UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Molecular mechanisms and inhibition of mRNA recruitment in eukaryotic and viral translation initiation

Permalink

https://escholarship.org/uc/item/3mp596bg

Author

Berry, Katherine Elizabeth

Publication Date

2010

Peer reviewed|Thesis/dissertation

Molecular mechanisms and inhibition of mRNA recruitment in eukaryotic and viral translation initiation

by

Katherine Elizabeth Berry

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Jennifer A. Doudna, Chair
Professor Matthew B. Francis
Professor James M. Berger

Fall 2010

Abstract

Molecular mechanisms and inhibition of mRNA recruitment in eukaryotic and viral translation

initiation

by

Katherine Elizabeth Berry

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Jennifer A. Doudna, Chair

The regulation of protein synthesis is critical in viral infection, cell death, and development. Most regulation of translation occurs during the rate-limiting step of translation initiation, the process by which the ribosome binds to and positions initiator tRNA and messenger RNA at the initiation (AUG) codon. An understanding of the detailed mechanisms of this initiation process is critical to our understanding of translational control in biology. This work discusses the molecular mechanisms by which messenger RNA is recruited to human 40S ribosomal subunits during translation initiation and positioned correctly in the mRNA binding cleft and our efforts to inhibit this process in the lifecycle of Hepatitis C virus.

Eukaryotic initiation factor 3 (eIF3) was previously thought to bind primarily to the solvent side of 40S ribosomal subunits, and the principal role of the eIF3j subunit during translation initiation was believed to be facilitating eIF3 binding to 40S subunits. The C-terminal domain of human eIF3j is now known to bind in the mRNA binding cleft and aminoacyl (A) site of the 40S ribosomal subunit. We utilized a recombinant biochemical system of human initiation factors to biophysically examine eIF3j's interaction with mRNA and interface binding factors to define the role of this eIF3 subunit in the translation initiation pathway. We demonstrate that eIF3j interacts directly or indirectly with eIF1A on the 40S subunit. We also show that eIF3j influences the interaction of mRNA with the 40S subunit's mRNA binding cleft during translation initiation, reducing mRNA's affinity for 40S subunits until $eIF2\text{-}tRNA_i^{met}\text{-}GTP$ (TC) is present. These biochemical observations help to explain why deletion of eIF3j from *S. cerevisiae* leads to a leaky scanning phenotype.

Hepatitis C virus (HCV) is a considerable global health problem for which new classes of therapeutics are needed. The HCV genomic RNA contains an internal ribosome entry site (IRES) in its 5′ untranslated region (UTR), the structure of which is essential for viral protein translation. We developed a high-throughput assay to identify compounds that selectively block translation

initiation from the HCV IRES. Rabbit reticulocyte lysate conditions were optimized to faithfully report on authentic HCV IRES-dependent translation relative to a 5′ capped mRNA control. Despite well-optimized *in vitro* translation conditions, no selective HCV IRES inhibitors were found in the end, as the vast majority of hits proved to be luciferase and general translation inhibitors. The analysis of these molecules, and the finding that a large fraction of false positives resulted from off-target effects, highlights the challenges inherent in screens for RNA-specific inhibitors.

The HCV IRES faces a challenge in that the majority of its structure binds to the solvent side of the 40S subunit, yet its initiation codon needs to reach the P site in the mRNA binding cleft, and must do so without the many cap-binding and scanning factors used by a cellular message. The IRES includes a predicted pseudoknot interaction near the AUG start codon, but the results of previous studies of its structure have been conflicting. Using mutational analysis coupled with activity and functional assays, we verified the importance of pseudoknot base pairings for IRES-mediated translation, and conducted a comprehensive study of the structural tolerance and functional contributions of the pseudoknot. Ribosomal toeprinting experiments show that the entirety of the pseudoknot element positions the initiation codon in the mRNA binding cleft of the 40S ribosomal subunit. Optimal spacing between the pseudoknot and the start site AUG resembles that between the Shine-Dalgarno sequence and the initiation codon in bacterial mRNAs. In addition, we validated the HCV IRES pseudoknot as a potential drug target using antisense 2′-OMe oligonucleotides. Initial steps have been taken to solve a high-resolution structure of this IRES pseudoknot domain.

dedicated to the many catalysts in my life

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Å	Angstrom
A	adenine
A_{260}	absorbance at 260 nanometers
A site	amino acyl site
aa	amino acid
AUG	initation codon: Adenosine-Uridine-Guanosine
ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
°C	degree Celsius
C	cytosine
Ci	curie
Cl	chloride
CTD	C-terminal domain
CrPV	cricket paralysis virus
Cryo-EM	Cryo-eletron microscopy
CSFV	classical swine fever virus
CV	column volume
Δ	delta (deletion)
Da	dalton(s)
dom	domain
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ddNTP	dideoxynucleotide triphosphate
DNase	deoxyribonuclease
DTT	dithiolthreitol
E site	exit site
EB	10 mM Tris-HCl pH 8.5
EDTA	ethylenediaminetetraacetic acid
eIF	eukaryotic initiation factor
EMCV	encephalomyocarditis virus
EtOH	ethanol
FF	firefly
FP	fluorescence polarization
G	guanosine
GDP	guanosine 5'-diphosphate
GEF	guanine exchange factor
GMPPNP	guanosine 5'-[β,γ-imido]triphosphate
GTP	guanosine 5'-triphosphate
g	gram(s)
h	hour(s)
H_2O	water
HCV	Hepatitis C virus

HDV	hepatisis δ virus (ribozyme)
HEPES	N-2-hydroxyethylpiperazine-N´-2-ethanesulfonic acid
HH	hammerhead (ribozyme)
HTS	high-throughput screen
IRES	internal ribosome entry site
IPTG	isopropyl-β-D-thiogalactoside
L	loop
LB	Luria-Bertani broth
k	kilo
K	potassium
K_D	dissociation constant
k_{off}	dissociation rate constant
k_{on}	association rate constant
$k_{on,obs}$	observed association rate constant
L	liter
λ	wavelength
m	milli
μ	micro
M	molar / moles per liter
MetRS	methionyl-tRNA synthetase
$Mg^{(2+)}$	magnesium
miR / miRNA	microRNA
min	minute(s)
mRNA	messenger RNA
MW	molecular weight
n	nanomolar
Na	sodium
NH_4	ammonium
NPI	normalized percent inhibition
nt	nucleotide(s)
NTD	N-terminal domain
NTP	nucleotide triphosphate
OAc	acetate
OD	optical density
ODU	optical density unit
oligo	oligonucleotide(s)
ORF	open reading frame
p	pico
P site	peptidyl site
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
RN	*Renilla*
RRM	RNA recognition motif
RRL	rabbit reticulocyte lysate
s	second

S	stem
SD	Shine-Dalgarno
SDS	sodium dodecyl sulfate
TAE	40 mM Tris-acetate pH 8.0, 1 mM EDTA
TBE	90 mM Tris-borate, 2 mM EDTA
TC	ternary complex (eIF2/Met-tRNA$_i$/GTP)
TCA	trichloroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TE	10 mM Tris-HCl pH 7.5, 1 mM EDTA
THE	33 mM Tris, 66 mM HEPES, 0.1 mM EDTA, pH 7.5
tRNA	transfer RNA
tRNA$_i$	initiator transfer RNA
U	uridine
uORF	upstream open reading frame
UTR	untranslated region
UV	ultraviolet
vol	volume
W	watt
WT	wild type

Chapter 1

Introduction

Introduction

1.1 Biological importance of translation control

The regulation of gene expression is critical to the survival of all organisms, allowing them to adapt to changes in cellular environments, maintain homeostasis, develop correctly and respond to pathogens. It has long been appreciated that a good deal of this gene expression regulation occurs at the level of transcription: how much messenger RNA (mRNA) is transcribed from a given locus of DNA. Over the last 40 years, it has been increasingly appreciated that all organisms utilize translational control as well, and that this is especially important for eukaryotes (Matthews et al. 2007). Translational control entails modulating the efficiency with which one or all mRNAs in the cell are translated into proteins by ribosomes and is one facet of post-transcriptional gene regulation.

The regulation of gene expression at the level of translation affords the cell an immediate, cytoplasmic response to a cellular stress or signal and allows for spatial, temporal and message-specific control. Translational control plays important roles in diverse biological functions, including viral infection, cancer and normal cellular processes such as in neuronal function, embryonic development and cellular differentiation (Gebauer & Hentze 2004). For example, neurons require new and localized protein synthesis of specific messages for synaptic plasticity and long-term memory formation (Costa-Mattioli et al. 2009); in contrast, failure to translationally repress a class of messages in neurons leads to Fragile X Mental Retardation syndrome (Sonenberg & Hinnebusch 2009). In addition, overexpression of many translation factors (e.g. eIF4E and eIF4G) is common in many cancer types (Thumma & Kratzke 2007; Silvera et al. 2010). Indeed, overexpression of the cap-binding protein eIF4E alone is tumorigenic in mice (Schneider & Sonenberg 2007). Furthermore, in embryonic development, localized translation helps establish polarity of the embryo (Sonenberg & Hinnebusch 2009).

To fully understand the diverse roles of translational control and regulation in biology, a detailed mechanistic understanding of the process of protein synthesis is critical. This body of work focuses on the molecular mechanisms that underlie both general cap-dependent translation initiation and a specific case of cap-independent, viral translation initiation in Hepatitis C virus infection, and where the differences in these mechanisms can be exploited for new viral therapeutics. The significance of this work requires an understanding both of the field's current knowledge of eukaryotic translation initiation mechanisms and of the infection, translation strategy and current and future therapeutic prospects for Hepatitis C virus. A survey of each of these areas will be provided throughout this introduction.

1.2 Translation in eukaryotes

The overall process of translation in both eukaryotes and prokaryotes consists of four broad phases: initiation, elongation, termination and recycling. Initiation is the process by which 80S ribosomes bind to and recognize AUG start codons in the ribosomal P site with the help of initiator tRNA (tRNA$_i$). During elongation, the mRNA code is read out by codon-anticodon base pairing in the small (40S) ribosomal subunit, while the peptidyl transferase center of the large (60S) ribosomal subunit catalyzes peptide-bond formation between amino acids charged onto cognate tRNAs. Termination involves recognition of a stop codon and release of the nascent polypeptide chain. During recycling, the small and large ribosomal subunits are split apart in preparation for additional rounds of initiation.

Prokaryotic translation initiation uses direct base pairing between nucleotides upstream of the start codon (the "Shine-Dalgarno" sequence) and in ribosomal RNA to directly place the start codon in the P site of the small ribosomal subunit upon mRNA binding (Shine & Dalgarno 1974; Yusupova et al. 2001); eukaryotic translation initiation proceeds through a more complex mechanism. In eukaryotes, initial interaction of mRNA with ribosomes does not generally place the start codon directly in the P site. Rather, the mRNA binds to ribosomal complexes initially via a 5′ 7-methyl-guanosine cap, with the start codon often substantially downstream from the ribosomal subunit. Start codon recognition proceeds through a "scanning" mechanism in which the small subunit of the ribosome and associated protein factors move along mRNA in a 5′ to 3′ direction until an AUG codon in a suitable nucleotide context is recognized in the P site via base pairing with initiator tRNA (Jackson et al. 2010). This scanning model was initially proposed by Marilyn Kozak to explain the almost exclusively monocistronic nature of eukaryotic mRNAs and the preference for initiating translation from the most 5′ AUG in an untranslated region (UTR) despite these AUGs falling in variable positions (Kozak 1978). This scanning model is now well established and predicts that, all else being equal, the correct initiation codon will be the most 5′ in the UTR, and therefore the first encountered as the small subunit scans along the message. In mammals, however, there is also an ideal nucleotide context for a start codon, known as the Kozak consensus sequence, which also influences the efficiency with which a start codon is recognized (Kozak 1986, 1991). This consensus sequence is GCC**R**CC<u>AUG</u>**G**, where R represents either purine base (A or G) and the most critical positions outside of the AUG (+4 and -3) are indicated in bold.

