (Martineau et al., 2004). FGF1A, FGF1C and FGF2 have similar activity to each other but much less actvity than the EMCV IRES in cell culture. The same IRES constructs electrotransferred into mouse muscle cells showed FGF1A to be much more active than FGF1C and similar to EMCV while the FGF2 IRES seemed to exhibit no activity at all. Clearly the context of available ITAFs must favor the translation of one mRNA over another. A very similar contextual difference is seen where IRES from muscle relevant genes, SMAD and utrophin, are active in myoblast C2C12 muscle cells but not active in 293T renal epithelial cells (Shiroki et al., 2002) or differentiated muscle cells (Miura et al., 2005). Other examples are the IRES from transcripts of the calcium channel proteins like Scamper, exhibiting tissue specific activity in kidney cells (De Pietri Tonelli et al., 2003) and the Nkx6.1 IRES being most active in β-cells (Watada et al., 2000). The Apaf1 IRES has been found to be more active in neuronal cell types, possibly due to the presence

of a neuronal isoform of PTB which seems to confer greater activity than PTB-1 (Mitchell et al., 2003) and it may be in those cells where the APAF IRES is most physiologically relevant. This correlates well with the developmental problems found in the brains of Apaf-1 knockout mice (Cecconi et al., 1998). In contrast, the IRES of HRV is repressed in neuronal cells due to the presence of mRNA binding protein DRBP76/NF90 (Merrill et al., 2006) whereas the similar poliovirus IRES is not (Gromeier et al., 2000). A list of IRES that are regulated can be found in the review by Komar and Hatzoglou (Komar and Hatzoglou, 2005). Whereas viral IRES might share a more universal context of translation regulation and therefore some similarity has been found, the larger number of cellular contexts that would require different regulation infers a large number of classes of IRES with many different sequence and structural components. Even saying there are regulatory classes of IRES may be too rigid as there may be a loose overlap of some mRNA translation regulation. For this reason in this thesis I focus on only one IRES structure, the X-linked Inhibitor of Apotosis (XIAP) IRES to begin a search for similar structures.

1.4 18S complementation and modular elements

It has been pointed out by Chappell et al. (Chappell et al., 2000) that partial IRES activity is still retained when segments of a 5' UTR ascribed to full IRES activity are partially deleted and therefore some elements which help to recruit ribosomes must still exist in the remaining sequence. In some UTRs, non-overlapping segments of sequence retain partial IRES activity suggesting different modules may synergistically act to provide full IRES activity *in vivo*. Non-overlapping fragments of the Kv1.4 IRES which each retained partial activity showed different patterns of activity when tested in a variety of cell types (Jang et al., 2004). This suggests distinct modular elements with different modes of regulation. Some

examples of postulated IRES elements are listed in Table 3. Chappel et al. (Chappell et al., 2000) had found an example of a distinct module with a 9 nt. motif in the Gtx mRNA, complementary to 18S rRNA that can function as a site for internal initiation of translation. This is an attractive model for ribosome recruitment for internal initiation as it parallels the function of the bacterial Shine-Delgarno sequence and 16S rRNA. This complementarity to the 18S rRNA is not new and has also been shown to be necessary for reattachment for scanning of the mRNA during "ribosome shunting" where the ribosome is stalled due to a complex structure in the mRNA and must disengage and then re-engage the mRNA on the 3' side of the structure (Yueh and Schneider, 2000). Similar motifs with IRES activity exhibiting 18S rRNA complementarity were found in a library of random nucleotides (Owens et al., 2001), in the plant potato virus Y RNA (Akbergenov et al., 2004) and have been suggested to be in YAP1 and TIF4631 transcripts of yeast (Zhou et al., 2001). Additional copies of the elements from either Gtx or the potato virus Y arranged in tandem produced additive increases of IRES activity. A segment in the 3'UTR of a hibiscus plant virus, although perhaps not an IRES, was shown to enhance translation due to an 18S rRNA complementation (Koh et al., 2002). This is not a universal truth about sequences complementary to 18S as several of the matches in the YAP1 and Tif4631 5' UTRs are in segments that confer no IRES activity (Zhou et al., 2001). The rRNA/mRNA interaction can not be too great as increasing the degree of complementarity increases the thermodynamic stability of this interaction and lowers the efficiency of translation, and if the degree of complemtarity is large enough, it can completely inhibit translation (Hu et al., 1999; Verrier and Jean-Jean, 2000). This differential ability to translate an mRNA based upon the rRNA and mRNA interactions as well as the interactions due to changes in the structure of ribosomal subunits from one cell type to another is postulated as an overall

method of translation control in "The ribosome filter hypothesis" of Mauro and Edelman (Mauro and Edelman, 2002). The last few years have seen the realization that rRNA is no longer just a scaffolding for the ribosomal proteins; the ribosome is a ribozyme and translation is now more RNA centered (Woese, 2001). It is therefore reasonable to believe with present day evidence that interactions between mRNA and rRNA can enhance translation initiation.

Relatively small IRES elements have also been synthetically derived using a bicistronic vector with 50 nucleotide long randomly generated sequences inserted between two reporter genes (Venkatesan and Dasgupta, 2001). Although two of the sequences that exhibited IRES activity showed no sequence homology with any transcripts in the public databases, the findings did show that a relatively small sequence is needed for IRES activity. In this case, the synthetic IRES are not complementary to 18S rRNA but were able to compete for the trans-activating factors utilized by the poliovirus IRES. This procedure was repeated again in yeast and 56 IRES functioning elements were found with 10 having significant matches in the yeast 18S rRNA (Zhou et al., 2003).

Fine mapping of the c-myc IRES sequence found a minimal 50 base sequence which was responsible for the bulk of the IRES activity (Cencig et al., 2004). The c-myc sequence did not seem to map to 18S rRNA and was also not dependent upon the secondary structure formed for activity. Within this element, two 14 nt. segments with an AX_6AC motif were chiefly responsible for ribosome recruitment reducing c-myc IRES activity to these modular units. The IRES found within the APC gene that possibly is responsible for the milder form of Adenomatosis Polyposis Coli is only 84 nucleotides long (Heppner Goss et al., 2002) and may be also representative of a modular unit with IRES activity.

In an Acyclovir resistant strain of Herpes Simplex Virus very low levels of thymidine kinase are translated by a small IRES which requires only 12 bases and contains a CUG start codon (Griffiths and Coen, 2005). The low level of thymidine kinase prevents its activation of Acyclovir but it is high enough to retain the virus's pathogenicity. As well as being a new IRES modular element, this also shows how very low levels of IRES translation initiation can be physiologically significant.

The above evidence supports that an IRES can be made of modular units that act synergistically with or without trans-acting factors to recruit a ribosome and enhance translation initiation. The overall structure of the UTR for these small units may be somewhat unimportant. A very stable structure like a large hairpin or a tertiary structure that would bury a modular recognition sequence would probably still have an effect on IRES activity. The modulation of the structure by mRNA binding proteins could therefore have a regulatory effect by changing the access to these small modules. As these short IRES sequences may be available elsewhere on mRNA sequences, they suggest the possibility of a much expanded proteome (Griffiths and Coen, 2005).

1.5 IRES structure

RNA motifs recognizable by RNA binding proteins are often made up of structure and/or some primary sequence. Some of the well known mRNA structural elements are the Iron Response Element, the SElenoCysteine Insertion Sequence, and the viral HIV-1 TAR. Structures can be preserved without preserving the primary sequence but proteins often bind onto specific sequence motifs and so to search for common attributes of known IRES sequences one would expect a mixture of these two characters, structures and some small sequence motifs being preserved. Although it is the 5' UTR sequence that is responsible for

recruiting the ribosome, the 3'UTR may play a similar role (Izquierdo and Cuezva, 2000) or synergistically enhance the IRES mediated translation of an mRNA (Dobrikova et al., 2003; Koh et al., 2003; Lopez de Quinto et al., 2002). It is understood that in 3'UTRs there is a role played by motifs that control translation (Mazumder et al., 2003). Virtually all the characterized IRES have been tested and function independently of their 3'UTRs, for this reason I am solely considering the 5' UTRs in this thesis.

There are several methods to experimentally determine RNA structure, such as x-ray crystallography, NMR, chemical and enzymatic structural probing, as well as mutational analysis (Kjems and Egebjerg, 1998). Structures that have been predicted using computer programs often are functionally tested for validity with small sequence mutations (Kanamori and Nakashima, 2001). X-ray crystallography and NMR, while more definitive in determining the structure, have limitations on either the type or length of the RNA molecules that can be investigated. RNA alone tends to form poor crystals that exhibit weak diffraction or heterogeneous samples (Ke and Doudna, 2004), possibly due to the dynamic folding of many RNA molecules. The only crystal structures of RNA, not complexed with proteins, and greater than 76 bases in length in the structure databases, are either ribonuclease P or group I intron ribozymes. The limitation for NMR is due to the severe spectral overlap of the four different nucleotides. These two methods have been used in examining small structurally stable motifs extracted from larger RNA molecules. Enzymatic and chemical probing are aids in determining a secondary structure of an RNA molecule but often have some ambiguous results due to the inherent difficulty in the experiments and interpreting the results. These wet-lab determinations of structure can also be supported with phylogenetic data comparisons where sequences that have changed have preserved the structure to preserve the function of the structure. The best data from all these approaches

are used to support the proposed model, but conflicting secondary structure models do arise in the literature, as has been evident for domain II of HCV (Honda et al., 1999; Lytle et al., 2002; Zhao and Wimmer, 2001).

We must also consider how much of the proposed mRNA structure is significant for IRES activity *in vivo* as compared to it in a naked state *in vitro* when it is derived. Messenger RNAs *in vivo* start folding as they are transcribed and are covered with the protein complexes required for translocation, splicing, stabilization, capping, polyadenylation, cellular localization, and translation. *In vivo* chemical probing of structures has been performed on relatively abundant RNAs like rRNA, telomerase RNA, snRNA (Mathews et al., 2004; Zaug and Cech, 1995) and could possibly be used to determine whether IRES structures determined *in vitro* exist *in vivo* as well.

1.5.1 Viral IRES Structures

Several viral IRES share similar RNA secondary structure suggesting that similar structures instead of specific sequence recognition sites are used to bind initiation factors used for cap-independent translation. This is a great aid in understanding the IRES mechanism overall but it must be remembered that ssRNA viruses have a life cycle in the cytoplasm and will not interact with the nucleoprotein complex like cellular mRNA produced in the nucleus. Therefore we would not expect the non-canonical protein factors to necessarily interact in exactly the same mechanism. Several groups have used characteristics of the viral IRES to separate them into three groups. Type I, which include entero- and rhinoviruses, translate poorly in rabbit reticulocyte lysates (RRL) and require the ribosome to bind and then scan downstream to a start codon 30 to 150 nucleotides away. Type II viral IRES include cardio- and apthoviruses, translate very efficiently in RRL,

encompass the AUG start codon and do not require scanning. Type I IRES are stimulated by the eIF4G cleavage products produced by the viral protease but Type II are not. Type III IRES are typified by Hepatitis A, do not translate at all in RRL, and encompass the AUG start codon. These classes, which may not definitively separates viral IRES types serve to show how some viruses use a similar mechanisms to initiate translation while others use distinctly separate mechanisms to arrive at the same end goal.

Comparative studies involving covariation analysis as well as enzymatic and chemical probing of structures demonstrated a conservation of structures between members of the Picornaviridae family, the enterovirus, Poliovirus and Coxsackievirus B3, and the human rhinovirus (Pilipenko et al., 1989b; Rivera et al., 1988). Similar studies found sequence and structural conservation between the Picornaviridae cardioviruses (EMCV, TMEV) and the apthovirus FMDV (Pilipenko et al., 1989a). Further examination of variants and structural probing of Poliovirus (Pilipenko et al., 1992), and mutational analysis of EMCV (Hoffman and Palmenberg, 1995; Kolupaeva et al., 1996), and FMDV (Lopez de Quinto and Martinez-Salas, 1997) have further refined the original structure models. The preservation of a stem structure in the base of region 3 of FMDV has been shown to be more important than the sequence that creates the structure (Martinez-Salas et al., 1996; Martinez-Salas et al., 2002). The central domain was re-analyzed with chemical and enzymatic probing altering slightly the previous structure model (Fernandez-Miragall and Martinez-Salas, 2003).

Several members of the Flaviridae family contain IRES sequences for which structures have been determined. In the Pestivirus genus, there is Bovine Viral Diarrhea Virus (BVDV), and Classic Swine Fever Virus (CSFV), in the Hepacivirus genus there is HCV and unclassified are the GBV-A,B and C viruses. Lemon and Honda (Lemon and Honda, 1997) reviewed the similarity and structural importance of their IRES structures. The IRES

structure that has been studied most is from the Hepatitis C virus (HCV) due to the virus's importance as a human pathogen. The initial secondary structure had been predicted using both enzymatic probing results as constraints on an early version of the MFOLD program (Zuker and Stiegler, 1981), along with comparative computer predicted models of the UTRs of BVDV and CSFV (Brown et al., 1992). There have been numerous studies adding to the refinement of the IRES structure (Lemon and Honda, 1997; Lyons et al., 2001; Odreman-Macchioli et al., 2001) arriving at two similar models (Honda et al., 1999; Lytle et al., 2002; Zhao and Wimmer, 2001) with some differences in domain II. Smaller stem loop motifs of the HCV IRES structure have had their tertiary structure determined using both x-ray crystallography (Kieft et al., 2002) and NMR (Collier et al., 2002; Klinck et al., 2000; Lukavsky et al., 2000). Recently the tertiary structure for domain II has been determined with NMR (Lukavsky et al., 2003), which finally resolves the conflicts of structural models proposed for this domain. It is important to note that the extensive chemical and enzymatic probing of the HCV IRES over 10 years was still not accurate due to the limitations of that approach. Using cryo-EM, studies of the HCV IRES in the 40S ribosome subunit, domain II, which is necessary for IRES activity specifically produced conformational changes to the 40S ribosome (Spahn et al., 2001). Segments III e and f have recently been shown to interact with ribosomal protein S5 in an IRES specific manner which does not seem to be important in cap-dependent translation initiation (Ray and Das, 2004), suggesting that interaction of IRES-containing RNA with the ribosome can be different than that of mRNA that does not contain an IRES, even without considering non-ribosomal protein factors.

The HCV IRES is an example of how both tertiary and secondary structure is important for IRES function. Conservation of specific stems in domain II preserves IRES function regardless of the sequence (Honda et al., 1999). Point mutations which alter the tertiary

structure of interdomain interactions (Kieft et al., 1999) and mutations to the pseudoknot (Wang et al., 1995) severely affect IRES activity. The importance of these domains have become clearer as domain III has been shown to directly interact with eIF3 and the 40S ribosomal subunit (Kieft et al., 2001) and cryo-EM studies have shown domains II and III wrapped on the 40S subunit both interacting with the E-site and creating a conformational change of the subunit itself (Spahn et al., 2001). Further studies of this interaction show that the HCV IRES structure sits in the same spot of the 40S normally occupied by initiation factor eIF4G (Siridechadilok et al., 2005) suggesting that the IRES structure has taken the place of eIF4G, which is not needed for IRES function *in vitro*.

The other known flaviviridae IRESes from GBV-B, BVDV and CSFV share similar secondary structures to HCV (Lemon and Honda, 1997). BVDV and CSFV structure models (Brown et al., 1992) were originally proposed by aligning four sequences, doing manual sequence covariation analysis and using the determine of base pairs as constraints on the FOLD program (Zuker and Stiegler, 1981). Mutation analysis has substantiated the CSFV pseudoknot (Fletcher and Jackson, 2002; Rijnbrand et al., 1997), stemloop IIIa (Fletcher and Jackson, 2002) as well as other sections of the structure with enzymatic probing (Kolupaeva et al., 2000). Only a little supportive work (Grassmann et al., 2005) has been done on the original model of the BVDV UTR structure (Brown et al., 1992). GBV-B was modeled comparing similar sequences to HCV and using those that might basepair as constraints in the MFOLD program, followed by mutational studies on the similar stems and loops in domains II and III (Rijnbrand et al., 2000), as well as NMR on specific domain III stemloops (Rijnbrand et al., 2004).

RNase P is an endoribonuclease that processes the 5' end of tRNA precursors to the correct length. It has been used to cleave the 5' UTR IRES containing regions of HCV,

BVDV, CrPV, EMCV and CSFV (Lyons and Robertson, 2003). This suggest that the structures of these IRESes may mimic a portion of a tRNA. Lyons and Robertson (Lyons and Robertson, 2003) postulate a model where these IRESes would occupy the E site of the ribosome, positioning it for the AUG start codon downstream. A structural element similar to tRNA, spatially situated so the start codon would be properly placed in the ribosome would greatly enhance our ability to find IRESes in the database, as effective search methods exist that can find these structures within genomic sequences (Lowe and Eddy, 1997; Tsui et al., 2003). These putative tRNA-like sequences may not reflect a complete tRNA structure as RNase P cleaves several different RNA structures (Gopalan et al., 2002), including a minimal stem structure of a tRNA molecule as a substrate (Zuleeg et al., 2001). A complete tRNA structure may also not be needed for affinity to the E site of the ribosome.

This proposed structural element used to direct the positioning of mRNA in the ribosome has an even greater role in CrPV IRES mediated translation where translation initiation occurs without eIF2 and an initiator Met-tRNA$_i$ in the P-site (Wilson et al., 2000b). The first activated tRNA starts translation on an Alanine codon at the A-site. The interactions in CrPV substituting for Met-tRNA would be different than those for a tRNA but still suggest structural elements that may be shared with other IRES. This mechanism may have evolved to overcome the antiviral response of the cell to shut off protein synthesis though PKR phosphorylation of eIF2α (Stark et al., 1998). Cellular stress leading to transcription of TNF-α and IFN-γ activates PKR as well through conserved RNA motifs in the UTRs of their mRNAs (Kaempfer, 2003), requiring cellular proteins to translate in this context in an alternative manner, possibly akin to CrPV.

Other Dicistroviridae also seem to be able to initiate translation without a methionine tRNA. A structural model was proposed using the intergenic IRES sequence upstream of

the capsid proteins from PSIV (Plautria Stali Intestine virus), TSV (Taura syndrome virus) and ABPV (Acute Bee Paralysis virus) and used to search nucleotide databases with a pattern searching program (Nishiyama et al., 2003). The key region of the model was shown experimentally to be a pseudoknot (PK1) 5' of the capsid coding region. This structure, even with relaxed search parameters was not found in extensive database searches strongly suggesting this functional structure to be distinct to Dicistroviridae.

There is a definite conservation of RNA structures between groups of viruses within their IRES sequences. The mutation analysis studies have shown the preservation of these structures in viral RNAs are more important for IRES function than the actual sequences.

1.5.2 Cellular IRES Structures

We know that viruses have evolved regulatory mechanisms borrowed from host cells and therefore it is felt that since some viruses have shared IRES RNA structure, some cellular IRES most probably have a shared structure/function relationship. At this time, the secondary structure has been derived for several cellular IRES with enzymatic and chemical probing of the transcripts of C-Myc (Le Quesne et al., 2001), L-Myc (Jopling et al., 2004), Apaf-1 (Mitchell et al., 2003), FGF-2 (Bonnal et al., 2003b), FGF1 (Martineau et al., 2004), Kv1.4 (Jang et al., 2004), Bag-1 (Pickering et al., 2004), Igf2 (Pedersen et al., 2002), cat-1 (Yaman et al., 2003), Mnt and MTG8a (Mitchell et al., 2005). Their structure/function relationships will be described below.

An early computational study has predicted a common Y structure (Le and Maizel, 1997) for cellular IRES based upon the comparison of several orthologs of Bip and FGF2 UTRs. This pattern had been adapted for the PATSEARCH program (Grillo et al., 2003) to annotate the UTRdb entries as putative IRES motifs and is used by the UTRscan webserver.

Of the 33677 entries in Release 20 of UTRdb, 7122 entries have been annotated as IRES, which is 21% of all the entries.

Some studies have proposed that mRNA binding proteins open up the natural RNA structure of the IRES and present single stranded RNA for other ITAFs or the small ribosomal subunit to bind. Therefore the structure would not be the "landing pad" for the ribosome but the 'attenuator' that is controlled by the ability of some proteins to change it. For example, the poly rC binding protein 1 (PCBP1) appears to open the Bag-1 IRES structure allowing PTB-1 to bind. Mutations which open up the binding region of the PCBP1 on the Bag-1 IRES seems to remove any requirement for this factor even enhancing IRES activity after PTB-1 is added (Pickering et al., 2004). Although PCBP1's role seems to be to open up the structure for PTB-1, it is not clear whether the structure is necessary for its binding. A similar mechanism was proposed for the ITAFs Unr and PTB in the APAF-1 IRES (Mitchell et al., 2003). In that case Unr opens up the secondary structure allowing PTB to bind by possibly exposing recognition sequences that were previously buried or double stranded. IRES activity regulated by structure is also described by the "zipper model" of the CAT-1 (cationic amino acid transporter) transcript. This IRES structure only attains its active form upon amino acid starvation and the translation of an upstream ORF (Yaman et al., 2003). Structural probing has shown RNA-RNA interactions which inhibit IRES activity are opened up by the scanning ribosome and perhaps by unknown ITAFS during translation of the upstream ORF.

The relevance of the c-myc IRES structure is still outstanding. Initially, full activity was assigned to a 394 base region upstream of the AUG whose structure was derived using chemical and enzymatic probing (Le Quesne et al., 2001). Mapping of the ribosome landing site by engineering several AUGs along the 5' UTR revealed that start codons that were at

least 220 bases upstream from the standard AUG had no effect on the downstream reporter ORF. Interestingly, domain 1 (-380 to -221) upstream of the proposed ribosome binding site could have a negative effect on IRES initiated translation if a couple of the dsRNA regions were disrupted, suggesting some structural requirements. Mutations that disrupted the predicted pseudoknots in position -185 to -347 only seemed to enhance IRES activity perhaps showing in the case of this IRES, that the pseudoknots attenuate IRES activity and proteins that would open them up would allow for IRES activity. More recent work has shown that the major portion of c-myc IRES activity can be minimized to a 50 base region from -143 to -94 (Cencig et al., 2004) and that the sequence and not the proposed structure was important for IRES activity (Cencig et al., 2004). *In vivo*, the overall structure of the UTR may provide a role for IRES regulation as increased levels of PCBP1 and PCBP2 enhance c-myc IRES activity (Evans et al., 2003), possibly opening up the structure for other ITAFs and the ribosome, as with Bag-1.

An interesting structural study has been performed on a segment of the Kv1.4 IRES where a series of deletion mutants have had their structures derived from enzymatic probing (Jang et al., 2004). A 0.2 kb fragment on the 3'end of the 1.2 kb Kv1.4 UTR has only about 25% percent of the activity of the full UTR but specific internal deletions raised the activity to that of the full UTR. In contrast, any deletions of the polypyrimidine tract seemed to remove all activity even if some of the original structural elements were mostly preserved. As the 1 kb region upstream to this segment also retains activity on its own without the polypyrimidine tract, the conclusion seems to be that several modular elements of sequence and structure contribute to overall IRES activity. Unfortunately the data are not clear on what defines the contributing elements other than the polypyrimidine tract.

FGF2 has a UTR with a unique G-quartet structure within its IRES (Bonnal et al., 2003b). Although deletion of this element cuts the relative IRES activity in half, deletion analysis has shown stem-loop II has the greatest effect on overall activity. Deletion or changes of the loop alone can achieve this effect. The entire stem is not required for activity and it is not clear whether only the sequence of the loop needs to be available as a single stranded region to preserve the effect of stem-loop II.

The minimal IRES sequence of FGF1A seems to be conserved over several mammalian species. Computationally predicted structures from the mouse, rat, cow and two primates sequences seem very similar to the empirically derived human sequence structure (Martineau et al., 2004). The most important element, DII, as shown by deletion analysis, retained more activity when mutations preserved the structure rather than changed the structure. This suggests the need for some structural elements for complete activity.

In all these studies on cellular IRES, only some show a possible preference for structure preservation to retain full activity, suggesting there is no need for an overall structure as seen in some viral IRES.

1.6 Databases and IRES

A searchable database of published IRES exists at http://ifr31w3.toulouse.inserm.fr/IRESdatabase/ (Bonnal et al., 2003a). Unfortunately the database has not been updated since 2002 and Table 1 includes all published cellular IRES sequences to May, 2006. Many of the IRES publications did not give exact coordinates to specific database entries for the UTR sequences used in their research. For this reason Table 1 includes the Genbank Identifier number (GI) and the position on the sequence of the minimal known IRES. Filling the gap left by the previous IRES database, a new online

database of IRES sequence has been created at http://www.iresite.org/IRESite_web.php (Mokrejs et al., 2006).

IRES structure data for 16 cellular IRES and 5 viral IRES sequences are available at Rfam (Griffiths-Jones et al., 2003), a database of non-coding RNA. For the most part, published IRES structures are used initially at Rfam, a covariance sequence model is also built from the known structure using UTR sequences from different transcript entries and known ortholog sequences. Where no published structure is known, the covariance models for the IRES structures have either been built from *de novo* structure predictions from energy minimization program predictions (RNAFOLD), or from a sequence alignment using the PFOLD program. It should be noted that if the alignment sequences do not give any covariance data, the seed structure would dominate the resulting structure model. This could possibly be the case for HCV, whose Rfam structure appears to be little different from its seed structure (Lytle et al., 2002) but dissimilar from the HCV NMR structure data (Lukavsky et al., 2003). For tertiary structure information of IRES elements, data can be found in the structure database at NCBI or RNAbase (Murthy and Rose, 2003), a specific RNA structure database which may include additional annotations not included in the standard 3-d structure files. At present there are 13 IRES related elements from GBV-B, Enterovirus, and mostly HCV. Tertiary structures can be converted into secondary structure diagrams using RNAVIEW (Yang et al., 2003).

There are other available datasets of IRES enriched sequences where microarray expression studies have been carried out on cells undergoing a perturbation that would turn down the amount of cap-dependent translation and allow for a greater amount of IRES mediated translation. Assuming that IRES elements under these conditions would more efficiently recruit ribosomes, experiments that isolated mRNA bound by multiple ribosomes

(polysomes) should show mRNAs with a greater likelihood of containing an IRES element. Johannes et al. (Johannes et al., 1999) isolated polysomes from poliovirus infected cells and evaluated any increased amount of bound mRNA compared to non-infected cells using a 10K human cDNA array. They found approximately 200 transcripts with a greater than two fold enrichment of polysome bound mRNA of approximately 7000 hits that produced an acceptable signal on the microarray. There are several other studies using microarray examination of polysome bound mRNA, cells with and without von Hippel-Lindau tumor suppressor protein (Galban et al., 2003), rapamycin's effect on translation (Grolleau et al., 2002), resting and mitogenically activated fibroblast (Zong et al., 1999), synchronized cells in the mitotic cycle (Qin and Sarnow, 2004), hypoxia (Blais et al., 2004), TRAIL induced apoptosis (Bushell et al., 2006) and yeast during cell-cycle arrest (Serikawa et al., 2003) or rapid growth (Arava et al., 2003) . Some research has combined the proteomic approach of measuring both the increase of protein expression using 2D gels and mass spectrometry as well as the mRNA levels (Grolleau et al., 2002; Ideker et al., 2001). This allows one to specifically discover transcripts that are regulated post-transcriptionally. Microarray studies of mRNA bound to mRNA binding proteins implicated as ITAFs will also yield a database of transcripts of which some may contain IRES.

Many more characteristics of the UTR sequences could be examined by comparing and crossing all of these datasets looking for sequence or predicted structural motifs that may be required for translation initiation in the periods of translation dysregulation.

1.7 RNA structure prediction and search software

RNA secondary structure is usually predicted computationally by biologists in two ways. The first would use thermodynamic values for the base-paring, base stacking and near

neighbour forces to calculate the most thermodynamically stable structure in the "minimum free energy" method. The second method is a comparative method that examines the covariation of several homologous sequences to predict sequence positions that would consistently form base pairs even if specific base pairs change from one homologous sequence to another. Covariation is seen as the most accurate method of structure prediction but requires many sequences with similar structures to do this properly. The most extensive structure prediction using this methods has been done by R. Guttell and colleagues for over 25 years on ribosomal RNA where crystallographic data of the ribosome have shown their structure models to be 97% accurate (Gutell et al., 2002). Covariation structure models have been predicted for viral IRES but unfortunately the relatively small number of similar sequences for cellular IRES does not allow this method to be used to fully define a structure. Covariation models can also produce evolutionarily preserved structures or specific basepairing that may rely on co-factors like ions or proteins. These structural elements would not be apparent when empirically examining naked RNA structures without these factors present. Variations and combinations of these two methodologies, thermodynamics and covariation, combined with other computational algorithms have increased the predictive ability of computer programs. For a review on the development of secondary structure prediction programs see Zuker (Zuker, 2000), Higgs (Higgs, 2000) and Mathews (Mathews, 2006; Mathews and Turner, 2006). There have been a few studies (Doshi et al., 2004; Dowell and Eddy, 2004; Ji et al., 2004; Perriquet et al., 2003) which have evaluated and compared the accuracy of some prediction software. In this section of the introduction and within the results section of the thesis I examine a few commonly used computer approaches to RNA structure or pattern predictions in more detail as it applies to IRES (Baird et al., 2006).

1.7.1 Single sequence structure prediction

It has been hard to predict non-coding regions of transcripts within genomic sequence leaving sequence annotation of UTR regions of messenger RNA to empirical evidence of cDNA, 5' RACE methods, and cap trapping methods (Carninci et al., 1996). Although distinct ordered structures may exist, the overall degree of secondary structure as measured by the lowest free energy of possible folds of non-coding regions of mRNA was not seen to be different from random sequence or from coding sequence (Clote et al., 2005; Rivas and Eddy, 2000; Workman and Krogh, 1999). There could be possible problems in assessing the actual energy status of the structure because tertiary interactions which would effect the overall energy calculations are not calculated or the global secondary structures effect on the actual folding are not considered with each prediction (Workman and Krogh, 1999). There has been some evidence to show that if one examines simpler putative hairpin structures within nematode genomic sequence, below a certain level of free energy, there are less of these sequences in UTR regions than intron and intergenic regions (Pervouchine et al., 2003) suggesting lower overall secondary structure in UTRs.

ED_SCAN (Le et al., 2003) looks for local energy differences within a sequence from a sliding window. Plotting the energy differences across a sequence has shown peaks for the Iron Response Elements (IRE) in ferritin RNA, Rev elements in HIV and let-7 RNAs in nematode and fruitfly (Le et al., 2003). This has the possibility of finding specific IRES secondary structure motifs within a UTR if such a thing exists.

MFOLD is a powerful "minimum free energy" program and probably the most commonly used software for predicting RNA secondary structures based upon the thermodynamics of the possible folds using a well defined empirically determined energy table (Mathews et al., 1999a). It has been evaluated (Doshi et al., 2004; Mathews et al.,

1999a) and most recently shown to average 41% accuracy of predicted basepairs for 16S and 23S RNA (Doshi et al., 2004). Studies on other datasets have shown an MFOLD accuracy of 56% (Dowell and Eddy, 2004). Users of this software tend to assume RNA folds will undertake their most thermodynamically stable form, the structure with the minimal free energy, but unfortunately in nature, if one examines functional RNAs like tRNA and RNase P RNA, this is not always true. Correct structures are believed to usually lie within 10% of the minimum energy value predicted (Shapiro et al., 2001) and therefore MFOLD and similar programs will also produce suboptimal folds to increase the likelihood of an accurate prediction. MFOLD has some common limitations in that it can not predict interloop base pairing, pseudoknots and, like all other programs, does not show non-canonical basepairing, with the exception of the G-U wobble, although non-canonical pairing energy values are calculated from the energy tables. PKNOT (Rivas and Eddy, 1999) takes the MFOLD algorithm a step further to allow pseudoknots but it does not allow user defined constraints on the predicted folds. The computational expense of inclusion of pseudoknots restricts the sequences examined by PKNOT to approximately 140 bases, which would be limiting for looking at IRES sequences. HOTKNOTS uses a heuristic method of solving the most stable structure with pseudoknots and takes larger sequences (Ren et al 2005). Pseudoknots were formally considered rare but several viral and cellular IRES contain them. A recent analysis of RNA structure topologies estimates pseudoknots in 90% of the sequences over 120 base in length (Kim et al., 2004). One of MFOLD's great advantages is that experimental or phylogenetic comparative data can be used as constraints in the structure prediction, greatly increasing the accuracy of the prediction.

Mathews (Mathews, 2004) has updated the energy calculations in his Windows program **RNASTRUCTURE** a derivative of MFOLD. It can incorporate enzymatic and chemical

modification probing of the structure as constraints showing an increase in accuracy by 26% to 86% over *in vivo* structure prediction (Mathews, 2004). The latest version also includes a partition function calculation that predicts the basepairing probabilities for all possible basepairs. These probabilities can then be used to estimate the confidence in predicted basepairs in low free energy structures. Knowing stems with strong probability of forming would be advantageous in the design of further experiments. A comparative survey of computationally predicted RNA structures of Flaviviridae (Thurner et al., 2004) reported that their structure prediction of the HCV IRES was similar to empirically derived published structures when using the McCaskill's partition algorithm (McCaskill, 1990). The Vienna package also offers a minimal free energy program, **RNAFOLD** (Hofacker, 2003), similar to MFOLD. Although these three programs use the same energy data table, they use different energy rules to calculate the structures. While RNAFOLD gives very similar results to MFOLD for the minimal free energy structure of a given sequence, RNASTRUCTURE produces a different lowest energy structure and a variety of different suboptimal folds.