The initiation phase of translation is generally found to be rate limiting in eukaryotic protein synthesis (Holcik & Sonenberg 2005; Sonenberg & Hinnebusch 2009). Because of this, initiation is a critical point for regulation of protein synthesis for the cell. Initiation of translation is quite complex in eukaryotes, requiring >30 polypeptides as factors to assist in the initiation process, as opposed to the 3 initiation factors required in prokaryotes (Sonenberg & Hinnebusch 2009). This additional complexity mirrors that of translation regulation in eukaryotes. In bacteria, translational control appears to center on controlling the accessibility of the Shine-Dalgarno sequence and initiation codon, where a single-stranded ribosome-binding site dramatically increases the translatability of a bacterial mRNA (Geissmann et al. 2009).

1.3 Mechanisms of general eukaryotic translation initiation

1.3.1 Introduction to eukaryotic initiation factors (eIFs)

As mentioned above, over 30 individual polypeptide chains, comprising at least 13 distinct protein complexes, participate in the process and regulation of eukaryotic translation initiation. Such proteins are referred to as eukaryotic initiation factors (eIFs). Here, these factors will be introduced one-by-one and their general function described. In the following sections, the overall assembly process and concerted action of these factors, as well as global regulatory mechanisms of protein synthesis, will be described.

<u>eIF1</u> is a single-subunit initiation factor which is a functional (but not structural) homolog of the bacterial initiation factor IF3 (Lomakin et al. 2003): it binds to the P site of the small ribosomal subunit, as determined by directed hydroxyl radical probing (Lomakin et al. 2003), and helps to select the correct start codon in the correct context in part by destabilizing complexes assembled at incorrect start codons (Pestova et al. 1998a; Pestova & Kolupaeva

3

2002). The core of eIF1's structure resembles other small RNA-binding proteins, while the N-terminal 30 amino acids are unstructured at least in free eIF1 (Fletcher et al. 1999).

eIF1A is both a structural and functional homolog of bacterial IF1 and binds in the A site of the 40S subunit, as determined by directed hydroxyl radical probing (Yu et al. 2009). Together with eIF1, eIF1A induces a conformational change in the 40S ribosomal subunit, which has been observed by cryo-electron microscopy and has been speculated to play a role in the transition from a scanning-competent "open" 40S conformation to a "closed" AUG-recognizing conformation (Passmore et al. 2007). In addition to the core oligonucleotide-binding (OB) domain, which is homologous to IF1, eIF1A also has an additional small domain as well as N- and C-terminal tails which are largely unstructured in the absence of 40S subunits (Battiste et al. 2000). These tails each influence start codon selection, but one promotes the process and the other opposes it (Fekete et al. 2007).

eIF2 consists of three protein subunits, referred to as alpha, beta and gamma. Along with a GTP nucleotide and a methionine-charged initiator tRNA (met-tRNA$_i^{met}$), eIF2 helps to form the so-called "ternary complex" (TC) which is critical for delivering initiator tRNA to the 40S ribosomal subunit. eIF2 is a GTPase; during a round of translation initiation, the GTP nucleotide is hydrolyzed to GDP, leaving a eIF2-GDP byproduct which must be recycled. Adjusting the concentration of active ternary complex by phosphorylating eIF2 is a point of major translational regulation for the cell (see section 1.3.6).

eIF2B is a five-subunit complex and functions as a guanine exchange factor (GEF) for the eIF2-GDP complex that is formed after a round of translation initiation. As a GEF, it stimulates the exchange of GDP for GTP to allow more active GTP-bound ternary complex to form (Williams et al. 2001). The importance of this GEF for cellular function is underscored by the fact that the inhibitory effect of eIF2α phosphorylation on active ternary complex stems from phosphorylated eIF2 acting as a tight-binding, competitive inhibitor of eIF2B GEF activity; indeed, much more of the eIF2B structure seems to be dedicated to inhibition by phosphorylated eIF2 than to GEF activity itself (Gomez et al. 2002).

eIF3 consists of thirteen individual protein subunits in humans, six of which are conserved in *S. cerevisiae*. eIF3 binds directly to the 40S ribosomal subunit, where it helps to prevent premature joining of the large subunit and to recruit other factors such as the ternary complex and cap-binding factors, and also assists in scanning (Hinnebusch 2006). In humans, eIF3 binds largely to the solvent side of the 40S subunit (Siridechadilok et al. 2005) and may extend the mRNA binding cleft by binding mRNA beyond the E site of the 40S (Pisarev et al. 2008).

eIF4A is a DEAD-box ATP-dependent RNA helicase which is thought to assist in unwinding secondary structure in mRNA and is needed for scanning through 5′ UTRs with significant structures (Linder 2003). The activity of eIF4A is stimulated by the homologous proteins eIF4B and eIF4H (Rogers et al. 2001; Pestova & Kolupaeva 2002).

eIF4E binds directly and specifically to the 5′ 7-methyl-guanosine cap a the 5′ end of eukaryotic mRNAs. The positively charged, aromatic cap moiety is recognized by π-π interactions with conserved tryptophans (Marcotrigiano et al. 1997). Since the majority of eukaryotic translation initiation is cap-dependent, this eIF4E-cap interaction is a primary determinant of mRNA recruitment to ribosomes, and is another critical point of regulation for translation levels in a cell (see section 1.3.6).

eIF4G is a central scaffolding protein that, in humans, interacts with eIF4E, eIF4A, eIF4G, as well as with poly(A) binding protein (PABP) which binds to the 3′ poly(A) tail of

messenger RNAs (Jackson et al.). It links the cap-binding protein eIF4E back to the 40S ribosomal subunit via its interaction with eIF3, helps to stimulate eIF4A helicase activity (Rogers et al. 2001), and aids in circularization of mRNAs via its interaction with PABP (Tarun & Sachs 1996). Underscoring its importance in translation, certain viruses, including poliovirus, encode a protease that specifically cleaves eIF4G between its N- and C- terminal domains to disrupt cap-dependent translation by the host (Lloyd 2006).

eIF5 is a GTPase activating factor (GAP) for eIF2, which stimulates eIF2's hydrolysis of GTP to GDP in the context of the ternary complex (Das et al. 2001; Paulin et al. 2001). Indeed, eIF5 has been shown to increase the GTPase activity of eIF2 by five orders of magnitude *in vitro* (Algire et al. 2005).

eIF5B is a structural homolog of the bacterial initiation factor IF2. In bacteria, IF2 is a GTPase that brings the initiator tRNA to the P site of 30S subunits and facilitates subunit joining. eIF5B is also a GTPase which binds in a similar position to ribosomes as IF2 (Unbehaun et al. 2007) and aids in subunit joining (Pestova et al. 2000; Shin et al. 2002).

1.3.2 Pathway and dynamics of ribosomal complex assembly

The initial steps of translation initiation such as mRNA binding, scanning and AUG recognition involve only the small 40S ribosomal subunit; the large 60S subunit will join to give an 80S ribosome only after the start codon has been recognized. A general outline of how the translation initiation pathway in humans is thought to proceed is presented in Figure 1.1. Many of these steps have been inferred from circumstantial evidence and the exact order of initiation factor binding to small ribosomal subunits – or to what extent there is a defined order in which factors will bind *in vivo* – is not yet known. The application of single molecule fluorescence approaches to observe the binding of these factors in real time will certainly lead to exciting insights about what steps are absolutely required before others. In yeast, the process of initiation factor binding may be more concerted than in humans, as many initiation factors are thought to form a stable multifactor complex off of the 40S (Asano et al. 2000).

As depicted in Figure 1.1, eIF3, eIF1 and eIF1A are likely the first factors to bind to 40S subunits during initiation in humans, as they have been shown to play a key role in ribosome recycling, and would therefore be present on 40S subunits as they are released from post-termination 80S ribosomes (Pisarev et al. 2007). The ternary complex is thought to bind next, along with its GAP, eIF5, to give a 43S complex. This entire 43S ribosomal complex will bind to mRNA via its 5′ 7-methyl-guanosine cap, which connects back to the 40S ribosomal subunit itself via eIF4E-eIF4G-eIF3 interactions. Once mRNA has bound, the entire ribosomal complex is thought to scan along the mRNA in a 5′ to 3′ direction until an AUG is recognized in the P site by codon-anticodon base pairing with the initiator tRNA (Kozak 1978, 2002). ATPases such as eIF4A, 4B and 4H are thought to assist the 40S ribosomal complex in scanning through regions of stronger secondary structures, as the eIF4A-dependence and amount of ATP required for scanning through a 5′ UTR is directly proportional to the degree of structure within the untranslated region of mRNA (Jackson 1991; Svitkin et al. 2001).
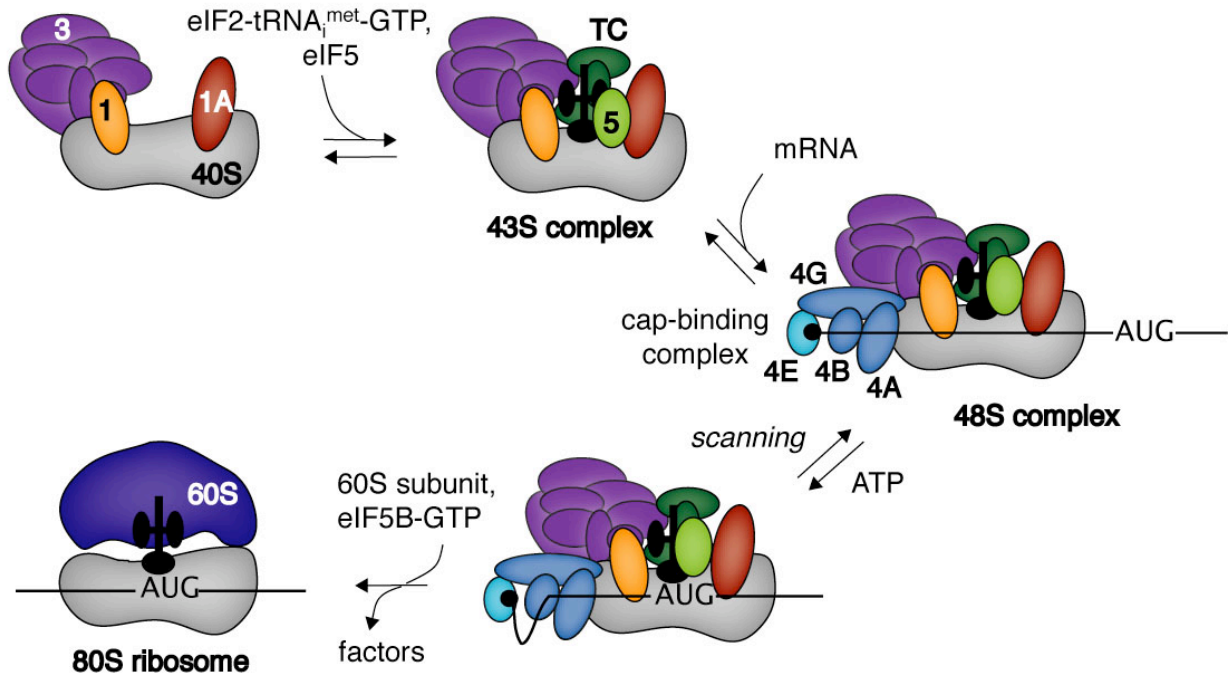
Figure 1.1: Pathway of cap-dependent translation initiation. Overall scheme of the general translation initiation pathway in humans. eIF1, 1A, and 3 each play a role in ribosome recycling and are likely the first factors to bind to the small (40S) ribosomal complex in the human initiation pathway. The ternary complex of eIF2-tRNA$_i^{met}$-GTP (TC) and eIF5 are subsequently recruited, generally before mRNA recruitment, except in times of limiting TC concentrations. mRNA recruitment happens principally through the cap-binding complex (eIF4E, 4A, 4G, and 4H). Scanning along the mRNA then allows the 40S ribosomal complex to locate an AUG initiation codon, triggering irreversible changes in the complex that allow interface factors to be released and the 60S subunit to join. The final product of translation initiation is an 80S ribosome, ready to enter elongation.

1.3.3 Mechanisms of AUG recognition

As the ribosomal complex scans along mRNA, recognition of an AUG leads to an irreversible process that allows for joining of the 60S ribosomal subunit. Fundamentally, this is accomplished by coupling base pairing between an AUG codon in mRNA and the anticodon of tRNA$_i$ to the hydrolysis of GTP by the ternary complex. Numerous initiation factors, such as eIF1, eIF1A, eIF2 and eIF5, play roles in either preventing GTP hydrolysis before AUG recognition, stimulating GTP hydrolysis upon recognition, or additionally stimulating GTP hydrolysis when the AUG is in a favorable Kozak consensus context (Algire & Lorsch 2006). The various positive and negative interactions that exist in the 43S ribosomal complex prior to and after start codon recognition are depicted in Figure 1.2; these interactions are discussed in greater detail below.

6
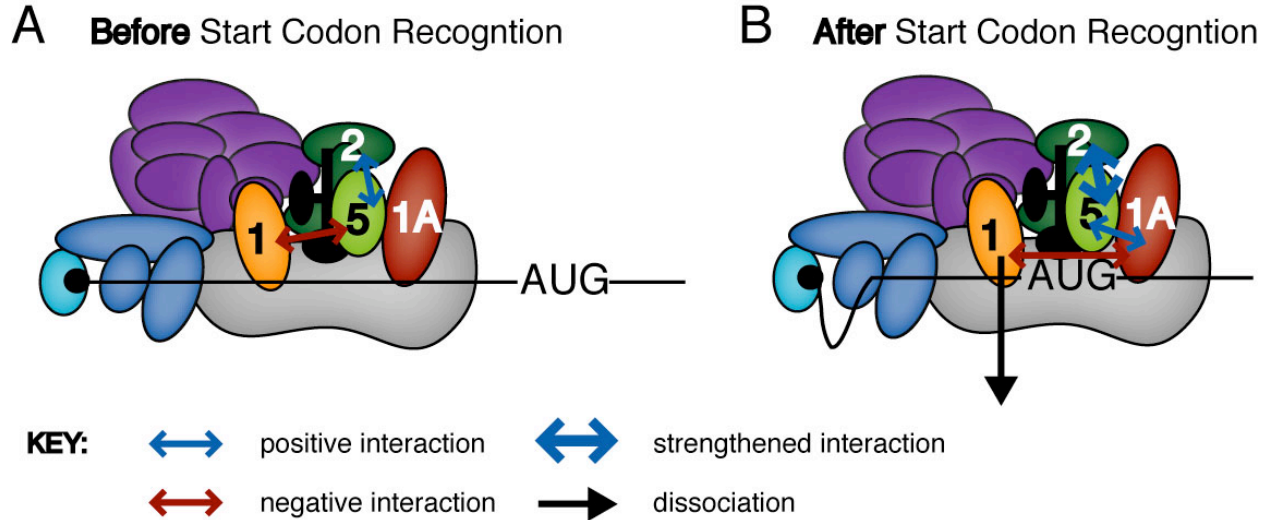
Figure 1.2: Changes in eIF interaction network involved in AUG recognition. Cartoons of pre-initiation ribosomal complexes with current models of known interactions between interface binding factors indicated with arrows. Significant changes in these interactions occur upon start codon recognition (*A* vs. *B*), allowing hydrolysis of GTP or release of inorganic phosphate (P_i) to be coupled to AUG recognition. *(A)* Before start codon recognition, eIF5 and eIF2 interact, but eIF1 and eIF5 NTD compete for binding in the P site. *(B)* After start codon recognition, a conformational change occurs between eIF1 and eIF1A, eIF1 dissociates, the eIF1A CTD interacts with eIF5 to strengthen the eIF5/eIF2 interaction, facilitating GTP hydrolysis and/or P_i release.

It is believed that eIF1 and eIF1A play key roles in maintaining an "open" and scanning-competent conformation of the 40S ribosomal subunit before a start codon is recognized. This has been structurally visualized as a conformational change in the mRNA binding cleft of the 40S subunit in the presence of eIF1 and 1A (Passmore et al. 2007). Prior to start codon recognition, eIF1 interacts with eIF5 and inhibits its GAP activity (Unbehaun et al. 2004; Valasek et al. 2004). Upon start codon recognition, a conformational change between eIF1 and eIF1A triggers eIF1 dissociation from the 40S subunit (Maag et al. 2005). The importance of eIF1 dissociation in AUG recognition is supported by observations that the propensity of initiation at noncognate UUGs by eIF1 mutants correlates well with weakened binding affinities between eIF1 and 40S subunits (Cheung et al. 2007). eIF1 dissociation is thought to effect eIF5 both directly and indirectly via eIF1A. The N-terminal domain (NTD) of eIF5 is structurally very similar to eIF1 (Conte et al. 2006) and may compete with eIF1 for binding to the P site (Nanda et al. 2009). eIF1 dissociation allows eIF5 to bind stably to the P site and also allows eIF1A's C-terminal tail to leave its position in the P site to interact with eIF5 (Maag et al. 2006; Yu et al. 2009). In addition to its eIF1- and eIF1A-dependent roles in start codon selection, eIF5 may directly participate in this process, as it has been observed that eIF5 mutants preferentially allow initiation from UUG over other near-cognate start codons (Huang et al. 1997). While start codon recognition is generally stated to stimulate GTP hydrolysis activity by eIF2, recent biochemical evidence suggests that, in fact, GTP may be pre-hydrolyzed as GDP-P_i on eIF2 and that start codon recognition actually triggers the irreversible dissociation of inorganic phosphate (Algire et al. 2005).

1.3.4 Contributions of yeast genetics to AUG recognition mechanisms

While biochemical experiments have and will continue to play key roles in deducing mechanistic details regarding interactions that lead to correct AUG recognition, these biochemical experiments have been greatly aided and expanded by *in vivo* genetics experiments in the model eukaryotic organism *S. cerevisiae* (Hinnebusch et al. 2007). Both forward genetics to screen for new initiation factors with roles in AUG recognition and reverse genetics to test specific hypotheses generated through biochemical experiments have been used to great effect.

Two key phenotypes used by the yeast community to probe AUG recognition are a leaky scanning (Gcd⁻) phenotype and a UUG misrecognition (Sui⁻) phenotype (Hinnebusch et al. 2007). The Sui⁻ phenotype is read out from an assay measuring how efficiently a reporter gene (*his4*) is translated when the wild type authentic AUG is changed to a less favorable, alternative start codon (UUG). Proteins whose mutants have Sui⁻ phenotypes play important roles in the specific recognition of an AUG. The Gcd⁻ phenotype is deduced from an assay that takes advantage of the natural *GCN4* translational control system in yeast (Hinnebusch 2005). Leaky scanning is determined by placing a reporter gene downstream of an upstream open reading frame (uORF) such that the reporter is only efficiently translated if there is significant leaky scanning through the uORF, even in a cell where the eIF2α kinase has been deleted, such that the leaky scanning does not arise from limiting TC levels. Mutants displaying a Gcd⁻ phenotype either begin scanning prematurely or fail to recognize AUGs as they are encountered. The opposite phenotype (Gcn⁻) can be found by starving cells for amino acids, which would normally stimulate leaky scanning on *GCN4*, and looking for mutants that do not display leaky scanning even under conditions of decreased TC concentrations (Hinnebusch 2005). These mutants provide information about the formation of TC itself, as well as the speed of scanning.

Table 1.1 summarizes the numerous translation initiation factors that have been implicated in AUG recognition by yeast genetics experiments (Hinnebusch et al. 2007; Lorsch & Dever 2010). Where known, compensatory mutations for each of these genetic defects are also listed, as these can provide additional hypotheses about interacting partners or mechanisms of defects. The wide distribution of factors contributing to AUG recognition in yeast (eIF1, eIF1A, eIF5 all subunits of eIF2, and many subunits of eIF3) underscores the fundamental importance of initiating translation at the correct start codon in order to avoid the translation of aberrant proteins.

Table 1.1: Contribution of yeast genetics to AUG recognition mechanisms. Summary of mutations in *S. cerevisiae* that result in Gcd⁻ or Sui⁻ phenotypes (described in text), and those mutations that can compensate for these phenotypes (Hinnebusch et al. 2007; Lorsch & Dever 2010).

Gcd⁻ (leaky scanning)		Sui⁻ (nonAUG initiation)	
Initiation Factor	Compensated by…	Initiation Factor	Compensated by…
eIF1 C-terminal tail		eIF1	eIF5 GAP function mutation eIF3c NTD mutants
eIF1A CTD	TC overexpression	eIF1A CTD	
eIF1A OB fold	Overexpression of mutant eIF1A		
eIF2 (3/3 subunits)		eIF2 (3/3)	
eIF2B (4/5 subunits)			
eIF3b			
eIF3c	TC overexpression	eIF3c NTD	
eIF3j	eIF1A overexpression		
		eIF4GII	eIF1 overexpression
eIF5		eIF5	eIF1 overexpression eIF3c NTD mutants
		tRNA$_i^{met}$ anticodon loop	

1.3.5 Mechanisms of subunit joining

After start codon recognition, eIF1 dissociation and P$_i$ release from the ternary complex, the activity of the GTPase eIF5B is required for the removal of interface binding factors and the joining of the 60S ribosomal subunit (Pestova et al. 2000). While the large subunit can join the small subunit in the absence of GTP hydrolysis by eIF5B, GTP hydrolysis is required for eIF5B to be released from the newly formed 80S ribosome and for catalysis of the first peptide bond (Shin et al. 2002). Mechanistically, there are data to suggest that eIF5B may interact with eIF1A on the 40S ribosomal subunit, aiding in eIF5B recruitment and stimulating eIF5B GTPase activity (Marintchev et al. 2003; Olsen et al. 2003; Acker et al. 2006), and that these two factors may dissociate together from the ribosome (Fringer et al. 2007). In addition to the activity of eIF5B, joining of the 60S subunit itself appears to be important for the displacement of certain interface binding factors, such as eIF2 (Unbehaun et al. 2004).

1.3.6 Mechanisms of translation regulation

As mentioned above, two major pathways of global regulation of translation levels in the cell center on the translation factors eIF2 and eIF4E. eIF2 phosphorylation is a key feature of cellular responses to many different stresses, from nutrient starvation to viral infection (Ron & Harding 2007). Different stresses activate one of four eIF2 kinases, which each phosphorylate the alpha subunit of eIF2 (Yamasaki & Anderson 2008). Phosphorylation of eIF2 inhibits the formation of active ternary complex and therefore downregulates overall translation within the cell (see description of eIF2B below). Phosphporylated eIF2α acts as a tight-binding, competitive inhibitor of eIF2B GEF activity, leading to decreased active ternary complex concentrations.

Interestingly, while eIF2 phosphorylation globally downregulates cellular protein synthesis, it can simultaneously stimulate the translation of select messages. Notable mRNAs that exemplify this phenomenon are those encoding the GCN4 transcription factor in yeast and the homologous ATF4 protein in mammals (Vattem & Wek 2004; Hinnebusch et al. 2007). The mechanism by which the translation of these messages is stimulated when ternary complex is depleted involves upstream open reading frames (uORFs) positioned in such a way that if the initial uORF is well translated under conditions of high ternary complex concentration, the main ORF is not effectively translated. Interestingly, many human mRNAs have uORFs (50%, compared to 13% in yeast), but the full regulatory roles of these abundant uORFs have yet to be determined (Jackson et al. 2010).

The interaction of eIF4E with 7-methyl-guanosine caps at the 5′ end of mRNAs is another critical point of regulation for translation levels in a cell. A class of small proteins called 4E-binding proteins (4E-BPs) can bind to eIF4E and inhibit its interaction with the rest of the translation machinery in a phosphorylation-state-dependent manner. Phospho-4E-BPs directly compete with eIF4G (see below) for binding with eIF4E (Marcotrigiano et al. 1999). These 4E-BPs are phosphorylated by the mammalian target of rapamycin (mTOR) kinase, which is stimulated by many diverse stresses to downregulate global translation (Yamasaki & Anderson 2008) and is activated during mitosis to downregulate translation (Sivan & Elroy-Stein 2008).