Whereas dynamic programming as demonstrated by MFOLD finds the mathematically optimal path through a matrix of all the possible base pairings formed by an RNA molecule, "Genetic Algorithm" (GA) selects structures from a large changing population of structures. The foldings with the highest degree of fitness, which usually have the lowest free energy, are selected, rearranged, and "mutated" for the next generation of structures until there is a convergence. A massive parallel implementation of this algorithm for several supercomputers has predicted 80 to 90% of the correct basepairing in the poliovirus 5' UTR (Shapiro et al., 2001). Until recently, this approach was limited to specific high performance machines, it has now been adapted to other parallel computers like Linux clusters.

1.7.2 Conserved structures in unaligned sequences

Although the overall structure is important for IRES like that in HCV, several cellular IRES show reduced activity when non-overlapping pieces of the sequences are tested for cap-independent translation suggesting that smaller modules may be structural motifs for RNA binding proteins. The cellular IRES secondary structures published so far do not share an overall similarity and if classes of IRES exhibit activity in different cellular context, they may only share a similar small structural motif. If this is the case, software capable of finding specific structural motifs in unaligned sequences may be more useful than finding global similarities.

The **DYNALIGN** (Mathews, 2005; Mathews and Turner, 2002) program is an adaptation of the Sankoff dynamic programming algorithm (Sankoff, 1985), which seeks similar structures between two RNA sequences using free energy minimization and not sequence identity. In contrast to the original Sankoff algorithm, DYNALIGN limits the comparison between only two sequences within user-defined parameters limiting the distance between aligned nucleotides to make the computations more manageable. As the program looks for a global structure alignment, the two sequences would be expected to have a similar length and overall structure. If one of the sequences has a known structure, it could possibly help to assess the accuracy of the prediction of an unknown sequence. It has not been shown whether DYNALIGN could predict a shared motif if several sequences were used in a pairwise comparison that did not share a global structure. It has been updated to include experimental and comparative data as constraints, as well as updated energy values (Mathews, 2005).

CARNAC (Computer Alignment of RNA by Cofolding) (Perriquet et al., 2003) takes two or more homologous sequences and using energy minimization, sequence similarity and

covariation devises a shared secondary structure model. It is available as a webserver at http://bioinfo.lifl.fr/carnac/carnac.html.

PFOLD (Knudsen and Hein, 1999; Knudsen and Hein, 2003) uses a comparative method using stochastic context-free grammars (SCFG) to predict a common structure. SCFG are a set of rules determined by probability from examining a set of sequences. PFOLD assumes that all the sequences are related through evolution and therefore can be aligned and share a common secondary structure. Input requires a structurally correct alignment of the sequences for accurate results. Since this is not possible without some structural data, a multiple alignment from a program like ClustalW with as many sequences as possible which are close in evolutionary distance will improve accuracy. PFOLD has been used to predict the IRES structures of FGF2, MNT, BIP, HSP-70 and BAG-1 in the RFAM database using UTRs from several transcripts or orthologs.

There are several other programs that can use unaligned sequences as inputs to generate structure predictions. RNAGA (Chen et al., 2000) uses a genetic algorithm to predict an overall structure from a group of structurally related sequences. The inputs for RNAProfile (Pavesi et al., 2004) are a group of sequences which 'mostly' share a conserved motif and the number of hairpins being sought. It gave good results for test cases of the iron response element (IRE), the SECIS motifs as well as hairpins within RNase P (Pavesi et al., 2004). ComRNA (Ji et al., 2004) like RNAProfile in the first step, produces all the stems that exist in each of the input sequence. If a critical number of sequences share a subset of similar stems the assembled compatible stems are used to construct a secondary structure profile and with refinement the best structures are reported.

1.7.3 Database Searching

Whereas DYNALIGN is able to search all the possible folds in comparing two sequences in order to find a structural similarity, this method would be overly taxing in terms of computational time and space to use it to search a complete database. One comparison takes about 15 minutes on a machine of moderate speed, therefore a database search of 20,000 entries would take 6 months. RSEARCH (Klein and Eddy, 2003) is designed to search a database with a sequence that has a known secondary structure. Scoring matrices analogous to the BLOSUM (Henikoff and Henikoff, 1993) protein matrices used by BLAST and FASTA were built with alignments of the small subunit rRNA and are called RIBOSUM. One scoring matrix is for the basepairing changes and one is for the base sequence changes. Comparison of the query sequence to the unknown sequence is done using an algorithm termed "profile stochastic context-free-grammar" (SCFG). A profile SCFG also termed "covariance model" is produced from the query sequence and each state of the comparison to an unknown sequence is taken care of in the model, whether it is a nucleotide match, a basepair match, a gap on either side, or end of stemloop. Although the computational expense is very high, this program seems closest to what a molecular biologist would use to find new sequences sharing a similar structure/function relationship with a query sequence.

Pattern matching programs serve the user by searching primary sequence databases for structural patterns that they have defined or already know to exist. Several programs have been developed with their own search mechanism and pattern description syntax: RNAMOT (Laferriere et al., 1994), PALINGOL (Billoud et al., 1996), PATSEARCH (Grillo et al., 2003), and RNAMOTIF (Macke et al., 2001). **RNAMOTIF** is a successor to the secondary structure pattern searching programs RNABOB (Eddy, 1996) and RNAMOT. It

is the most powerful pattern-searching program now available because of its extensive and flexible descriptor scripting language. It can find pseudoknots as well as triplexes and G-quartets that are found in the FGF2 IRES (Bonnal et al., 2003b). It has been used successfully to predict Rev response elements in lentivirus (Lesnik et al., 2002), small RNA molecules in the *E. coli* genome (Chen et al., 2002) and tRNA genes (Tsui et al., 2003). It has also been coupled with an "evolutionary algorithm" starting with a loose structure descriptor with no sequence constraints and successfully picked up Iron Response Elements and Signal Recognition Particles in a wide variety of organisms (Fogel et al., 2002).

ERPIN (Easy RNA Profile IdentificatioN) (Lambert et al., 2004) takes pattern searching to a higher level of complexity by adding sequence alignments to the structure annotation creating a profile for each residue position. Due to the expertise required to prepare a correct RNA sequence alignment and parameter tuning, only pre-prepared patterns are available on the ERPIN web server for searching user submitted sequence. An e-value (Expect value) analogous to the value produced by BLAST, is calculated to estimate finding the pattern by chance with that score or higher in the same dataset (Lambert et al., 2005). The IRES from type 1 poliovirus is one of approximately 50 RNA genes or elements available as search patterns.

1.8 Thesis outline

The focus of this thesis is to search for a common overall RNA secondary structure between some cellular IRESes. It is organized in the following order of experimental objectives:

1) Compare the published IRES sequence data to 5' UTR databases to identify differences in UTR length, degree of secondary structure, presence of upstream AUG codons, GC content or presence of tRNA-like motifs.
2) To evaluate RNA motif/structure predicting and search programs in their ability to identify structures responsible for IRES function using the known HCV IRES structure as a test sequence.
3) To predict a structure for the XIAP IRES using enzymatic probing.
4) To use the RNA secondary structure model of the XIAP IRES to search for new and possibly similar IRESes to XIAP.
5) To test these predicted IRESes for IRES activity
6) Derive an empirical structure model of any new IRESes and compare them to the XIAP IRES structure.

Chapter 2

Materials and Methods

2.1 Construction and manipulation of databases

All measurements for length, AUG content, %GC content of databases and parsing of databases were done with Perl (Practical Extraction and Report Language) scripts using Bioperl (Stajich et al., 2002) modules. Databases from Ensembl (Clamp et al., 2003), NCBI RefSeq (Pruitt et al., 2003) and UTRdb (Pesole et al., 2002) were downloaded from their respective sites and converted to fasta formats.

2.2 Bioinformatic software

The converted Matthews Correlation Coefficient (MCC) as presented by Stormo's group in equation 5 (Gorodkin et al., 2001; Ji et al., 2004) was used as a RNA structure comparative metric between the actual and predicted sequences. CT structure files were used for comparison. RNA structures for all figures have been produced with RnaViz2 (De Rijk et al., 2003).

2.3 RNA secondary structure determination

RNA secondary structures were determined using enzymatic probing with RNase T1, RNase A, RNase T2, and RNase V1 according to the protocol supplied with the Ambion enzymes. RNA was *in vitro* transcribed using the MaxiShortScript kit (Ambion) following the manufacturer's protocol. All RNA structure was examined from both ends. Radioactive 5'end labeling was performed on uncut RNA with Ambion's KinaseMax Kit. Instead of 3'

end-labeling, cold, RNase digested RNA was reverse transcribed using a fluorescent tagged primer from the 3' end and the reaction was separated on a Licor IR^2 DNA Sequencer alongside a control sequence reaction of the same sequence. RNase cut sites were used as constraints in either MFOLD (Zuker, 2003), version 3.2 or RNASTRUCTURE 4.2 (Mathews et al., 2004) to predict secondary structure models.

2.4 Structure database search

A human 5' UTR database was generated from the ENSEMBL genome annotation site for NCBI build 34. Exact duplicate UTRs were removed along with shorter UTRs that fully overlapped other entries. The database was supplemented with full-length UTRs of a published IRES dataset (Baird et al., 2006). CT structure files were converted to Stockholm format and used to search the UTR database using the default parameters of RSEARCH on a grid of 6 Sun workstations.

2.5 Cell Culture and Transfection Reagents

Human embryonic kidney (HEK) 293T and 293A cells were maintained in standard conditions in Dulbecco's modified Eagle's medium supplemented with heat-inactivated 10% fetal calf serum, 2 mM L-glutamine, and 1% antibiotics (100 units/ml penicillin-streptomycin). Transient transfections were performed using Lipofectamine Plus reagent (Invitrogen) according to the manufacturer's protocol. Briefly, cells were seeded at a density of 6 x 10^5 cells in 2 ml of culture medium per well in 6-well plates and were transfected 24 h later in serum-free Opti-MEM medium (Invitrogen) with 2 μg of DNA and 4 μl of lipid per well. The transfection mixture was replaced 3 h later with fresh Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2 mM glutamine, and 1% antibiotics.

2.6 Bicistronic constructs

The pβ-Gal/CAT bicistronic vector containing the minimal human XIAP IRES was built on a pcDNA3 (Invitrogen) backbone and described previously as pβgal/5'(-162)/CAT (Holcik et al., 2003). The empty vector was modified to include an NheI site upstream of the original XhoI site to enable directional cloning of UTR sequences. The new UTRs were isolated by reverse transcribing total RNA from HEK 293T cells using the First-Strand cDNA Synthesis kit (Amersham Biosciences) with random hexamer primers and then PCR amplifying the UTRs with the following specific primers: AQP4-Nhe-F cgctagcAAGGACAGTTTGGATAAT, AQP4-Xho-R cctcgagCCACCATGATGTTCTCT, ELG1-Nhe-F cgctagcACTTTTGGTGGGCATTTA, ELG1-Xho-R cctcgagCTGGGCGGGGAATA, PCD10-Nhe-F cgctagCCTCAGTTGCTGGTAAG, PCD10-Xho-R cctcgagAAGCCAACTACAGTTGAA, VEGF-D-Nhe-F cgctagcACTTCTCTGCATTTTCT, VEGF-D-Xho-R cctcgagTTTCAATATCCACTGATT, ZINCF-Nhe-F cgctAGCCTGAGAAGATGATGC, ZINCF-Xho-R cctcgaGCTTTCCCACATTCACA. The bicistronic construct with the ATF4 5' UTR was provided as a gift from Dr. Jamie Blais. Constructs with the CMV promoter removed were created by digesting with BglII and religating. The XIAP-PPT construct with the -47 to -1 sequence removed was previously described (Holcik et al., 1999). The AQP4-PPT construct was created using the AQP4-Nhe-F primer and the AQP4-59-Xho-R primer CCTCGAGAGAAACAAATCAGCA and cloned into the pβ-Gal/CAT bicistronic vector. PCR primers for the randomly picked UTRs are given in Table 11 of Appendix II. Bicistronic constructs containing the IRESes from Apaf-1, DAP5, c-myc and Hiap2 were previously described (Henis-Korenblit et al., 2000; Marash and Kimchi, 2005; Nevins et al.,

2003; Warnakulasuriyarachchi et al., 2004). Monocistronic constructs of XIAP and Apaf-I were constructed by digesting with NotI, which excises the β-Gal ORF, and religating.

2.7 ß-Gal, CAT and NeoPTII analysis

Transiently transfected human embryonic kidney cells, HEK 293T or HEK 293A, were harvested in PBS 24 h post-transfection in CAT enzyme-linked immunosorbent assay (ELISA) kit lysis buffer (Roche), and cell extracts were prepared using the protocol provided by the manufacturer. ß-Galactosidase (ß-Gal) enzymatic activity in cell extracts was determined by the spectrophotometric assay using ONPG (*o*-nitrophenyl-ß-D-galactopyranoside) (MacGregor et al., 1991), and the Chloramphenicol-Acetyl Transferase (CAT) levels were determined by using the CAT ELISA kit (Roche) and the protocol provided by the manufacturer. Neomycin phosphotransferase II levels were detected using neomycin phosphotransferase II ELISA (Agdia) following the protocol of the manufacturer. Relative IRES activity was determined as a ratio of CAT to ß-Gal.

2.8 Quantitative RT-PCR

Total RNA was isolated from cells that were previously transfected with the β-Gal/5' UTR/CAT bicistronic constructs as described above using either the Absolutely RNA kit (Stratagene) or RNeasy kit (Qiagen). For quantitative RT-PCR, reverse transcription was carried out using the First-Strand cDNA Synthesis kit (Amersham Biosciences) with NotI-$d(T)_{18}$ primers. The quantitative PCR was performed using the QuantiTect SYBR green PCR kit (Qiagen) and analyzed on an ABI Prism 7000 sequence detection system using the ABI Prism 7000 SDS Software. Quantitative PCRs were carried out to detect β-galactosidase using primers 5'-ACTATCCCGACCGCCTTACT-3' and 5'-

CTGTAGCGGCTGATGTTGAA-3' and CAT RNA using primers 5'-GCGTGTTACGGTGAAAACCT-3' and 5'-GGGCGAAGAAGTTGTCCATA-3' as described previously (Holcik et al., 2005).

Chapter 3

Results

3.1 Bioinformatics examination of UTRs

Databases of UTR sequences can be obtained from several sources. Using the web-based ENSMART in Ensembl (Clamp et al., 2003) now known as BioMart, UTR datasets from a variety of genomes can be made with a user defined number of flanking nucleotides. As almost all genes produce several transcripts, one must still filter out redundant 5' UTRs when creating a database. A pre-filtered non-redundant database, UTRdb (Pesole et al., 2002) can be found at http://bighost.area.ba.cnr.it/BIG/UTRHome/ where UTRs with greater than 90% sequence overlap and 95% nucleotide identity within the overlapping region (Grillo et al., 1996) have been removed. The UTRdb site also has available UTR databases created from the NCBI Reference Sequence (RefSeq) project (Pruitt et al., 2003). These RefSeq datasets are redundant. The RefSeq database from NCBI can also be used to produce a dataset from the transcript entries but requires users to write their own program/script to pull out the UTR sequences. These datasets allow one to compare the overlaps in the UTR sequence characteristics believed to have IRES elements as discovered by different means versus a complete set of a genome's UTRs.

In this work, the complete annotated 5' UTR datasets of the human genome from Ensembl, UTRdb and RefSeq are compared with an IRES dataset of full-length UTRs made from the entries described in Table 1. Exact duplicate entries from both the Ensembl and RefSeq datasets have been removed. A second RefSeq dataset has been made by selecting

Table 1. Reported Cellular IRES

Gene Name	Function	Organism	GI (IRES seq. pos.)	Ref.
α-CaM Kinase II	Alpha subunit of Ca-calmodulin-dependent kinase II	Rat	203208 (280-431)	(Pinkstaff et al., 2001)
ABETA	Amyloid β A4 precursor protein	Human	341201 (899-1049)	(Qin and Sarnow, 2004)
AML1/RUNX1	Transcription factor	Human	2944212 (8498-10040)	(Pozner et al., 2000)
Antp	Antennapedia - Homeotic gene	Drosophilia	16648361 (1-1709)	(Oh et al., 1992)
Apaf-1	Proapoptotic factor	Human	2330014 (1-580)	(Coldwell et al., 2000)
APC	Adenomatosis Polyposis Coli gene	Human	182396 (487–570)	(Heppner Goss et al., 2002)
ARC	Cytoskeleton association protein	Rat	854413 (1–200)	(Pinkstaff et al., 2001)
AT1R	Angiotensin II type 1 receptor - G protein coupled receptor	Human	18490885 (1-275)	(Martin et al., 2003)
BAG-1 p36	Anti-apoptotic factor	Human	1143475 (1–413)	(Coldwell et al., 2001)
Bcl-2	Anti-apoptotic factor	Human	179366 (313-1461)	(Sherrill et al., 2004)
BiP	ER protein chaperone	Human	1143491 (1-225)	(Macejak and Sarnow, 1991)
ßPix-b_L	Pak-interacting exchange factor isoform b	Mouse	37788384 (1-375)	(Rhee et al., 2004)
Cat-1	Cationic amino acid transporter	Rat	18542255 (1–273)	(Fernandez et al., 2001)
c-Jun	Transcription factor	Chicken	212221 (500-815)	(Sehgal et al., 2000)
c-Myb	Transcription factor	Human	45502012 (1-202)	(Mitchell et al., 2005)
c-Myc	Transcription factor	Human	11493193 (2501-2881,4506-4523)	(Stoneley et al., 1998)
Connexin26	Gap junction protein	Human	1762120 (1472-1631,4780-4804)	(Lahlou et al., 2005)
Connexin32	Gap junction protein	Human	974140 (404-529,884-903)	(Hudder and Werner, 2000)
Connexin43	Gap junction protein	Rat	45593193 (1–232)	(Schiavi et al., 1999)
CCND1	Cyclin D1	Human	22788696 (1380–1591)	(Shi et al., 2005)
Cyr61	Intracellular signalling	Human	2791897 (1-226)	(Johannes et al., 1999)
DAP5	Translation initiation factor	Human	1903413 (1-357)	(Henis-Korenblit et al., 2000)
Dendrin	putative modulator of the postsynaptic cytoskeleton	Rat	1752674 (1-151)	(Pinkstaff et al., 2001)
eIF4G	Translation initiation factor	Human	21655145 (341-538)	(Johannes and Sarnow, 1998)
ERα	Estrogen Receptor α	Human	182192 (293-814)	(Barraille et al., 1999)
FMR1	RNA binding protein	Human	1668818 (13698-13964)	(Chiang et al., 2001)
FGF1a	Fibroblast growth factor 1 A - angiogenic factor	Human	178226 (1-360)	(Martineau et al., 2004)

Table 1. Reported Cellular IRES (continued)

Gene Name	Function	Organism	GI (IRES seq. pos.)	Ref.
FGF1a	Fibroblast growth factor 1 A angiogenic factor	Mouse	4321971 (865–1238)	(Martineau et al., 2004)
FGF1b	Fibroblast growth factor 1 B angiogenic factor	Human	9125828 (35-185)	(Martineau et al., 2004)
FGF1c	Fibroblast growth factor 1 C angiogenic factor	Human	Poorly defined	(Martineau et al., 2004)
FGF1d	Fibroblast growth factor 1 D angiogenic factor	Human	21595686 (10-93)	(Martineau et al., 2004)
FGF2	Fibroblast growth factor	Human	31361 (486-809)	(Bonnal et al., 2003b; Vagner et al., 1995b)
HAP4	transcriptional activator	Yeast	3762 (228-503)	(Iizuka et al., 1994)
hairless	Transcription repressor	Fly	157621 (686-1072)	(Maier et al., 2002)
Hiap2	Anti-apoptotic protein	Human	34367137 (1313-1465)	(Warnakulasuriyarachchi et al., 2004)
HIF-1α	Transcription factor	Mouse	12857319 (1-287)	(Lang et al., 2002)
HnRNPA/B	Heterogeneous nuclear riboprotein A/B	Human	33872877 (1-227)	(Qin and Sarnow, 2004)
Hsp70	Heat Shock protein	human	184416 (274-491)	(Rubtsova et al., 2003) contradicts (Yueh and Schneider, 2000)
Hsp70	Heat Shock Protein	Fly (Dm)	157720 (1514–1757)	(Hernandez et al., 2004)
Hsp101	Heat Shock Protein	Plant	4584956 (290-438)	(Dinkova et al., 2005)
IGF-II leader 1	Insulin-like Growth factor II- UTR leader 1	Human	26190552 (130580-130698, 139619-139838, 141156-141399, 153434-153443)	(Teerink et al., 1995)
IGF-II leader 2	Insulin-like Growth factor II- UTR leader 2	Human	6453816 (125-755)	(Pedersen et al., 2002)
IGF-IR	Growth factor receptor	Rat	204774 (413-1358)	(Giraud et al., 2001)
Kv1.4	Voltage-gated potassium channel	Mouse	26331157 (1-1201)	(Negulescu et al., 1998)
La1	RNA binding protein – more abundant transcript	Human	511006 (240-345,2329-2340)	(Carter and Sarnow, 2000)
La1'	RNA binding protein - less abundant transcript	Human	511006 (698-886,2329-2340)	(Carter and Sarnow, 2000)
LEF-1	Lymphoid Enhancer Factor	Human	22858703 (1523-2703)	(Jimenez et al., 2005)
L-myc	Lung –myc	Human	188906 (224–431,796-807)	(Jopling et al., 2004)
MAP2	Cytoskeleton associated protein	Rat	987493 (1–370)	(Pinkstaff et al., 2001)
Mnt	MAX binding protein - Transcriptional repressor	Human	1841919 (35-215)	(Stoneley et al., 2001)
MS	Methionine Synthase	Human	1763268 (126-397)	(Oltean and Banerjee, 2005)
MTG8a (RUNX1T1)	Transcription Factor	Human	940399 (1-411)	(Mitchell et al., 2005)

Table 1. Reported Cellular IRES (continued)

Gene Name	Function	Organism	GI (IRES seq. pos.)	Ref.
MYCHEX1	Upstream open reading frame on c-myc transcript	Human	11493193 (2223-2306)	(Nanbru et al., 2001)
Myt2	Myelin transcription factor 2	Rat	2246660 (997-1158)	(Kim et al., 1998)
Nap1L1	Nucleosome assembly protein 1-like 1	Human	461207 (16-142)	(Qin and Sarnow, 2004)
NBS1	Nijmegen breakage syndrome allele	Human	Undefined	(Maser et al., 2001)
Neurogranin (RC3)	neural-specific regulator of CaMKII activity	Rat	924645 (4217–4478)	(Pinkstaff et al., 2001)
Nkx6.1	Homeodomain transcription factor	Mouse	11118686 (2988-3959)	(Watada et al., 2000)
N-myc	Neuronal myc - Transcription factor	Human	11692795 (1-324)	(Jopling and Willis, 2001)
Notch2	Intercellular Signalling receptor	Fly	1622786 (1-238)	(Lauring and Overbaugh, 2000)
NPM1	Nucleophosmin	Human	2745708 (1222-1320)	(Qin and Sarnow, 2004)
NRF	NF-κB repressing factor - Transcription factor	Human	7406601 (1-656)	(Oumard et al., 2000)
ODC	Ornithine decarboxylase	Rat	205803 (1-426)	(Pyronnet et al., 2000)
P150(TIF4631)	Translation initiation factor homologue of eIF4G	yeast	1323279 (14130–14641)	(Zhou et al., 2001)
P27(Kip1)	Cyclin-dependent kinase inhibitor	Mouse	532771 (1-221)	(Miskimins et al., 2001)
PKCδ	Protein Kinase C delta	Rat	206180 (7-365)	(Morrish and Rumsby, 2002)
PITSLRE	Cyclin dependent kinase	Human	507159 (946-1128)	(Cornelis et al., 2000; Tinton et al., 2005)
PP2Cβ	Protein Phosphatase 2Cβ	Rat	12666526 (1-400)	(Seroussi et al., 2001)
Reaper	Pro-apoptotic protein	Fly (Dm)	476009 (1-174)	(Hernandez et al., 2004)
Runx2 Type I	Runt-related transcription factor 2-UTR2	Mouse	391766 (1-1018)	(Xiao et al., 2003)
Runx2 Type II	Runt-related transcription factor 2 – UTR1	Mouse	3901257 (1-207)	(Xiao et al., 2003)
Scamper	Calcium channel	Dog	21553346 (268-368)	(De Pietri Tonelli et al., 2003)
Smad5	Mediator of bone morphogenetic protein signalling	Human	4433529 (259-360)	(Shiroki et al., 2002)
SNM1	Homolog of yeast "sensitivity to nitrogen mustard" gene	Human	577302 (1-921)	(Zhang et al., 2002)
TIE2	Tyrosine kinase with immunoglobulin-like and EGF-like domains 1	Human	28411198 (28326-27857)	(Park et al., 2005)
TFIID	Transcriptional activator	Yeast	172898 (86-275)	(Iizuka et al., 1994)

Table 1. Reported Cellular IRES (continued)

Gene Name	Function	Organism	GI (IRES seq. pos.)	Ref.
TRKB	Neurotrophin receptor – Tropomyosin-rleated tyrosine kinase	Human	18369868 (3761-4040)	(Dobson et al., 2005)
Ubx	Ultrabithorax – homeotic gene	Fly	8794 (3518–4115)	(Hart and Bienz, 1996)
Unr	Upstream of N-*ras*	Human	52220548 (1-468)	(Cornelis et al., 2005)
Utr	Utrophin	Mouse	74144053 (195-704)	(Miura et al., 2005)
V1br	Vasopressin V1b receptor	Rat	945040 (35-544)	(Rabadan-Diehl et al., 2003)
VEGF-A *iresA*	Vascular endothelial growth factor - A	Human (mouse)	4154290 (2864-3403) (1134964 (1218-2234))	(Huez et al., 2001; Stein et al., 1998)
VEGF-A *iresB*	Vascular endothelial growth factor - A	Human	4154290 (2363-2863)	(Huez et al., 2001; Stein et al., 1998)
Vimentin	Structural protein	Human	2437834 (1758-1902)	(Qin and Sarnow, 2004)
XIAP	Apoptosis inhibitor	Human	28290426 (306-409)	(Holcik et al., 1999)
YAP1	Yes-associated Protein 1 transcriptional activator	Yeast	4797 (207–333)	(Zhou et al., 2001)

Table 2. Reported Viral IRES

Virus	Viral name - product	Viral host	GI (IRES seq. pos.)	Ref.
Dicistroviridae; *Cripavirus*				
CrPV	Cricket Paralysis Virus - ORF1 - nonstructural proteins	Insect	8895506 (1-711)	(Wilson et al., 2000a)
CrPV	ORF2 - structural proteins	Insect	8895506 (6025-6216) contains ccu start codon needed for pseudoknot	(Wilson et al., 2000a)
DCV	Drosophila C virus IRES1	Insect	2388672 (1-801)	(Johnson and Christian, 1998)
DCV	Drosophila C virus IRES2	Insect	2388672 (6080-6266) contains ccu start codon needed for pseudoknot	(Johnson and Christian, 1998; Kanamori and Nakashima, 2001)
PSIV	Plautia Stali Intestine virus - capsid protein	Insect	2344756 (5949-6195)	(Sasaki and Nakashima, 1999)
RhPV	Rhopalosiphum padi ORF2	Insect	2911298 (6327-7112) gca start codon	(Domier et al., 2000)
TRV	Triatoma virus - 5' UTR	Insect	6003484 (1-551)	(Czibener et al., 2005)
TRV-IGR	Triatoma virus -intergenic region	Insect	6003484 (5934-6111)	(Czibener et al., 2005)
TSV	Taura syndrome virus - capsid protein	Shrimp	`contains ccu start codon needed for pseudoknot	(Hatakeyama et al., 2004)
Flaviviridae; *Hepacivirus*				
HCV	Hepatitis C virus	Human	12831192 (1-344)	(Tsukiyama-Kohara et al., 1992)
Flaviviridae; *Pestivirus*				
BVDV	Bovine viral diarrhea virus	Cow	9836967 (1-385)	(Poole et al., 1995)
CSFV/ HoCV	Classical Swine Fever Virus / Hog Cholera Virus	Pig	12584212 (1-376)	(Rijnbrand et al., 1997)
Flaviviridae; Unclassified				
GBV-B	Hepatitis virus insolated B from patient GB	Primates	13162187 (23-448)	(Grace et al., 1999)
Herpesviridae; *Rhadinovirus* ds (dsDNA)				
KSHV	Karposi-sarcoma-associated herpesvirus v-flip	Human	2065526 (123206-122709)	(Bieleski and Talbot, 2001; Grundhoff and Ganem, 2001; Low et al., 2001)
Retroviridae				
F-MuLV	Friend murine leukemia virus glygag and gag polyprotein	Mouse	61544 (1-357) 61544 (1-621)	(Berlioz and Darlix, 1995)
F-MuLV	Friend Murine Leukemia virus - envelope gene	Mouse	61544 (5492-5780)	(Deffaud and Darlix, 2000a)

Table 2. Reported Viral IRES (continued)

Virus	Viral name - product	Viral host	GI (IRES seq. pos.)	Ref.
HaMSV	Harvey murine sarcoma virus -VL30	Rat	207672 (25-543)	(Berlioz et al., 1995)
HTLV-1	Human T-cell Lymphotropic virus 1 - R and partial U5 region	Human	221866 (354-621) no start codon	(Attal et al., 1996)
MoMuLV	Moloney Murine Leukemia virus	Mouse	331973 (912-1040)	(Vagner et al., 1995a)
RSV	Rous Sarcoma virus - gag	Chicken	2801459 (230-382)	(Deffaud and Darlix, 2000b)
RSV-src	Rous Sarcoma virus - src	Chicken	2801459 (7066-7131) no proof that full spliced UTR exists.	(Deffaud and Darlix, 2000b)
SIV	Simian Immunodeficiency Virus	Primate	334657 (507-1043)	(Ohlmann et al., 2000)
Picornaviridae; *Aphthovirus*				
FMDV	Foot and Mouth Disease Virus	Mammals	61076 (252-716)	(Kuhn et al., 1990)
Picornaviridae; *Cardiovirus*				
EMCV	Encephalomyocarditis virus	Human	9626692 (260-836)	(Jang et al., 1988)
TMEV	Theiler's murine encephalomyelitis virus	Mouse	62039 (31-1070)	(Pilipenko et al., 1994)
Picornaviridae; *Enterovirus*				
CVB3	Coxsackievirus B3	Human	54399970 (1-745)	(Yang et al., 1997)
EV71	Enterovirus 71 strain BrCr	Human	1171120 (1-746)	(Thompson and Sarnow, 2003)
PV	Poliovirus	Human	61127 (34-750)	(Pelletier and Sonenberg, 1988)
Picornaviridae; *Hepatovirus*				
HAV	Hepatitis A virus	Human	329585 (150-720)	(Brown et al., 1994)
Picornaviridae; *Rhinovirus*				
HRV	Human Rhinovirus	Human	221708 (20-625)	(Borman and Jackson, 1992)
Picornaviridae; *Teschovirus*				
PTV-1	Porcine teschovirus serotype 1 strain Talfan	Pig	13111645 (146-434)	(Pisarev et al., 2004)
Lentivirus; Primate lentivirus group				
HIV	Human immunodeficiency virus type 1 gag	Human	4558520 (335–808) note: resides in CDS	(Buck et al., 2001)
Luteoviridae; *Polerovirus*				
PLRV	Potato Leafroll virus	Plant (potato)	222301 (1513-1728) includes 211 bases of CDS	(Jaag et al., 2003)
Potyviridae; *Potyvirus*				
CPMV	Cowpea mosaic virus	Plant	58910 (161-514)	(Thomas et al., 1991)
PVY	Potato virus Y	Plant	61450 (1-187)	(Levis and Astier-Manifacier, 1993)
TEV	Tobacco etch virus	Plant	335201 (2-147)	(Carrington and Freed, 1990)

Table 2. Reported Viral IRES (continued)

Virus	Viral name - product	Viral host	GI (IRES seq. pos.)	Ref.
Poxviridae; *Avipoxvirus* (dsDNA)				
REV	Avian reticuloendotheliosis virus type A	Bird	28927668 (400-974)	(Lopez-Lastra et al., 1997)
Tobamovirus				
crTMV	Tobacco mosaic virus (Crucifer) - cpgene	Plant	488713 (5456-5606)	(Ivanov et al., 1997)
Totiviridae; *Giardiavirus* (dsRNA virus)				
GLV	Glardia lamblia virus	*Glardia lamblia*	1352866 (1-369)	(Garlapati and Wang, 2005)

Note for Table 1 and 2: sequences are either the minimal sequence of the fully functioning IRES or the beginning of the known 5' UTR. The 3'end of each sequence includes position +3, the start codon. Some errors in the published positions have been corrected upon communications with the authors where possible.

Table 3. Reported minimal IRES modules

Gene/virus	Full name and product	host	Sequence	Ref.
ARC-1	Active rRNA Complementary sequence to rice 18S rRNA	Plant	AUACUCCCCC	(Akbergenov et al., 2004)
c-Myc (minimal IRES)	Transcription factor	Human	GGGGACTTTGCACTGGAACTTAC AACACCCGAGCAAGGACGCGACTCT	(Cencig et al., 2004)
Gtx	Homeobox	Mouse	CCGGCGGGT	(Chappell et al., 2000)
Rbm3	glycine-rich RNA-binding protein	Mouse	UUUAUAAUUUCUUCUUCCAGAA	(Chappell and Mauro, 2003; Chappell et al., 2001)
synthetic	Short synthetic nucleotides transcribed within an IRES reporter construct	Human/ yeast	5 positive 50mers/ 56 positive 18mers	(Venkatesan and Dasgupta, 2001; Zhou et al., 2003)
HSV	Herpes Simplex Virus thymidine kinase mutant	Human	CCGUGCUGGCGU CUG is start codon	(Griffiths and Coen, 2005)

only the entries that have been commented in their feature tables as having fully reviewed 5' UTRs. The remaining entries in RefSeq would have been generated without a curator and may not represent full-length UTRs.