1.4 Introduction to Hepatitis C virus: infection and treatment

Hepatitis C virus causes a blood-borne infection estimated to infect ~180 million people worldwide, with especially high infection rates in northern Africa, making the virus a considerable global health problem (Suzuki et al. 2007; Webster et al. 2009). Six genotypes of HCV are known and can vary up to 35% in sequence identity across the genome (Webster et al. 2009). Left untreated, ~75% of HCV infections become chronic and lead to liver cirrhosis and carcinoma in ~20% of cases (Webster et al. 2009). The current standard of care in HCV treatment is combination of pegylated interferon-α and ribavirin, which has limited efficacy and causes significant side effects (McHutchinson et al. 2006). In particular, these treatments show lower efficacy for genotype 1 (40-50%) than genotypes 2 and 3 (70-80%), and are only effective for about 50% of patients overall (Webster et al. 2009).

Given the problems with the current standard of care, many pharmaceutical companies have been interested in developing new HCV therapeutics, and new viral protease and polymerase inhibitors appear promising in clinical studies (Jarvis 2010). Despite these promising new therapeutics on the horizon, there is always the risk of drug-resistant mutant HCV strains emerging; as is common in viruses, the genome is replicated by a polymerase without proof-reading activity, leading to a high mutation rate and an increased ability to evolve under selective pressure, such as that introduced by drug treatment. This is especially problematic as HCV is known to exist as a population of closely related quasispecies during infection, allowing favorable genomic diversity to be rapidly amplified (Suzuki et al. 2007). Thus, new drugs and new drug targets are still needed for HCV.

HCV is a positive-sense, single stranded RNA virus of the family Flaviviridae (Webster et al. 2009). The virus primarily infects hepatocytes, and interactions between HCV envelope proteins and low-density-lipoprotein (LDL) receptors may contribute to hepatocyte-cell entry (Bartenschlager et al. 2010). Replication of HCV occurs cytoplasmically and in close conjunction with the endoplasmic reticulum (ER), as all viral proteins are either directly or indirectly associated with membranes that arise from the ER (Bartenschlager et al. 2010). The

most conserved regions of HCV's 9600 nucleotide (nt) RNA genome are the 5′ and 3′ UTRs (Figure 1.3), which serve to recruit the RNA polymerase to the 3′ ends of each strand. In addition, the 5′ UTR of the positive-sense genomic strand recruits the cell's translational machinery to a single, long open reading frame. The long polypeptide that is translated is subsequently cleaved into individual proteins by a combination of viral and cellular proteases (Suzuki et al. 2007). Nucleotides 40-340 of the positive-sense RNA form a specific secondary and tertiary RNA structure essential for viral protein synthesis (Tsukiyama-Kohara et al. 1992; Kieft et al. 1999). This is one of the most conserved regions of the HCV genome, as the sequence is held under selective pressure both to correctly form conserved structures in both the 5′ UTR of the positive-sense strand and in 3′ UTR of the negative-sense strand; the latter structure is necessary for RNA replication. Because effective translation from the positive-sense 5′ UTR is required for efficient infection and this region tolerates so little mutational flexibility, it is an attractive target for new drugs against the virus.
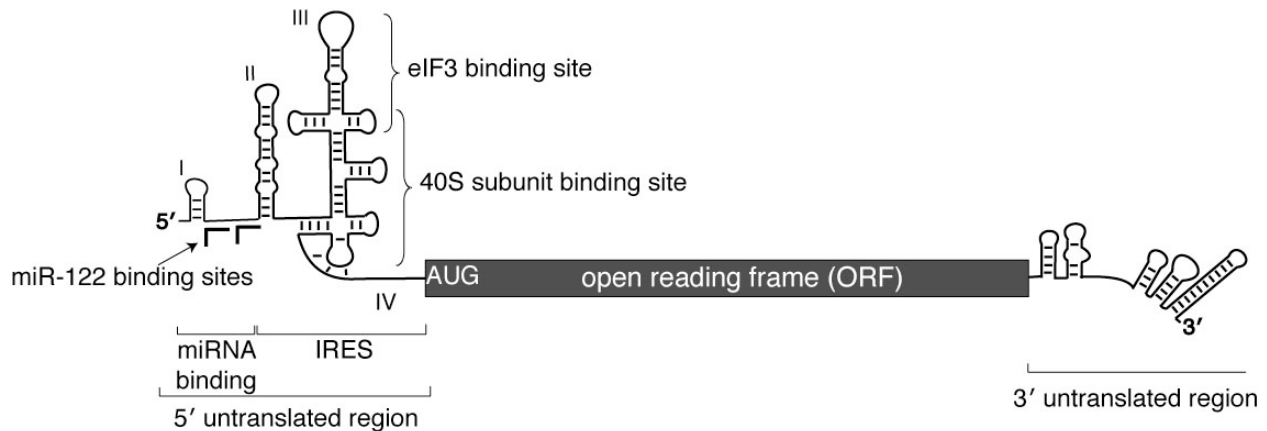
Figure 1.3: Structures and binding sites within HCV genomic RNA. The HCV genomic RNA encodes a single, long open reading frame and contains very highly conserved 5′ and 3′ UTRs. Conserved RNA structures in the 3′ UTR aids in recruitment of the RNA replication machinery, while those in the 5′ UTR bind directly to 40S ribosomal subunits and eIF3 to stimulate translation of the ORF. Recently discovered liver-specific miRNA binding sites are also located in the 5′ UTR and greatly stimulate viral propagation.

1.5 Translational control by internal ribosome entry sites (IRESs)

1.5.1 Introduction to IRESs

As described above, the majority of translation initiation in eukaryotic cells is cap-dependent, requiring the 5′ 7-methyl-guanosine cap for efficient recruitment of mRNA to ribosomal complexes. Over the last two decades, specific examples of mRNA sequences in the 5′ UTR have been identified that can stimulate translation internally and without the need for a free 5′ end or a 5′ cap. Because of this internal recruitment of ribosomal complexes, these sequences have been termed "internal ribosome entry sites," or IRESs. An appealing model for the biological roles of IRESs is that they could allow particular, privileged messages to continue to be efficiently translated even under cellular conditions in which global, cap-dependent protein synthesis is downregulated (Baird et al. 2006; Sonenberg & Hinnebusch 2009; Gilbert). Such

privileged messages could be viral mRNAs or cellular messages encoding proteins important in biological processes such as apoptosis or stress responses.

Typical 5′ UTRs in eukaryotes range from 50-120 nts (Suzuki et al. 2000). The first IRES sequence was identified in poliovirus mRNA in response to a conundrum about how mRNAs from picornaviruses could be efficiently translated, despite being uncapped and having 5′ UTRs ranging from 600-1200 nts with multiple out-of-frame, upstream AUGs (Pelletier et al. 1988; Pelletier & Sonenberg 1988). Since this original discovery, more than 80 cellular IRESs and 55 viral IRESs have been identified (Baird et al. 2006; Mokrejs et al. 2006). It should be noted that some, though not all, of these discovered cellular IRESs have turned out to be artifacts of the assays used to identify them; putative IRES sequences can serve as cryptic promoters and increase reporter gene activity through spurious means when transfection of bicistronic reporter DNA is utilized (Kozak 2005; Gilbert 2010). Therefore the use of RNA reporters and/or very sensitive northern blotting is required to unambiguously identify an IRES.

1.5.2 Diversity of IRESs

Of the >55 viral IRESs that have been discovered (Baird et al. 2006), several distinct classes can be identified, for which there is a great deal of diversity in terms of the mechanism of translation initiation by the IRES and the specific initiation factors required (Figure 1.4). On one side of the IRES mechanistic spectrum fall IRESs from picornaviruses (including poliovirus and EMCV) which require all initiation factors used in cap-dependent translation except for eIF4E and the N-terminal half of eIF4G (Balvay et al. 2009). Importantly, eIF4G is proteolytically cleaved during viral infection, giving the viral IRES an advantage over endogenous, cap-dependently translated messages (Lloyd 2006). On the other extreme of this spectrum is the cricket paralysis virus intergenic (CrPV IGR) IRES, which requires no initiation factors beyond the 40S ribosomal subunit for start codon recognition (Wilson et al. 2000; Jan & Sarnow 2002). Indeed, not even the ternary complex is required, as the structure of the CrPV IGR IRES actually mimics the AUG/initiator tRNA anticodon loop in the P site (Costantino et al. 2008) and can directly enter elongation. The IRES possessed by Hepatitis C virus (HCV) falls between the extremes of IRESs from viruses such as EMCV and CrPV in terms of translation factor requirements (Figure 1.4). The HCV IRES and related classical swine fever virus (CSFV) IRES only require eIF3 and the eIF2 ternary complex for AUG recognition, although eIF5 and eIF5B still facilitate subunit joining (Pestova et al. 1998b). Recent evidence suggests an intriguing alternative pathway for HCV translation initiation in which the need for eIF2 could be bypassed during times of stress by eIF5B-mediated recruitment of initiator tRNA (Pestova et al. 2008; Terenin et al. 2008).

There is also diversity in viral IRESs in terms of whether scanning-based mechanisms are used even after internal recruitment of 40S ribosomal subunits (Figure 1.4). IRESs in the HCV and CrPV classes are capable of placing their start codon directly into the P site and thus do not require scanning, nor do they recruit the necessary factors to be able to scan. Picornavirus IRESs, on the other hand, still recruit many scanning-associated factors and do frequently need to scan from their initial point of internal recruitment to place the initiation codon in the P site (Jang 2006).
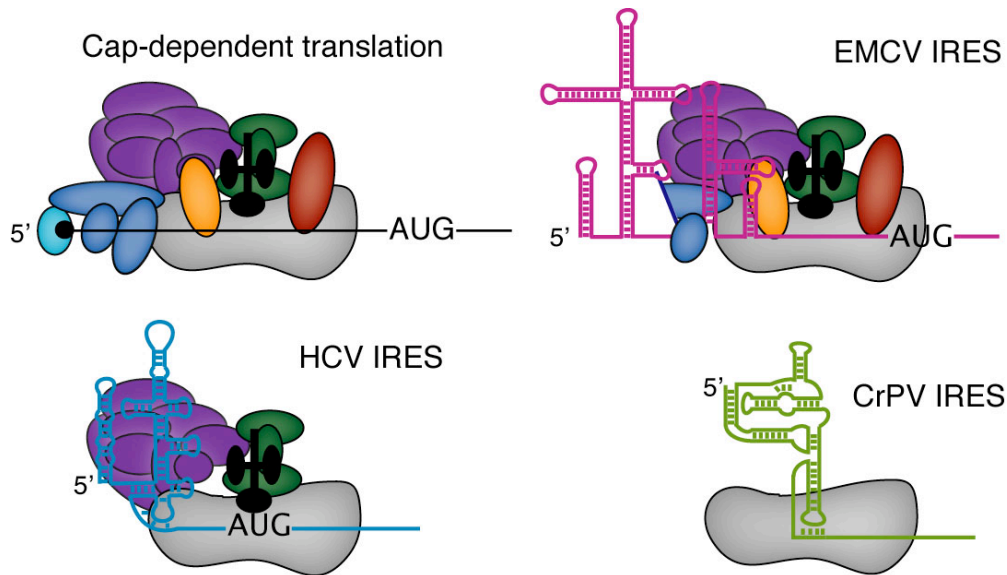
Figure 1.4: Diversity in translation initiation mechanism of viral IRESs. Comparison of initiation factor- and scanning- requirements for cap-dependent translation, encephalomyocarditis virus (EMCV), hepatitis C virus (HCV), and cricket paralysis virus (CrPV) IRESs. The EMCV IRES shares many similarities with poliovirus IRES, and HCV and classical swine fever virus (CSFV) have very similar IRES structures. Translation factors are depicted with the same coloring scheme as in Figure 1.1.