3.1.1 UTR length

It is believed that transcripts with shorter 5' UTRs are translated more efficiently and that the bulk of mRNAs have relatively short 5' UTRs. It is also assumed that UTRs containing an IRES would tend to be longer as it would be a method of translation initiation required for transcripts which would not be translated as efficiently by "cap-dependent" translation initiation. To see if this is true, the overall length distribution of human 5' UTRs represented by the RefSeq, UTRdb and Ensembl datasets were compared to the lengths of mammalian 5' UTRs with IRESes (Figure 1). Although the number of entries in the IRES dataset is much smaller, the lengths of the characterized UTRs with IRES are overall much longer. It should be noted that the median of all UTR lengths is somewhere between 156-172 bases in length which would mean that 50% of all genes have UTRs that length or longer and could contain a minimal IRES sequence. The 3rd quartile is over 280 bases which means 25% of all genes have UTRs which are quite large. The reviewed RefSeq is very similar in size as the complete RefSeq but somewhat shorter than the UTRdb, which has garnered many more entries than the RefSeq dataset.

3.1.2 UTR degree of RNA secondary/tertiary structure

It is common for introductions of studies on IRES to say that they exist in UTRs with a high degree of structure. It is not clear how this has ever been proven but one might assume that UTRs with a level of structure where a 43S ribosomal complex is having difficulty

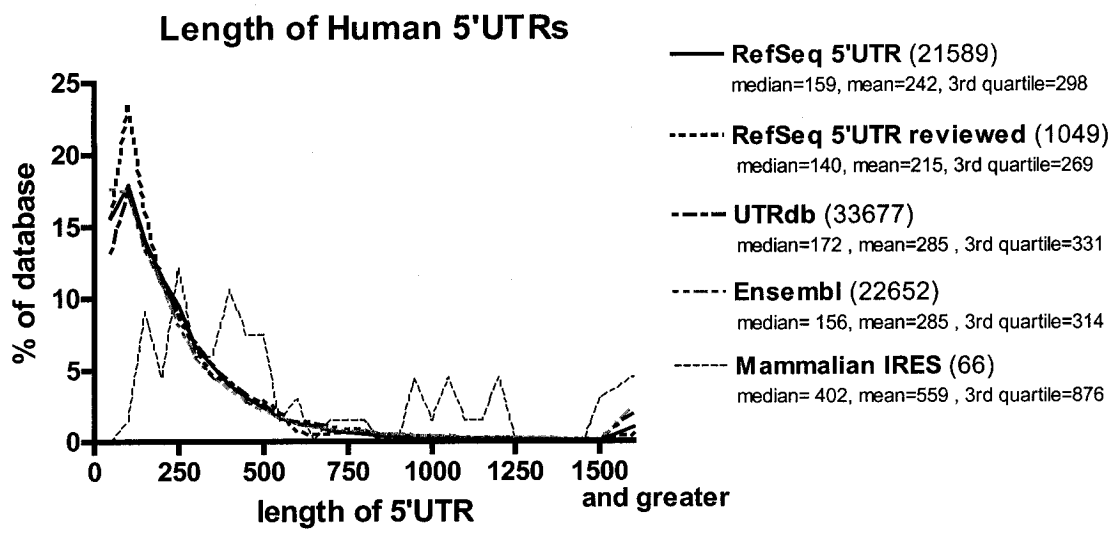


Figure 1. The lengths of 5' UTRs of human transcripts from non-redundant datasets. The frequency of different lengths of all the UTRs in each dataset is placed in increasing bins of 50 nucleotides and plotted as a percentage of the total number of UTRs in each database. A non-redundant dataset of 5' UTRs from UTRdb, RefSeq, fully reviewed UTRs of RefSeq by a curator and the mammalian transcripts containing IRES in their 5' UTR from Table 1, Appendix I are compared. The legend gives the median, mean and 3rd quartile values in each dataset with the total number of UTRs in each dataset in brackets.

scanning from the cap site to the start codon might benefit from initiating a scan from an IRES which would be further downstream and close to the AUG start codon. A simple measure of structure stability of a UTR can be the level of GC content. Whereas all combinations of basepairings are possible in RNA structures, the GC basepair has three hydrogen bonds instead of the one or two found in other basepairs. In Figure 2, the proportion of GC content in UTRs is compared as a percent of each of the datasets. The Ensembl dataset, which is not shown, gave similar results to RefSeq. Although it is possible that there is a slight tendency for IRES to have a higher GC content, this may only appear so because of the small dataset giving a spike in the results. It is evident that there are IRES with GC content over the complete range seen within all UTRs of the human genome. The % GC content of the IRES dataset seems quite similar to the overall GC distribution of UTRs (Figure 2A) suggesting that the degree of RNA secondary structure stability as correlated to the amount of GC pairing possible is no different than for most UTRs. A similar result is obtained if the secondary structure-scanning algorithm of Rivas and Eddy (2000) is used instead of %GC content (Figure 2B). This does not negate the possiblity that there could be secondary structure elements within a 5' UTR that impede translation but the overall degree of structure in UTRs with IRES does not appear different than is found for all human 5' UTRs.

3.1.3 Upstream AUG start codons

Another possible reason for a transcript to be inefficiently translated would be if there were AUG start codons upstream of the "normal" start codon. Although leaky scanning of the ribosome through upstream AUGs can still efficiently initiate at the main start site (Wang and Rothnagel, 2004), the ribosome can scan to these sites, initiate assembly into the

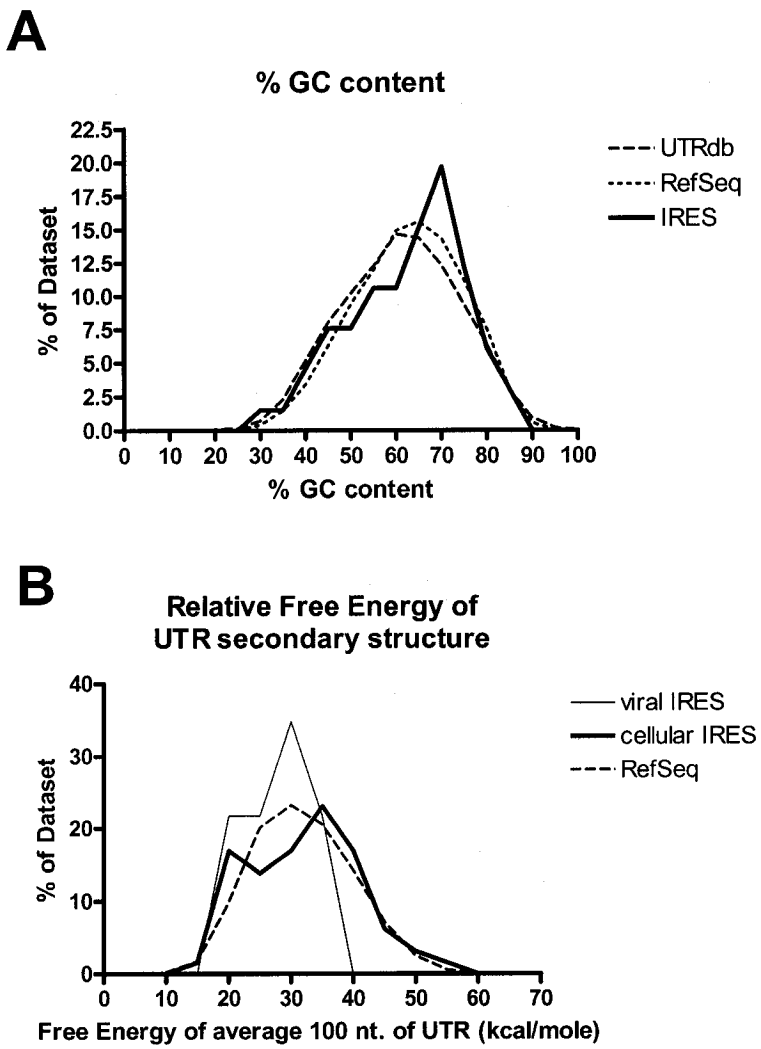


Figure 2. The distribution of the % GC content of human 5' UTRs. (A)The degree of RNA folding and structure stability can be partially assessed by the percentage of possible G and C basepairing within the UTR sequence and therefore the % of Gs and Cs within the sequence. The % GC content of mammalian UTRs with published IRES are compared to the human 5' UTRs from UTRdb and RefSeq and show a similar distribution. Plotted values are grouped in bins of 5. (B) Using a sliding window secondary structure prediction algorithm of Rivas and Eddy (2000) based on thermodynamics, the relative free energy of the predicted structure is compared for IRES UTRs and the human RefSeq 5' UTRs >100 nucleotides in length. A window of 100 bases is used, sliding every 5 bases. The average energy values for all the windows in each UTR is calculated and used as a measure of secondary structure.

80S ribosome to start translation of a small upstream open reading frame (uORF) and then fall off at the next stop codon. For example, two uORFs are responsible for translational regulation of ATF4 during basal and stress conditions. Both uORFs are translated in normal conditions and the ribosome falls off the transcript after the second uORF. During stress, the lack of eIF2-GTP due to phosphorylation of eIF2α causes the ribosome to take longer to become competent to reinitiate making it skip the second uORF and re-initiate at the ATF4 coding sequence (Vattem and Wek, 2004), similar to what is seen in yeast with GCN4 (Hinnebusch, 1997). An IRES close to the 3' end of the UTR could circumvent the false starts and provide a way to translate the proper coding sequence. To see if there is a difference between UTRs with IRES and the overall 5' UTRs in the human genome, the datasets were scanned for upstream AUGs. The data are represented in Figure 3 and show a subset of transcripts with no upstream AUGs in all the datasets in Figure 3a and therefore upstream AUGs are not universally present in 5' UTRs containing an IRES. Overall the trend in the number of upstream AUGs found in UTRs containing an IRES is very similar to the complete dataset of all 5' UTRs. Perhaps the upstream AUGs confer no regulatory importance if the surrounding sequence is not favorable and the AUG is not recognized by a scanning ribosome as a strong start codon (Kozak, 2005). Around 10% of all the upstream AUGs would be considered strong start codons (RNNAUGG, where R is a purine) as defined by Kozak (Kozak, 2005) whereas 41% to 45 % of the coding region start codons would be considered strong start codons. Therefore, there is an importance to the sequence surrounding AUGs as strong start codons are less likely to be found upstream of the accepted start codon. There are a slightly lower percentage of strong upstream AUGs (6.3 %) in UTRs containing an IRES whereas one might expect there to be more if the IRES mechanism was used to overcome a 5' UTR that made the transcript less efficiently

translated but the difference may also be due to the smaller sample size of UTRs containing an IRES.

If the number of upstream AUGs was occurring randomly at a normal rate with no significant selective pressure, then longer UTRs would have more, and shorter ones less, with a normal distribution centered on 1 in ever triplet of bases for every 64 bases of the sequence. In Figure 3b, the overall frequency in all datasets is less than expected from random sequence. Therefore, there is some selection pressure on having a lower number of upstream AUGs but there is still a subset of approximately 10% of the UTRs with more AUGs than expected randomly. There is a slightly higher ratio of upstream AUGs in UTRs containing an IRES than those without, but the trend of the frequency is very similar to the complete 5' UTR database suggesting that the frequency of upstream AUGs is not that significant for UTRs with an IRES.

3.1.4 IRES and tRNA patterns

UTRdb has annotated the sequences that contain patterns that match possible RNA consensus structure/sequences of known regulatory elements such as the Iron Response Element (IRE), the Histone 3'UTR stem-loop structure (HSL3) or even IRES elements as adapted from the computationally predicted IRES structure of Le and Maizel (Le and Maizel, 1997). In Release 20 of UTRdb, there are ~34,000 human 5' UTR sequences, and the IRES pattern is found ~7,000 times, which is about 20% of all the entries. Although the number of mRNA containing IRES elements could be quite abundant, there is doubt that the pattern actually detects IRES. Using the same pattern to search the cellular IRES from Table 1, the pattern search program, RNAMOTIF, finds the pattern in about 21% of the

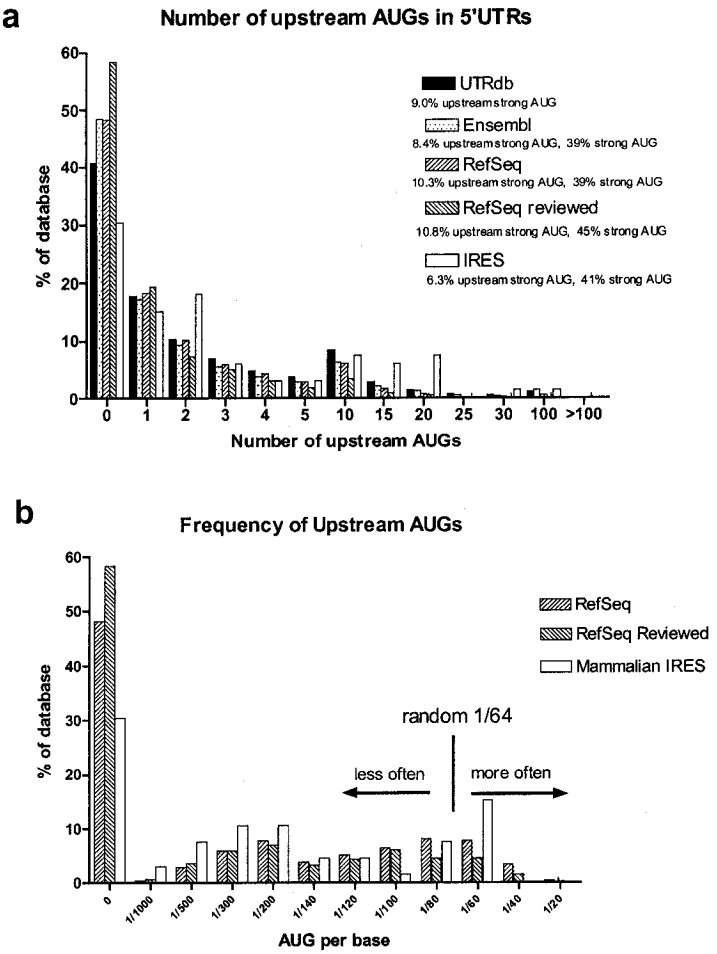


Figure 3. The number and frequency of upstream AUGs. (a) The number of AUGs found upstream of the natural start codon in 5' UTRs is compared in several non-redundant 5' UTR datasets. Data comes from human transcripts from RefSeq (21589, Release 13), RefSeq 5' UTRs fully reviewed by NCBI staff (1049) and the 66 mammalian UTRs containing published IRES from Table 1. Data bins are represented as % of the total entries in each dataset. Kozak describes "strong" AUGs, which would more efficiently initiate translation, as having the pattern RNNAUGG (Kozak 2005). (b) The frequency of upstream AUGs in several non-redundant dataset. A histogram in several asynchronous bins containing the relative frequency of upstream AUGs as a percentage of each dataset compares the 5' UTR from human transcripts in RefSeq, and fully reviewed RefSeq UTRs.

entries showing that this pattern is no more common in UTRs containing known IRES as in all UTRs.

As RNAMOTIF has been used successfully to find tRNA sequences (Tsui et al., 2003), I examined if this program can pick out those tRNA-like regions believed to be in IRESes as exhibited by viral IRESes from CrPV, HCV, EMCV, and BVDV. Using the original descriptor of Tsui et al. (Tsui et al., 2003) for eukaryotic tRNA but with no introns, RNAMOTIF finds only 1 putative tRNA structure in the 78 published eukaryotic sequences and only 23 sequences in the 33677 sequences in UTRdb. By relaxing the descriptor so that only 2 of the 8 specific nucleotides of the script are required, over 60 % of the IRES data shows a pattern match to this "loose" descriptor, as seen in Figure 4. As many of these matches gave structures with positive free energy values, which would mean they are not thermodynamically stable, a third tRNA script was made with cutoff matches over -10 kcal/mol. This value is calculated using the energy tables of Mathews et al. (Mathews et al., 1999a) with the efn2 program included in the RNAMOTIF package. It must be noted that the loose tRNA descriptor did not find a match in CrPV, which is known to interact with the ribosome and initiate translation without need of an initiator Met-tRNA$_i$ (Wilson et al., 2000b). Although there is no proof that cellular IRES can initiate translation without an initiator Met-tRNA$_i$ this tRNA-like pattern is notably found in more UTRs containing an IRES than in the complete set of 5' UTRs suggesting some possible shared structure between a subset of IRESes.

3.2 Evaluation of RNA software

In examining the makeup of UTRs containing known IRES, several current software programs were tested to see if they could aid in the detection of new IRES. These software

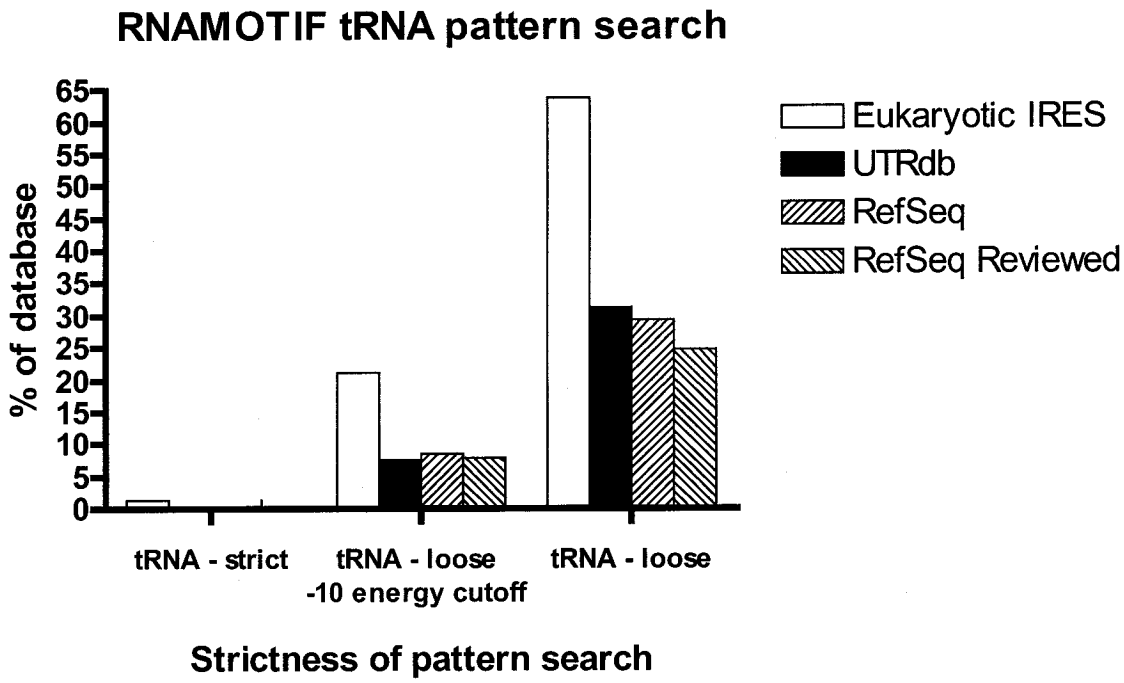


Figure 4. RNAMOTIF pattern matches to tRNA descriptors of different levels of stringency within 5' UTR human sequences. Using the eukaryote tRNA descriptor of Tsui et al. 2003 without introns, the number of pattern matches is shown as a percentage of the total number of UTRs in either UTRdb, RefSeq, fully reviewed UTRs in RefSeq, or published IRES from eukaryote transcripts in Table 1 (see datasets in Figure 1). The first pattern is strict due to the specific sequences required in the pattern matches as well as the overall tRNA structure. The second pattern only requires 2 of the 8 specific base sequences, which are standard in tRNAs as well as a maximal -10.0 kCal/mol free energy cut-off for the putative matching RNA

programs were not given a rigorous comparison with well-used test datasets but more of a "trial by fire" test using known IRES structures and sequences available at the time. This will still yield some information that can be incorporated into methodology of RNA structure prediction as well as new information on IRES sequence/structure characteristics. Sequence data for the IRESes used in the testing came from the database entries given in Table 1 for cellular IRESes and Table 2 for viral IRESes. Structure files, in the connect (CT) format, were generated for the following experimentally derived IRES secondary structures: HCV (Baird et al., 2006), CSFV (Brown et al., 1992; Fletcher and Jackson, 2002), Apaf-1 (Mitchell et al., 2003), FGF-2 (Bonnal et al., 2003b), Bag-1 (Pickering et al., 2004), and Mnt (Mitchell et al., 2005). The connect files were created using MFOLD by adding the constraints from the published structure diagrams until all the basepairings matched the published structures. Pseudoknots for HCV and CSFV were added manually by editing the CT files directly.

3.2.1 ED_SCAN

In Figure 2, there did not seem to be an overall difference in the %GC content of UTRs with an IRES and those without. Therefore, UTRs with an IRES possibly have no difference in overall structure stability from UTRs that do not have IRES. This does not preclude local areas of stability in a UTR that might reduce the efficiency of translation and, therefore, require an alternate mechanism of translation initiation, such as an IRES. An example of local structures that inhibit translation can be seen in synthetic gene constructs that contain a stable hairpin, which can hinder the scanning of a ribosome. These hairpins, although GC rich would not necessarily change the overall stability of the UTR or % GC content significantly. Small local structure within a UTR could also possibly be binding sites

for proteins like ITAFs, which may have an effect on translation. The Iron Response Element, although not an ITAF binding site, is an example of a small hairpin that the IRE-binding protein binds to and affects translation of ferritin mRNAs. For these reasons, ED_SCAN (Le et al., 2003) was tested as it looks for local energy differences within a sequence from a sliding window. The energy difference (E_{diff}) is represented by the lowest energy for a possible fold within the window subtracted from the lowest energy possible for a fold that has been constrained to prohibit any basepairing that has been used in the lowest energy fold. The greater the E_{diff} of a window, the greater the order of the structure. In order to compare the E_{diff} across the sequence, a Z score (Zscr) is used that incorporates the E_{diff} mean of all the windows analyzed and the standard deviation (Le et al., 2003).

As a control, ED_SCAN was first tested on both the ferritin gene looking for the iron response elements, and on the let-7 microRNA genes in genomic nematode sequences which duplicate the results from the original paper. The correct results were produced as shown in Figure 5a which matches the output shown by Le et al. (Le et al., 2003). ED_SCAN was then used to predict small structures within the UTR of the known IRES structures of HCV, Apaf-1, CSFV or c-myc. As an example, the results from HCV are shown in Figure 5b where the highest peak did not map to any particular structure. The peaks within the UTR segments were also not significantly higher than would be expected with random sequence, with Zscr <3. Therefore in these four IRES structures a notable local structure could not be detected by ED_SCAN.

3.2.2 MFOLD

Since MFOLD has been often used by researchers for *de novo* prediction of IRES

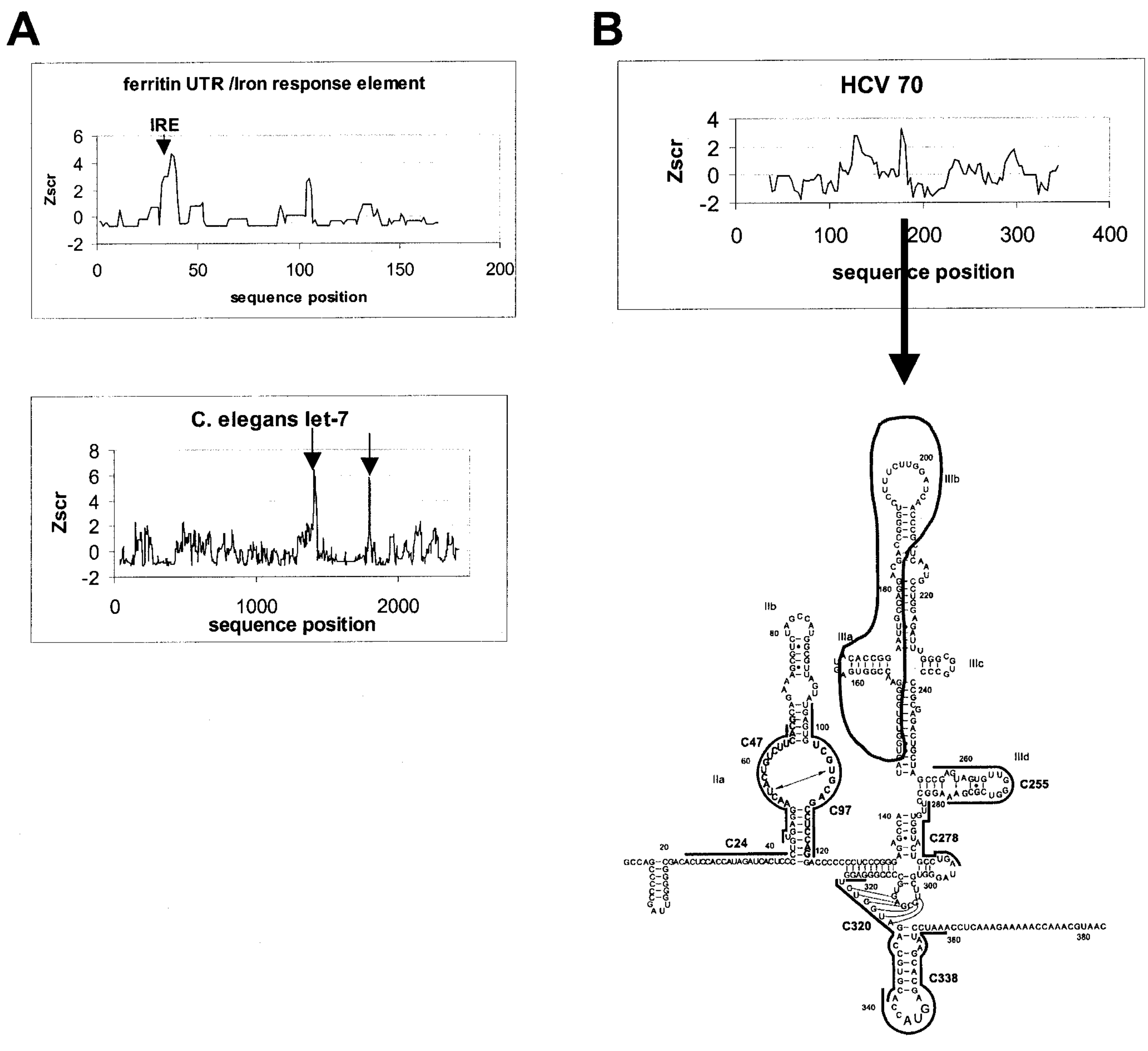


Figure 5. ED_SCAN detects local regions of structure in (A) The iron response element in the ferritin gene UTR and the let-7 microRNA in genomic C. elegans sequence are found as high peaks denoted by down arrows. The Zscr is plotted in the middle of a window of 70 bases as it slides across the sequence. (B) The Zscr plot of HCV UTR containing the structure shown below from a sliding window of 70 bases. The highest peak corresponds to the circled region in the HCV structure.

structures (Negulescu et al., 1998; Oumard et al., 2000; Schiavi et al., 1999; Sehgal et al., 2000; Sherrill et al., 2004; Van Eden et al., 2004; Xiao et al., 2003) MFOLD was used to test its accuracy with the most accurately defined IRES structure which is the IRES from HCV as well as the empirically derived APAF1 (Mitchell et al., 2003) IRES secondary structure. The secondary structure model used for the HCV IRES is based upon the model presented by Puglisi et. al. (Puglisi et al., 2001) incorporating updated basepairing data in domain II (Lukavsky et al., 2003) and IIId (Klinck et al., 2000) from NMR results as well as crystallographic studies of domain III (Kieft et al., 2002). Tertiary structure data can be converted to secondary structure using RNAVIEW (Yang et al., 2003).

In Figure 6 and 7, the empirically determined structures of the 5' UTR of HCV and APAF-1 are presented with the basepairs and single stranded residues correctly predicted by MFOLD denoted with grey circles. The energy states of the predicted structures were calculated by MFOLD using efn2. The RNA structure version of the Matthews Correlation Coefficient (MCC), as presented by Stormo's group (Gorodkin et al., 2001; Ji et al., 2004), was used as a comparative metric between the actual and predicted sequences as it takes into account both the sensitivity and accuracy of the prediction, by including both true and false positive predicted basepairing, as well as true and false negative basepairings. Of a possible MCC from 0 to 1, with 1 being a perfect match, the best MFOLD prediction for HCV is 0.59 (Figure 6A) and the prediction with the lowest energy has an MCC of 0.39 (Figure 6B). Although these predictions are not that accurate, it should be noted that the structure predicted with the lowest free energy was not the best. All of the base pairs and single stranded regions that were correctly predicted in the 36 structures produced by MFOLD, with the suboptimality parameter set at 50%, are shown in Figure 6C. No more structures were produced at 50% suboptimality then at 10%. This is a cursory examination of the

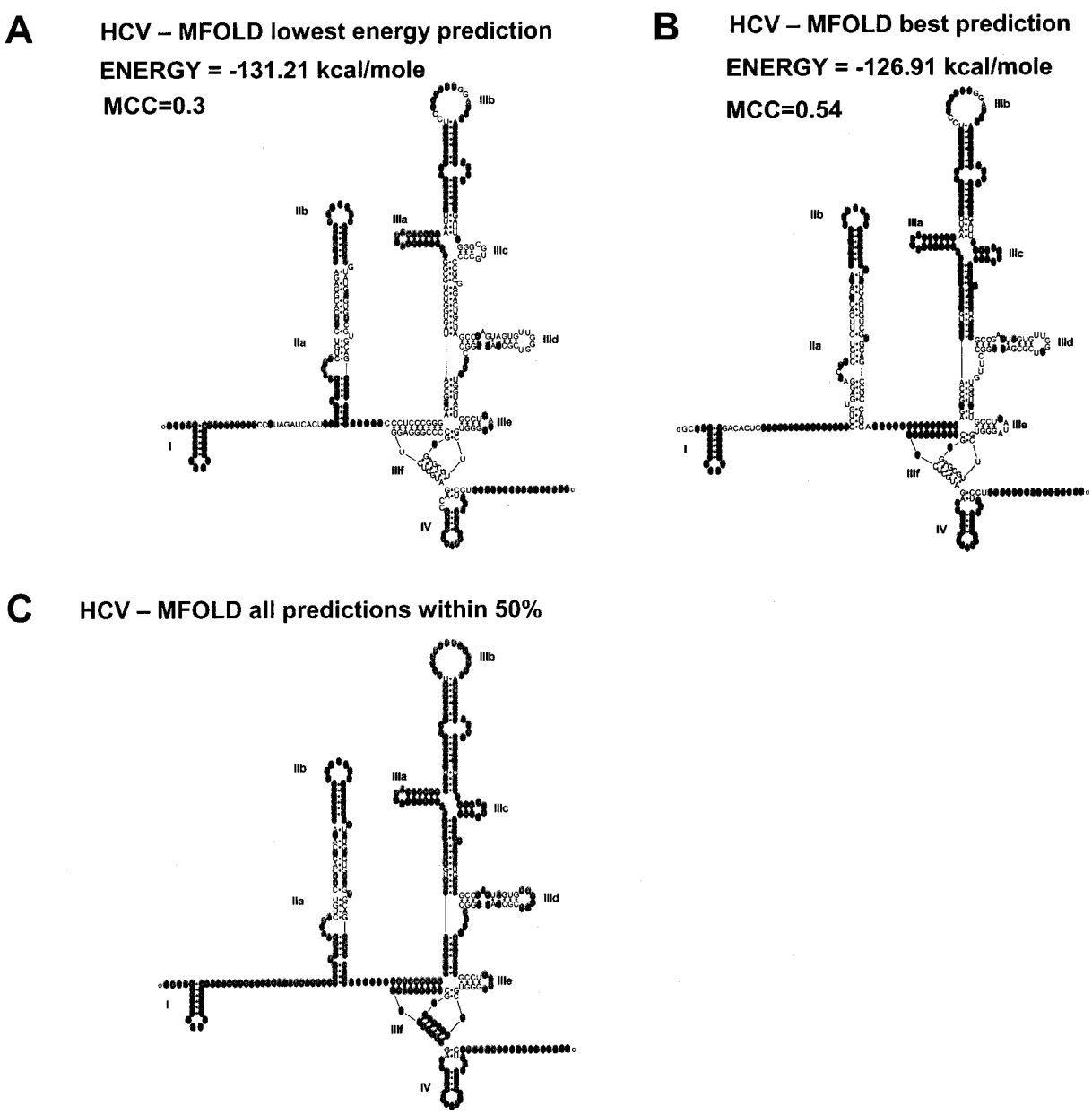


Figure 6. The correctly predicted RNA secondary structure of HCV IRES by MFOLD. (a) The empirically predicted structure of the HCV IRES is shown with correctly predicted basepairs and non-pairing bases with grey background of the lowest energy prediction when using MFOLD with no constraints. (b) The best predicted structure was not the most thermodynamically stable fold and the predicted basepairs and non-pairing bases shown are in grey. (c) All of the correctly predicted basepairs and non-pairing bases from all of the 36 predicted sub-optimal folds using a 50% sub-optimal parameter in MFOLD. Structures for figures have been produced with RnaViz2 (De Rijk et al., 2003).