While the general mechanisms and, in many cases, the structures of viral IRESs are well understood, there has been considerably less mechanistic work conducted thus far on cellular IRESs. It does not appear that most cellular IRESs have large, defined secondary or tertiary structures as are known for viral IRESs; this may make their mechanisms more difficult to elucidate. Indeed, many cellular IRESs may turn out to be small stretches of primary and/or secondary mRNA structure that can help to recruit of translation factors and therefore provide a competitive advantage for their messages under conditions in which ribosomes and translation factors are limiting. For instance, a class of verified cellular IRESs in yeast contains A-rich stretches in the 5′ UTR, which help to recruit PABP internally and boost translation efficiency (Gilbert et al. 2007). Many IRESs may be stimulated by recruiting IRES trans-acting factors (ITAFs), such as the RNA binding proteins PTB and La, which are proteins that are not required for standard cap-dependent translation, but that stimulate specific IRESs (Meerovitch et al. 1989; Pilipenko et al. 2000). The importance of ITAFs for IRES translation is suggested by the observation that certain human IRESs, which are not translated efficiently in rabbit reticulocyte lysate (RRL), are stimulated by the addition of a small amount of HeLa extract, presumably due to the addition of a human-specific protein (Dorner et al. 1984).

1.5.3 Mechanism of HCV IRES translation

The HCV IRES adopts an ion-dependent structure with specific secondary and tertiary structural elements (Kieft et al. 1999). The RNA is capable of binding directly and independently to 40S ribosomal subunits and eukaryotic initiation factor 3 (eIF3; Pestova et al. 1998b; Kieft et al. 2001). Assembly studies suggest that the IRES binds first to 40S subunits and that this complex then recruits eIF3 and the ternary complex (TC) of met-tRNA$_i^{met}$-eIF2-GTP (Ji et al.

13

2004; Otto & Puglisi 2004). With help of eIF5 and eIF5B, GTP hydrolysis by TC allows for the 60S ribosomal subunit to join, forming an elongation-competent 80S ribosome (Pestova et al. 1998b; Locker et al. 2007). As mentioned above, it has recently been observed that the IRES may facilitate an alternative eIF2-independent pathway of translation under conditions of increased eIF2a phosphorylation in response to infection-stimulated PKR (Pestova et al. 2008; Terenin et al. 2008).

  Significant progress has been made in mapping various functions of the IRES to specific structural domains (Figure 1.3). The basal portion of domain III is required for 40S recruitment (Kieft et al. 2001) while regions in domain IV and domain II are also contacted by the 40S subunit, but do not contribute greatly to the binding affinity of the IRES for the 40S subunit (Kolupaeva et al. 2000; Kieft et al. 2001; Spahn et al. 2001). The apical IIIabc domain contributes to the affinity of the IRES for eIF3 (Sizova et al. 1998; Kieft et al. 2001). Domain II binds in the tRNA exit site leading to a conformational change in 40S subunits (Spahn et al. 2001), and promoting eIF5-mediated GTP hydrolysis (Locker et al. 2007).

  Recent discoveries have implicated miR-122, a liver-specific microRNA, in the life cycle of HCV, and this may play a large role in the liver-cell specificity of HCV infection. Two copies of miR-122 are able to bind upstream of the IRES in the 5′ UTR of the HCV genome in the region known as dom I (Figure 1.3; Jopling et al. 2005; Jopling et al. 2008). The binding of this miRNA to dom I of the HCV 5′ UTR greatly enhances viral propagation (Jopling et al. 2005), and part of this effect may stem from stimulating translation by the IRES (Henke et al. 2008). This HCV RNA-miRNA interaction is critical enough for viral infectivity that sequestration of miR-122 with antisense oligonucleotides leads to substantially reduced viral loads in HCV-infected chimpanzees, and therefore has strong therapeutic potential (Lanford et al. 2010).

  Substantial structural information is available for the HCV IRES, both alone and bound to the 40S ribosomal subunit or to eIF3. The cryo-EM reconstruction of the HCV IRES bound to 40S subunits revealed several interesting features (Figure 1.5A; Spahn et al. 2001). First of all, the majority of the IRES binds on the solvent surface of the 40S, along the back of the platform domain, with dom II of the IRES reaching around toward the interface surface of the 40S and into the E site. Dom II is known to be important for stable positioning of the mRNA in the binding cleft (Pestova et al. 1998b; Kolupaeva et al. 2000) and may interact directly or indirectly with the mRNA in the E site. The RNA leading to the AUG and ORF should emanate from the body of the IRES from the backside of the platform, and likely travels the same path through the mRNA binding cleft as has been seen in prokaryotic crystal structures (Yusupova et al. 2001) and has been observed via cross-linking in eukaryotic complexes (Pisarev et al. 2008). Intriguingly, the 40S subunit conformational change induced by dom II of the IRES (Spahn et al. 2001) is similar to that induced by eIF1 and eIF1A (Passmore et al. 2007), suggesting common mechanisms between cap-dependent and -independent translation initiation.

A **40S**-**HCV IRES** reconstruction,
viewed from interface surface
(Spahn et al., 2001)

B **HCV IRES**-**eIF3** reconstruction
(Siridechadilok et al., 2005)

~90°

C **40S**-**HCV IRES**-**eIF3** model,
viewed from side/solvent surface
(Siridechadilok et al., 2005)
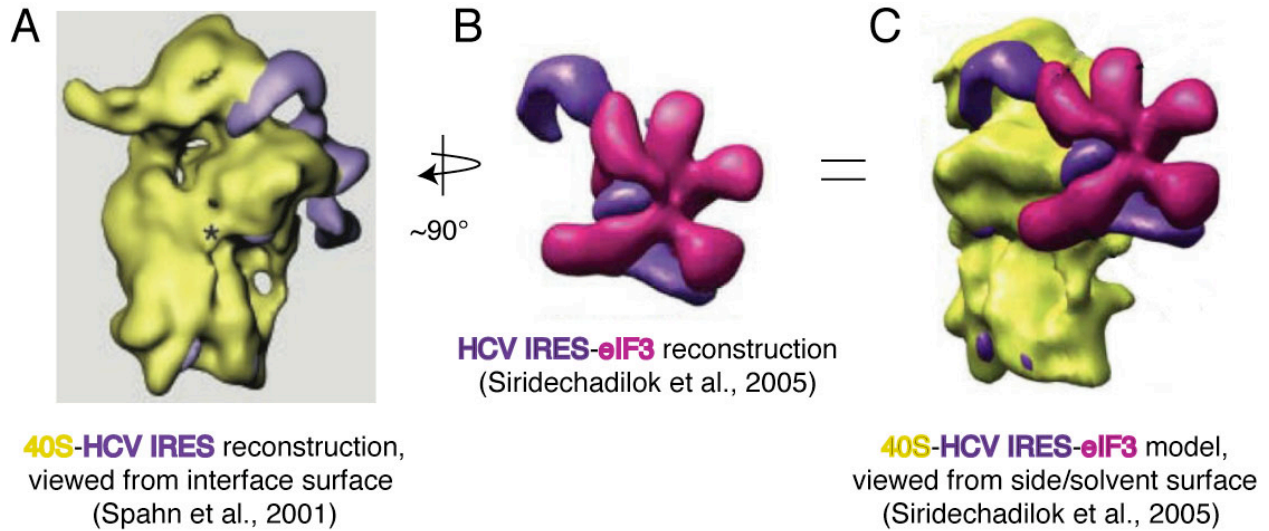
Figure 1.5: Structural models for 40S/HCV IRES and HCV IRES/eIF3 complexes. Cryo-electron microscopy reconstructions for the binary complexes of *(A)* 40S/HCV IRES and *(B)* HCV IRES/eIF3, and *(C)* a model for the ternary complex of 40S/HCV IRES/eIF3 all together. Solvent and interface (60S-binding) surfaces of the 40S subunit are indicated. Due to the fact that the HCV IRES can bind to 40S ribosomal subunits and eIF3 either independently or simultaneously, it is sensible that eIF3 will bind in the same general location on the 40S subunit as the HCV IRES; both are to bind predominantly to the solvent surface of the 40S subunit. Note: representations of structures are taken directly from others' publications (Spahn et al. 2001; Siridechadilok et al. 2005).

A cryo-EM reconstruction of the HCV IRES bound to eIF3 has also been solved (Figure 1.5B; Siridechadilok et al. 2005). This structure suggests a fairly large surface of interaction between the IRES and eIF3, perhaps more than expected from biochemical analysis of the interaction (Sizova et al. 1998; Kieft et al. 2001). The IRES was observed to populate three discrete conformations when bound to eIF3, underscoring the conformational flexibility of the IRES. Of these three conformations, the most prevalent was nearly identical to the IRES conformation when bound to 40S subunits (Spahn et al. 2001), allowing a ternary complex of the HCV IRES, eIF3, and 40S subunits to be modeled (Figure 1.5C; Siridechadilok et al. 2005). The lack of steric clashes between 40S and eIF3 in this model suggests that minimal conformational changes are required from the IRES when binding to both 40S subunits and eIF3 simultaneously. This model also suggests that, even during cap-dependent translation in the absence of the IRES, eIF3 binds largely to the solvent side of the 40S subunit, behind the platform domain and next to the E site. Indeed, cross-linking studies suggest that eIF3 may help to extend the mRNA binding cleft beyond the E site, with two subunits of human eIF3 cross-linking to mRNA 5′ of the start codon (Pisarev et al. 2008).

The structure of the HCV IRES RNA has also been well studied at high resolution; NMR spectroscopic and x-ray crystallographic methods have been combined in a divide-and-conquer approach to solve the structure of various domains and sub-domains of the HCV IRES (Figure 1.6). Consistent with the sharp bend observed in cryo-EM reconstructions, crystal and NMR structures of domain II show a large degree of curvature at a bulge in subdomain IIa. The structures of many of the helical regions of dom III have been solved, but a structure is known

for only one of the three multi-helical junctions in dom III, and there is no known structure of the IIIef fourway junction or its associated pseudoknot. RNase probing experiments show that both double- and single-stranded specific RNases cleave the pseudoknot region, consistent with a high degree of conformational flexibility of the RNA (Wang et al. 1995; Kolupaeva et al. 2000; Fletcher et al. 2002).
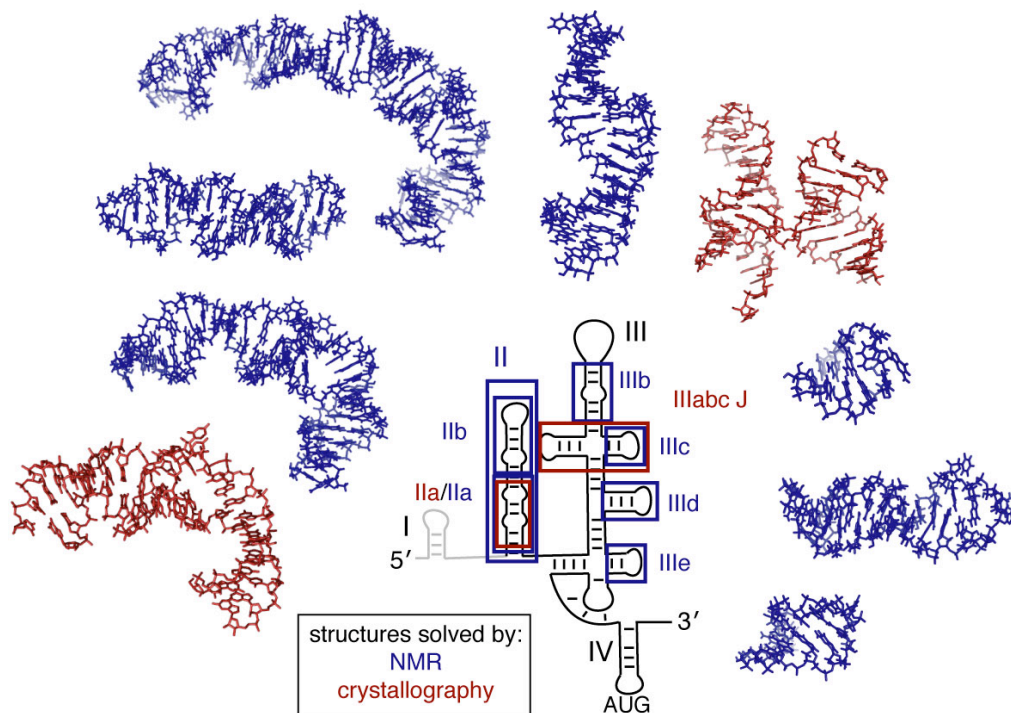
Figure 1.6: Solved structures of HCV IRES domains. The structures of individual domains of the HCV IRES that have been solved at high resolution are mapped onto the overall secondary structure of the IRES, with lowest energy NMR structures displayed in blue, and structures solved using x-ray crystallography shown in red. No high-resolution structural information is yet known for dom I, dom IV, the IIId junction, or the IIIef pseudoknot domain. References: IIIb (Collier et al. 2002), IIIc (Rijnbrand et al. 2004), JIIIabc (Kieft et al. 2002), IIId and IIIe (Lukavsky et al. 2000), II by NMR (Lukavsky et al. 2003), IIa by crystallography (Zhao et al. 2008).

1.6. Efforts to identify HCV IRES inhibitors

Due to the considerable interest in the HCV IRES as a drug target (Gallego & Varani 2002; McHutchinson et al. 2006), numerous groups have used diverse strategies to develop inhibitors against the IRES, from binding-based approaches, to functional screens, to nucleic-acid based and peptide-based inhibitors. These strategies take advantage of approaches that have been applied to RNA drug targets in general (Hermann & Westhof 2000; Bodoor et al. 2009).

At least two efforts have been made to screen for small molecules that can interact directly with the entire IRES or a region thereof in hopes that such small molecules might disrupt a key conformation of the IRES. A mass spectrometry-based screen was conducted for small molecules that could bind to dom IIa of the HCV IRES, the conformation of which is thought to be critical for IRES activity (Seth et al. 2005). This approach identified a class of benzimidazoles that showed good activity against an HCV IRES replicon assay without showing any significant

toxicity using one particular (MTT) assay. No published data, however, have directly validated that these compounds specifically inhibit the translation from the HCV IRES. Alternatively, it is possible that they could inhibit general translation or another process to which the replicon assay is highly sensitive (see appendix IB). Despite the lack of validation that the activity of this class of molecules stems from a direct and specific effect on HCV IRES-driven translation, biophysical characterization using FRET and NMR has shown that these benzimidazoles do indeed bind to a particular loop in dom II RNA and that their binding disrupts the overall conformation of the RNA (Parsons et al. 2009; Paulsen et al. 2010).

Another binding-based approach to look for HCV IRES inhibitors took advantage of *in vitro* evolution system based on mRNA display to evolve cyclic peptides that could bind to HCV IRES RNA. Despite preliminary reports that a cyclic peptide was identified which bound tightly and specifically to the HCV IRES and could inhibit *in vitro* translation from the IRES (Litovchick & Szostak 2008), these findings could not be repeated in our laboratory (see appendix 1A) and this report has since been retracted (Litovchick & Szostak 2009).

A functional *in vitro* translation screen for cap-dependent and HCV IRES inhibitors has been conducted in Krebs extracts (Novac et al. 2004). This screen utilized a bicistronic message, possessing a firefly luciferase driven by a 5′ cap and a renilla luciferase driven by an internal HCV IRES sequence. Results from bicistronic messages are not necessarily straightforward as the two ORFs are physically coupled and so the translation of the two ORFs is not fully independent. This screen yielded new classes of general translation inhibitors, but no HCV IRES-specific inhibitors were reported (Novac et al. 2004).

A conceptually similar screen was conducted by ISIS pharmaceuticals using bicistronic reporter RNAs. While this screen appears to have identified a biaryl guanidine scaffold which initially showed ~2.5-fold selectivity for IRES- over cap-dependent translation, fairly extensive structure-activity relationship (SAR) studies were not able to improve this specificity to even 10-fold (Jefferson et al. 2004). In addition, very minimal validation data are shown to prove the specificity or mechanism of action of the compounds.

Certain groups have approached the challenge of designing HCV IRES inhibitors by trying to target putative IRES trans-acting factors for the IRES, such as the La protein. A peptide from the C-terminus of La has been shown to act as a dominant negative version of La, competing with the full-length protein for binding to the IRES RNA and blocking assembly of 48S complexes (Pudi et al. 2005). Interestingly, a peptide from the N-terminus of La also inhibits HCV IRES translation *in vitro* and *in vivo*, but does not compete with La for binding with the IRES (Izumi et al. 2004). Instead, the N-terminal La peptide seems to bind to other putative ITAFs PTB and PCBP (Fontanes et al. 2009).

Antisense nucleic acid-based strategies have also been very popular to target the HCV IRES. Aptamers that bind to the IIId loop or to the AUG-containing dom IV, with or without an attached ribozyme domain, have showed efficient inhibition of IRES translation activity (Kikuchi et al. 2005; Romero-Lopez et al. 2007; Romero-Lopez et al. 2009). Very short (6-10mer) peptide nucleic acids (PNAs) that target these two regions also inhibit the IRES (Alotte et al. 2008). DNAzymes targeted against different regions of the IRES have also been shown to be effective inhibitors *in vitro* and in cell culture (Roy et al. 2008). In addition, siRNAs targeted against dom II of the IRES have been shown to specifically degrade IRES RNA from all six genotypes in Huh7 cells (Prabhu et al. 2006).

The nucleic acid-based inhibitor approach has progressed the furthest with an antisense oligonucleotide which blocks 40S binding by binding to the IIId loop of the HCV IRES (Tallet-

Lopez et al. 2003). Although this nucleic acid-based candidate (ISIS 14803) reached phase I clinical trials, clinical development was halted due to limited on-target activity and aminotransferase flares (McHutchison et al. 2006). Whether this will prevent development of other nucleic acid-based inhibitors of the IRES remains to be seen. While certain nucleic acid-based drugs have proved effective (Nimjee et al. 2005; Alvarez-Salas 2008), efforts to develop small, drug-like, molecules as IRES inhibitors would no doubt provide many advantages in the clinic.

Chapter 2

Impact of eIF3j in recruitment of mRNA by human ribosomes

*A portion of the work presented in this chapter has been previously published as part of the following paper: Fraser, C.S., Berry, K.E., Hershey, J.W., Doudna, J.A. (2007). eIF3j is Located in the Decoding Center of the Human 40S Ribosomal Subunit. *Molecular Cell* 26, 811-819.

*Dr. Christopher Fraser performed the eIF3j-BABE probing experiments and the published eIF3j and eIF1A anisotropy experiments. Katherine Berry performed the published mRNA anisotropy experiments and all of the unpublished work (Preliminary Results) described.

2.1 Introduction

As discussed in Chapter 1, eukaryotic translation initiation is a complex multi-step process that allows for accurate initiation of translation from the correct AUG start codon and regulation of global protein synthesis levels. Eukaryotic initiation factor 3 (eIF3) plays a central role in this process, as it binds directly to the 40S ribosomal subunit, and helps to prevent premature joining of the large subunit. In addition, eIF3 helps to recruit other factors such as the ternary complex and cap-binding factors, and to facilitate scanning (Hinnebusch 2006). eIF3 consists of thirteen individual protein subunits in humans, six of which are conserved in *S. cerevisiae* (Hinnebusch 2006). In humans, eIF3 is likely to be one of the first factors to bind to 40S ribosomal subunits during translation initiation as, along with eIF1 and eIF1A, it has been shown to play a key role in ribosome recycling, and would therefore be present on 40S subunits as they are released from post-termination 80S ribosomes (Pisarev et al. 2007). A cryo-electron microscopy (cryo-EM) model suggests that the majority of eIF3 binds on the solvent side of the 40S subunit (Siridechadilok et al. 2005). eIF3 may extend the mRNA binding cleft by binding mRNA beyond the E site of the 40S subunit, as it is known to interact nonspecifically with RNA (Pisarev et al. 2008).

A fairly large surface of interaction is predicted between eIF3 and the 40S subunit from the cryo-EM model (Siridechadilok et al. 2005), but biochemical evidence suggests that one of the thirteen subunits of eIF3, eIF3j, plays an especially large role in facilitating binding between the eIF3 complex and 40S subunit. This j subunit of eIF3 is the most loosely associated to the rest of the eIF3 complex and is not always stoichiometrically bound to eIF3 (Valasek et al. 2001b; Fraser et al. 2004). *In vitro*, eIF3 purified without the j subunit does not stably associate with 40S subunits through a sucrose gradient (Fraser et al. 2004), although certain single stranded nucleic acids can help alleviate this binding defect (Kolupaeva et al. 2005). It has also been shown that HCR1, the eIF3j homolog in the yeast *S. cerevisiae*, assists in eIF3 recruitment to 40S subunits *in vivo* (Nielsen et al. 2006), although this result has since been questioned (ElAntak et al. 2010). Deletion of HCR1 in *S. cerevisiae* leads to slow growth, a late stage 40S ribosome biosynthesis defect, and a translation initiation defect (Valasek et al. 2001a). Human eIF3j can rescue the translation initiation defect, but not the 40S biogenesis defect, when heterologously expressed in *hcr1Δ* yeast strains (Valasek et al. 2001a). This underscores conservation between the yeast and human eIF3 complexes. Regulation of eIF3j-mediated recruitment of eIF3 to 40S subunits is known to contribute to the regulation of translation levels in human cells. For instance, the C-terminal 16 amino acids of eIF3j are removed by caspase-3 cleavage during apoptosis (Bushell et al. 2000), and the resulting j subunit recruits eIF3 much less efficiently to 40S ribosomal subunits (Fraser et al. 2004). During T cell lymphocyte activation, increased association of eIF3j with the eIF3 complex leads to increased eIF3/40S subunit association, contributing to upregulation of protein synthesis levels (Miyamoto et al. 2005).

Due to the key role eIF3j plays in recruitment of eIF3 to 40S subunits, there was interest in understanding where eIF3j binds on the 40S ribosomal subunit. Dr. Chris Fraser utilized directed hydroxyl radical probing to site-specifically examine the positions of eIF3j amino acids when bound to the 40S subunit (Fraser et al. 2007). For this technique, nine single-cysteine positions in eIF3j were labeled with an Fe(II)-EDTA moiety. The positions of ribosomal RNA (rRNA) cleavages induced from Fe(II)-EDTA at each position in eIF3j/40S subunit complexes were mapped by primer extension reactions and modeled onto the structure of the homologous 30S bacterial small ribosomal subunit (Figure 2.1). It was quite a surprise to find that eIF3j does

not bind on the solvent side of the 40S ribosomal subunit with the bulk of eIF3, but rather on the interface surface in the mRNA binding cleft. Specifically, the C-terminal half of eIF3j binds in the decoding center of the 40S subunit, while the N-terminal amino acids do not induce rRNA cleavages and most likely bind to the rest of eIF3 via an interaction with the eIF3b RRM (Nielsen et al. 2006; ElAntak et al. 2007).
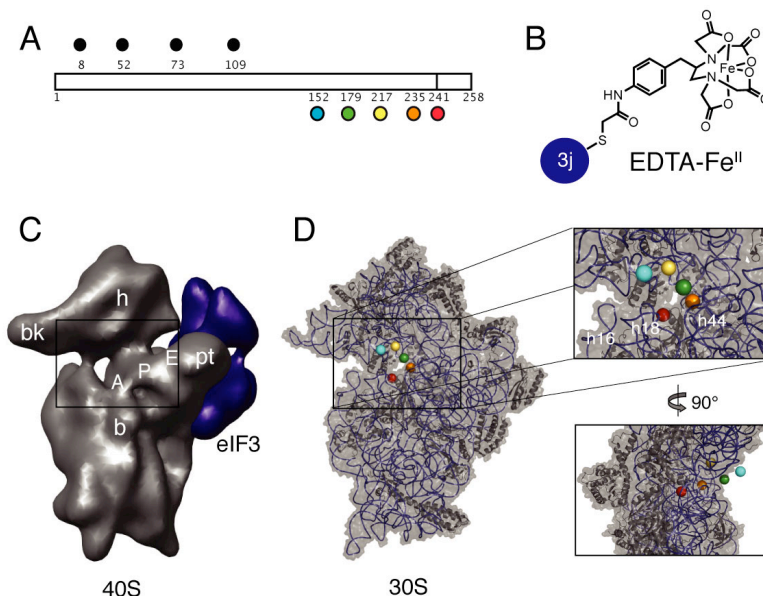
Figure 2.1: Mapping the position of eIF3j on the 40S Subunit. *(A)* Schematic of eIF3j indicating positions of single cysteine mutations introduced for Fe(II)-EDTA attachment sites. Position 241 indicates the site of cleavage by caspase-3. *(B)* Depiction of cysteine-conjugated iron EDTA moiety, utilized in site-directed hydroxyl radical probing. *(C)* Model of eIF3 bound to the 40S subunit based on cryo-EM reconstructions of eIF3-HCV-IRES and HCV-IRES-40S complexes (Siridechadilok et al. 2005). The 40S is shown in dark grey and eIF3 is shown in purple. Landmarks for the 40S subunit are indicated: A, A site; P, P site; E, E site; bk, beak; b, body; pt, platform; and h, head. The boxed area indicates the corresponding area boxed in the 30S subunit view shown in *D. (D)* Modeled positions of eIF3j amino acids, colored as in *A*, in the crystal structure of the *T. thermophilus* 30S (Ogle et al. 2001) with detailed views of the decoding site (right). Helices 16 and 44 of the 30S subunit are indicated as h16 and h44 respectively.