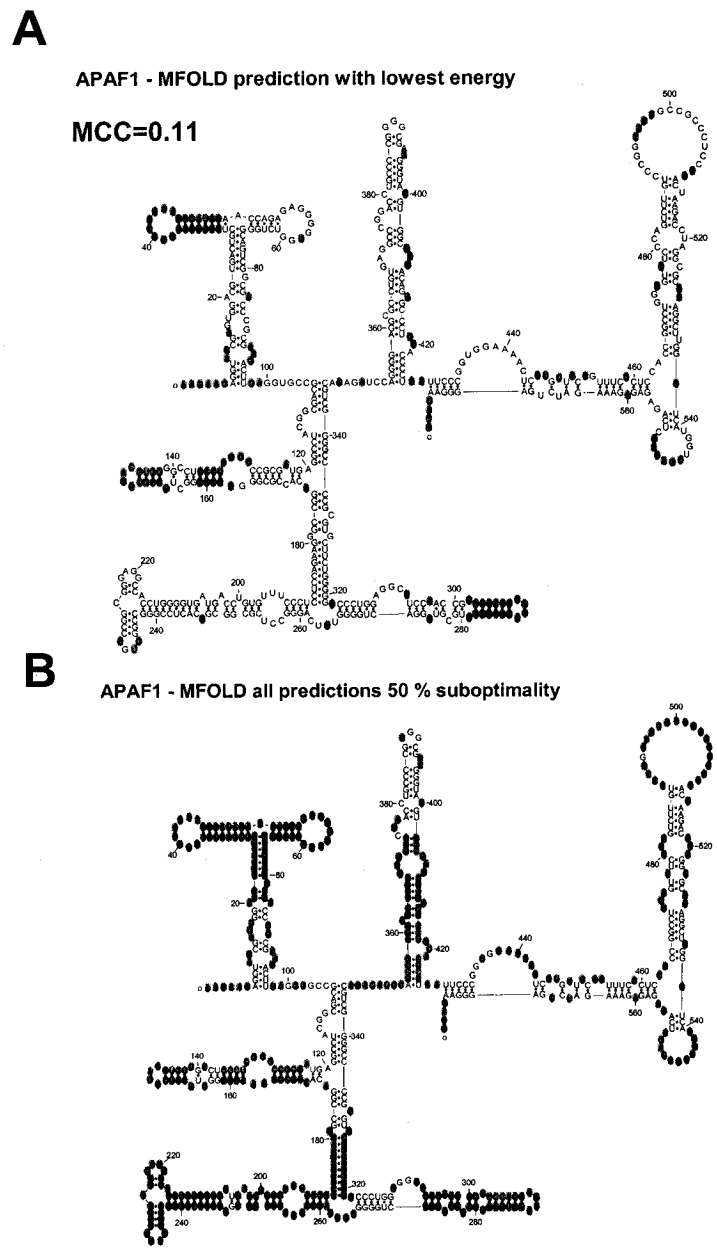


Figure 7. The correctly predicted RNA secondary structure of APAF1 IRES by MFOLD. (A) The empirically predicted IRES structure of Apaf-1 (Mitchell et al., 2003) is shown with the MFOLD correctly predicted basepairs or non-pairing bases shadowed in grey for the most thermodynamically stable structure. (B) All of the non-pairing bases or basepairs that were correctly predicted from all of the sub-optimal folds predicted using a 50% suboptimal parameter in MFOLD are presented as grey shadows behind each base. This stresses the need for some empirically derived constraints for large RNA fold predictions when using MFOLD.

folding space considered by this program and is overall quite good, with some exceptions. The NMR results of domain II (Lukavsky et al., 2003) and of domain IIId (Klinck et al., 2000) indicate several non-canonical basepairings suggesting that these regions can not be expected to be properly predicted by any secondary structure prediction program. In Figure 6 these basepairs are aligned together but are not connected as a basepair. Several basepairs in domain IIIe, IIId and IIa are not predicted in the set of low free energy secondary structures predicted by MFOLD. The structure predictions of the larger UTR of Apaf-1 in Figure 7 fairs less well compared to the experimentally predicted structure than was seen with the HCV IRES structure prediction. This experimental structure also has not been as rigorously examined in as many studies as the HCV structure. Much of the experimental structure is not predicted correctly with the 50 % suboptimality parameter as can be seen in Figure 7B. In this case, either the empirically derived model of the Apaf-1 IRES is poor or the program has failed in accurately predicting the structure.

3.2.3 DYNALIGN

DYNALIGN predicts a common global structure based on similar folding energies between two sequences. The example comparison in Figure 8, uses the CSFV and HCV IRESes which are known to have similar structures (Brown et al., 1992; Lemon and Honda, 1997) even though the sequence identity is low. Using Blast2seq (Tatusova and Madden, 1999) which uses blast to do a pairwise comparison of two sequences, there is no significant match between the HCV and CSFV sequence even with expectation values set at 100. The HCV IRES structure is predicted quite accurately with an MCC of 0.75. One could assume if the HCV structure was correctly predicted then the prediction of CSFV would also be accurate. As the structure of CSFV (Brown et al., 1992; Fletcher and Jackson, 2002) is

DYNALIGN

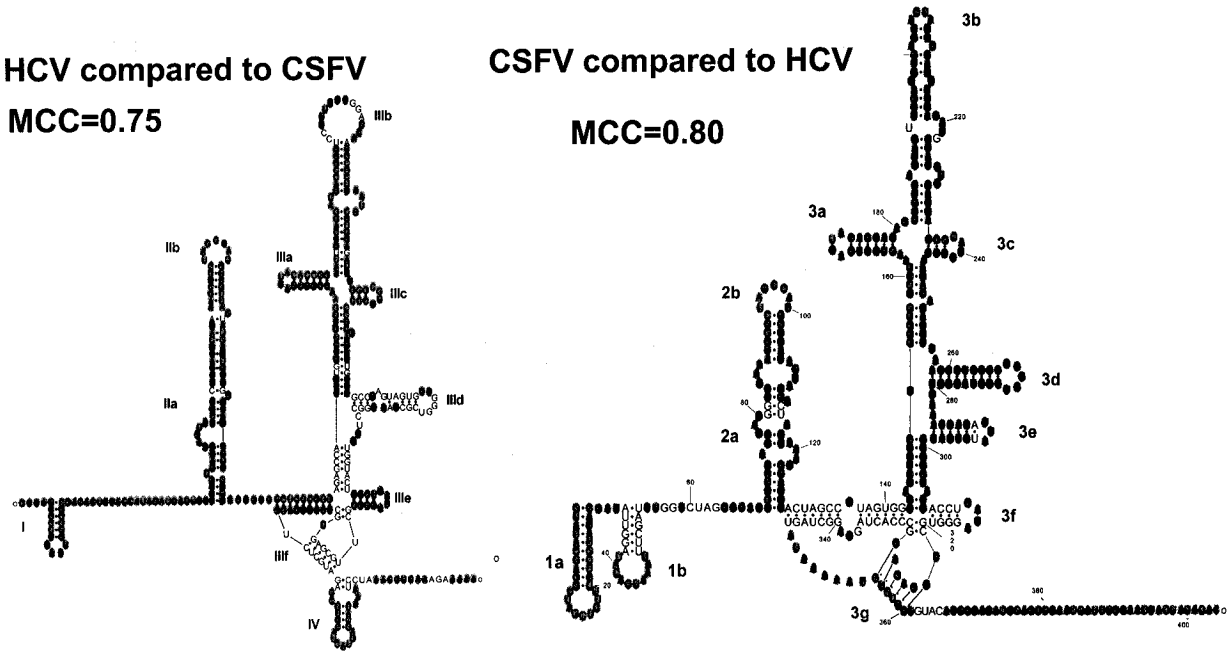


Figure 8. Simultaneous prediction of HCV and CSFV IRES secondary structure with DYNALIGN. Two phylogenetically related sequences like HCV and CSFV are well predicted by DYNALIGN. The correctly predicted basepairs and non-pairing bases are shown with grey background. The Matthews Correlation Coefficient (MCC) (Gorodkin et al., 2001) is used to measure the accuracy of the prediction when compared to the empirically derived structures.

known, the accuracy of the prediction can be calculated and it also has a high MCC value of 0.8. Although the minimization of free energy for de novo structure prediction is not always accurate, the added constraint of using a second sequence, which shares a similar overall structure, but not necessarily primary sequence, greatly improves secondary structure predictions.

3.2.4 CARNAC

CARNAC (Perriquet et al., 2003) is able to predict partially correct structures with only two sequences but it would need more to produce a complete structure. Using the structurally similar HCV, CSFV and BVDV IRESes sequences (Brown et al., 1992; Lemon and Honda, 1997) as an example dataset, CARNAC predicted folds of HCV and CSFV RNA were compared to the known structures. The MCC of the CARNAC predicted structure for the HCV IRES was 0.282 when compared to the known HCV structure (as seen in Figure 6) and 0.587 for the predicted structure of CSFV. As the accuracy of these predictions was so low, CARNAC was not considered for further use in this project.

3.2.5 Pfold

To examine the accuracy of Pfold (Knudsen and Hein, 2003), several of the predicted IRES structures of FGF2, MNT and BAG-1 in the Rfam (Griffiths-Jones et al., 2003) database which were predicted by Pfold were compared to their empirically derived structures. The Pfold generated structures were predicted using available orthologs. If we compare the published structures of FGF-2 (Bonnal et al., 2003b), Bag-1 (Pickering et al., 2004) and Mnt (Mitchell et al., 2005) which have been determined by enzymatic/chemical probing to the Pfold generated structures, we find a poor correlation in structure models with an MCC of 0.11 for FGF2, 0.25 for MNT and 0.00 for BAG-1. Assuming the

experimentally derived structures are somewhat correct, Pfold has performed very poorly and will not be considered for further use in this project. These results are also a cautionary note to anyone using the Rfam structures that have been derived from Pfold.

3.2.6 ComRNA

The ComRNA (Ji et al., 2004) program takes unaligned sequences as an input. With this program, pseudoknots are also considered, which can be important for IRES structures. This is a command line program written in C++, available as source code and the output is single lines of sequence with a single line stem pattern description underneath. Ji *et al.* (Ji et al., 2004) feel the upper limit of allowable input would be 20 sequences of less than 300 nucleotides each. Using only the HCV, CSFV and BVDV UTRs as input sequences, which are from 372 to 403 bases in length, the computations were too great to produce useful results after 24 hours of processing time on a Sun Fire V20z computer with an AMD Opteron 248 (2.2 GHz) processor and 8 Gigabytes of memory. As many UTRs have much longer sequences, this program is too computationally intense to be of further use.

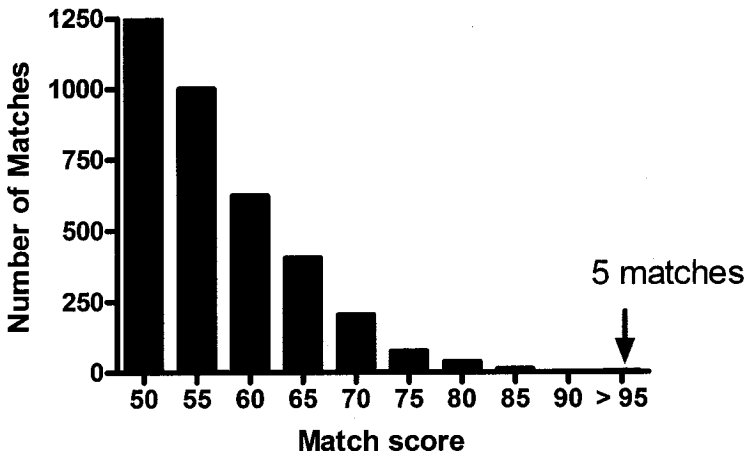
3.2.7 RSEARCH

RSEARCH (Klein and Eddy, 2003) searches primary sequence data with a secondary structure looking for both structural and sequence similarity. The RSEARCH program was tested in a search for similar UTRs, using the well-studied HCV IRES structure that incorporated the latest NMR and crystallography data (Baird et al., 2006). A database was compiled of human 5' UTR sequences from RefSeq (Pruitt et al., 2003) NCBI build 35 and the IRES data from Tables 1 and 2. Exact duplicates of database entries were removed from the dataset. RSEARCH outputs a sequence and structural alignment of the query structure and the RSEARCH-predicted structure of the database entry for each match that is above a

user-set cutoff score. A distribution of the RSEARCH match scores using the HCV IRES structure as the test query, is shown in Figure 9. The top hits were the unclassified Flaviridae Hepatitis GB isolate B, GbvB (Grace et al., 1999), CSFV, and the BVDV 5' UTRs, all known to have similar structures to the HCV IRES. Neither the CSFV nor BVDV 5' UTRs have enough similar primary sequence to the HCV IRES to be found using the pairwise BLAST program, Blast2Seq, as mentioned in the DYNALIGN results section, and were therefore identified due to structural similarity. GbvB, a hepatitis B isolate, has approximately 80 % nucleotide identity over 50 basepairs that matches the HCV pseudoknot.

The top matching human 5' UTR from "pleckstrin homology domain containing, family G (with RhoGef domain) member 4", PLEKHG4, has a region with high structural similarity to domain III of HCV. The structural alignment output of RSEARCH is shown in Figure 10. Large non-matching gaps in the alignment are signified as square brackets within the sequence lines. There are 6 gaps in this alignment, which map to each of the loops (IIIa, IIIb, IIIc IIId, IIIe and the pseudoknot of IIIf) on domain III of HCV (see Figure 6 for HCV structure). Most notable is the large 225 nucleotide gap in stemloop IIIe. This loop is not likely to occur in order to preserve a structure similar to the HCV IRES and therefore this high match score is suspect as being an artifact.

A **Distribution of RSEARCH scores for HCV IRES structure search of human 5'UTR and IRES database**



B

Score	RNA source
701.52	HCV
133.29	GbvB
119.64	BVDV
97.32	CSFV
95.70	pleckstrin homology domain containing, family G (with RhoGef domain) member 4 (PLEKHG4)

Figure 9. The distribution of RSEARCH match scores from a search of a human 5' UTR and IRES sequence database with the HCV IRES structure. (A) A histogram of the number of matches to the HCV IRES structure for a given RSEARCH match score plotted in bins of 5 above the score of 50. The number of matches includes multiple matches to the same UTR. There were no matches with a score of 90 to 95. (B) The match score and identity of the matching 5' UTR sequence.

```
>NM_015432

  Plus strand results:

 Query = 125 - 323, Target = 1759 - 2313
 Score = 95.70

     {{{{{.{{{{{[-[[[[(((((((((((((,,<<<<~~~~~~>>>><<<<-<<<<<--<<
 125 CCUCC.CGGGAGAGCCAUAGUGGUCUGCGGAACCGG*[ 8]*CCGGAAUUGCCAGGACGA 185
     +C+CC CGGG GA+ C++ G+GGUCUGCGG  CC+G      C+GG++ UG+ +G+  G+
   1 ACGCCgCGGGUGAGGCGC-GCGGUCUGCGGGGCCCG*[14]*CGGGGC-UGGCGGGCUGU 66

     -<<<<<__~~~~~~>>>>>->>----->>>>>->>>>,<<<~~~~~>>>))).)-)))).)
 186 CCGGGUCC*[13]*ACCCGCUCAAUGCCUGGAGAUUUGGG*[5]*CCCCCG.CGAGAC.U 248
     CCGGG+CC       +CCCGC+C  UG+C+ + G ++ GG+      +CCCCG CGAGAC U
  67 CCGGGCCC*[36]*GCCCGCACCUUGUCUUCCGCGCCGGC*[8]*GCCCCGuCGAGACcU 159

     ))).)<<<~~~~~~>>>,,,,]]]]-]]<<<~~~~~~~>>><<~~~~~~>>}}.}}}}}}
 249 GCU.AGCC*[22]*GGCCUUGUGGUACUGCC*[  6]*GGUGC*[ 9]*GCCC.CGGGAG 322
     +CU +G+C       G+CC  G+G +AC GCC        GGUGC       GCCC CGGG+G
 160 GCUcGGGC*[48]*GCCCCGGCGGCACUGCC*[225]*GGUGC*[81]*GCCCaCGGGCG 554

     }
 323 G 323
     +
 555 U 555
```

Figure 10. RSEARCH predicted structure alignment of HCV IRES query structure to human PLEKHG4 5' UTR. Output alignment shows query structure in bracket notation on first line. Open and close brackets of one type, basepair together. The second line is the query sequence. The third line shows the sequence identity. Plus signs signify nucleotide changes where the structure is preserved and dots signify gaps in the alignment. The fourth line is the target database entry sequence; the numbers on the end of this line do not correspond to the correct sequence position in the UTR but is correctly noted above the alignment. Large gaps in the alignment are noted with square brackets on the sequence lines. The AUG start codon for HCV starts at position 342 and over 200 basepairs upstream for PLEKHG4 at position 2536.

3.2.8 ERPIN server

The one IRES pattern in the ERPIN (Lambert et al., 2004) list of predefined patterns, is from type 1 poliovirus and was used to search the IRES dataset from Table 1 and 2. This pattern was not found within the IRES database created from Table 1 and also was not found in 10 of 25 complete type 1 poliovirus genome sequences from Genbank. This shows that this particular IRES pattern is quite restrictive and not useful for finding unknown IRES sequences.

As mentioned in the introduction, some viral sequences have been postulated to have tRNA-like sequences in their IRES structure that allow them to interact with the ribosome and, in the case of CrPV, to start translation without the need of a eIF2/GTP/Met-tRNA$_i$ complex. For that reason, a tRNA pattern from ERPIN was used to search the IRES dataset. Functionally significant or not, the pattern for tRNA type 1 was found in VEGF -A (mouse) UTR at about -1200 upstream of the start codon with an e-value of 10^{-6}. A tRNA structure has not been considered in IRES sequences so far upstream from an AUG and may only be coincidental as it is not clear what mechanism this could be part of. It is also not preserved in the human VEGF-A 5' UTR.

3.3 IRES search

3.3.1 Stressed IRES

It has become more apparent that different IRESes may be functionally more relevant in one particular cell context than another (Nevins et al., 2003). Initially I thought that if a particular IRES could be found that was distinctly regulated during a cellular stress, which set it apart from other IRESes, then the possibility existed that a specific structure or binding

site motif could be identified that was responsible for that regulation. Therefore as only a cursory examination of different cell contexts I measured the activity of various IRES sequences using UV irradiation or overexpression of the eIF4G family member DAP5 as stressors.

UV irradiation is known to inhibit the level of translation within a cell through the phosphorylation of eIF2α (Deng et al., 2002; Wu et al., 2002) and a simple Bradford assay of protein levels of UV irradiated cells shows this (Figure 11A). Several IRES that were subcloned into the β-Gal/UTR/CAT bicistronic construct were transiently transfected into HEK 293T cells. After 24 hrs., the cells were treated with UV irradiation. In my hands, each of the IRESes exhibited increased activity after UV irradiation (Figure 11B, 11C). Although some IRESes, like DAP5 or HIAP2, showed a greater increase than others, even the control empty vector exhibited some relative increase in translation of the 2nd open reading frame of the bicistronic construct, possibly due to increased ribosomal readthrough.

Death associated protein 5, DAP5 (also known as NAT1 or p97), is proteolytically cleaved into a protein fragment p86 during apoptosis which is believed to affect translation and IRES activity (Henis-Korenblit et al., 2000; Henis-Korenblit et al., 2002). Overexpression of the p86 protein fragment increased the relative IRES activity of the DAP5 and Apaf-1 IRESes but not XIAP (Figure 12A). The empty bicistronic vector also increased the amount of CAT protein expressed relative to β-Gal activity suggesting a change to the overall translation read through between the two open reading frames. Monocistronic constructs of the 5' UTRs of XIAP and Apaf-1 in front of the CAT open reading frame exhibited a decrease in overall translation as did the empty vector when co-expressed with p86 (Figure 12B). The IRES may not be contributing enough activity relative to the cap-dependent translation initiation to be noticed in the monocistronic constructs.

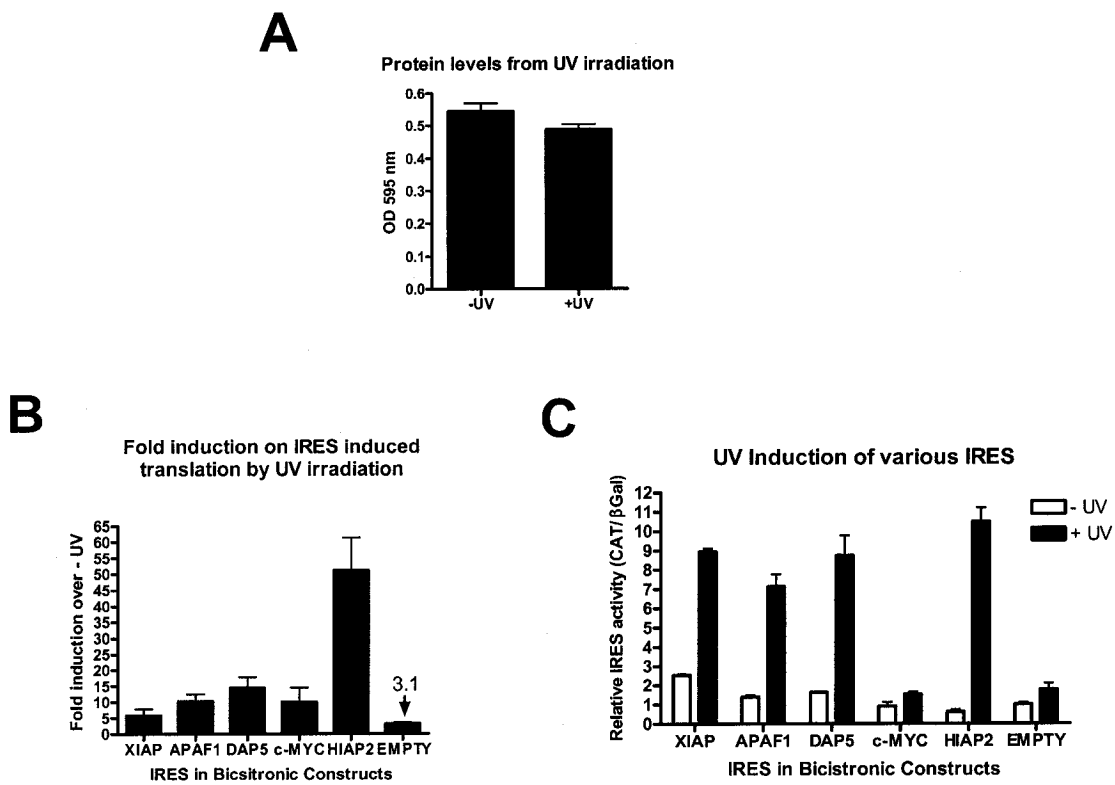


Figure 11. UV induction of IRES activity in bicistronic constructs. (A) Total protein levels of 6 samples with and without UV irradiation as measured by the Bradford assay. (B) Fold induction of relative IRES activity after UV irradiation as measured by ß-Gal activity over CAT protein levels. IRESes are subcloned between a β-Gal and CAT open reading frame in a bicistronic construct. (C) Relative IRES activity with and without UV induction of IRESes from the XIAP, Apaf-1, c-myc, DAP5 and HIAP2 5' UTRs. The XIAP samples were diluted 10 fold for these results.

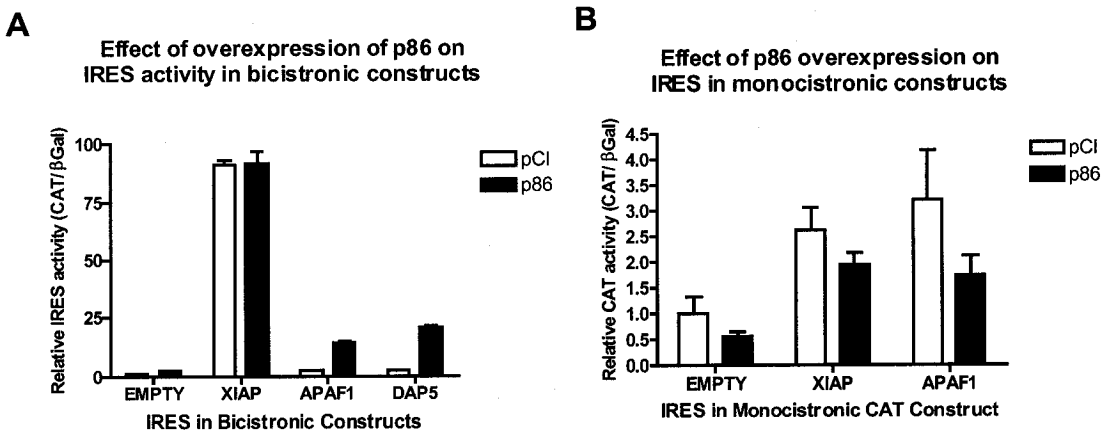


Figure 12. The effect of the DAP5 protein fragment (p86) overexpression on IRES activity. (A) Bicistronic vectors containing either the XIAP, Apaf-1, or DAP5 IRESes were co-transfected with either an empty expression vector pCI or an vector expressing the DAP5 protein fragment, p86. Values are relative to the empty bicistronic vector cotransfected with the empty expression vector set to 1. (B) Monocistronic constructs with either XIAP or Apaf-1 UTRs 5' of the CAT open reading frame were co-transfected with either an empty expression vector, pCI, or a vector expressing the DAP5 protein fragment, p86. Values are relative to the monocistronic CAT vector without any UTR co-transfected with the empty expression vector set to 1.

In both UV irradiation and p86 overexpression, there is a global change in the translation mechanism of the cell. This change is not always consistent with the artificial bicistronic construct control vector (Figure 11B, 12A). In order to find some structural commonality between IRES, I initially proposed to look for specific IRES elements that would function in specific cell contexts. The trouble in doing that would be the separation of specific IRES function from overall translational changes or particulars of the model system. For that reason, it was decided to focus on one IRES and look for structurally related IRESes.

3.3.2 Determination of the XIAP IRES secondary structure

XIAP (X-linked Inhibitor of Apoptosis Protein) is one of the most important cellular inhibitors of apoptosis due to its binding to and direct inhibition of specific caspases (Holcik and Korneluk, 2001). Holcik et al. (Holcik et al., 1999) demonstrated that XIAP is translated by an IRES. This method of cap-independent translation would be absolutely necessary to maintain the anti-apoptosis protection provided by the XIAP protein while cells are under some form of stress (Holcik et al., 2000b; Holcik et al., 1999; Nevins et al., 2003; Ungureanu et al., 2006; Yoon et al., 2006). It is at these times that cap-dependent translation initiation is compromised. Therefore, it is reasonable to believe that a similar functioning IRES exists that enables the expression of cellular genes involved in the regulation of cellular growth, proliferation and death.

The minimum XIAP 5' UTR sequence required for full IRES activity from has been determined to exist between nucleotides -162 to -34 relative to the start codon in the 5' UTR of XIAP mRNA (Holcik and Korneluk, 2000; Holcik et al., 2003). To determine the secondary structure of the XIAP IRES element, the sequence from -162 to +3 of the XIAP

5' UTR was *in vitro* transcribed and probed with RNases T1, V1 and T2. The structure of the sequence from -162 to +50 of the XIAP IRES as it is cloned into the bicistronic reporter plasmid pβgal/(-162)/CAT was also examined to determine if flanking sequences would affect the XIAP IRES structure. The cut sites for all RNases were the same with both RNAs. Therefore, the CAT ORF sequence has no influence on the XIAP IRES structure as expected, since the XIAP IRES exhibits strong activity in this reporter construct. The relative sensitivities of sites to these RNases are presented in Table 4 (Appendix I). The RNase cut sites (Table 4, Appendix I) were used as constraints (Table 4, Appendix I) in the MFOLD structure prediction webserver (Zuker, 2003) to predict a secondary structure for the XIAP IRES (Figure 13).

3.3.3 XIAP IRES structure search of human 5' UTR sequences.

A database of human 5' UTR sequences was compiled from the Ensembl database (NCBI build 34) using Ensmart (now known as BioMart) (Clamp et al., 2003). The sequences were filtered for redundancy by removing exact matches, and if the same 5' UTR sequence existed within 5' UTRs of different lengths, the shorter UTR was discarded. The 5' UTRs from published cellular and viral IRESes were added from an in-lab database (Baird et al., 2006) to the Ensembl created database. The RSEARCH program (Klein and Eddy, 2003) was successfully used on a search for similar UTRs using the well-studied HCV IRES structure as reported above in section 3.2.4. Also as mentioned, RSEARCH found structurally similar UTRs to the HCV IRES that had no significant primary sequence similarity.

After confirming the utility of the search process using the HCV IRES structure, the XIAP IRES structure (Figure 13) was next used as a query to search the same dataset and

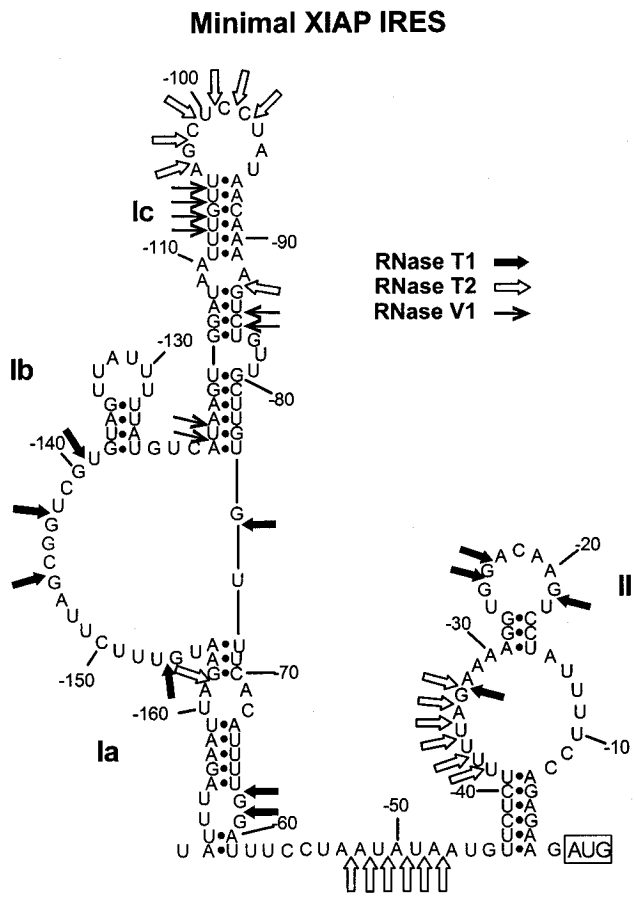


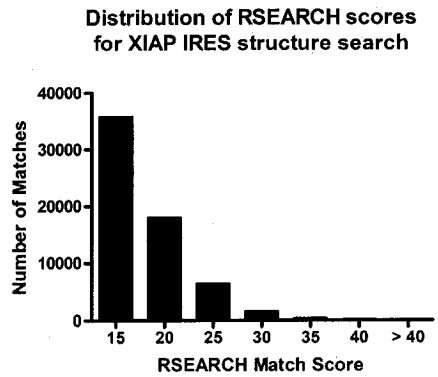
Figure 13. The proposed XIAP IRES RNA secondary structure model. The sites sensitive to RNase T1, which cuts after single stranded guanine nucleotides, RNase T2, which cuts after single stranded nucleotides, and RNase V1, which cuts within regions of double stranded RNA, are shown on the derived structural model. These cut sites were used as constraints in the MFOLD (Zuker, 2003) program to predict the secondary structure of the minimal XIAP IRES sequence.

the distribution of the match scores are displayed in Figure 14A. The top matches of the XIAP IRES structure were mouse and rat orthologs with scores of 173 and 144 respectively (Figure 14B). The next highest match score for a known IRES was NRF, with a score of 35.5. There were approximately 80 matches with scores higher than 35 with the highest score, after the XIAP IRES orthologs, being a match to the 5' UTR from ELG1, an uncharacterized gene, with a score of 47.5.

3.3.4 The 5' UTRs from NRF, AQP4, and ELG1 exhibit IRES activity.

The 5' UTR of NRF was shown previously to contain IRES activity (Oumard et al., 2000), confirming that the search could identify functional IRES. Five additional 5' UTRs were selected from the list of potential structure matches to test for IRES activity, and these are denoted by an asterisk in Figure 14B. Several criteria were used to select from the list of 5' UTRs matches to be tested for IRES activity. As all characterized IRES elements are located just upstream of the AUG, only matches that were at the 3'end of the UTRs were chosen. Structure matches that had non-aligning inserts greater than 100 bases within the structural alignments were also not chosen for further testing. The sequences for each of the five chosen 5' UTRs were amplified by RT-PCR using RNA isolated from HEK 293T cells as a template, and the resulting products were subcloned into the β-Gal/CAT bicistronic reporter vector (Holcik et al., 1999). The 5' UTR from ATF4, which has been shown to have no IRES activity (J. Blais, personal communication), was also cloned into the β-Gal/CAT bicistronic vector and serves as a negative control. These plasmids were subsequently transfected into HEK 293T cells, and relative IRES activity was determined by assessing the levels of β–galactosidase and CAT expression. Of the five UTRs tested, only

A



B

SCORE	GENE	Gene description
332.24	XIAP	170 bases 5'UTR
173.09	MIAP3	Mouse homolog of XIAP
144.07	RIAP	Rat homolog of XIAP
47.49	ELG1 *	uncharacterized
44.37	CMH9	titin isoform novex -3; connectin; CMH9, included; cardiomyopathy, dilated 1G (autosomal dominant)
44.18	HIF-2 α	Hypoxia-inducible factor 2 alpha
41.70	HIAP1	Inhibitor of Apoptosis protein 1 (on minus strand)
41.60	VEGF-D *	Vascular endothelial growth factor D precursor
38.29	AQP4 *	Aquaporin 4 mercurial insensitive water channel
36.00	PCD10 *	programmed cell death 10; apoptosis -related protein 15
35.50	NRF	NF-kappaB repressing factor is a constitutive transcriptional silencer of the multifunctional cytokine interferon -beta
35.17	ZNF146 *	Zinc finger protein 146

Figure 14. The distribution of RSEARCH match scores from a search of a human 5' UTR and IRES sequence database with the XIAP IRES structure. (A) A histogram of the number of matches to the XIAP IRES structure for a given RSEARCH match score plotted in bins of 5 above the score of 15. The number of matches includes multiple matches to the same UTR. (B) A selection of the top match scores above the score of 35 showing the score and identity of the matching 5' UTR sequence.

the ELG1 and Aquaporin 4 (AQP4) UTRs exhibit IRES activity, although this activity is much lower than that exhibited by the XIAP IRES (Figure 15A).