The fact that eIF3j binds in the decoding center of 40S subunits suggests that this protein may play roles in translation initiation other than only recruiting the eIF3 complex to the 40S subunit, and specifically, that it may interact with other interface-binding factors or influence the ribosome's interaction with mRNA. We utilized a recombinant biochemical system of human initiation factors to quantitatively examine eIF3j's interaction with mRNA and interface binding factors in order to define the role of this eIF3 subunit in the translation initiation pathway. We verify the location of eIF3j's binding by showing that it binds anticooperatively with eIF1A and find that the presence of eIF3j on the 40S subunit decreases the association of mRNA until the eIF2/Met-tRNA$_i$ ternary complex (TC) is present. Efforts to dissect the mechanism and consequences of these interactions are discussed.

2.2 Methods

2.2.1 Purification of 40S subunits

For each preparation, HeLa extract (150 ml; ~50 g cells) was thawed and supplemented with the following (final concentrations): 100 mM KCl, 5 mM Mg(OAc)$_2$, 0.5% sodium deoxycholate, 0.5% Triton X-100, 1x protease inhibitor cocktail. The resulting lysate was clarified by centrifugation at 20,000xg for 20 min at 4°C. The lysate was then layered onto two sucrose cushions in Ti45 Beckman ultracentrifuge tubes so that the final tube contains: 24 ml lysate; 24 ml cushion buffer containing 0.5 M sucrose; 17 ml cushion buffer containing 0.7 M sucrose (Cushion buffer: 20 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM Mg(OAc)$_2$, 1 mM DTT). The lysate was centrifuged in a Beckman Ti45 rotor at 28,000 rpm for 12 h at 4°C. Following centrifugation, the ribosome pellet was washed briefly with H$_2$O and resuspended in ribosome buffer (Ribosome buffer: 20 mM HEPES-KOH pH 7.5, 10% glycerol, 100 mM KCl, 5 mM Mg(OAc)$_2$, 1 mM DTT). The 80S ribosomes were further purified using a sephacryl S-400 HR (16/60) column (GE Healthcare). The column was run at 1 ml/min and 3 ml fractions were collected. The fractions containing 80S ribosomes were pooled and concentrated using an Amicon Ultra to a final concentration of between 100 and 200 (A$_{260}$) ODU/ml. A final concentration of 1.5 mM puromycin (Sigma) was added and the ribosomes were incubated on ice for 10 min and then 30°C for 10 min. Following incubation, the ribosomes were chilled on ice for 3 min and then layered over 10-40% (w/v) linear sucrose gradients containing separation buffer (Separation buffer: 20 mM HEPES-KOH pH 7.5, 300 mM KCl, 1 mM Mg(OAc)$_2$, 1 mM DTT) and centrifuged in a Beckman SW-32 rotor at 29,000 rpm for 12 h at 4°C. After centrifugation, each gradient was fractionated using an ISCO gradient fractionator, and the absorbance profile at 254 nm was monitored. Fractions were collected and the 40S ribosomal subunits and 60S ribosomal subunits were pooled and dialyzed into 2L storage buffer (Storage buffer: 20 mM HEPES-KOH pH 7.5, 75 mM KCl, 3 mM Mg(OAc)$_2$, 10 % glycerol, 1 mM DTT). The ribosomal subunits were then concentrated using an Amicon Ultra to between 70-100 ODU/ml, frozen in liquid nitrogen and stored at -80°C.

Hints: The rationale for purifying 80S ribosomes first and then splitting them apart before a final purification step is that the interface surfaces of the 40S and 60S subunits are particularly delicate and so it is best to keep them protected in the 80S ribosome as long as possible.

2.2.2 Purification of eIF1 and eIF1A

The expression vectors for eIF1 and eIF1A (Fraser et al. 2007) were transformed into *E.coli* BL21 (DE3) cells and grown in 1.5L LB medium supplemented with 50 μg/ml kanamycin at 37°C until the absorbance at 600 nm reached 0.5. Protein expression was induced with 1 mM IPTG, and the cells were grown at 30°C for an additional 4 h. Cells were recovered by centrifugation and resuspended in lysis buffer (Lysis buffer: 20 mM HEPES-KOH pH 7.5, 400 mM KCl, 10 mM β-mercaptoethanol and 1x EDTA-free protease inhibitor cocktail (Roche)) supplemented with 5 mM imidazole. Resuspended cells were lysed by sonication and then centrifuged at 20,000xg for 20 min at 4°C. The resulting supernatant was applied to 5 ml of Superflow Ni^{2+} resin (Qiagen) and then washed in 50 ml lysis buffer supplemented with 10 mM imidazole. The proteins were eluted from the column with 15 ml lysis buffer supplemented with imidazole (200 mM).

To remove the MBP fusion protein, 1 mg of recombinant His$_6$-tagged TEV protease was added and dialyzed over night at 4°C with the purified protein into cleavage buffer (Cleavage buffer: 20 mM HEPES-KOH pH 7.5, 10% glycerol, 1 mM DTT), supplemented with KCl (50 mM). Cleaved samples were filtered through a 0.2 μm syringe filter and applied to a Mono-S (10/10) column (GE healthcare) equilibrated in cleavage buffer supplemented with 50 mM KCl. eIF1 or eIF1A were eluted using a 120 ml linear gradient of KCl (50 mM to 500 mM) in cleavage buffer. The fractions containing eIF1 or eIF1A were pooled, concentrated to 2 mg/ml using an Amicon Ultra concentrator (Millipore) and frozen and stored in small aliquots at -80°C. This method of purification results in highly pure proteins with the addition of two amino acids (G and H) on the N-terminus, as a result of the expression construct.

Hints: If these proteins are not soluble at 50 mM KCl overnight before ion exchange chromatography, it is useful to dialyze into 100 mM KCl during overnight TEV cleavage, and dilute the protein 2-fold immediately before a hard spin and loading of the column.

2.2.3 Purification of eIF3

Post-nuclear HeLa cell lysates were a kind gift from Eva Nogales and Robert Tjian (HHMI/UC Berkeley). For each preparation, 400 ml HeLa extract (~150 g cells) was thawed and supplemented with the following (final concentrations): 10% glycerol, 1 mM EDTA, 1 mM EGTA, 50 mM NaF, 50 mM beta-glycerol phosphate (βGP). The lysate was then centrifuged, to remove mitochondria, in a Beckman JA-10 rotor at 7000 rpm for 20 min at 4°C. Following centrifugation, the supernatant (S10) was brought to 400 mM salt with the addition of KCl and was allowed to stir 2 min on ice. The lysate was then centrifuged in Beckman Ti45 rotor at 44,000 rpm for 4 h at 4°C. Following centrifugation, the middle 4/5 of the lysate was removed, avoiding the pellet and lipids at the top. This supernatant was brought to 40% ammonium sulfate with the addition of a saturated solution, was allowed to stir 1 hour at 4°C, and was subsequently centrifuged at 10,000xg for 20 min at 4°C. This pellet (the A cut) was saved at -20°C. To the A cut supernatant was added solid ammonium sulfate to a final concentration of 70%. The solution was allowed to stir 1 hour at 4°C, and was subsequently centrifuged at 10,000xg for 20 min at 4°C. This pellet (the B/C cut) was saved at -80°C for eIF2 purification (see below). The A cut pellet was resuspended in Buffer X (20 mM HEPES-KOH, pH 7.5, 10% glycerol, 100 mM KCl, 1 mM EDTA, 1 mM EGTA, 50 mM NaF, 50 mM βGP, 1 mM DTT) and dialyzed into Buffer X for 2.5 h. The solution was then centrifuged at 20,000xg for 25 min at 4°C, passed through a 0.45 μm filter, and applied to a Mono-Q (10/10) column (GE healthcare) equilibrated in Buffer X with 10% Buffer Y (20 mM HEPES-KOH, pH 7.5, 1 M KCl, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 50 mM NaF, 50 mM βGP, 1 mM DTT). eIF3 was eluted using a 120 ml linear gradient between 200 mM and 600 mM KCl, and 3 ml fractions were collected. The fractions containing eIF3 were pooled, dialyzed for 2 h into Buffer X with 5% Buffer Y, and applied to a Mono-S (10/10) column (GE healthcare) equilibrated in the same buffer. eIF3 was eluted using a 120 ml linear gradient of KCl (50 mM to 400 mM), and 2 ml fractions were collected. The fractions containing eIF3 were adjusted to 75 mM KCl with Buffer X and concentrated using an Amicon Ultra concentrator to a final concentration of between 1.4 and 2.5 mg/ml, frozen in liquid nitrogen and stored at -80°C.

2.2.4 Purification of eIF2

The B/C cut from eIF3 purifications (see above) was stored at -80°C, and the precipitate was resuspended in 50 ml Buffer B (20 mM HEPES-KOH pH 7.5, 1 mM DTT, 10% glycerol)

containing 50 mM KCl and dialyzed in 2 L of the same buffer for 2.5 h at 4°C. Following dialysis the lysate was passed through a 0.2μm filter and loaded onto a MonoQ (10/10) column (GE Healthcare) equilibrated in Buffer B containing 50 mM KCl. The column was eluted with a 120 ml gradient from 50-500 mM KCl in Buffer B at 2 ml/min, collecting 3 ml fractions. The fractions containing eIF2 were identified by SDS-PAGE and were pooled, dialyzed into Buffer B containing 50 mM KCl for 2.5 h at 4°C and loaded onto a Mono S (10/10) column equilibrated in the same buffer. The column was eluted with a 120 ml gradient from 50 to 500 mM KCl in Buffer B at 2 ml/min, collecting 3 ml fractions. The fractions containing eIF2 were pooled, dialyzed into Buffer C (Buffer C: 20 mM HEPES-KOH pH 7.5, 10% glycerol, 1 mM DTT, 100 mM KCl) containing 100 mM potassium phosphate pH 7.5 for 2.5 h at 4°C and then loaded onto a CHT5-1 ceramic hydroxyapatite column (Biorad). The column was eluted with a 50 ml 100-400 mM potassium phosphate gradient in Buffer C at 2 ml/min, collecting 2.5 ml fractions. This purification step was found to be essential in order to separate a minor contaminating protein, ABC50. The fractions containing eIF2 were pooled, concentrated using an Amicon Ultra to 1.5 ml and loaded onto a Superose 6 (16/60) column equilibrated in Buffer B containing 150 mM KCl. The column was run at 0.3 ml/min and 1 ml fractions were collected. The fractions containing eIF2 were pooled and concentrated using an Amicon Ultra to a final concentration of between 3 and 6 mg/ml and frozen in liquid nitrogen and stored at -80°C.

Hints: There is much less eIF2 in these extracts than eIF3. It is advisable to collect ammonium sulfate cuts from two or more eIF3 preps before continuing with an eIF2 prep. Protein levels throughout the prep are low enough that western blotting with an eIF2alpha antibody may be necessary to follow the protein over columns.

2.2.5 Purification of eIF3j

eIF3j was cloned, expressed using baculovirus infection of High-5 cells, and purified by Dr. Chris Fraser, as described (Fraser et al. 2007). Later experiments showed that eIF3j could also be effectively expressed in *E.coli* BL21 (DE3) from a pSV272 (His-MBP-TEV) or pet23a (His-TEV) expression vector, and purified with a nickel affinity column, ortho-nickel column following TEV cleavage, Mono-S ion exchange column, and S75 size-exclusion column, following standard procedures.