To ensure that the IRES activity exhibited by these two 5' UTRs is not due to the function of a cryptic promoter within the cloned UTR sequence, the CMV promoter was excised from the vector. In such a promoterless construct any expression of CAT must be due to the presence of cryptic promoter activity within the cloned UTR sequence. The β-Gal and CAT activity was measured and is expressed relative to the levels of the neomycin phosphotransferase II (NPTII), which is also contained on the bicistronic vector (but driven by an SV40 promoter, and therefore expressed independently from the β–Gal/CAT bicistronic RNA); therefore NPTII expression levels serve as a control for transfection efficiency. Transient transfections were performed in HEK 293A cells because HEK 293T cells contain the NPTII gene. None of the UTRs tested displayed any cryptic promoter activity (Figure 15B). The UTRs with IRES activity were also tested for spurious splicing events by performing quantitative RT-PCR (qRT-PCR) analysis on mRNA isolated from transfected cells using oligonucleotide primers specific to each of the β-Gal and CAT open reading frames (Holcik et al., 2005). No splicing events were detected as the ratio of CAT to β-Gal mRNA remained the same for all of the constructs tested (Figure 15C). Therefore, these data indicate that the ELG1 and AQP4 5' UTRs exhibit *bona fide* IRES activity.

3.3.5 RNA secondary structure determination of the AQP4 and ELG1 5' UTRs, and comparison to the XIAP IRES structure.

To determine if the IRES-containing 5' UTRs of AQP4 and ELG1 were identified by the RSEARCH program due to actual structural similarities between these UTRs and the XIAP IRES, the secondary structures of the ELG1 and AQP4 5' UTR sequences were

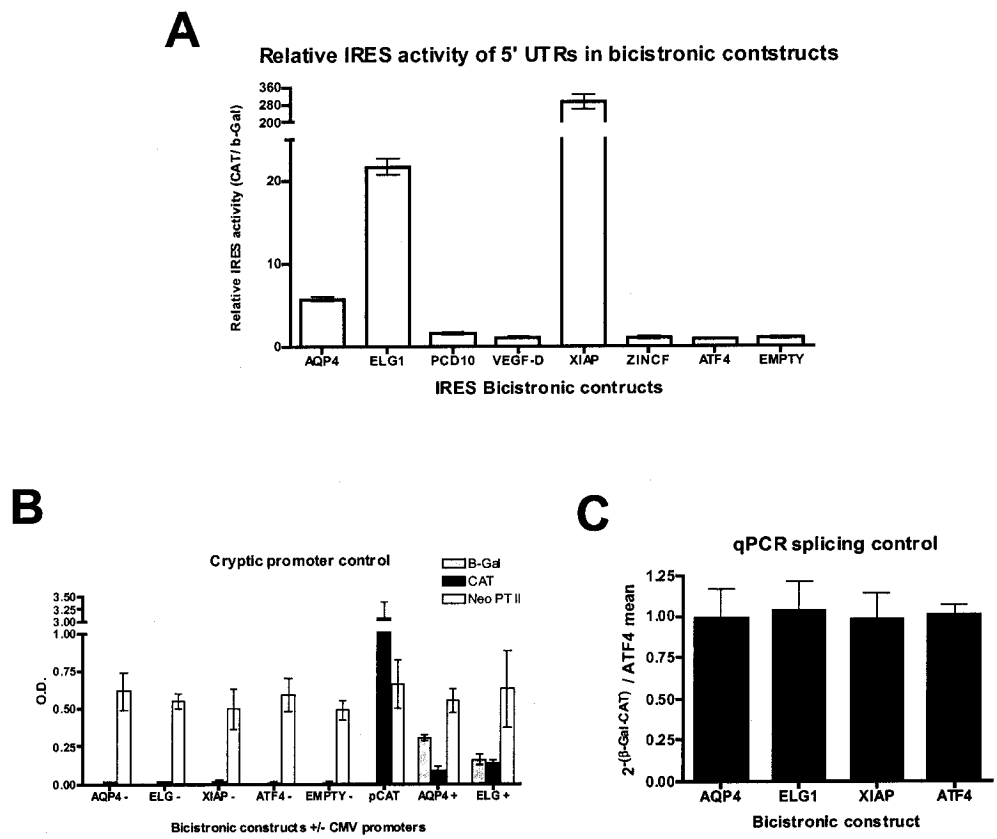


Figure 15. Testing the IRES activity of the five selected 5' UTRs using bicistronic reporter constructs. (A) The IRES activity of each 5' UTR in β-Gal/UTR/CAT bicistronic construct measured as CAT protein levels relative to β-Gal activity. All values are relative to the empty vector, which is set to 1. The 5' UTR of ATF4 and the empty vector are negative controls. (B) Testing for cryptic promoter activity within the UTRs. Promoter activity was tested in bicistronic constructs in which the CMV promoter that drives expression of the bicistronic message was excised. These constructs were transfected into 293A cells and CAT and β-Gal expression was determined. A monocistronic plasmid in which CAT gene expression is driven by the CMV promoter was used as a positive control, as well as the bicistronic constructs of AQP4 and ELG1 that contain the CMV promoter. The neomycin phosphotransferase II (Neo PTII) gene, which also resides on the transfected plasmids, was used as a transfection control. ELISA measurements of Neo PTII protein levels are presented as O.D. values. (C) The AQP4 and ELG1 bicistronic messages were tested for spurious splicing events using quantitative RT-PCR of amplicons within the β-Gal and CAT ORFs. The fold differences of the CAT and β-Gal amplicons was calculated as $2^{-[Ct(\beta\text{-Gal}) - Ct(CAT)]}$ and was plotted relative to the ATF4 negative splicing control, which has no IRES activity. Values represent the mean of 9 biological samples and error bars represent +/- S.E.M. of 9 biological samples.

determined by RNase probing. The secondary structure models that were determined for the AQP4 and ELG1 UTRs and the RSEARCH sequence/structure matches with the XIAP IRES structure are shown in Figure 16. Surprisingly, the secondary structure of ELG1 IRES does not match the structure of XIAP IRES, as was predicted by RSEARCH program (Fig. 16A). The sequences of ELG1 that RSEARCH attributes to the basepaired stems of domain I or II of the XIAP IRES do not come together in the empirically determined structure. Closer examination of ELG1 sequence shows that it contains 35 AUG repeats, which could theoretically help it to form many structures, such as that predicted by RSEARCH. However, the AQP4 5' UTR structure contains a domain that exhibits similarity to the structure observed in the XIAP IRES domain II. Notably, a double-stranded sequence in the AQP4 5' UTR that contains a polypyrimidine tract is similar to the structure of a polypyrimidine tract in the XIAP IRES (Figure 16B). This polypyrimidine tract has been shown previously to be important for XIAP IRES activity and upon deletion of these sequences IRES activity is lost (Holcik et al., 1999). Therefore it was necessary to determine if this polypyrimidine tract is also important for AQP4 IRES activity. To this end, the sequence between -59 to -3 within the AQP4 5' UTR was deleted in the bicistronic vector, and the resulting construct was then tested for IRES activity. The polypyrimidine tract is absolutely required for the IRES activity of both AQP4 and XIAP (Figure 17).

3.3.6 IRES activity of 10 random UTRs.

Although the AQP4 5' UTR showed some structural similarity and the functionally important polypyrimidine tract relative to the XIAP IRES, there is a possibility that if one were to test the IRES activity of 10 randomly picked 5' UTRs over a certain length, that 40% would exhibit some IRES activity as seen by the previous results, which found 2 of 5 UTRs

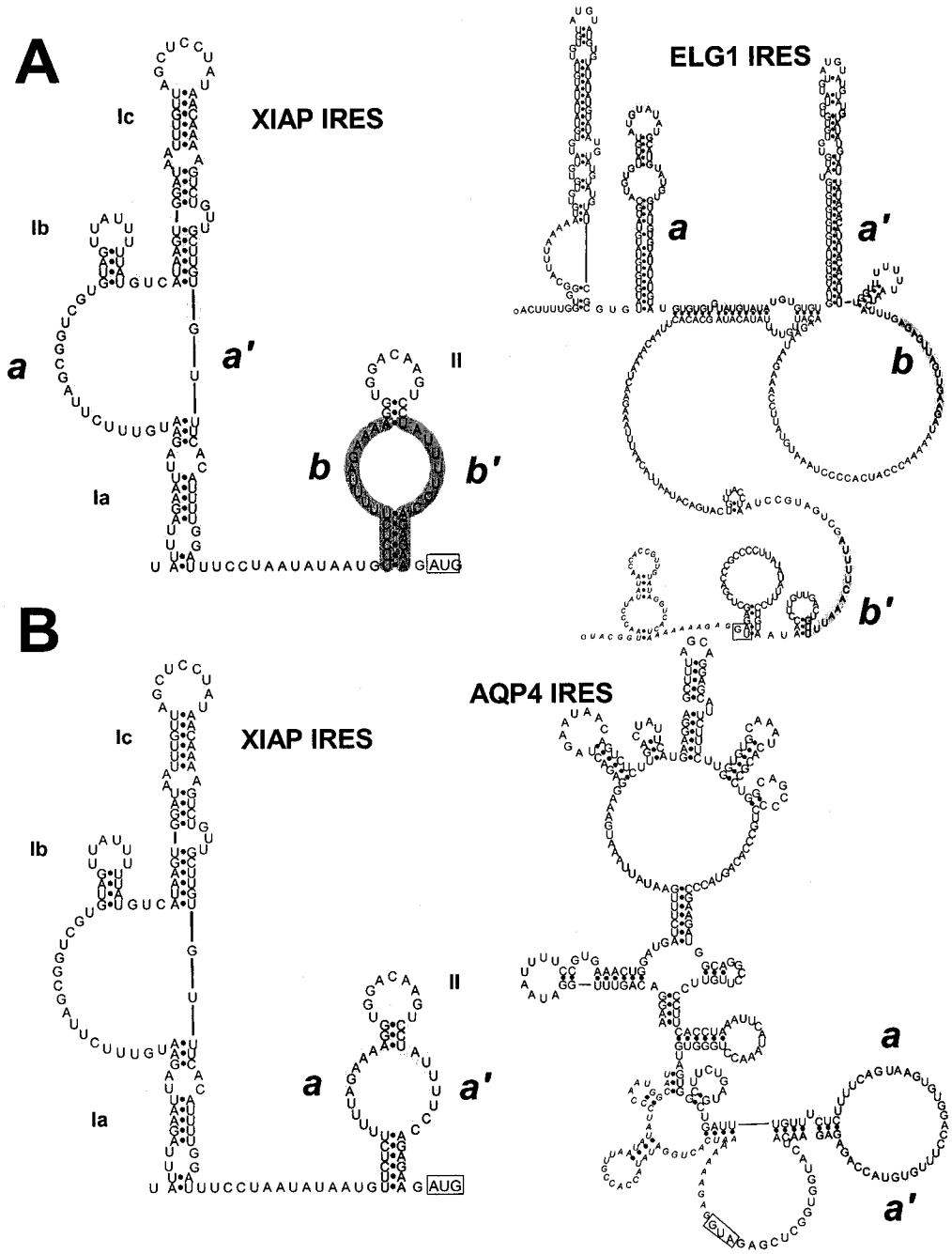


Figure 16. Comparison of the RNA secondary structures of the ELG1, AQP4, and XIAP IRES. The RNA secondary structure models of AQP4 and ELG1 were determined with enzymatic probing using RNase T1, T2, V1 and A. The enzyme cut sites were used as constraints for an RNA secondary structure prediction program and are available in the supplemental data. (A) The XIAP IRES structure is compared to the ELG1 IRES structure. The RSEARCH-predicted matching sequence and structure between the XIAP and ELG1 IRES are outlined in gray and labeled as a/a' and b/b' on each structure. (B) The XIAP IRES structure is compared to the AQP4 IRES structure showing the RSEARCH-predicted matching sequence and structure in gray and labeled as a/a' on each structure.

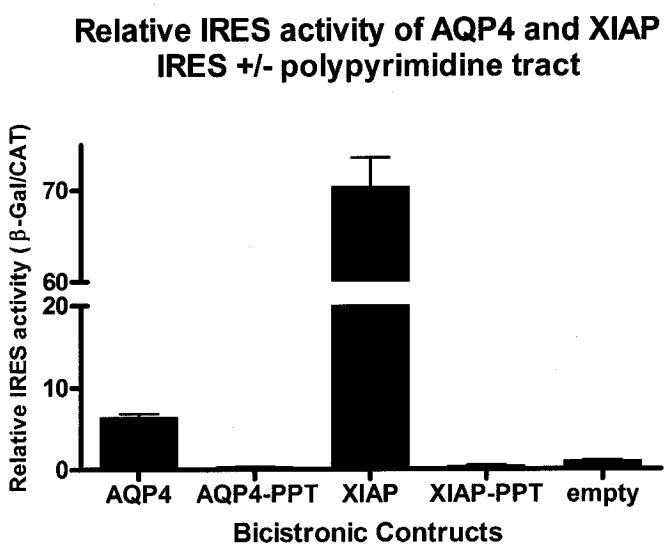


Figure 17. Deletion of the polypyrimidine tracts in XIAP and AQP4 abrogates IRES activity. Bicistronic constructs deleting sequence from -47 to -1 in the XIAP IRES (XIAP-PPT) and -50 to -3 in the AQP4 IRES (AQP4-PPT), which removes the matching polypyrimidine tracts, were tested for IRES activity. IRES activity is expressed as CAT protein levels relative to β-Gal activity. Values are given relative to the empty bicistronic vector, which was set to 1.

tested containing an IRES (Figure 15). Although this would not be a sample size large enough to be statistically significant, this hypothesis was tested. Using the human 5' UTR database from UTRdb Release 21 (Pesole et al., 2002) , 10 UTRs between the sizes of 350 and 1000 nucleotides were randomly chosen for testing using a perl script. The size range was chosen for ease of cloning into the β-Gal/CAT bicistronic vector but long enough to contain an IRES equivalent to the size of the HCV IRES. UTRs were rejected that contained NheI and XhoI restriction sites contained within the bicistronic vector. Only UTRs that were well represented in the NCBI EST database with greater than 5 entries, and that exist in more than one cell type, were used to avoid problems isolating rare or tissue specific transcripts. The UTRdb entry number for each UTR label is given in Table 7 in Appendix III.

Although this was a cursory experiment and needs further validation, at least 4 of the 9 UTRs exhibit approximately 3 fold or greater IRES activity over the empty bicistronic vector (Figure 18). It is important to note that the UTRs should be tested for cryptic promoter activity or aberrant splicing to qualify these UTRs as truly having IRES activity. The results do suggest that IRESes are common and that the search for IRESes similar to the XIAP IRES may have found AQP4 and ELG1 IRES by chance.

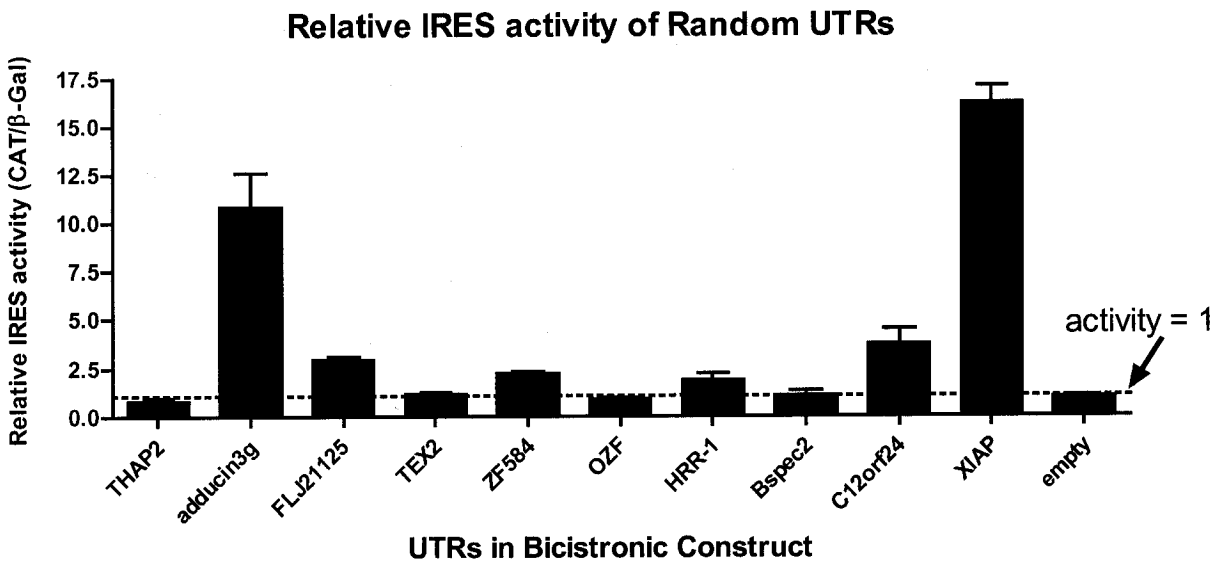


Figure 18. Relative IRES activity of randomly pick 5' UTRs. Each UTR was cloned by RT-PCR and inserted into the β-Gal/UTR/CAT bicistronic construct, transfected into HEK 293A cells and tested for IRES activity after 24 hrs. Relative IRES activity was determined as a ratio of CAT protein levels to ß-Gal enzymatic activity. All values are relative to the empty vector, which is set to 1. Values represent the mean of 9 biological samples and error bars represent+/- S.E.M. of 9 biological samples.

Chapter 4

Discussion

4.1 UTR characteristics

The most notable difference in cellular IRES sequence from the genome UTR datasets is that IRES containing UTRs are generally longer (Figure 1). It should be also noted that the median length of 5' UTRs is over 150 nucleotides, longer than commonly stated dogma of "usually <100 bases" (Lewin, 2006). As minimal IRES sequences can be approximately 150 nucleotides in length (XIAP's minimal IRES is from position -162 to -34 (Holcik and Korneluk, 2000; Holcik et al., 2003)) or less, one conclusion from this comparison is that half of all UTRs are long enough to include an IRES. The cloning of random UTRs (section 3.3.6; Figure 18) where 4 of the 9 UTRs exhibited IRES activity suggests that IRESes are more common than the previously believed 7 to 10 % of all transcripts (Johannes and Sarnow, 1998).

Commonly IRESes are referred to being in UTRs that are long and have extensive secondary structure. Although viral IRESes are known to have specific structures, neither the stability nor the structure of IRES containing UTRs have been proven to be greater than the majority of 5' UTRs. The %GC content of the published IRESes in Table 1 has an overall similar distribution as all the 5' UTRs annotated in human genes (Figure 2) suggesting that the degree of structure may not be a general characteristic of UTRs with IRES sequences.

Of the approximately 23,000 human genes in the human genome, around half of them have AUGs upstream of the defined AUG start codon as seen in Fig. 3a and 3b. Upstream AUGs are common in UTRs and although UTRs with IRES possibly have more upstream AUGs than found in the normal distribution, UTRs with IRES containing no upstream AUGs are common as well. Therefore no generalities can be made concerning upstream AUGs in 5' UTRs containing IRES. Although many of the upstream AUGs will most likely be in a context relatively close to the Kozak consensus sequence (Suzuki et al., 2000) this alone does not seem to confer the properties of a proper start codon (Peri and Pandey, 2001). Approximately 10% of upstream AUGs would be considered "strong" (Kozak, 2005) and efficiently initiate translation while almost 50% of the standard initiation codons are considered "strong" (Figure 3a). Perhaps the ribosome starts to translate with the first AUG in the proper context of the surrounding folded RNA structure, bound proteins and/or a continuous open reading frame. Possibly the start codon is decided upon by a "pioneer" scan (Ishigaki et al., 2001) of the mRNA or an RNA binding protein (McBratney and Sarnow, 1996). Ribosomes can reinitiate translation on downstream AUGs efficiently through leaky scanning (Wang and Rothnagel, 2004) and many of the upstream open reading frames (uORFs) are preserved through evolution (Crowe et al., 2006), but the important context may have more to do with the distance of the uORFs and AUGs from each other and the relative amount of eIF4F complex available in the cell (Poyry et al., 2004).

4.2 RNA software

Several programs that could aid in the prediction of or search for cellular IRESes were evaluated with IRES data. Although ED_SCAN could successfully find areas of local structural significance with the let-7 micro-RNA or Iron Response element, no significant

areas of local structure were found in the IRES of HCV (Figure 5), Apaf-1, CSFV or c-myc. The Zscr values were below 3 which is typical of random sequence (Le et al., 2003). It is unknown whether any significant local structures exist in the cellular IRES, Apaf-1 and c-myc, but the global structures for HCV and CSFV are known to be important but were not evident with this software.

MFOLD (Zuker, 2003) is the most commonly used program for *de novo* structure prediction and its accuracy has been measured at 41% (Doshi et al., 2004), 56% (Dowell and Eddy, 2004) or 73 % (Mathews et al., 1999b) depending on the dataset and method of measuring accuracy. The modified Matthews Correlation Coefficient was used in this work (Gorodkin et al., 2001) as a measurement of accuracy because it took into consideration correctly and falsely predicted base pairs. The lowest energy structure of the HCV IRES had an MCC of 0.39 (Figure 6a) where a perfectly predicted structure would have a value of 1. Notably, one of the suboptimal structures (Figure 6b) produced a better prediction of MCC = 0.54 which suggests that the most stable structure is either not taken in nature (Shapiro et al., 2001) or other forces of the tertiary structure have an effect on the secondary structure that are not taken into account. In looking at all the basepairs that are correctly predicted in all of the suboptimal folds (Figure 6c, 7b) some are missed by them all. The stems that are missed in the suboptimal structures in APAF-1, or basepairs in HCV, would suggest that MFOLD is not considering all possible stems in its search space and, therefore, would never be able to predict a correct structure for these molecules. The examples shown in Figures 6 and 7 suggest that predictions should not be attempted until there is at least some wet-lab experimental data available for the structure. Virtually all the IRES secondary structures that have been empirically derived using chemical or enzymatic probing have used MFOLD or a similar program with probing results as constraints to determine the final structures (Bonnal

et al., 2003b; Jang et al., 2004; Jopling et al., 2004; Le Quesne et al., 2001; Mitchell et al., 2003; Pickering et al., 2004).

The most accurate prediction of the HCV IRES structure was with the DYNALIGN program resulting in an MCC of 0.75 while the CSFV was predicted with an MCC of 0.80. DYNALIGN does not use sequence alignment to make its prediction but energy minimization between the two structures. As it works on global structures, it is not known if it would be able to find local similar structures shared between the two sequences. If one structure was known already, the accuracy of its prediction would possibly signify the correct prediction of an unknown sequence structure.

CARNAC (Perriquet et al., 2003) which uses energy minimization, sequence similarity and covariation to predict a shared global structure only predicted the HCV IRES with an MCC of 0.28 and the CSFV IRES with 0.589 (section 3.2.4). As these two structures, and the BVDV IRES, have known similar structures, one would expect a better prediction. As it is not known if cellular IRES, if any, share a structure, this program would not be useful without knowing already that an overall structure is important, and that it would be conserved in orthologs through evolution. Pfold (Knudsen and Hein, 2003) also predicts a common structure from a group of structurally related sequences. In section 3.2.5, the Pfold predicted structures of the three IRESes in the Rfam database, FGF2, MNT and BAG-1, do not match the respective empirically proven structures (Bonnal et al., 2003b; Mitchell et al., 2005; Pickering et al., 2004). There are several possibilities for the bad predictions: not enough co-variation data in the orthologs used, the empirically derived structures are incorrect or there is not a preserved structure, as the most important parameter is the primary sequence recognition sites.

The structure prediction examples given are not rigorous in the assessment of the predictive powers of these programs. Although only anecdotal, they give a frame of reference to the study of IRES sequences and the search for possible shared structural elements. Although RNA secondary structure prediction has benefited from a depth of intellect, as is exhibited by the algorithms that exist today, accuracy of *de novo* secondary structure prediction of large RNA molecules depends somewhat on good fortune. Protein binding, along with the dynamics of RNA folding as the mRNA is transcribed, may produce a structure different than the *in vitro* defined structure. However some *in vitro* transcribed IRES sequences still function in cell free translation experiments. It is important to have either empirically derived structural data or covariance data from truly related structures as additional inputs for structure prediction. In the case of cellular IRESes, the diversity of regulatory contexts may be the biggest hindrance to a global understanding of the IRES mechanisms held within the primary sequence of UTRs containing IRES elements. If there was a wish list of program functions to help define IRES structure/sequence/function from primary sequence, one might be the ability to take all the sequences and do a "cluster" motif search, similar to microarray clustering assuming that IRESes will separate into groups. There is also quite a bit of deletion and directed mutation analysis data that alters the activity of IRESes. These data seem underutilized within current prediction software. Could these data be used, similarly to structure probing data, in that a high score could be placed on sequences that were crucial to activity and a low score on sequences that showed no effect when altered? The apparent modular elements within some IRESes as exhibited by partial activity of non-overlapping fragments of the UTRs suggest that there is no need for an similar overall structure in every cellular IRES sequence.

4.3 IRES search

4.3.1 HCV search

In section 3.2.7, the efficacy of the RSEARCH program was tested by using the well-characterized structure of the HCV IRES to search a database of human 5' UTRs and UTRs known to contain IRES. This search effectively identified IRESes from GbvB, CSFV, and BVDV (Figure 9B), which have been previously shown to have structural similarity to the HCV IRES (Brown et al., 1992; Grace et al., 1999; Lemon and Honda, 1997). This exercise established that structurally similar UTRs could be found using the RSEARCH program, validating it as an approach to look for other IRES structures.

The one high scoring match to the human PLEKHG4 5' UTR is perplexing. Although the structural alignment to domain III of HCV is quite strong, there are a couple of anomalies which suggest this match may not be significant. First, the two gaps at the analogous IIId and IIIe stemloops are quite large and the proposed structure may not form as RSEARCH has predicted (Figure 10). Second, the published IRESes so far have been fairly close to the AUG start codons and this structure is over 200 bases upstream of the known start codon. However, the structural alignment to domain IIIb of the HCV IRES is quite good (Figure 10). Domain IIIb of HCV is known to bind directly to eIF3, the adapter protein which normally binds the 40S ribosomal subunit and eIF4G (Kieft et al., 1999; Siridechadilok et al., 2005). This subjugates the need for the eIF4F complex during translation initiation of the viral genome and may play a role in PLEKHG4 translation. The putative structural homology to domain IIIb is found in many of the lower scoring matches. A mutation in the 5' UTR of the PLEKHG4 gene has been shown to be responsible for autosomal dominant cerebellar ataxia through aberrant protein aggregation (Ishikawa et al.,

2005). Although the mutation is 16 bases upstream of the AUG start codon and not with the structural alignment region, this match to the HCV IRES structure suggests a possible mechanism of translation regulation that bears further study.

4.3.2 XIAP search

In section 3.3, the secondary structure of the XIAP IRES (Figure 13) was empirically determined to help identify novel cellular IRESes by searching for human 5' UTR sequences that display a structural similarity to the XIAP IRES. The search of the human 5' UTR database using the structure of the XIAP IRES resulted in the identification of five 5' UTRs in which the region with structural similarity was in the correct position (at the 3' end of the UTR, proximal to the AUG codon) and orientation to exhibit IRES activity. Characterization of the IRES activity of these identified sequences showed that the 5' UTRs of AQP4 and ELG1 promote IRES-dependent translation (Figure 15). However, empirical determination of the structure of the AQP4 and ELG1 5' UTRs and subsequent comparison to the structure of the XIAP IRES (Figure 16) revealed that the structures of these sequences share little similarity. A polypyrimidine tract in the XIAP IRES and AQP4 IRES was similar in both sequences and found to be absolutely necessary for IRES activity. These results suggest that unlike the viral IRES, the overall structure of cellular IRESes are not an important factor in determining IRES activity, but perhaps small sequence motifs and the cohort of proteins which bind them may define IRES activity.

It has been recognized that both viral and cellular IRES do not share primary sequence homology, and therefore it is not possible to identify mRNAs that harbour IRES elements by comparison of sequence data using search programs such as BLAST. This realization has presented a barrier to the genome-wide identification of IRES, resulting in the

identification of mRNAs that contain IRES in a piecemeal fashion – a message is suspected to be translated under conditions that repress cap-dependent translation and the 5' UTR is subsequently tested for IRES activity. However, several studies have highlighted the importance of secondary structure for the function of viral IRESes (Brown et al., 1991; Kieft et al., 1999; Odreman-Macchioli et al., 2001) and some viral IRESes share structural homology, suggesting that an underlying determinant for function of viral IRESes is the presence of specific structural elements (Le et al., 1998). For this reason, it was hypothesized that cellular IRESes may also share structural homology, and that comparison of the secondary structure of 5' UTRs may be a means to identify novel IRESes on a genomic scale.

The search of the 5' UTR of the human genome (supplemented with 5' UTR known to exhibit IRES activity from all species and viruses) with the structure of the XIAP IRES did indeed result in the identification of RNA sequences that exhibit IRES activity. Of the top matches, two were the 5' UTRs of the mouse XIAP ortholog (MIAP) and the rat XIAP ortholog (RIAP) that both display IRES activity (Holcik et al., 1999) (M. Holcik, unpublished), and one, NRF, that has previously been shown to have IRES activity (Oumard et al., 2000). This also includes two others (ELG1 and AQP4) that also display IRES function. AQP4, a water channel-forming protein expressed in the brain, has recently been shown to have a role in edema during eclampsia, as its protein levels are elevated during pregnancy (Quick and Cipolla, 2005); however, mRNA levels have not been shown to change (Quick and Cipolla, 2005), suggesting a possible IRES function. Interestingly, the search did not identify the 5' UTRs of p27(Kip1) (Miskimins et al., 2001) or Bcl-xL, both of which express IRES activity (Yoon et al., 2006)(matching scores were below the threshold of significance, between 16 and 17). Yoon et al. (Yoon et al., 2006) recently showed that the

IRES activity of XIAP, p27(Kip1), and Bcl-xL are specifically and severely impaired in cells harbouring mutations in the pseudouridinase gene Dkc1. This mutation was shown to prevent the proper pseudouridylation of rRNA and is thought to disrupt ribosome structure (Yoon et al., 2006). The IRES-dependent translation of the XIAP, p27(Kip1), and Bcl-xL messages are specifically sensitive to these changes, as global translation is not affected. It was therefore hypothesized that these IRESes may share some common feature, such as secondary structure, that is sensitive to changes in ribosome architecture (Yoon et al., 2006). Based on the results of this thesis, the secondary structures of these IRES elements may not be similar, and that other factors, such as changes in the association of *trans*-acting factors with the modified ribosome or direct association of the IRES with the ribosome, may account for the coordinated reduction of XIAP, p27(Kip1), and Bcl-xL IRES activity in Dkc1 mutant cells.

Comparison of the empirically-determined secondary structures of the ELG1 and AQP4 5' UTRs to the secondary structure of the XIAP IRES showed only limited homology, with only small regions of AQP4 displaying some similarity within the overall structure. This suggests that cellular IRESes do not share a globally similar structure, as do viral IRES. No overall structural similarity has been found among the 11 known cellular IRES structures mentioned in the introduction. Whereas a virus needs to compete efficiently to get its proteins expressed within an infected cell, cellular transcripts with an IRES are more likely translationally regulated, with several *trans*-acting factors contributing differently, depending upon the cellular context. Studies of the HCV IRES have shown it to directly interact with the ribosome to induce conformational changes (Spahn et al., 2001), and this interaction appears to occur where eIF4G normally is bound (Siridechadilok et al., 2005). Cricket paralysis virus IRES has an RNA structure which allows it to interact with

the ribosome by mimicking an eIF2/GTP/Met-tRNA$_i$ complex (Wilson et al., 2000b). The pattern for the capsid IRES of Dicistroviridae built from empirical data of CrPV and PSIV did not find any matches in eukaryotic mRNAs using the pattern searching program 'Scan for Matches' (Dsouza et al., 1997; Hatakeyama et al., 2004). This suggests a structure/function mechanism for these Dicistroviridae viral IRESes, which is distinct from eukaryotic IRESes and other viral IRES, in their interaction with ribosomes. The importance of structures such as these is yet to be shown for any cellular IRES.

Important to note, is the presence of a polypyrimidine tract that was folded into a stem-loop structure in both the XIAP and AQP4 IRES. It was previously found that the polypyrimidine tract is absolutely necessary for XIAP IRES activity (Holcik and Korneluk, 2000), and therefore, it was tested whether deletion of this sequence in the AQP4 5' UTR would also abrogate IRES function. Indeed, deletion of the AQP4 5' UTR polypyrimidine tract causes a complete loss of IRES activity. Because this small region is required for IRES activity, short motifs may be critical determinants of IRES function. Subsequent experiments addressing this premise, by Dr. Stephen Lewis in our lab, used RNA-affinity chromatography to pull down RNA interacting proteins. He has shown that AQP4, XIAP and ELG1 bind similar IRES *trans*-acting factors (ITAFs) (Appendix III). All three bind La autoantigen, which has been previously shown to be necessary for XIAP IRES activity (Holcik and Korneluk, 2000). AQP4 and XIAP, which share the polypyrimidine tract, also share binding to PTB. It has been found that a repeated PTB binding motif can induce IRES activity (Mitchell et al., 2005). Moreover, searches for 5' UTRs that harbour this PTB-binding motif have resulted in the identification of novel IRES elements (albeit not XIAP or AQP4) (Mitchell et al., 2005). These findings, suggest that the recruitment of a particular

group of mRNA binding proteins (ITAFs) is the critical factor in cellular IRES-mediated translation and not global structure.

4.4 Summary

The standard translation initiation mechanism is inhibited or impeded during mitosis, cell stress or viral infection. During these times the efficiency of translation of a subset of mRNAs is elevated relative to the total population of transcripts. One mechanism by which this is achieved is through the recruitment of ribosomes by IRES elements in the 5' UTR. Viral IRES sequences have shown some RNA structural conservation between different viral families necessary for translation initiation. Recent viral IRES/ribosome studies suggest these structures allow the bypassing of other cofactors involved in standard translation initiation by direct interactions with ribosome, as seen for HCV RNA or the tRNA structure of CrPV. Is has been postulated that there are shared sequence/structural motifs between cellular IRESes as well, allowing subsets of mRNAs to be more efficiently translated. Perhaps these shared motifs are not required to hijack the translation process as virus IRESes do but are required to retain enough translation efficiency to be biologically relevant in that cellular context. The non-canonical protein factors, the ITAFs, which have been shown to be involved in IRES function are all involved in normal mRNA processing and their binding may not be so specific as to require a particular stem-loop, however, the IRES structure may dictate whether an ITAF is able to bind. So far only PTB has shown some structural preference for a binding sequence motif in a double stranded context (Mitchell et al., 2005). Other ITAFs like PCBP open up the structure for PTB to bind (Mitchell et al., 2003). Do the cellular IRESes have structures that prevent ribosome binding unless specific ITAFs are present? And would these structures need to be preserved? Some mRNA

binding proteins are needed to specifically recruit the small ribosomal subunit to the transcript while others change or open up the structure to allow ITAFs and the ribosome to bind.

The possibility of different classes of IRESes and of cellular IRES diversity may exist for different changes to the translational machinery, different forms of initiation factors, different factors and different ribosome assemblies. Even in the studies of IRESes there is a need to avoid becoming dogmatic about a single mechanism of translation initiation, one universal ITAF, one structure, one way to recruit a ribosome. Diversity of regulation seems to work well for the cell and even better for whole organisms. The design may not seem intelligent but it is good enough.

References

Akbergenov, R., Zhanybekova, S., Kryldakov, R. V., Zhigailov, A., Polimbetova, N. S., Hohn, T., and Iskakov, B. K. (2004). ARC-1, a sequence element complementary to an internal 18S rRNA segment, enhances translation efficiency in plants when present in the leader or intercistronic region of mRNAs. Nucleic Acids Res *32*, 239-247.

Ali, I. K., McKendrick, L., Morley, S. J., and Jackson, R. J. (2001). Truncated initiation factor eIF4G lacking an eIF4E binding site can support capped mRNA translation. Embo J *20*, 4233-4242.

Anwar, A., Ali, N., Tanveer, R., and Siddiqui, A. (2000). Demonstration of functional requirement of polypyrimidine tract-binding protein by SELEX RNA during hepatitis C virus internal ribosome entry site-mediated translation initiation. J Biol Chem *275*, 34231-34235.

Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *100*, 3889-3894.

Attal, J., Theron, M. C., Taboit, F., Cajero-Juarez, M., Kann, G., Bolifraud, P., and Houdebine, L. M. (1996). The RU5 ('R') region from human leukaemia viruses (HTLV-1) contains an internal ribosome entry site (IRES)-like sequence. FEBS Lett *392*, 220-224.

Back, S. H., Shin, S., and Jang, S. K. (2002a). Polypyrimidine tract-binding proteins are cleaved by caspase-3 during apoptosis. J Biol Chem *277*, 27200-27209.

Back, S. H., Kim, Y. K., Kim, W. J., Cho, S., Oh, H. R., Kim, J. E., and Jang, S. K. (2002b). Translation of polioviral mRNA is inhibited by cleavage of polypyrimidine tract-binding proteins executed by polioviral 3C(pro). J Virol *76*, 2529-2542.

Baird, S. D., Turcotte, M., Korneluk, R. G., and Holcik, M. (2006). Searching for IRES. RNA *12*, 1755-1785.

Barraille, P., Chinestra, P., Bayard, F., and Faye, J. C. (1999). Alternative initiation of translation accounts for a 67/45 kDa dimorphism of the human estrogen receptor ERalpha. Biochem Biophys Res Commun *257*, 84-88.

Barreau, C., Paillard, L., and Osborne, H. B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res *33*, 7138-7150.

Belsham, G. J., and Sonenberg, N. (1996). RNA-protein interactions in regulation of picornavirus RNA translation. Microbiol Rev *60*, 499-511.

Berlioz, C., and Darlix, J. L. (1995). An internal ribosomal entry mechanism promotes translation of murine leukemia virus gag polyprotein precursors. J Virol *69*, 2214-2222.

Berlioz, C., Torrent, C., and Darlix, J. L. (1995). An internal ribosomal entry signal in the rat VL30 region of the Harvey murine sarcoma virus leader and its use in dicistronic retroviral vectors. J Virol *69*, 6400-6407.

Bernstein, J., Sella, O., Le, S. Y., and Elroy-Stein, O. (1997). PDGF2/c-sis mRNA leader contains a differentiation-linked internal ribosomal entry site (D-IRES). J Biol Chem *272*, 9356-9362.

Bhattacharyya, S., and Das, S. (2005). Mapping of secondary structure of the spacer region within the 5'-untranslated region of the coxsackievirus B3 RNA: possible role of an apical GAGA loop in binding La protein and influencing internal initiation of translation. Virus Res *108*, 89-100.

Bieleski, L., and Talbot, S. J. (2001). Kaposi's sarcoma-associated herpesvirus vCyclin open reading frame contains an internal ribosome entry site. J Virol *75*, 1864-1869.

Bieleski, L., Hindley, C., and Talbot, S. J. (2004). A polypyrimidine tract facilitates the expression of Kaposi's sarcoma-associated herpesvirus vFLIP through an internal ribosome entry site. J Gen Virol *85*, 615-620.

Billoud, B., Kontic, M., and Viari, A. (1996). Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence database. Nucleic Acids Res *24*, 1395-1403.

Blais, J. D., Filipenko, V., Bi, M., Harding, H. P., Ron, D., Koumenis, C., Wouters, B. G., and Bell, J. C. (2004). Activating transcription factor 4 is translationally regulated by hypoxic stress. Mol Cell Biol *24*, 7469-7482.

Bohl, F., Kruse, C., Frank, A., Ferring, D., and Jansen, R. P. (2000). She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. Embo J *19*, 5514-5524.

Bonnal, S., Boutonnet, C., Prado-Lourenco, L., and Vagner, S. (2003a). IRESdb: the Internal Ribosome Entry Site database. Nucleic Acids Res *31*, 427-428.

Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A. C., and Vagner, S. (2003b). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. J Biol Chem *278*, 39330-39336.

Bonnal, S., Pileur, F., Orsini, C., Parker, F., Pujol, F., Prats, A. C., and Vagner, S. (2005). Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. J Biol Chem *280*, 4144-4153.

Borman, A., and Jackson, R. J. (1992). Initiation of translation of human rhinovirus RNA: mapping the internal ribosome entry site. Virology *188*, 685-696.

Brown, E. A., Zajac, A. J., and Lemon, S. M. (1994). In vitro characterization of an internal ribosomal entry site (IRES) present within the 5' nontranslated region of hepatitis A virus RNA: comparison with the IRES of encephalomyocarditis virus. J Virol *68*, 1066-1074.

Brown, E. A., Day, S. P., Jansen, R. W., and Lemon, S. M. (1991). The 5' nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. J Virol *65*, 5828-5838.

Brown, E. A., Zhang, H., Ping, L. H., and Lemon, S. M. (1992). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. Nucleic Acids Res *20*, 5041-5045.

Brown, E. C., and Jackson, R. J. (2004). All five cold-shock domains of unr (upstream of N-ras) are required for stimulation of human rhinovirus RNA translation. J Gen Virol *85*, 2279-2287.

Buck, C. B., Shen, X., Egan, M. A., Pierson, T. C., Walker, C. M., and Siliciano, R. F. (2001). The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. J Virol *75*, 181-191.

Bushell, M., Stoneley, M., Kong, Y. W., Hamilton, T. L., Spriggs, K. A., Dobbyn, H. C., Qin, X., Sarnow, P., and Willis, A. E. (2006). Polypyrimidine Tract Binding Protein Regulates IRES-Mediated Gene Expression during Apoptosis. Mol Cell *23*, 401-412.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., *et al.* (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics *37*, 327-336.

Carrington, J. C., and Freed, D. D. (1990). Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. J Virol *64*, 1590-1597.

Carter, M. S., and Sarnow, P. (2000). Distinct mRNAs that encode La autoantigen are differentially expressed and contain internal ribosome entry sites. J Biol Chem *275*, 28301-28307.

Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A., and Gruss, P. (1998). Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. Cell *94*, 727-737.

Cencig, S., Nanbru, C., Le, S. Y., Gueydan, C., Huez, G., and Kruys, V. (2004). Mapping and characterization of the minimal internal ribosome entry segment in the human c-myc mRNA 5' untranslated region. Oncogene *23*, 267-277.

Chappell, S. A., and Mauro, V. P. (2003). The internal ribosome entry site (IRES) contained within the RNA-binding motif protein 3 (Rbm3) mRNA is composed of functionally distinct elements. J Biol Chem *278*, 33793-33800.

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

Chappell, S. A., Owens, G. C., and Mauro, V. P. (2001). A 5' Leader of Rbm3, a Cold Stress-induced mRNA, Mediates Internal Initiation of Translation with Increased Efficiency under Conditions of Mild Hypothermia. J Biol Chem *276*, 36917-36922.

Chen, J. H., Le, S. Y., and Maizel, J. V. (2000). Prediction of common secondary structures of RNAs: a genetic algorithm approach. Nucleic Acids Res *28*, 991-999.

Chen, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J., and Blyn, L. B. (2002). A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. Biosystems *65*, 157-177.

Cherry, S., Doukas, T., Armknecht, S., Whelan, S., Wang, H., Sarnow, P., and Perrimon, N. (2005). Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. Genes Dev *19*, 445-452.

Chiang, P. W., Carpenter, L. E., and Hagerman, P. J. (2001). The 5'-untranslated region of the FMR1 message facilitates translation by internal ribosome entry. J Biol Chem *276*, 37916-37921.

Cho, S., Kim, J. H., Back, S. H., and Jang, S. K. (2005). Polypyrimidine tract-binding protein enhances the internal ribosomal entry site-dependent translation of p27Kip1 mRNA and modulates transition from G1 to S phase. Mol Cell Biol *25*, 1283-1297.

Clamp, M., Andrews, D., Barker, D., Bevan, P., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., *et al.* (2003). Ensembl 2002: accommodating comparative genomics. Nucleic Acids Res *31*, 38-42.

Clemens, M. J. (2001). Translational regulation in cell stress and apoptosis. Roles of the eIF4E binding proteins. J Cell Mol Med *5*, 221-239.

Clemens, M. J., Bushell, M., Jeffrey, I. W., Pain, V. M., and Morley, S. J. (2000). Translation initiation factor modifications and the regulation of protein synthesis in apoptotic cells. Cell Death Differ *7*, 603-615.

Clote, P., Ferre, F., Kranakis, E., and Krizanc, D. (2005). Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. RNA *11*, 578-591.

Coldwell, M. J., Mitchell, S. A., Stoneley, M., MacFarlane, M., and Willis, A. E. (2000). Initiation of Apaf-1 translation by internal ribosome entry. Oncogene *19*, 899-905.

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

Collier, A. J., Gallego, J., Klinck, R., Cole, P. T., Harris, S. J., Harrison, G. P., Aboul-Ela, F., Varani, G., and Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. Nat Struct Biol *9*, 375-380.

Cornelis, S., Tinton, S. A., Schepens, B., Bruynooghe, Y., and Beyaert, R. (2005). UNR translation can be driven by an IRES element that is negatively regulated by polypyrimidine tract binding protein. Nucleic Acids Res *33*, 3095-3108.

Cornelis, S., Bruynooghe, Y., Denecker, G., Van Huffel, S., Tinton, S., and Beyaert, R. (2000). Identification and characterization of a novel cell cycle-regulated internal ribosome entry site. Mol Cell *5*, 597-605.

Costa-Mattioli, M., Svitkin, Y., and Sonenberg, N. (2004). La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro. Mol Cell Biol *24*, 6861-6870.

Cote, C. A., Gautreau, D., Denegre, J. M., Kress, T. L., Terry, N. A., and Mowry, K. L. (1999). A Xenopus protein related to hnRNP I has a role in cytoplasmic RNA localization. Mol Cell *4*, 431-437.

Crowe, M. L., Wang, X. Q., and Rothnagel, J. A. (2006). Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. BMC Genomics *7*, 16.

Cuesta, R., Laroia, G., and Schneider, R. J. (2000). Chaperone hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initiation complexes. Genes Dev *14*, 1460-1470.

Czibener, C., Alvarez, D., Scodeller, E., and Gamarnik, A. V. (2005). Characterization of internal ribosomal entry sites of Triatoma virus. J Gen Virol *86*, 2275-2280.

De Pietri Tonelli, D., Mihailovich, M., Schnurbus, R., Pesole, G., Grohovaz, F., and Zacchetti, D. (2003). Translational control of Scamper expression via a cell-specific internal ribosome entry site. Nucleic Acids Res *31*, 2508-2513.

De Rijk, P., Wuyts, J., and De Wachter, R. (2003). RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics *19*, 299-300.

Deffaud, C., and Darlix, J. L. (2000a). Characterization of an internal ribosomal entry segment in the 5' leader of murine leukemia virus env RNA. J Virol *74*, 846-850.

Deffaud, C., and Darlix, J. L. (2000b). Rous sarcoma virus translation revisited: characterization of an internal ribosome entry segment in the 5' leader of the genomic RNA. J Virol *74*, 11581-11588.

Deng, J., Harding, H. P., Raught, B., Gingras, A. C., Berlanga, J. J., Scheuner, D., Kaufman, R. J., Ron, D., and Sonenberg, N. (2002). Activation of GCN2 in UV-irradiated cells inhibits translation. Curr Biol *12*, 1279-1286.

Dever, T. E. (2002). Gene-specific regulation by general translation factors. Cell *108*, 545-556.

Dinkova, T. D., Zepeda, H., Martinez-Salas, E., Martinez, L. M., Nieto-Sotelo, J., and de Jimenez, E. S. (2005). Cap-independent translation of maize Hsp101. Plant J *41*, 722-731.

Dobrikova, E., Florez, P., Bradrick, S., and Gromeier, M. (2003). Activity of a type 1 picornavirus internal ribosomal entry site is determined by sequences within the 3' nontranslated region. Proc Natl Acad Sci U S A *100*, 15125-15130.

Dobson, T., Minic, A., Nielsen, K., Amiott, E., and Krushel, L. (2005). Internal initiation of translation of the TrkB mRNA is mediated by multiple regions within the 5' leader. Nucleic Acids Res *33*, 2929-2941.

Domier, L. L., McCoppin, N. K., and D'Arcy, C. J. (2000). Sequence requirements for translation initiation of Rhopalosiphum padi virus ORF2. Virology *268*, 264-271.

Doshi, K. J., Cannone, J. J., Cobaugh, C. W., and Gutell, R. R. (2004). Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. BMC Bioinformatics *5*, 105.

Dowell, R. D., and Eddy, S. R. (2004). Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics *5*, 71.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol *3*, 195-205.

Dsouza, M., Larsen, N., and Overbeek, R. (1997). Searching for patterns in genomic data. Trends Genet *13*, 497-498.

Eddy, S. R. (1996). RNABOB: Pattern searching for RNA secondary structures. http://www.genetics.wustl.edu/eddy/software/, *2006*.

Ehrenfeld, E. (1996). Initiation of Translation by Picornavirus RNAs. In Translational Control, J. W. B. Hershey, M. B. Mathews, and N. Sonenberg, eds. (Cold Spring Harbor, Cold Spring Harbor Laboratory Press), pp. 549-574.

Evans, J. R., Mitchell, S. A., Spriggs, K. A., Ostrowski, J., Bomsztyk, K., Ostarek, D., and Willis, A. E. (2003). Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. Oncogene *22*, 8012-8020.

Evdokimova, V., Ovchinnikov, L. P., and Sorensen, P. H. (2006). Y-box binding protein 1: providing a new angle on translational regulation. Cell Cycle *5*, 1143-1147.

Fernandez, J., Yaman, I., Mishra, R., Merrick, W. C., Snider, M. D., Lamers, W. H., and Hatzoglou, M. (2001). Internal ribosome entry site-mediated translation of a mammalian mRNA is regulated by amino acid availability. J Biol Chem *276*, 12285-12291.

Fernandez-Miragall, O., and Martinez-Salas, E. (2003). Structural organization of a viral IRES depends on the integrity of the GNRA motif. RNA *9*, 1333-1344.

Fletcher, S. P., and Jackson, R. J. (2002). Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. J Virol *76*, 5024-5033.

Fogel, G. B., Porto, V. W., Weekes, D. G., Fogel, D. B., Griffey, R. H., McNeil, J. A., Lesnik, E., Ecker, D. J., and Sampath, R. (2002). Discovery of RNA structural elements using evolutionary computation. Nucleic Acids Res *30*, 5310-5317.

Galban, S., Fan, J., Martindale, J. L., Cheadle, C., Hoffman, B., Woods, M. P., Temeles, G., Brieger, J., Decker, J., and Gorospe, M. (2003). von Hippel-Lindau protein-mediated repression of tumor necrosis factor alpha translation revealed through use of cDNA arrays. Mol Cell Biol *23*, 2316-2328.

Gallego, J. (2002). Internal initiation of translation by viral and cellular IRESs--a new avenue for specific inhibition of protein synthesis? Curr Opin Drug Discov Devel *5*, 777-784.

Garlapati, S., and Wang, C. C. (2005). Structural elements in the 5'-untranslated region of giardiavirus transcript essential for internal ribosome entry site-mediated translation initiation. Eukaryot Cell *4*, 742-754.

Gingras, A. C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem *68*, 913-963.

Giraud, S., Greco, A., Brink, M., Diaz, J. J., and Delafontaine, P. (2001). Translation initiation of the insulin-like growth factor I receptor mRNA is mediated by an internal ribosome entry site. J Biol Chem *276*, 5668-5675.

Gooding, C., Kemp, P., and Smith, C. W. (2003). A novel polypyrimidine tract-binding protein paralog expressed in smooth muscle cells. J Biol Chem *278*, 15201-15207.

Gopalan, V., Vioque, A., and Altman, S. (2002). RNase P: variations and uses. J Biol Chem *277*, 6759-6762.

Gorodkin, J., Stricklin, S. L., and Stormo, G. D. (2001). Discovering common stem-loop motifs in unaligned RNA sequences. Nucleic Acids Res *29*, 2135-2144.

Grace, K., Gartland, M., Karayiannis, P., McGarvey, M. J., and Clarke, B. (1999). The 5' untranslated region of GB virus B shows functional similarity to the internal ribosome entry site of hepatitis C virus. J Gen Virol *80 (Pt 9)*, 2337-2341.

Gradi, A., Svitkin, Y. V., Imataka, H., and Sonenberg, N. (1998a). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. Proc Natl Acad Sci U S A *95*, 11089-11094.

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

Grassmann, C. W., Yu, H., Isken, O., and Behrens, S. E. (2005). Hepatitis C virus and the related bovine viral diarrhea virus considerably differ in the functional organization of the 5' non-translated region: implications for the viral life cycle. Virology *333*, 349-366.

Griffiths, A., and Coen, D. M. (2005). An unusual internal ribosome entry site in the herpes simplex virus thymidine kinase gene. Proc Natl Acad Sci U S A *102*, 9667-9672.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Grillo, G., Attimonelli, M., Liuni, S., and Pesole, G. (1996). CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. Comput Appl Biosci *12*, 1-8.

Grillo, G., Licciulli, F., Liuni, S., Sbisa, E., and Pesole, G. (2003). PatSearch: A program for the detection of patterns and structural motifs in nucleotide sequences. Nucleic Acids Res *31*, 3608-3612.

Grolleau, A., Bowman, J., Pradet-Balade, B., Puravs, E., Hanash, S., Garcia-Sanz, J. A., and Beretta, L. (2002). Global and specific translational control by rapamycin in T cells uncovered by microarrays and proteomics. J Biol Chem *277*, 22175-22184.

Gromeier, M., Lachmann, S., Rosenfeld, M. R., Gutin, P. H., and Wimmer, E. (2000). Intergeneric poliovirus recombinants for the treatment of malignant glioma. Proc Natl Acad Sci U S A *97*, 6803-6808.

Grundhoff, A., and Ganem, D. (2001). Mechanisms governing expression of the v-FLIP gene of Kaposi's sarcoma-associated herpesvirus. J Virol *75*, 1857-1863.

Gutell, R. R., Lee, J. C., and Cannone, J. J. (2002). The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol *12*, 301-310.

Han, B., Dong, Z., and Zhang, J. T. (2003). Tight control of platelet-derived growth factor B/c-sis expression by interplay between the 5'-untranslated region sequence and the major upstream promoter. J Biol Chem *278*, 46983-46993.

Hart, K., and Bienz, M. (1996). A test for cell autonomy, based on di-cistronic messenger translation. Development *122*, 747-751.

Hatakeyama, Y., Shibuya, N., Nishiyama, T., and Nakashima, N. (2004). Structural variant of the intergenic internal ribosome entry site elements in dicistroviruses and computational search for their counterparts. RNA *10*, 779-786.

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

Henikoff, S., and Henikoff, J. G. (1993). Performance evaluation of amino acid substitution matrices. Proteins *17*, 49-61.

Henis-Korenblit, S., Strumpf, N. L., Goldstaub, D., and Kimchi, A. (2000). A novel form of DAP5 protein accumulates in apoptotic cells as a result of caspase cleavage and internal ribosome entry site-mediated translation. Mol Cell Biol *20*, 496-506.

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

Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., Kroger, A., Schirmbeck, R., Reimann, J., and Hauser, H. (2001). Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. Nucleic Acids Res *29*, 3327-3334.

Heppner Goss, K., Trzepacz, C., Tuohy, T. M., and Groden, J. (2002). Attenuated APC alleles produce functional protein from internal translation initiation. Proc Natl Acad Sci U S A *99*, 8161-8166.

Hernandez, G., Vazquez-Pianzola, P., Sierra, J. M., and Rivera-Pomar, R. (2004). Internal ribosome entry site drives cap-independent translation of reaper and heat shock protein 70 mRNAs in Drosophila embryos. RNA *10*, 1783-1797.

Higgs, P. G. (2000). RNA secondary structure: physical and computational aspects. Q Rev Biophys *33*, 199-253.

Hinnebusch, A. G. (1997). Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. J Biol Chem *272*, 21661-21664.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res *31*, 3429-3431.

Hoffman, M. A., and Palmenberg, A. C. (1995). Mutational analysis of the J-K stem-loop region of the encephalomyocarditis virus IRES. J Virol *69*, 4399-4406.

Holcik, M. (2004). Targeting translation for treatment of cancer--a novel role for IRES? Curr Cancer Drug Targets *4*, 299-311.

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

Holcik, M., and Korneluk, R. G. (2001). XIAP, the guardian angel. Nat Rev Mol Cell Biol *2*, 550-556.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat Rev Mol Cell Biol *6*, 318-327.

Holcik, M., Sonenberg, N., and Korneluk, R. G. (2000a). Internal ribosome initiation of translation and the control of cell death. Trends Genet *16*, 469-473.

Holcik, M., Gordon, B. W., and Korneluk, R. G. (2003). The internal ribosome entry site-mediated translation of antiapoptotic protein XIAP is modulated by the heterogeneous nuclear ribonucleoproteins C1 and C2. Mol Cell Biol *23*, 280-288.

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

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

Holcik, M., Graber, T., Lewis, S. M., Lefebvre, C. A., Lacasse, E., and Baird, S. (2005). Spurious splicing within the XIAP 5' UTR occurs in the Rluc/Fluc but not the betagal/CAT bicistronic reporter system. RNA *11*, 1605-1609.

Honda, M., Beard, M. R., Ping, L. H., and Lemon, S. M. (1999). A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. J Virol *73*, 1165-1174.

Hu, M. C., Tranque, P., Edelman, G. M., and Mauro, V. P. (1999). rRNA-complementarity in the 5' untranslated region of mRNA specifying the Gtx homeodomain protein: evidence that base- pairing to 18S rRNA affects translational efficiency. Proc Natl Acad Sci U S A *96*, 1339-1344.

Hudder, A., and Werner, R. (2000). Analysis of a Charcot-Marie-Tooth disease mutation reveals an essential internal ribosome entry site element in the connexin-32 gene. J Biol Chem *275*, 34586-34591.

Huez, I., Bornes, S., Bresson, D., Creancier, L., and Prats, H. (2001). New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. Mol Endocrinol *15*, 2197-2210.

Hundsdoerfer, P., Thoma, C., and Hentze, M. W. (2005). Eukaryotic translation initiation factor 4GI and p97 promote cellular internal ribosome entry sequence-driven translation. Proc Natl Acad Sci U S A *102*, 13421-13426.

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science *292*, 929-934.

Iizuka, N., Najita, L., Franzusoff, A., and Sarnow, P. (1994). Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. Mol Cell Biol *14*, 7322-7330.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L. E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-617.

Ishikawa, K., Toru, S., Tsunemi, T., Li, M., Kobayashi, K., Yokota, T., Amino, T., Owada, K., Fujigasaki, H., Sakamoto, M., *et al.* (2005). An autosomal dominant cerebellar ataxia linked to chromosome 16q22.1 is associated with a single-nucleotide substitution in the 5'

untranslated region of the gene encoding a protein with spectrin repeat and Rho guanine-nucleotide exchange-factor domains. Am J Hum Genet *77*, 280-296.

Ivanov, P. A., Karpova, O. V., Skulachev, M. V., Tomashevskaya, O. L., Rodionova, N. P., Dorokhov Yu, L., and Atabekov, J. G. (1997). A tobamovirus genome that contains an internal ribosome entry site functional in vitro. Virology *232*, 32-43.

Izquierdo, J. M., and Cuezva, J. M. (2000). Internal-ribosome-entry-site functional activity of the 3'-untranslated region of the mRNA for the beta subunit of mitochondrial H+-ATP synthase. Biochem J *346 Pt 3*, 849-855.

Jaag, H. M., Kawchuk, L., Rohde, W., Fischer, R., Emans, N., and Prufer, D. (2003). An unusual internal ribosomal entry site of inverted symmetry directs expression of a potato leafroll polerovirus replication-associated protein. Proc Natl Acad Sci U S A *100*, 8939-8944.

Jackson, R. J., Hunt, S. L., Reynolds, J. E., and Kaminski, A. (1995). Cap-dependent and cap-independent translation: operational distinctions and mechanistic interpretations. Curr Top Microbiol Immunol *203*, 1-29.

Jacobs, J. L., and Dinman, J. D. (2004). Systematic analysis of bicistronic reporter assay data. Nucleic Acids Res *32*, e160.

Jang, G. M., Leong, L. E., Hoang, L. T., Wang, P. H., Gutman, G. A., and Semler, B. L. (2004). Structurally distinct elements mediate internal ribosome entry within the 5'-noncoding region of a voltage-gated potassium channel mRNA. J Biol Chem *279*, 47419-47430.

Jang, S. K., Krausslich, H. G., Nicklin, M. J., 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*, 2636-2643.

Ji, Y., Xu, X., and Stormo, G. D. (2004). A graph theoretical approach for predicting common RNA secondary structure motifs including pseudoknots in unaligned sequences. Bioinformatics *20*, 1591-1602.

Jimenez, J., Jang, G. M., Semler, B. L., and Waterman, M. L. (2005). An internal ribosome entry site mediates translation of lymphoid enhancer factor-1. RNA *11*, 1385-1399.

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.

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 U S A *96*, 13118-13123.

Johnson, K. N., and Christian, P. D. (1998). The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. J Gen Virol *79 (Pt 1)*, 191-203.

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

Jopling, C. L., Spriggs, K. A., Mitchell, S. A., Stoneley, M., and Willis, A. E. (2004). L-Myc protein synthesis is initiated by internal ribosome entry. RNA *10*, 287-298.

Kaempfer, R. (2003). RNA sensors: novel regulators of gene expression. EMBO Rep *4*, 1043-1047.

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.

Kanamori, Y., and Nakashima, N. (2001). A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. RNA *7*, 266-274.

Ke, A., and Doudna, J. A. (2004). Crystallization of RNA and RNA-protein complexes. Methods *34*, 408-414.

Keene, J. D., and Tenenbaum, S. A. (2002). Eukaryotic mRNPs may represent posttranscriptional operons. Mol Cell *9*, 1161-1167.

Kieft, J. S., Zhou, K., Jubin, R., and Doudna, J. A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. RNA *7*, 194-206.

Kieft, J. S., Zhou, K., Grech, A., Jubin, R., and Doudna, J. A. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. Nat Struct Biol *9*, 370-374.

Kieft, J. S., Zhou, K., Jubin, R., Murray, M. G., Lau, J. Y., and Doudna, J. A. (1999). The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. J Mol Biol *292*, 513-529.

Kim, J. G., Armstrong, R. C., Berndt, J. A., Kim, N. W., and Hudson, L. D. (1998). A secreted DNA-binding protein that is translated through an internal ribosome entry site (IRES) and distributed in a discrete pattern in the central nervous system. Mol Cell Neurosci *12*, 119-140.

Kim, N., Shiffeldrim, N., Gan, H. H., and Schlick, T. (2004). Candidates for novel RNA topologies. J Mol Biol *341*, 1129-1144.

Kim, Y. K., Hahm, B., and Jang, S. K. (2000). Polypyrimidine tract-binding protein inhibits translation of bip mRNA. J Mol Biol *304*, 119-133.

Kjems, J., and Egebjerg, J. (1998). Modern methods for probing RNA structure. Curr Opin Biotechnol *9*, 59-65.

Klein, R. J., and Eddy, S. R. (2003). RSEARCH: Finding homologs of single structured RNA sequences. BMC Bioinformatics *4*, 44.

Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A., and Aboul-Ela, F. (2000). A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. RNA *6*, 1423-1431.

Knudsen, B., and Hein, J. (1999). RNA secondary structure prediction using stochastic context-free grammars and evolutionary history. Bioinformatics *15*, 446-454.

Knudsen, B., and Hein, J. (2003). Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Res *31*, 3423-3428.

Koh, D. C., Liu, D. X., and Wong, S. M. (2002). A six-nucleotide segment within the 3' untranslated region of hibiscus chlorotic ringspot virus plays an essential role in translational enhancement. J Virol *76*, 1144-1153.

Koh, D. C., Wong, S. M., and Liu, D. X. (2003). Synergism of the 3'-untranslated region and an internal ribosome entry site differentially enhances the translation of a plant virus coat protein. J Biol Chem *278*, 20565-20573.

Kolupaeva, V. G., Hellen, C. U., and Shatsky, I. N. (1996). Structural analysis of the interaction of the pyrimidine tract-binding protein with the internal ribosomal entry site of encephalomyocarditis virus and foot-and-mouth disease virus RNAs. RNA *2*, 1199-1212.

Kolupaeva, V. G., Pestova, T. V., and Hellen, C. U. (2000). Ribosomal binding to the internal ribosomal entry site of classical swine fever virus. RNA *6*, 1791-1807.

Komar, A. A., and Hatzoglou, M. (2005). Internal ribosome entry sites in cellular mRNAs: mystery of their existence. J Biol Chem *280*, 23425-23428.

Kozak, M. (2001). New ways of initiating translation in eukaryotes? Mol Cell Biol *21*, 1899-1907.

Kozak, M. (2003). Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation. Gene *318*, 1-23.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13-37.

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.

Laferriere, A., Gautheret, D., and Cedergren, R. (1994). An RNA pattern matching program with enhanced performance and portability. Comput Appl Biosci *10*, 211-212.

Lahlou, H., Fanjul, M., Pradayrol, L., Susini, C., and Pyronnet, S. (2005). Restoration of functional gap junctions through internal ribosome entry site-dependent synthesis of endogenous connexins in density-inhibited cancer cells. Mol Cell Biol *25*, 4034-4045.

Lambert, A., Legendre, M., Fontaine, J. F., and Gautheret, D. (2005). Computing expectation values for RNA motifs using discrete convolutions. BMC Bioinformatics *6*, 118.

Lambert, A., Fontaine, J. F., Legendre, M., Leclerc, F., Permal, E., Major, F., Putzer, H., Delfour, O., Michot, B., and Gautheret, D. (2004). The ERPIN server: an interface to profile-based RNA motif identification. Nucleic Acids Res *32*, W160-165.

Lang, K. J., Kappel, A., and Goodall, G. J. (2002). Hypoxia-inducible factor-1alpha mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. Mol Biol Cell *13*, 1792-1801.

Lauring, A. S., and Overbaugh, J. (2000). Evidence that an IRES within the Notch2 coding region can direct expression of a nuclear form of the protein. Mol Cell *6*, 939-945.

Le Quesne, J. P., Stoneley, M., Fraser, G. A., and Willis, A. E. (2001). Derivation of a structural model for the c-myc IRES. J Mol Biol *310*, 111-126.

Le Sommer, C., Lesimple, M., Mereau, A., Menoret, S., Allo, M. R., and Hardy, S. (2005). PTB regulates the processing of a 3'-terminal exon by repressing both splicing and polyadenylation. Mol Cell Biol *25*, 9595-9607.

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

Le, S. Y., Liu, W. M., and Maizel, J. V., Jr. (1998). Phylogenetic evidence for the improved RNA higher-order structure in internal ribosome entry sequences of HCV and pestiviruses. Virus Genes *17*, 279-295.

Le, S. Y., Chen, J. H., Konings, D., and Maizel, J. V., Jr. (2003). Discovering well-ordered folding patterns in nucleotide sequences. Bioinformatics *19*, 354-361.

Lemon, S. M., and Honda, M. (1997). Internal Ribosome Entry Sites within the RNA genomes of Hepatitis C Virus and other Flavivirues. Seminars in Virology *8*, 274-288.

Lesnik, E. A., Sampath, R., and Ecker, D. J. (2002). Rev response elements (RRE) in lentiviruses: an RNAMotif algorithm-based strategy for RRE prediction. Med Res Rev *22*, 617-636.

Levis, C., and Astier-Manifacier, S. (1993). The 5' untranslated region of PVY RNA, even located in an internal position, enables initiation of translation. Virus Genes *7*, 367-379.

Lewin, B. (2006). updated ed. of Genes VIII edn (Upper Saddle River, NJ, Pearson Education Inc., Prentice Hall).

Lewis, S. M., and Holcik, M. (2005). IRES in distress: translational regulation of the inhibitor of apoptosis proteins XIAP and HIAP2 during cell stress. Cell Death Differ *12*, 547-553.

Ling, J., Morley, S. J., and Traugh, J. A. (2005). Inhibition of cap-dependent translation via phosphorylation of eIF4G by protein kinase Pak2. Embo J *24*, 4094-4105.

Liu, Z., Dong, Z., Han, B., Yang, Y., Liu, Y., and Zhang, J. T. (2005). Regulation of expression by promoters versus internal ribosome entry site in the 5'-untranslated sequence of the human cyclin-dependent kinase inhibitor p27kip1. Nucleic Acids Res *33*, 3763-3771.

Lopez de Quinto, S., and Martinez-Salas, 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.

Lopez de Quinto, S., Saiz, M., de la Morena, D., Sobrino, F., and Martinez-Salas, E. (2002). IRES-driven translation is stimulated separately by the FMDV 3'-NCR and poly(A) sequences. Nucleic Acids Res *30*, 4398-4405.

Lopez-Lastra, M., Gabus, C., and Darlix, J. L. (1997). Characterization of an internal ribosomal entry segment within the 5' leader of avian reticuloendotheliosis virus type A RNA and development of novel MLV-REV-based retroviral vectors. Hum Gene Ther *8*, 1855-1865.

Low, W., Harries, M., Ye, H., Du, M. Q., Boshoff, C., and Collins, M. (2001). Internal ribosome entry site regulates translation of Kaposi's sarcoma-associated herpesvirus FLICE inhibitory protein. J Virol *75*, 2938-2945.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res *25*, 955-964.

Lukavsky, P. J., Kim, I., Otto, G. A., and Puglisi, J. D. (2003). Structure of HCV IRES domain II determined by NMR. Nat Struct Biol *10*, 1033-1038.

Lukavsky, P. J., Otto, G. A., Lancaster, A. M., Sarnow, P., and Puglisi, J. D. (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. Nat Struct Biol *7*, 1105-1110.

Lyons, A. J., and Robertson, H. D. (2003). Detection of tRNA-like structure through RNase P cleavage of viral internal ribosome entry site RNAs near the AUG start triplet. J Biol Chem *278*, 26844-26850.

Lyons, A. J., Lytle, J. R., Gomez, J., and Robertson, H. D. (2001). Hepatitis C virus internal ribosome entry site RNA contains a tertiary structural element in a functional domain of stem-loop II. Nucleic Acids Res *29*, 2535-2541.

Lytle, J. R., Wu, L., and Robertson, H. D. (2002). Domains on the hepatitis C virus internal ribosome entry site for 40s subunit binding. RNA *8*, 1045-1055.

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

MacGregor, G. R., Nolan, G. P., Fiering, S., Roederer, M., and Herzenberg, L. A. (1991). Use of E. coli lacZ (beta -galactosidase) as a reporter gene. In Methods in molecular biology, E. J. Murray, and J. M. Walker, eds. (Clifton, N.J., Humana Press Inc.), pp. 217-235.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res *29*, 4724-4735.

Maier, D., Nagel, A. C., and Preiss, A. (2002). Two isoforms of the Notch antagonist Hairless are produced by differential translation initiation. Proc Natl Acad Sci U S A *99*, 15480-15485.

Marash, L., and Kimchi, A. (2005). DAP5 and IRES-mediated translation during programmed cell death. Cell Death Differ *12*, 554-562.

Martin, M. M., Garcia, J. A., McFarland, J. D., Duffy, A. A., Gregson, J. P., and Elton, T. S. (2003). Translation of the human angiotensin II type 1 receptor mRNA is mediated by a highly efficient internal ribosome entry site. Mol Cell Endocrinol *212*, 51-61.

Martineau, Y., Le Bec, C., Monbrun, L., Allo, V., Chiu, I. M., Danos, O., Moine, H., Prats, H., and Prats, A. C. (2004). Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNAs. Mol Cell Biol *24*, 7622-7635.

Martinez-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.

Martinez-Salas, E., Quinto, S. L., Ramos, R., and Fernandez-Miragall, O. (2002). IRES elements: features of the RNA structure contributing to their activity. Biochimie *84*, 755-763.

Maser, R. S., Zinkel, R., and Petrini, J. H. (2001). An alternative mode of translation permits production of a variant NBS1 protein from the common Nijmegen breakage syndrome allele. Nat Genet *27*, 417-421.

Mathews, D. H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA *10*, 1178-1190.

Mathews, D. H. (2005). Predicting a set of minimal free energy RNA secondary structures common to two sequences. Bioinformatics *21*, 2246-2253.

Mathews, D. H. (2006). Revolutions in RNA secondary structure prediction. J Mol Biol *359*, 526-532.

Mathews, D. H., and Turner, D. H. (2002). Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. J Mol Biol *317*, 191-203.

Mathews, D. H., and Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol *16*, 270-278.

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999a). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol *288*, 911-940.

Mathews, D. H., Burkard, M. E., Freier, S. M., Wyatt, J. R., and Turner, D. H. (1999b). Predicting oligonucleotide affinity to nucleic acid targets. RNA *5*, 1458-1469.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A *101*, 7287-7292.

Mauro, V. P., and Edelman, G. M. (2002). The ribosome filter hypothesis. Proc Natl Acad Sci U S A *99*, 12031-12036.

Mazumder, B., Seshadri, V., and Fox, P. L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci *28*, 91-98.

McBratney, S., and Sarnow, P. (1996). Evidence for involvement of trans-acting factors in selection of the AUG start codon during eukaryotic translational initiation. Mol Cell Biol *16*, 3523-3534.

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers *29*, 1105-1119.

Merrill, M. K., Dobrikova, E. Y., and Gromeier, M. (2006). Cell-Type-Specific Repression of Internal Ribosome Entry Site Activity by Double-Stranded RNA-Binding Protein 76. J Virol *80*, 3147-3156.

Millard, S. S., Vidal, A., Markus, M., and Koff, A. (2000). A U-rich element in the 5' untranslated region is necessary for the translation of p27 mRNA. Mol Cell Biol *20*, 5947-5959.

Miskimins, W. K., Wang, G., Hawkinson, M., and Miskimins, R. (2001). Control of cyclin-dependent kinase inhibitor p27 expression by cap-independent translation. Mol Cell Biol *21*, 4960-4967.

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

Mitchell, S. A., Spriggs, K. A., Coldwell, M. J., Jackson, R. J., and Willis, A. E. (2003). The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. Mol Cell *11*, 757-771.

Mitchell, S. A., Spriggs, K. A., Bushell, M., Evans, J. R., Stoneley, M., Le Quesne, J. P., Spriggs, R. V., and Willis, A. E. (2005). Identification of a motif that mediates polypyrimidine tract-binding protein-dependent internal ribosome entry. Genes Dev *19*, 1556-1571.

Miura, P., Thompson, J., Chakkalakal, J. V., Holcik, M., and Jasmin, B. J. (2005). The utrophin A 5'-untranslated region confers internal ribosome entry site-mediated translational control during regeneration of skeletal muscle fibers. J Biol Chem *280*, 32997-33005.

Mokrejs, M., Vopalensky, V., Kolenaty, O., Masek, T., Feketova, Z., Sekyrova, P., Skaloudova, B., Kriz, V., and Pospisek, M. (2006). IRESite: the database of experimentally verified IRES structures (www.iresite.org). Nucleic Acids Res *34*, D125-130.

Morrish, B. C., and Rumsby, M. G. (2002). The 5' untranslated region of protein kinase Cdelta directs translation by an internal ribosome entry segment that is most active in densely growing cells and during apoptosis. Mol Cell Biol *22*, 6089-6099.

Murthy, V. L., and Rose, G. D. (2003). RNABase: an annotated database of RNA structures. Nucleic Acids Res *31*, 502-504.

Nanbru, C., Prats, A. C., Droogmans, L., Defrance, P., Huez, G., and Kruys, V. (2001). Translation of the human c-myc P0 tricistronic mRNA involves two independent internal ribosome entry sites. Oncogene *20*, 4270-4280.

Negulescu, D., Leong, L. E., Chandy, K. G., Semler, B. L., and Gutman, G. A. (1998). Translation initiation of a cardiac voltage-gated potassium channel by internal ribosome entry. J Biol Chem *273*, 20109-20113.

Nevins, T. A., Harder, Z. M., Korneluk, R. G., and Holcik, M. (2003). Distinct regulation of internal ribosome entry site-mediated translation following cellular stress is mediated by apoptotic fragments of eIF4G translation initiation factor family members eIF4GI and p97/DAP5/NAT1. J Biol Chem *278*, 3572-3579.

Nishiyama, T., Yamamoto, H., Shibuya, N., Hatakeyama, Y., Hachimori, A., Uchiumi, T., and Nakashima, N. (2003). Structural elements in the internal ribosome entry site of Plautia stali intestine virus responsible for binding with ribosomes. Nucleic Acids Res *31*, 2434-2442.

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V., *et al.* (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci U S A *100*, 14229-14234.

Odreman-Macchioli, F., Baralle, F. E., and Buratti, E. (2001). Mutational analysis of the different bulge regions of hepatitis C virus domain II and their influence on internal ribosome entry site translational ability. J Biol Chem *276*, 41648-41655.

Oh, S. K., Scott, M. P., and Sarnow, P. (1992). Homeotic gene Antennapedia mRNA contains 5'-noncoding sequences that confer translational initiation by internal ribosome binding. Genes Dev *6*, 1643-1653.

Ohlmann, T., Lopez-Lastra, M., and Darlix, J. L. (2000). An internal ribosome entry segment promotes translation of the simian immunodeficiency virus genomic RNA. J Biol Chem *275*, 11899-11906.

Oltean, S., and Banerjee, R. (2005). A B12-responsive Internal Ribosome Entry Site (IRES) Element in Human Methionine Synthase. J Biol Chem *280*, 32662-32668.

Oumard, A., Hennecke, M., Hauser, H., and Nourbakhsh, M. (2000). Translation of NRF mRNA is mediated by highly efficient internal ribosome entry. Mol Cell Biol *20*, 2755-2759.

Owens, G. C., Chappell, S. A., Mauro, V. P., and Edelman, G. M. (2001). Identification of two short internal ribosome entry sites selected from libraries of random oligonucleotides. Proc Natl Acad Sci U S A *98*, 1471-1476.

Park, E. H., Lee, J. M., Blais, J. D., Bell, J. C., and Pelletier, J. (2005). Internal translation initiation mediated by the angiogenic factor Tie2. J Biol Chem *280*, 20945-20953.

Pavesi, G., Mauri, G., Stefani, M., and Pesole, G. (2004). RNAProfile: an algorithm for finding conserved secondary structure motifs in unaligned RNA sequences. Nucleic Acids Res *32*, 3258-3269.

Pedersen, S. K., Christiansen, J., Hansen, T. O., Larsen, M. R., and Nielsen, F. C. (2002). Human insulin-like growth factor II leader 2 mediates internal initiation of translation. Biochem J *363*, 37-44.

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.

Peri, S., and Pandey, A. (2001). A reassessment of the translation initiation codon in vertebrates. Trends Genet *17*, 685-687.

Perriquet, O., Touzet, H., and Dauchet, M. (2003). Finding the common structure shared by two homologous RNAs. Bioinformatics *19*, 108-116.

Pervouchine, D. D., Graber, J. H., and Kasif, S. (2003). On the normalization of RNA equilibrium free energy to the length of the sequence. Nucleic Acids Res *31*, e49.

Pesole, G., Liuni, S., Grillo, G., Licciulli, F., Mignone, F., Gissi, C., and Saccone, C. (2002). UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002. Nucleic Acids Res *30*, 335-340.

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

Pickering, B. M., Mitchell, S. A., Spriggs, K. A., Stoneley, M., and Willis, A. E. (2004). Bag-1 internal ribosome entry segment activity is promoted by structural changes mediated by poly(rC) binding protein 1 and recruitment of polypyrimidine tract binding protein 1. Mol Cell Biol *24*, 5595-5605.

Pilipenko, E. V., Blinov, V. M., Chernov, B. K., Dmitrieva, T. M., and Agol, V. I. (1989a). Conservation of the secondary structure elements of the 5'-untranslated region of cardio- and aphthovirus RNAs. Nucleic Acids Res *17*, 5701-5711.

Pilipenko, E. V., Viktorova, E. G., Guest, S. T., Agol, V. I., and Roos, R. P. (2001). Cell-specific proteins regulate viral RNA translation and virus-induced disease. Embo J *20*, 6899-6908.

Pilipenko, E. V., Blinov, V. M., Romanova, L. I., Sinyakov, A. N., Maslova, S. V., and Agol, V. I. (1989b). Conserved structural domains in the 5'-untranslated region of picornaviral genomes: an analysis of the segment controlling translation and neurovirulence. Virology *168*, 201-209.

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.

Pilipenko, E. V., Gmyl, A. P., Maslova, S. V., Belov, G. A., Sinyakov, A. N., Huang, M., Brown, T. D., 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.

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

Pisarev, A. V., Chard, L. S., Kaku, Y., Johns, H. L., Shatsky, I. N., and Belsham, G. J. (2004). Functional and structural similarities between the internal ribosome entry sites of hepatitis C virus and porcine teschovirus, a picornavirus. J Virol *78*, 4487-4497.

Poole, T. L., Wang, C., Popp, R. A., Potgieter, L. N., Siddiqui, A., and Collett, M. S. (1995). Pestivirus translation initiation occurs by internal ribosome entry. Virology *206*, 750-754.

Poyry, T. A., Kaminski, A., and Jackson, R. J. (2004). What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? Genes Dev *18*, 62-75.

Pozner, A., Goldenberg, D., Negreanu, V., Le, S. Y., Elroy-Stein, O., Levanon, D., and Groner, Y. (2000). Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. Mol Cell Biol *20*, 2297-2307.

Proud, C. G. (2005). eIF2 and the control of cell physiology. Semin Cell Dev Biol *16*, 3-12.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2003). NCBI Reference Sequence project: update and current status. Nucleic Acids Res *31*, 34-37.

Puglisi, J. D., Kim, I., Lukavsky, P., Otto, G., Lancaster, A., and Sarnow, P. (2001). Structural and functional investigation of the hepatitis C virus IRES. Nucleic Acids Res Suppl, 263.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). A cell cycle-dependent internal ribosome entry site. Mol Cell *5*, 607-616.

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

Qin, X., and Sarnow, P. (2004). Preferential translation of internal ribosome entry site-containing mRNAs during the mitotic cycle in mammalian cells. J Biol Chem *279*, 13721-13728.

Quick, A. M., and Cipolla, M. J. (2005). Pregnancy-induced up-regulation of aquaporin-4 protein in brain and its role in eclampsia. Faseb J *19*, 170-175.

Rabadan-Diehl, C., Volpi, S., Nikodemova, M., and Aguilera, G. (2003). Translational regulation of the vasopressin v1b receptor involves an internal ribosome entry site. Mol Endocrinol *17*, 1959-1971.

Ray, P. S., and Das, S. (2004). Inhibition of hepatitis C virus IRES-mediated translation by small RNAs analogous to stem-loop structures of the 5'-untranslated region. Nucleic Acids Res *32*, 1678-1687.

Rhee, S., Yang, S. J., Lee, S. J., and Park, D. (2004). betaPix-b(L), a novel isoform of betaPix, is generated by alternative translation. Biochem Biophys Res Commun *318*, 415-421.

Rijnbrand, R., Abell, G., and Lemon, S. M. (2000). Mutational analysis of the GB virus B internal ribosome entry site. J Virol *74*, 773-783.

Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J., 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.

Rijnbrand, R., Thiviyanathan, V., Kaluarachchi, K., Lemon, S. M., and Gorenstein, D. G. (2004). Mutational and structural analysis of stem-loop IIIC of the hepatitis C virus and GB virus B internal ribosome entry sites. J Mol Biol *343*, 805-817.

Rivas, E., and Eddy, S. R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol *285*, 2053-2068.

Rivas, E., and Eddy, S. R. (2000). Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. Bioinformatics *16*, 583-605.

Rivera, V. M., Welsh, J. D., and Maizel, J. V., Jr. (1988). Comparative sequence analysis of the 5' noncoding region of the enteroviruses and rhinoviruses. Virology *165*, 42-50.

Rubtsova, M. P., Sizova, D. V., Dmitriev, S. E., Ivanov, D. S., Prassolov, V. S., and Shatsky, I. N. (2003). Distinctive properties of the 5'-untranslated region of human hsp70 mRNA. J Biol Chem *278*, 22350-22356.

Sanderbrand, S. A., Tautz, N., Thiel, H. J., Ochs, K., Beck, E., and Niepmann, M. (2000). Translation from the internal ribosome entry site of bovine viral diarrhea virus is independent of the interaction with polypyrimidine tract-binding protein. Vet Microbiol *77*, 215-227.

Sankoff, D. (1985). Simultaneous solution of the RNA folding alignment and protoseqauence problems. SIAM J Appl Math *45*, 810-825.

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.

Schiavi, A., Hudder, A., and Werner, R. (1999). Connexin43 mRNA contains a functional internal ribosome entry site. FEBS Lett *464*, 118-122.

Sehgal, A., Briggs, J., Rinehart-Kim, J., Basso, J., and Bos, T. J. (2000). The chicken c-Jun 5' untranslated region directs translation by internal initiation. Oncogene *19*, 2836-2845.

Sella, O., Gerlitz, G., Le, S. Y., and Elroy-Stein, 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.

Serikawa, K. A., Xu, X. L., MacKay, V. L., Law, G. L., Zong, Q., Zhao, L. P., Bumgarner, R., and Morris, D. R. (2003). The Transcriptome and Its Translation during Recovery from Cell Cycle Arrest in Saccharomyces cerevisiae. Mol Cell Proteomics *2*, 191-204.

Seroussi, E., Shani, N., Ben-Meir, D., Chajut, A., Divinski, I., Faier, S., Gery, S., Karby, S., Kariv-Inbal, Z., Sella, O., *et al.* (2001). Uniquely conserved non-translated regions are involved in generation of the two major transcripts of protein phosphatase 2Cbeta. J Mol Biol *312*, 439-451.

Shapiro, B. A., Bengali, D., Kasprzak, W., and Wu, J. C. (2001). RNA folding pathway functional intermediates: their prediction and analysis. J Mol Biol *312*, 27-44.

Sherrill, K. W., Byrd, M. P., Van Eden, M. E., and Lloyd, R. E. (2004). BCL-2 translation is mediated via internal ribosome entry during cell stress. J Biol Chem *279*, 29066-29074.

Shi, Y., Sharma, A., Wu, H., Lichtenstein, A., and Gera, J. (2005). Cyclin D1 and c-myc internal ribosome entry site (IRES)-dependent translation is regulated by AKT activity and enhanced by rapamycin through a p38 MAPK- and ERK-dependent pathway. J Biol Chem *280*, 10964-10973.

Shimazaki, T., Honda, M., Kaneko, S., and Kobayashi, K. (2002). Inhibition of internal ribosomal entry site-directed translation of HCV by recombinant IFN-alpha correlates with a reduced La protein. Hepatology *35*, 199-208.

Shiroki, K., Ohsawa, C., Sugi, N., Wakiyama, M., Miura, K., Watanabe, M., Suzuki, Y., and Sugano, S. (2002). Internal ribosome entry site-mediated translation of Smad5 in vivo: requirement for a nuclear event. Nucleic Acids Res *30*, 2851-2861.

Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science *310*, 1513-1515.

Sonenberg, N., and Dever, T. E. (2003). Eukaryotic translation initiation factors and regulators. Curr Opin Struct Biol *13*, 56-63.

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

Spellman, R., and Smith, C. W. (2006). Novel modes of splicing repression by PTB. Trends Biochem Sci *31*, 73-76.

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., *et al.* (2002). The Bioperl toolkit: Perl modules for the life sciences. Genome Res *12*, 1611-1618.

Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, R. D. (1998). How cells respond to interferons. Annu Rev Biochem *67*, 227-264.

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

Stoneley, M., and Willis, A. E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene *23*, 3200-3207.

Stoneley, M., Spencer, J. P., and Wright, S. C. (2001). An internal ribosome entry segment in the 5' untranslated region of the mnt gene. Oncogene *20*, 893-897.

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

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

Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA *9*, 644-647.

Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., *et al.* (2000). Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. Genomics *64*, 286-297.

Svitkin, Y. V., Gradi, A., Imataka, H., Morino, S., and Sonenberg, N. (1999). Eukaryotic initiation factor 4GII (eIF4GII), but not eIF4GI, cleavage correlates with inhibition of host cell protein synthesis after human rhinovirus infection. J Virol *73*, 3467-3472.

Tatusova, T. A., and Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett *174*, 247-250.

Teerink, H., Voorma, H. O., and Thomas, A. A. (1995). The human insulin-like growth factor II leader 1 contains an internal ribosomal entry site. Biochim Biophys Acta *1264*, 403-408.

Tenenbaum, S. A., Lager, P. J., Carson, C. C., and Keene, J. D. (2002). Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. Methods *26*, 191-198.

Tenenbaum, S. A., Carson, C. C., Atasoy, U., and Keene, J. D. (2003). Genome-wide regulatory analysis using en masse nuclear run-ons and ribonomic profiling with autoimmune sera. Gene *317*, 79-87.

Thomas, A. A., ter Haar, E., Wellink, J., and Voorma, H. O. (1991). Cowpea mosaic virus middle component RNA contains a sequence that allows internal binding of ribosomes and that requires eukaryotic initiation factor 4F for optimal translation. J Virol *65*, 2953-2959.

Thompson, S. R., and Sarnow, P. (2003). Enterovirus 71 contains a type I IRES element that functions when eukaryotic initiation factor eIF4G is cleaved. Virology *315*, 259-266.

Thurner, C., Witwer, C., Hofacker, I. L., and Stadler, P. F. (2004). Conserved RNA secondary structures in Flaviviridae genomes. J Gen Virol *85*, 1113-1124.

Tinton, S. A., Schepens, B., Bruynooghe, Y., Beyaert, R., and Cornelis, S. (2005). Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: roles

of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2alpha. Biochem J *385*, 155-163.

Tsui, V., Macke, T., and Case, D. A. (2003). A novel method for finding tRNA genes. RNA *9*, 507-517.

Tsukiyama-Kohara, K., Iizuka, N., Kohara, M., and Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. J Virol *66*, 1476-1483.

Ungureanu, N. H., Cloutier, M., Lewis, S. M., de Silva, N., Blais, J. D., Bell, J. C., and Holcik, M. (2006). Internal ribosome entry site-mediated translation of Apaf-1, but not XIAP, is regulated during UV-induced cell death. J Biol Chem *281*, 15155-15163.

Vagner, S., Waysbort, A., Marenda, M., Gensac, M. C., Amalric, F., and Prats, A. C. (1995a). Alternative translation initiation of the Moloney murine leukemia virus mRNA controlled by internal ribosome entry involving the p57/PTB splicing factor. J Biol Chem *270*, 20376-20383.

Vagner, S., Gensac, M. C., Maret, A., Bayard, F., Amalric, F., Prats, H., and Prats, A. C. (1995b). Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. Mol Cell Biol *15*, 35-44.

Vagner, S., Touriol, C., Galy, B., Audigier, S., Gensac, M. C., Amalric, F., Bayard, F., Prats, H., and Prats, A. C. (1996). Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. J Cell Biol *135*, 1391-1402.

Van Eden, M. E., Byrd, M. P., Sherrill, K. W., and Lloyd, R. E. (2004). Translation of cellular inhibitor of apoptosis protein 1 (c-IAP1) mRNA is IRES mediated and regulated during cell stress. RNA *10*, 469-481.

Vattem, K. M., and Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A *101*, 11269-11274.

Venkatesan, A., and Dasgupta, A. (2001). Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements. Mol Cell Biol *21*, 2826-2837.

Verrier, S. B., and Jean-Jean, O. (2000). Complementarity between the mRNA 5' untranslated region and 18S ribosomal RNA can inhibit translation. RNA *6*, 584-597.

Wang, C., 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.

Wang, X. Q., and Rothnagel, J. A. (2004). 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res *32*, 1382-1391.

Wang, Z., Weaver, M., and Magnuson, N. S. (2005). Cryptic promoter activity in the DNA sequence corresponding to the pim-1 5'-UTR. Nucleic Acids Res *33*, 2248-2258.

Warnakulasuriyarachchi, D., Cerquozzi, S., Cheung, H. H., and Holcik, M. (2004). Translational induction of the inhibitor of apoptosis protein HIAP2 during endoplasmic reticulum stress attenuates cell death and is mediated via an inducible internal ribosome entry site element. J Biol Chem *279*, 17148-17157.

Watada, H., Mirmira, R. G., Leung, J., and German, M. S. (2000). Transcriptional and translational regulation of beta-cell differentiation factor Nkx6.1. J Biol Chem *275*, 34224-34230.

Wilson, J. E., Powell, M. J., Hoover, S. E., and Sarnow, P. (2000a). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. Mol Cell Biol *20*, 4990-4999.

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

Woese, C. R. (2001). Translation: in retrospect and prospect. RNA *7*, 1055-1067.

Wollerton, M. C., Gooding, C., Robinson, F., Brown, E. C., Jackson, R. J., and Smith, C. W. (2001). Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). RNA *7*, 819-832.

Workman, C., and Krogh, A. (1999). No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. Nucleic Acids Res *27*, 4816-4822.

Wu, S., Hu, Y., Wang, J. L., Chatterjee, M., Shi, Y., and Kaufman, R. J. (2002). Ultraviolet light inhibits translation through activation of the unfolded protein response kinase PERK in the lumen of the endoplasmic reticulum. J Biol Chem *277*, 18077-18083.

Xiao, Z. S., Simpson, L. G., and Quarles, L. D. (2003). IRES-dependent translational control of Cbfa1/Runx2 expression. J Cell Biochem *88*, 493-505.

Yaman, I., Fernandez, J., Liu, H., Caprara, M., Komar, A. A., Koromilas, A. E., Zhou, L., Snider, M. D., Scheuner, D., Kaufman, R. J., and Hatzoglou, M. (2003). The zipper model of translational control: a small upstream ORF is the switch that controls structural remodeling of an mRNA leader. Cell *113*, 519-531.

Yang, D., Wilson, J. E., Anderson, D. R., Bohunek, L., Cordeiro, C., Kandolf, R., and McManus, B. M. (1997). In vitro mutational and inhibitory analysis of the cis-acting translational elements within the 5' untranslated region of coxsackievirus B3: potential targets for antiviral action of antisense oligomers. Virology *228*, 63-73.

Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003). Tools for the automatic identification and classification of RNA base pairs. Nucleic Acids Res *31*, 3450-3460.

Yoon, A., Peng, G., Brandenburg, Y., Zollo, O., Xu, W., Rego, E., and Ruggero, D. (2006). Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. Science *312*, 902-906.

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

Zaug, A. J., and Cech, T. R. (1995). Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. RNA *1*, 363-374.

Zhang, X., Richie, C., and Legerski, R. J. (2002). Translation of hSNM1 is mediated by an internal ribosome entry site that upregulates expression during mitosis. DNA Repair (Amst) *1*, 379-390.

Zhao, W. D., and Wimmer, E. (2001). Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. J Virol *75*, 3719-3730.

Zhou, W., Edelman, G. M., and Mauro, V. P. (2001). Transcript leader regions of two Saccharomyces cerevisiae mRNAs contain internal ribosome entry sites that function in living cells. Proc Natl Acad Sci U S A *98*, 1531-1536.

Zhou, W., Edelman, G. M., and Mauro, V. P. (2003). Isolation and identification of short nucleotide sequences that affect translation initiation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *100*, 4457-4462.

Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999). Messenger RNA translation state: the second dimension of high-throughput expression screening. Proc Natl Acad Sci U S A *96*, 10632-10636.

Zuker, M. (2000). Calculating nucleic acid secondary structure. Curr Opin Struct Biol *10*, 303-310.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res *9*, 133-148.

Zuleeg, T., Hansen, A., Pfeiffer, T., Schubel, H., Kreutzer, R., Hartmann, R. K., and Limmer, S. (2001). Correlation between processing efficiency for ribonuclease P minimal substrates and conformation of the nucleotide -1 at the cleavage position. Biochemistry *40*, 3363-3369.

Appendix I

RNase sites and MFOLD constraints

Table 4 RNase sensitive sites for the XIAP IRES structure.
Position is given from either the 5' of the UTR sequence or relative to the natural AUG start site. Specific RNase cut sites for RNase T1, V1 and T2 are given with 1=strong sensitivity to 5=no RNase cut.

num	5'	num	T1	V1	T2
1	T	-170			
2	A	-169			
3	T	-168			
4	T	-167			
5	T	-166			
6	A	-165			
7	G	-164			
8	A	-163			
9	A	-162			
10	T	-161			
11	T	-160			1
12	A	-159			
13	G	-158	5		1
14	A	-157			
15	A	-156			
16	T	-155			
17	G	-154	1		
18	T	-153			
19	T	-152			
20	T	-151			
21	C	-150			
22	T	-149			
23	T	-148			
24	A	-147			
25	G	-146	1	2	2
26	C	-145		2	2
27	G	-144	2	2	2
28	G	-143	1	2	2
29	T	-142		2	2
30	C	-141		2	2
31	G	-140	1		
32	T	-139			
33	G	-138	5		
34	T	-137			
35	A	-136			
36	G	-135	2		
37	T	-134			
38	T	-133			
39	A	-132			
40	T	-131			
41	T	-130			
42	T	-129			
43	T	-128			
44	T	-127			
45	A	-126			
46	T	-125			2
47	G	-124	5		2
48	T	-123			4
49	C	-122		2	2
50	A	-121		1	4
51	T	-120		4	4
52	A	-119			1
53	A	-118			1
54	G	-117	5	5	5
55	T	-116		5	4
56	G	-115	5	5	5
57	G	-114	2	5	5
58	A	-113			4
59	T	-112		2	4
60	A	-111			3
61	A	-110			3
62	T	-109			3
63	T	-108		1	3
64	T	-107		1	3
65	G	-106	5	2	
66	T	-105		2	
67	T	-104			
68	A	-103			2
69	G	-102	2		2
70	C	-101			1
71	T	-100			1
72	C	-99			1
73	C	-98			1
74	T	-97			
75	A	-96			
76	T	-95			
77	A	-94			2
78	A	-93			1
79	C	-92			2
80	A	-91			
81	A	-90			
82	A	-89			
83	A	-88			
84	G	-87	5		2
85	T	-86			
86	C	-85		1	
87	T	-84		1	
88	G	-83	5		
89	T	-82			
90	T	-81			
91	G	-80	5		
92	C	-79			
93	T	-78		1	
94	T	-77		2	
95	G	-76	5		
96	T	-75			
97	G	-74	1		
98	T	-73			
99	T	-72			
100	T	-71			
101	C	-70			
102	A	-69			
103	C	-68			
104	A	-67			
105	T	-66			
106	T	-65			
107	T	-64			
108	T	-63			
109	G	-62	1		
110	G	-61	1		
111	A	-60			
112	T	-59			
113	T	-58			
114	T	-57			
115	C	-56		2	2
116	C	-55		1	
117	T	-54		1	
118	A	-53		1	
119	A	-52			1
120	T	-51			1
121	A	-50			1
122	T	-49			1
123	A	-48			1
124	A	-47			1
125	T	-46			1
126	G	-45	2		1
127	T	-44			
128	T	-43			
129	C	-42		2	
130	T	-41		2	
131	C	-40		1	
132	T	-39		1	
133	T	-38		2	
134	T	-37			1
135	T	-36			1

136	T	-35			1
137	a	-34			1
138	G	-33	1		1
139	A	-32			1
140	A	-31			
141	A	-30			
142	A	-29			
143	G	-28	5		
144	G	-27	5		

145	T	-26			
146	G	-25	1		
147	G	-24	1		
148	A	-23			
149	C	-22			
150	A	-21			
151	A	-20			
152	G	-19	1		
153	T	-18			

154	C	-17			
155	C	-16			
156	T	-15			
157	A	-14			
158	T	-13			2
159	T	-12			2
160	T	-11		2	2
161	T	-10			2
162	C	-9		1	

163	c	-8		1	
164	A	-7		1	
165	G	-6	5	2	
166	A	-5			
167	G	-4	5		2
168	A	-3			
169	A	-2			
170	G	-1	5		

Table 5. MFOLD constraints for XIAP IRES structure prediction.
Constraints used to aid in predicting the RNA structure. P =prohibit basepairing, f = force basepairing, 1st digit is the sequence position, 3 digit is the number of nucleotides to continue the constraint.

MFOLD Constraints
p 13 0 1
p 17 0 1
p 25 0 1
p 28 0 1
p 31 0 1
f 50 0 1
p 68 0 6
f 86 0 2
p 109 0 2
p 119 0 6
p 134 0 6
p 146 0 2
p 152 0 1

Table 6. RNase sensitive sites for the AQP4 IRES structure

Position is given from the 5' of the UTR The AUG and CAT coding sequence is shown in gray. Specific RNase cut sites for RNase T1, V1, T2 and A are given with 1=strong sensitivity to 5=no RNase cut

num	5'	T1	V1	T2	A
1	a				
2	a				
3	g				
4	g				
5	a				
6	c				
7	a				
8	g		1	5	
9	t		2	3	
10	t		2	3	
11	t		1	3	
12	g		2	4	
13	g		4	4	
14	a		4	2	
15	t		4	2	
16	a		2	1	2
17	a		4	1	
18	t		5	1	
19	t		3	3	
20	t		4	4	
21	t		5	5	
22	c		1	5	
23	c		1	5	
24	g	3	4	5	
25	t	4	3	3	1
26	g	1	5	4	3
27	a	5	5	4	5
28	a	5	5	5	5
29	a	5	5	5	5
30	c	5	4	5	5
31	t	5	1	5	5
32	g	5	1	5	5
33	g	4	4	5	5
34	a	5	5	3	5
35	t	5	4	3	1
36	g	1	4	4	4
37	a	4	5	4	5
38	t	5	5	4	4
39	c	5	1	4	5
40	t	5	1	4	5
41	t	5	1	4	5
42	t	5	1	4	3
43	g	3	1	3	5
44	a	5	4	4	5
45	a	5	5	4	5
46	t	5	1	3	1
47	a	5	4	3	2
48	t	5	4	3	3
49	t	5	3	3	2
50	a	5	5	3	4
51	a	5	5	3	5
52	a	5	5	3	5
53	t	5	5	3	5
54	g	1	4	4	5
55	a	4			5
56	a	5			5
57	a	5			5
58	g	5			5
59	g	5			5
60	a	5			5
61	g	5	1	5	5
62	a	5	1	5	5
63	c	5		5	4
64	t	5	1	5	5
65	a	5	1	2	5
66	g	1	3	1	5
67	a	4	5	1	1
68	a	5	5	1	1
69	t	5	5	2	1
70	a	5	5	1	
71	a	5	5	5	
72	c	5	5	5	
73	a	5	5	5	
74	g	5	1	5	
75	t	5	5	5	
76	c	5	1	5	
77	t	5	5	5	
78	c	5	1	5	
79	t	5	5	5	
80	t	5	5	5	
81	g	1	5	5	
82	a	5			
83	c	5			
84	t	5			
85	a	5			
86	t	5			
87	t	5			
88	c	5			
89	a	5			
90	t	5			
91	g	3			
92	a	5			
93	a	5			
94	g	4			
95	g	4			
96	a	5			
97	g	4			
98	c	5			
99	t	5			
100	t	5		4	
101	t	5		2	
102	a	5		1	
103	g	2		3	
104	c	4			
105	a	5			
106	g	5			
107	g	5			
108	a	5			
109	g	5			
110	c				
111	a			2	
112	t			4	
113	t			3	
114	c			3	
115	t			3	
116	t		4		
117	t				
118	c				
119	t				
120	t				
121	g				
122	g				
123	t				
124	g				
125	t				
126	g		4		3
127	c		5	1	2
128	a		5		
129	a		5		
130	a		5		
131	t		5		
132	c		5		
133	a		5		
134	c		5		
135	g		5		
136	c				
137	c				
138	t				
139	g				
140	g				
141	c				
142	a				
143	g				
144	c			2	
145	c			2	
146	c				
147	c				
148	t				
149	g			2	
150	c				
151	c				
152	c				
153	a				
154	c				
155	a				
156	g				
157	t				
158	a				
159	c				
160	c				
161	c				
162	g				
163	a				
164	a				
165	g				
166	a				
167	t				
168	g				
169	g				
170	c				
171	a				
172	g				
173	g				
174	c				
175	c				
176	t				
177	t				
178	g				
179	t				
180	t				
181	c				
182	c				
183	c				
184	t				
185	t				
186	c				
187	a				
188	c				
189	c				

190	t				
191	a				
192	a				
193	a			1	
194	t			3	3
195	t			2	2
196	c			1	1
197	a			1	1
198	t			3	2
199	a			1	
200	a			3	
201	a			4	
202	c				
203	c				
204	t				
205	g				
206	g				
207	g				
208	t				
209	g				
210	t				
211	a				
212	g				
213	t				
214	g			4	3
215	g				
216	c				
217	t				
218	t		4	3	2
219	c				
220	t				

221	g				
222	a				
223	t				
224	g				
225	c				
226	t				
227	g	5			
228	a				
229	t				
230	t				
231	t				
232	g				
233	t				
234	t				
235	t				
236	c				
237	t				
238	c				
239	t				
240	t			3	
241	t				
242	t				
243	c				
244	a				1
245	g				
246	t				1
247	a				
248	a				
249	g	2			
250	t				
251	g	2			

252	t				
253	g	1			
254	g	1			
255	a				
256	c		3	4	
257	c		3	2	
258	t			4	
259	t			2	
260	t			3	
261	g	1		2	
262	t			1	3
263	g	1			
264	t				1
265	a				
266	c				
267	c			3	
268	a			3	
269	g	2		4	
270	a			3	
271	g	2			
272	a				
273	g	5			
274	a				
275	a				
276	c				
277	a				
278	t				
279	c				
280	a				
281	t				
282	g	1			
283	g	4			

284	t				
285	g	1			
286	g	1			
287	c				
288	t				
289	c				
290	g	1			
291	a				
292	g	1			
293	a				
294	t				
295	g	1			
296	g	1		1	
297	a			1	
298	g			1	
299	a			1	
300	a			1	
301	a			1	
302	a			1	
303	a				
304	a				
305	a				
306	t		2		
307	c		2		
308	a				
309	c				
310	t				
311	g				1
312	g				1
313	a		2		
314	t		1		4
315	a		2		

316	t		1		3
317	a				
318	c				
319	c				
320	a				
321	c			1	
322	c			1	
323	g	1			
324	t				
325	t				
326	a	1			
327	a				
328	t				
329	a				
330	t				
331	a				1
332	t				
333	c		1		
334	c		1		
335	c				
336	a				
337	a				
338	t				
339	g		1		
340	g		1		
341	c		1		
342	a		1		
343	t				

Table 7. MFOLD constraints for AQP4 structure.
Constraints used to aid in predicting the RNA structure. P =prohibit basepairing, F = force basepairing, 1st digit is the sequence position, 3rd digit is the number of nucleotides to continue the constraint.

F 8 0 5
P 14 0 5
F 22 0 2
P 25 0 2
F 31 0 2
P 35 0 2
F 39 0 5
P 47 0 8
F 61 0 2
F 64 0 1
P 66 0 5
F 74 0 1
F 76 0 1
P 102 0 2
P 111 0 1
P 127 0 1
P 144 0 2
P 149 0 1
P 193 0 1
P 195 0 5
P 218 0 1
P 244 0 1

P 246 0 1
P 249 0 1
P 251 0 1
P 253 0 2
P 257 0 1
P 259 0 1
P 261 0 4
P 268 0 1
P 270 0 1
P 282 0 1
P 285 0 2
P 290 0 1
P 292 0 1
P 295 0 8
F 306 0 2
P 311 0 2
F 313 0 4
P 321 0 3
P 326 0 1
P 331 0 1
F 333 0 2
F 339 0 4

Table 8. RNase sensitive sites for the ELG1 IRES structure.
Position is given from the 5' end of the UTR The AUG and CAT coding sequence is shown in gray. Specific RNase cut sites for RNase T1, V1, T2 and A are given with 1=strong to 5= no cut.

num	5'	T1	V1	T2	A
1	a				
2	c				
3	t				
4	t				
5	t				
6	t				
7	g		3	5	
8	g		1	5	
9	t		2	5	
10	g	1	3	5	
11	g		2	5	
12	g		5	5	
13	c		5	1	
14	a		4	1	
15	t		4	1	
16	t		4	1	
17	t		4	1	
18	a		4	1	
19	a		5	5	
20	a			5	
21	a			5	
22	a		1	5	
23	t		3		
24	g				
25	t				
26	g				
27	t				
28	g	4			
29	t				
30	g				
31	t		4	5	
32	a		4	4	
33	t			4	
34	g	1			
35	t				
36	g				
37	t				
38	a				
39	t		1		
40	a		1		
41	t		1		
42	a		1		
43	t		1	5	
44	g	3	4	5	
45	t		5	3	
46	a		2		
47	t		3	5	
48	g	3	5	5	
49	t		5	1	
50	g	1	4	1	
51	t			1	
52	a		5	5	
53	t		5	5	
54	g	4	3		
55	t		3		
56	a			3	
57	t		3	3	
58	g	3	5	4	
59	t		5	1	
60	g	1	5	1	
61	t				
62	a				
63	t				
64	a				
65	t				
66	a				
67	t				
68	g	4			
69	t				
70	a				
71	t				
72	a				
73	t				
74	g	5			
75	t				
76	a				
77	t				
78	g	5			
79	t				
80	a				
81	t				
82	g	4	3		
83	t				
84	a				
85	t				
86	c				
87	g	3			
88	c				
89	g	2			
90	t				
91	g	2			
92	t				
93	a				
94	t				
95	g				
96	t				
97	g				
98	t				
99	g				
100	t				
101	a				
102	t				
103	g				
104	t				
105	a				
106	t				
107	g		1		
108	c		3		
109	a				
110	t				
111	g	1		3	
112	t			1	
113	g	1		3	
114	t				
115	a				
116	t		4		
117	g		1		
118	t		4	4	
119	a			1	
120	t			1	
121	g	1		2	
122	t			2	
123	a			1	
124	t				
125	a				
126	t				
127	g		3		
128	c				
129	a				
130	t				
131	g				
132	t			3	
133	a			1	
134	t			1	
135	g			3	
136	t			1	
137	g				
138	t				
139	a				
140	t			4	
141	g	1		2	
142	t				
143	g	1			
144	t				
145	a				
146	t				
147	a				
148	t				
149	a				
150	t				
151	g				
152	t				
153	a				
154	t				
155	g				
156	t				
157	g				
158	t				
159	g				
160	t				
161	g				
162	t				
163	a				
164	t				
165	g				
166	t				
167	a				
168	t				
169	a				
170	t				
171	g				
172	t				
173	g				
174	t				
175	g				
176	t				
177	g				
178	t				
179	a				
180	t				
181	g				
182	t				
183	g				
184	t				
185	a				
186	t				
187	g				
188	t				
189	g				

190	t				
191	g				
192	t				
193	g				
194	t				
195	g				
196	t				
197	a				
198	t				
199	g				
200	t				
201	g				
202	t				
203	g				
204	t				
205	g				
206	t				
207	g	1			
208	t				
209	a				
210	t				
211	g				
212	t				
213	a				2
214	t				
215	g	1		3	
216	t				
217	a				
218	t				3
219	g	1		3	
220	t				
221	a			4	2
222	t				
223	g	4			
224	t				
225	a				
226	t				
227	a				
228	t				
229	g				
230	t				
231	a				
232	t				
233	t				

234	a				
235	t				
236	a				
237	c				
238	a				
239	c				
240	a				
241	t				
242	a				
243	t				
244	a		4		
245	c		2		
246	a		2		
247	c		2		
248	a		3		
249	t		4		
250	a		4		
251	t		4		
252	t		4		
253	g	4	3		
254	g	4	4		
255	t		3		
256	t		4		
257	t		4		
258	t		4	4	
259	t		4	5	
260	t		5		
261	t		5		
262	a		3		
263	a		1		
264	t		1		
265	c		3		
266	a		3		
267	t		2	3	
268	t			3	
269	t			2	
270	g	2		2	
271	a			2	
272	g	1		2	
273	a			3	
274	g	1		2	
275	t			2	2
276	t			2	1
277	a				

278	g	1			
279	t				
280	t				
281	g	2			
282	a				
283	a				
284	g	1		3	
285	a			4	
286	t			3	1
287	a				
288	a				
289	a				
290	a				
291	a				
292	c				
293	c		4		
294	c		3		
295	a		4		
296	t		3		
297	c				
298	a				
299	c		2		
300	c		2		
301	c		2		
302	c		4		
303	t		4		
304	a		3		
305	a		3		
306	a		3		
307	t		4		
308	g	2			
309	t				4
310	a				3
311	t				
312	t				
313	c				3
314	c			3	2
315	a			3	
316	a			3	
317	a				
318	g	1			
319	a				
320	a				
321	t				3

322	a				2
323	a				
324	g	1			
325	a				
326	a				
327	c				
328	a				3
329	t				
330	t				
331	g	1			
332	t				
333	t				
334	t				
335	t				
336	a				
337	t				
338	a				
339	c				
340	a				
341	t				
342	a				
343	g				
344	c				
345	a		3		
346	c				
347	a		3		
348	c		3		
349	t				
350	t				4
351	a				3
352	a				4
353	c				3
354	a				
355	a				
356	a				
357	a				
358	t				
359	c				
360	a				2
361	a				
362	g	1			
363	a				
364	a				
365	a				

366	t				
367	t				2
368	t				2
369	a				
370	a				
371	c	4			2
372	a				
373	t				2
374	t				2
375	a				
376	a				
377	t				
378	a				
379	c				
380	a				
381	g	1			
382	t				
383	a		3		
384	c		2		
385	t		2		
386	g		2		
387	t				
388	t				
389	a				
390	c				
391	c				
392	t				
393	a				
394	a				
395	t				
396	c				
397	c				
398	g	1			
399	t				1
400	a				
401	g	1			
402	t				
403	c				
404	g	2			
405	a				
406	t				
407	t				
408	t				
409	t				2

410	c				
411	a				
412	a				
413	a				
414	t				
415	t				
416	t				
417	t		1		
418	g		1		
419	t				
420	c	1			
421	a				
422	g	1			
423	t				
424	t				
425	g	1			
426	t				
427	t				
428	c		3		
429	c		3		
430	a		3		
431	a				

432	t				
433	a				
434	a		3		
435	t		3		
436	g		4		
437	t		3		
438	c		1		
439	c		1		
440	t		1		
441	t		3		
442	t		3	4	
443	a			2	
444	t			2	3
445	a			2	3
446	t			2	3
447	a				
448	t		3	3	
449	t		4		
450	c				
451	c		3		
452	c		3		
453	c		3		

454	g				
455	c				
456	c				
457	c				
458	a				
459	g	3			
460	c				
461	t				
462	c				
463	g				
464	a				
465	g	1			
466	a				
467	t				
468	g	1			
469	g	1		2	
470	a			2	
471	g				
472	a			2	
473	a			2	
474	a				
475	a			2	

476	a				
477	a				
478	a				
479	t		2		
480	c				1
481	a				
482	c				
483	t				
484	g				1
485	g				1
486	a		2		
487	t		1		
488	a		2		
489	t		1		
490	a				
491	c				
492	c				
493	a				
494	c			1	
495	c			1	
496	g	1			
497	t				

498	t				
499	g	1			
500	a				
501	t				
502	a				
503	t				
504	a				1
505	t				
506	c				
507	c				
508	c				
509	a				
510	a				
511	t				
512	g	1			
513	g	1			
514	c				
515	a				
516	t				

Table 9. MFOLD constraints forELG1 structure.

Constraints used to aid in predicting the RNA structure. P =prohibit basepairing, f = force basepairing, 1st digit is the sequence position, 3rd digit is the number of nucleotides to continue the constraint.

F 8 0 2
P 10 0 1
F 11 0 1
P 13 0 6
F 22 0 1
P 34 0 1
F 39 0 5
F 46 0 1
F 49 0 3
P 59 0 1
P 89 0 1
P 91 0 1
F 107 0 1
P 111 0 3
F 117 0 1
P 119 0 5
P 133 0 2
P 136 0 1
P 141 0 1
P 143 0 1
P 207 0 1
P 213 0 1
P 215 0 1
P 219 0 1

P 222 0 1
F 245 0 3
F 263 0 1
P 267 0 1
P 269 0 4
P 274 0 3
P 278 0 1
P 281 0 1
P 284 0 1
P 286 0 1
P 299 0 3
P 308 0 1
P 314 0 1
P 318 0 1
P 322 0 1
P 324 0 1
P 331 0 1
P 360 0 1
P 362 0 1
P 367 0 2
P 371 0 1
P 373 0 1
P 381 0 1
F 384 0 3

P 398 0 2
P 401 0 1
P 404 0 1
P 409 0 2
F 417 0 2
P 420 0 1
P 422 0 1
P 425 0 1
F 438 0 1
P 443 0 4
P 465 0 1
P 468 0 3
P 472 0 2
P 475 0 1
F 479 0 1
P 480 0 1
P 484 0 2
F 486 0 4
P 494 0 3
P 499 0 1
P 504 0 1
P 512 0 2

Appendix II

Primers for RT-PCR of random UTRs

Table 10. UTR label, UTRdb number, and description of random UTRs

utr label	utrdb#, description
Pap4	5HSA090623 PAP associated domain containing 4
THAP2	5HSA037253 hypothetical protein DKFZp564I0422
adducin3g	5HSA079674 adducin 3 (gamma), transcript variant 1
FLJ21125	4-5HSA080832 cDNA DKFZp686B0870
TEX2	5HSA076674 uncharacterized hypothalamus protein HT008
ZF584	5HSA076594 zinc finger protein 584
OZF	5HSA006254 OZF
HRR-1	5HSA008745 farnesol receptor HRR-1
Bspec2	5HSA051764 beta-spectrin 2 isoform 2
C12orf24	5HSA008965 clone 23733

Table 11. RT-PCR primers used for cloning random UTRs.

utr label	Upper Primer	Lower Primer
Pap4	gctag CAC AGA TTT ATC AAG C	ctcgag TTT TTA AGG TAG CAA TTA
THAP2	gctag CCG CTG CCA AAG AAA C	ctcgag TTC CCT TTC CCA CTC ATT
adducin3g	gctag CAG TGC GGA GCA GGA TGG	ctcga GGC CAG CCC GGG CAC ATG
FLJ21125	gctagc GCG GCC GGC TTG AAC T	ctcga GCT GGG AGC ACA GGG GAG
TEX2	gcta GCT AAA GTC GGA AAT CA	ctcga GGA TCT TCT TAG ACA CCA
ZF584	gcta GCG GGC CAC GGG AGT T	ctcga GGA CCG TGC GGG CAG AAC
OZF	gctag CGG AAG AAG CCT GAG AAG	ctcgag CTG AAT GGA AGC TTT CCC
HRR-1	gctag CCT CCT CCT CAC CTC A	ctcgag CCA AAT TTT TCA ATT GAA AT
Bspec2	gctag CGG CAA CCG TGG CAT	ctcga GTA ACG CCC CGG AGG ACT
C12orf24	gctag CCA CGT GCT TCG GTC CGT	ctcga GAC AGC ACG CGT GCG TCA

Appendix III

Isolation of IRES binding proteins

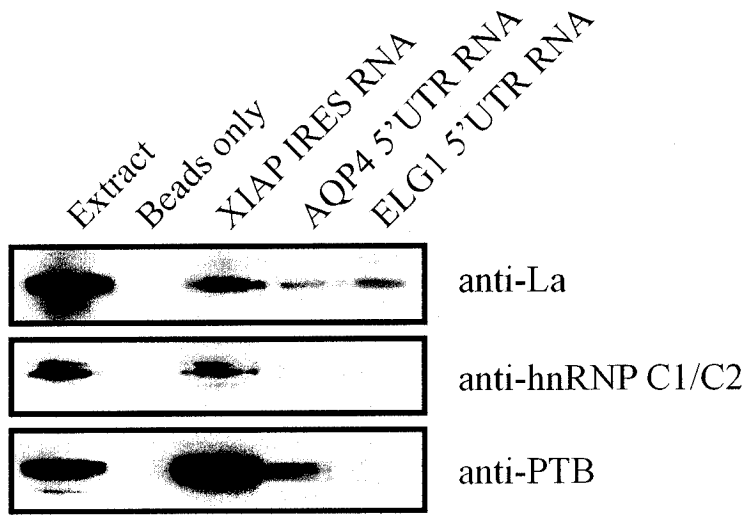


Figure I. RNA-affinity chromatography isolation of XIAP, AQP4, and ELG1 IRES-binding proteins. Pre-cleared protein extracts from HEK293T cells were incubated with either agarose beads coated with XIAP IRES RNA, agarose beads coated with AQP4 5'UTR RNA, agarose beads coated with ELG1 5'UTR RNA, or agarose beads alone. After protein binding, beads were washed extensively, pelleted, and proteins were eluted by boiling and resolved by SDS-PAGE. Following separation by SDS-PAGE, proteins were transferred to PVDF membrane and probed with anti-La A1, anti-hnRNP C1/C2, and anti-PTB antibodies.

- *Courtesy of Dr. Stephen Lewis*

Curriculum Vitae

STEPHEN BAIRD

20 JULIAN AVE • OTTAWA, ONTARIO K1Y 0S5
PHONE (613) 729-4310 (HOME)
EMAIL: SBAIRD57@YAHOO.CA

EXPERIENCE

2000-present Apoptosis Research Centre/University of Ottawa

PhD candidate

- Along with many bioinformatics applications I have used the following techniques: RT-PCR, RNA sequencing, RNA structure, Mass spectrometry with both proteins and RNA using Ciphergen MALDI-TOF, tissue culture, Northern blots, Western blots. ELISA assays, statiscal analysis

1990 - present Children's Hospital of Eastern Ontario Research Instititute. Ottawa, Ont.

Research Technician

- Responsible as a bioinformatics resource and system administrator of three SUN workstation and a network of PCs, Macs, network printers in the lab of Dr. R. Korneluk, Dr. A. MacKenzie, (Aegera Oncology Inc until 2000) and the Apoptosis Research Centre. Network administrator for U.of Ottawa subnet in the CHEO Research Instititute.

1989 University of Ottawa Ottawa, Ont.

Research Assistant

Subcloning and sequencing of genomic soybean nodulin clones in the lab of Dr. D. Johnson..

1984 National Research Council Ottawa, Ont.

Guest Worker/Research Assistant

worked in the laboratory of Dr. V. Seligy applying genetic engineering techniques to the cloning of cellulase genes of *Bacillus* species into *E. coli* and *S. cerisiae*..

1983 July-Dec National Research Council Ottawa, Ont.

Guest Worker

Worked in the lab of Dr. A.W. Kahn studying the cellulase enzyme complex of mutant strains of *Trichoderma reesei*

1983 Feb.-June University of Waterloo Waterloo, Ont.

Laboratory Assistant

Laboratory Assistant worked in the laboratory of Dr. B. Glick on a project concerned with the transfer of cellulase genes into *Azotobacter* species

TEACHING

- Canadian Genetic Diseases Network Bioinformatics Workshops.
  Bioinformatics - Ottawa. August 9-19, 2000,
  Genomics – Montreal. August 20 – 25, 2001,
  Bioinformatics – Vancouver. February 18 – March 2, 2002,
  Genomics – Montreal. October 28 – November 2, 2002,
  Bioinformatics – Ottawa. May 26 – June 7, 2003,
  Bioinformatics – Toronto. February 13 - 25, 2006.
- *Teaching Assistant* CSI4154 "Algorithms in Bioinformatics" U.of Ottawa Fall 2002.
- *Lecturer / course developement* BPS4104 "Bioinformatics" U.of Ottawa Fall 2001.
- *Supervision of Summer Students.* 2000 – 2006. Directed summer students (James Kennedy, Sylvie Balabanian, Jonathon Greenwood, Zachary King) in both bioinformatics and wet-lab projects.

EDUCATION

2000-2006 University of Ottawa Ottawa, Ont.

- Ph.D. Human and Molecular Genetics
- Thesis: Searching for Internal Ribosome Entry Sites.

2000 University of British Columbia/ University of Toronto

- Certificate in Bioinformatics
- 5 weeks of intensive courses in Bioinformatics. Genomics and Proteomics and software developement
- Advanced training organized by the Canadian Genetic Disease Network

1986-1988 University of Ottawa Ottawa, Ont.

- M.Sc. in Biology.
- Thesis: The Molecular Isolation and Characterization of Endo-ß-1,4-glucanase Genes from *Bacillus polymyxa* and *Bacillus circulans*.

1981-1984 University of Waterloo Waterloo, Ont.

- Bachelor of Independent Studies

SCHOLARSHIPS AND AWARDS

2001-2005	Ontario Graduate Scholarship University of Ottawa Excellence Scholarship
2002-2003	Ontario Graduate Scholarship University of Ottawa Excellence Scholarship
2003-2004	Ontario Graduate Scholarship University of Ottawa Excellence Scholarship Royal Arch Mason Medical Research Bursary BMI Graduate Student Poster Day "Chairman's Award"
2004-2005	Ontario Graduate Scholarship University of Ottawa Excellence Scholarship BMI Graduate Student Poster Day "1st place Ph.D. Micro and Immunology"

RECENT PRESENTATIONS

S. Baird, M. Turcotte, M. Holcik. *Discovery of novel internal ribosome entry site motifs from searches with the XIAP IRES structure.* Translational Control. Cold Spring Harbor, New York. Sept. 6 - 10, 2006.

S. Baird, M. Turcotte, M. Holcik and R.G. Korneluk. *Searching for IRES.* CIHR National Student Research Poster Competition. Winnipeg, Alberta. June 7, 2006

S. Baird, M. Turcotte, R.G. Korneluk and M. Holcik. *Searching for IRES Structurally Similar to the XIAP IRES.* RNA 2005, Banff, Alberta, May 24 - 28, 2005.

S. Baird. *Bioinformatics and the Biologist.* invited speaker - OCRI I.T. in Healthcare. Ottawa. Dec. 14, 2004

S. Baird, M. Turcotte, M. Holcik and R.G. Korneluk. *Searching for IRES.* The Canadian Genetic Diseases Network Annual Scientific Meeting, Talisman Inn, Kimberly ON. May 27 - 30, 2004.

S. Baird, M. Turcotte, R.G. Korneluk and M. Holcik. *Structural determination of Internal Ribosome Entry Sites with currently available software.* RNA 2004, U. of Wisconsin, Madison, Wisconsin, June 1 - 6, 2004.

PATENTS

Robert G. Korneluk, Alex E. MacKenzie, **Stephen Baird.** US Pat. # : 5,919,912 (July 6, 1999). MAMMALIAN IAP ANTIBODIES & DIAGNOSTIC KITS.

Robert G. Korneluk, Alex E. MacKenzie, **Stephen Baird,** Peter Liston. (1995) US Pat. # : 6,156,535 (Dec. 5, 2000), 6,541,457 (April 1, 2003), 6,689,562 (Feb. 10, 2004), 6,977,158 (Dec. 20, 2005), 7,067,281 (June 27, 2006). MAMMALIAN IAP GENE FAMILY, PRIMERS, PROBES, AND DETECTION METHODS.

Robert G. Korneluk, Alex E. MacKenzie, Peter Liston, **Stephen Baird,** Ben Tsang, Christine Pratt. US Pat. # : 6,107,041 (Aug. 22, 2000), 7,087,584 (Aug 8/2006). DETECTION AND MODULATION OF IAPs AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE DISEASE.

Robert G. Korneluk, Alex E. MacKenzie, Peter Liston, **Stephen Baird,** B. Tsang, Christine Pratt. US Pat. # : 6,300,492 (Oct. 9, 2001). MODULATION OF IAPS FOR THE DIAGNOSIS AND ANTISENSE TREATMENT OF PROLIFERATIVE DISEASE.

Robert G. Korneluk, Alex E. MacKenzie, Peter Liston, **Stephen Baird,** B. Tsang, Christine Pratt . US Pat. # : 6,133,437 (Oct. 17, 2000). MODULATION OF IAPS FOR THE TREATMENT OF PROLIFERATIVE DISEASES.

Robert G. Korneluk; Eric LaCasse, **Stephen Baird**, Martin Holcik, Sean Young. US Pat. # :6,673,917 (Jan. 6, 2004). ANTISENSE IAP NUCLEIC ACIDS AND USES THEREOF.

Robert G. Korneluk, Alex MacKenzie, **Stephen Baird**, Peter Liston. US Pat. # : 6,656,704 (Dec. 2, 2003). MAMMALIAN APOPTOSIS INHIBITOR PROTEIN GENE FAMILY, PRIMERS, PROBES & DETECTION METHODS.

Robert G. Korneluk, Herman H. Cheung, Peter Liston, Martin Holcik and **Stephen Baird** (pending) METHODS AND REAGENTS FOR INHIBITING ER STRESS-INDUCED APOPTOSIS. ATTORNEY DOCKET NO.: 50385/002001

PUBLICATIONS

Baird, S. D., Lewis, S. M., Turcotte, M., and Holcik, M. (2006). A search for structurally similar cellular IRES. Nucleic Acids Research, under revision.

Baird, S. D., Turcotte, M., Korneluk, R. G., and Holcik, M. (2006). Searching for IRES. RNA *12*, 1755-1785.

Holcik, M., Graber, T., Lewis, S. M., Lefebvre, C. A., Lacasse, E., and **Baird, S.** (2005). Spurious splicing within the XIAP 5' UTR occurs in the Rluc/Fluc but not the betagal/CAT bicistronic reporter system. Rna *11*, 1605-1609.

Hu, Y., Cherton-Horvat, G., Dragowska, V., **Baird, S.**, Korneluk, R. G., Durkin, J. P., Mayer, L. D., and LaCasse, E. C. (2003). Antisense oligonucleotides targeting XIAP induce apoptosis and enhance chemotherapeutic activity against human lung cancer cells in vitro and in vivo. Clin Cancer Res *9*, 2826-2836.

Brown, E. G., Liu, H., Kit, L. C., **Baird, S.**, and Nesrallah, M. (2001). Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. Proc Natl Acad Sci U S A *98*, 6883-6888.

LaCasse, E. C., **Baird, S.**, Korneluk, R. G., and MacKenzie, A. E. (1998). The inhibitors of apoptosis (IAPs) and their emerging role in cancer. Oncogene *17*, 3247-3259.

Chen, Q., **Baird, S. D.**, Mahadevan, M., Besner-Johnston, A., Farahani, R., Xuan, J., Kang, X., Lefebvre, C., Ikeda, J. E., Korneluk, R. G., and MacKenzie, A. E. (1998). Sequence of a 131-kb region of 5q13.1 containing the spinal muscular atrophy candidate genes SMN and NAIP. Genomics *48*, 121-127.

Lafreniere, R. G., Kibar, Z., Rochefort, D. L., Han, F. Y., Fon, E. A., Dube, M. P., Kang, X., **Baird, S.**, Korneluk, R. G., Rommens, J. M., and Rouleau, G. A. (1997). Genomic structure of the human GT334 (EHOC-1) gene mapping to 21q22.3. Gene *198*, 313-321.

Lafreniere, R. G., Rochefort, D. L., Kibar, Z., Fon, E. A., Han, F., Cochius, J., Kang, X., **Baird, S.**, Korneluk, R. G., Andermann, E., *et al.* (1996). Isolation and characterization of GT335, a novel human gene conserved in Escherichia coli and mapping to 21q22.3. Genomics *38*, 264-272.

Liston, P., Roy, N., Tamai, K., Lefebvre, C., **Baird, S.**, Cherton-Horvat, G., Farahani, R., McLean, M., Ikeda, J. E., MacKenzie, A., and Korneluk, R. G. (1996). Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes. Nature *379*, 349-353.

Phelan, C. M., Larsson, C., **Baird, S.**, Futreal, P. A., Ruttledge, M. H., Morgan, K., Tonin, P., Hung, H., Korneluk, R. G., Pollak, M. N., and Narod, S. A. (1996). The human mammary-derived growth inhibitor (MDGI) gene: genomic structure and mutation analysis in human breast tumors. Genomics *34*, 63-68.

Mahadevan, M. S., **Baird, S.**, Bailly, J. E., Shutler, G. G., Sabourin, L. A., Tsilfidis, C., Neville, C. E., Narang, M., and Korneluk, R. G. (1995). Isolation of a novel G protein-coupled receptor (GPR4) localized to chromosome 19q13.3. Genomics *30*, 84-88.

Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., **Baird, S.**, Besner-Johnston, A., Lefebvre, C., Kang, X., and et al. (1995). The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. Cell *80*, 167-178.

Mahadevan, M. S., Amemiya, C., Jansen, G., Sabourin, L., **Baird, S.**, Neville, C. E., Wormskamp, N., Segers, B., Batzer, M., Lamerdin, J., and et al. (1993). Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. Hum Mol Genet *2*, 299-304.

Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'Hoy, K., **Baird, S.**, Sabourin, L., Lennon, G., and et al. (1992). Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. Nat Genet *1*, 261-266.

Baird, S. D., and Johnson, D. A. (1991). Determination of the sizes of DNA fragments cloned into M13 without DNA purification. Biotechniques *10*, 180.

Baird, S. D., Hefford, M. A., Johnson, D. A., Sung, W. L., Yaguchi, M., and Seligy, V. L. (1990a). The Glu residue in the conserved Asn-Glu-Pro sequence of two highly divergent endo-beta-1,4-glucanases is essential for enzymatic activity. Biochem Biophys Res Commun *169*, 1035-1039.

Baird, S. D., Johnson, D. A., and Seligy, V. L. (1990b). Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus circulans. J Bacteriol *172*, 1576-1586.

Seligy, V.L., **S. Baird**, R. Bernier, M. Desrocher, M.J. Dove, A. Lo, R.M. Mackay and G.E. Willick. (1987) Functional mapping and expression of a Bacillus subtilis endo-1,4-ß-D-glucanase gene in Escherichia coli and the Yeasts Saccharomyces cervisiae and Schwanniomyces alluvius. in, Biological Research on Industrial Yeasts. G.G. Stewart et al. (ed.). Baton Rouge: CRC Press.

MacKay, R. M., Lo, A., Willick, G., Zuker, M., **Baird, S.**, Dove, M., Moranelli, F., and Seligy, V. (1986). Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene. Nucleic Acids Res *14*, 9159-9170.

Mackay, R. M., **Baird, S.**, Dove, M. J., Erratt, J. A., Gines, M., Moranelli, F., Nasim, A., Willick, G. E., Yaguchi, M., and Seligy, V. L. (1985). Glucanase gene diversity in prokaryotic and eukaryotic organisms. Biosystems *18*, 279-292.

Kahn, A.W., Chin, A., and **Baird, S.** (1985) Use of charcoal to minimize end product inhibition in enzymatic hydrolysis of cellulose. Biotech Letters 7: 447-450.