2.2.6 Preparation of human initiator tRNA

The initiator tRNA was *in vitro* transcribed with T7 polymerase using a hammerhead ribozyme construct, after digestion with BstN1, as previously reported (Batey et al. 2001). After 3 h at 37°C, the ribozyme was activated with the addition of 60 mM $MgCl_2$, the reaction was heated to 95°C for 2 min and cooled on ice 2 min (repeated twice). The cleaved tRNA was purified on a denaturing acrylamide gel (12%; 29:1), eluted into water, phenol/chloroform/IAA extracted (25:24:1, Sigma), and ethanol precipitated. The purified tRNA was allowed to refold by bring it to 2 mM $MgCl_2$, heating for 2 min at 80°C, and cooling for 2 min on ice (repeated twice). The folded RNA was charged *in vitro* using a purified tRNA synthetase (a gift from the Szostak lab) as previously described (Pestova and Hellen, 2001).

Hints: The detailed charging protocol was as follows:

MetRS was expressed and purified fresh for each charging reaction. MetRS was expressed in BL21 (PLysS) *E. coli* cells, induced with 1 mM IPTG for 4 h at 37°C. A single nickel affinity column was performed and the MetRS eluted was dialyzed into MetRS storage

buffer (MetRS storage buffer: 20 mM Tris-HCl pH 7.4, 60 mM NH4Cl, 50% glycerol), concentrated to ~23 mg/ml, and stored at -20°C until use.

Trial charging reactions (80 μL) were conducted, with similar conditions as below, but with varied concentration of MetRS (1, 3, and 8 μM) and charging times (0, 9, 15, 30, 45 min), since % charging can start to decline if the system is allowed to reach equilibrium. 5 μL aliquots were precipitated into cold TCA (5%) and filtered on Millipore MF-membrane filter (0.45 μm). Filters were dried and counts were measured by scintillation counting to determine ideal charging conditions.

The large scale reaction (2 ml) contained HEPES-KOH (50 mM), MgCl2 (20 mM), NH_4Cl (80 mM), DTT (5 mM), creatine phosphate (20 mM), ATP (10 mM), creatine phosphate kinase (280 μg/ml), methionine (10 mM), [^{35}S]-methionine (30 μL; 1175 Ci/mmol), BSA (400 μg/ml), MetRS (172 μg/ml; purified fresh), uncharged tRNA (228 μg/ml), MetRS storage buffer (516 μL) and was incubated at 30°C for 30 min (or whatever time was identified as ideal in trial reactions). Following incubation, aliquots (500 μL) were mixed with NaOAc (50 μL; 3 M, pH 5.2) and phenol/chloroform/isoamyl alcohol (500 μL; Sigma). Following vortexing and centrifugation, the aqueous layer was removed and extracted with chloroform (500 μL). Following vortexing and centrifugation, the aqueous layer was removed and RNA was ethanol precipitated, incubated at -80°C, and pelleted at high speed in a microcentrifuge for 10 min at 4°C. The supernatant was removed and the pellet was resuspended in NaOAc (0.1 M, pH 5.2, 300 μL total). Aliquots (150 μL) were loaded onto G25 syringe columns (1 ml; in 0.1M NaOAc, pH 5.2), and spun for 4 min at 1,000xg at 4°C. (Note: maintaining acidic conditions once tRNA is charged is critical to preventing hydrolysis of the unstable aminoacyl bond!). Concentration of tRNA was checked by A_{260}, and aliquots (100, 300, and 500 pmol) were flash frozen, lyophilized, and stored at -80°C until use. To determine degree of charging, a sample was counted by scintillation counting, as for small-scale reactions above. The amount of 35S (pmol) was compared to the amount of tRNA spotted (pmol), after assuming that scintillation counting is only 50% efficient for 35S. (Note: key parameters for these calculations are the ratio of cold:radiolabled methionine present in reactions, the μCi/μL of the [^{35}S]-methionine, and recalling that a pCi/pmol is equal to a cpm/pmol.)

2.2.7 Fluorescent labeling of eIF3j and eIF1A

For protein modification, single-cysteine versions of eIF3j and eIF1A were used, and the thiols of the cysteine residues were site-specifically reacted with fluorescein-5-maleimide.

Following dialysis into degassed modifcation buffer under nitrogen, eIF3j (200 μl; 70 μM) or eIF1A (200 μl; ~100 μM) was mixed with a 25-fold excess fluorescein-5-maleimide in modification buffer (Modification buffer: 50 mM HEPES-KOH, pH 7.0, 300 mM NaCl, 10% glycerol) and incubated at room temperature for 1 h. Following conjugation, eIF3j and eIF1A proteins were dialyzed over night at 4°C into buffer B containing 100 mM KCl, frozen in liquid nitrogen and stored at -80°C. All fluorescein-5-maleimide conjugated proteins were verified by mass spectrometry (data not shown).

Hints: The pH of modification buffer is essential for the maleimide to react with cysteine residues, but not with lysines or other nucleophiles. Modification reactions were performed under nitrogen in degassed buffer in order to avoid the need for DTT, the thiols of which would quench the maleimide. An alternative approach would be to modify proteins in the presence of the reducing agent TCEP, which does not react with maleimide groups.

2.2.8 Fluorescent labeling of model mRNA

The 20 nt RNA with the sequence 5′-AAGGAGGUAAAAUGUUUGCU-3′ was chemically synthesized by IDT and HPLC purified on a NucleoPac PA-100 (9x250) ion exchange column (Dionex) with a Waters HPLC system. The RNA was eluted with a lithium perchlorate gradient (from 20 mM to 200 mM) in buffer containing 20 mM sodium acetate in 9:1 water:acetonitrile, pH 6.5.

The 20mer RNA oligo was modified at the 5′ end using a two-step procedure in which the first step introduced a phosphorothioate moiety to the 5′ hydroxyl using ATP-gamma-S and a kinase, and the second step reacted this group with a maleimide-conjugated fluorescein dye. Because this RNA was chemically synthesized, the 5′ end was a free hydroxyl group and no phosphatase reaction was necessary. The RNA was heated at 65ºC for 5 min and cooled on ice. The kinase reaction (50 μL) was conducted in kinase buffer with annealed RNA (1.4-6.0 nmol), ATP-gamma-S (2 mM), RNasin inhibitor (1 μL; Fisher), T4 polynucleotide kinase (100 units; New England Biolabs) and was reacted at 37ºC for 3 h. The mixture was allowed to cool to room temperature, and ATP-gamma-S was removed with a Microspin G25 spin column (GE Healhcare). The resulting solution was carried on to the maleimide reaction without further purification. The phosophorothioated RNA was reacted with 2.3 mM fluorescein-maleimide (Molecular Probes) in a buffer of 100 mM HEPES-KOH pH 7.5:acetonitrile (1:1) for 4 h at 50ºC. The reaction mixture was extracted with phenol:chloroform:isoamyl alcohol (25:24:1; Sigma), then with chloroform alone, and the RNA was ethanol precipitated from the aqueous layer. A control reaction in which phosphorothioated RNA was omitted showed that the two rounds of organic extraction followed by ethanol precipitation removed all free fluorescein-5-maleimide. Thus, all fluorescein in the product, determined by fluorescence intensity (λ_{ex} = 490, λ_{em} = 518) compared to a standard curve, was assumed to be conjugated to RNA. The concentration of RNA was determined from the A_{260} of the product. These calculations showed that this labeling procedure results in ~50% recovery and ~80% labeling of the RNA.

For labeling of the model RNA with pyrene, the 20mer (Sequence: 5′-AAGGAGGUAAAAUGUUUGCU-3′) was synthesized with a 3′ amino group, with a 6-carbon linker, by Dharmacon. Amine-containing RNA (15 nmol) was reacted with N-(1-pyrenebutanoyl)cysteic acid, succinimidyl ester potassium salt (800 nmol; Molecular Probes) in 750 μM sodium tetraborate, pH 8.5, 14% DMSO (100 μL; final concentrations). The reaction was incubated at room temperature, protected from light, for 6 hr with occasional gentle vortexing, after which the RNA was extracted with phenol/chloroform/IAA, then extracted with chloroform alone, and then ethanol precipitated. After washing of the pellet, the RNA was resuspended in 20 mM HEPES-KOH pH 7.5 (100 μL; 70% was recovered), mixed with an equal volume of formamide loading dye, and purified on a 19% urea-PAGE gel, on which the pyrene-labeled RNA was retarded relative to the unlabeled RNA. The top (pyrene-labeled) band was cut, crushed, and eluted into DEPC water overnight at 4°C. RNA was extracted with phenol/chloroform/IAA, ethanol precipitated, and resuspended in DEPC water (32 μL). 10% of the starting RNA was recovered as pyrene-labeled, and another 8% of the starting RNA was recovered as unlabeled. The rest was lost to degradation or gel extraction.

2.2.9 Anisotropy-based binding experiments

Binding experiments with either fluorescein-labeled RNA or fluorescein-labeled eIF3j were conducted using a Wallac 1420 Multilabel Counter plate reading instrument (Perkin Elmer;

in Berger lab). The final concentration of the labeled species, indicated with an asterisk, was limiting (5-15 nM), and the concentration of 40S ribosomal subunits were increased in each reaction. The concentrations of other factors were >95% saturating. [eIF1] and [eIF1A] were 800 nM; [eIF3j] in *RNA experiments was 2.5 μM; [RNA] in *eIF3j experiments was 10 μM; the eIF2-tRNA-GMPNP was included at concentrations ranging from 350 nM for <30 nM 40S to 2 μM for 500 nM 40S ribosomes. Binding reactions (> 25 μL) were set up in buffer A (20 mM HEPES-KOH, 75 mM KCl, 2 mM Mg(OAc)$_2$, 10% glycerol, 1 mM DTT), heated for 1 min at 30ºC, and allowed to reach equilibrium at room temperature before the anisotropy was measured. Total fluorescence was also measured using the plate reader in order to account for quantum yield effects. Multiple repeated measurements showed that equilibrium had been reached. Plotting and fitting the data to obtain K_d values was conducted in one of two ways depending on whether the measured interaction was in the "tight-binding" regime (where the K_d is less than 10 times the final concentration of the labeled species) or in the "weak-binding" regime (where the K_d is more than 10 times the final concentration of the labeled species). In the tight-binding regime, we were able to reach saturation of binding by 1 μM 40S ribosomal subunits, so the maximum anisotropy of the complex could be directly measured. In the weak-binding regime, the maximum anisotropy of the complex could not be directly measured, but the difference in anisotropies between the factor at each concentration of 40S ribosome and that of the free factor could be fit to a simple hyperbolic binding isotherm to an estimated saturation point for the anisotropy. For each regime, the fraction bound was then calculated (incorporating quantum yield changes between the free and bound fluorophore) and fit to the solution of a quadratic equation describing an equilibrium reaction as previously reported (Maag & Lorsch 2003). Each binding experiment was repeated at least three times. The K_d values reported are the averages and the errors reported are the mean deviations. Control experiments showed that the anisotropy of the small molecule fluorescein-5-maleimide (quenched with beta-mercaptoethanol to prevent conjugation) was not affected by the presence of 40S ribosomes or BSA.

Hints: Because of the size of eIF3j (35 kDa), the fluorescence of conjugated fluorescein is already substantially depolarized before 40S subunit binding, based on the lifetime of its fluorescence (see helpful discussion in the Molecular Probes Handbook – Note 1.4). This makes the change in FP signal from eIF3j upon 40S subunit binding much smaller than from the ~6 kDa model mRNA. In order to use fluorescein FP for eIF3j-binding, various single-cysteine positions on eIF3j were tested, and S152C was found to be ideal. This position in eIF3j most likely possesses local motion, which is removed upon 40S subunit binding.

2.2.10 Filter binding experiments

For filter binding experiments, RNA was 5′ end labeled using γ-[^{32}P]-ATP and the binding reactions were set up in buffer A. For each experiment, four filters were utilized. From top to bottom, these were: a Tuffryn filter (Pall Corporation) to collect aggregates, a nitrocellulose membrane with 0.45 μm pores (Scleicher and Schuell) to bind protein-RNA complexes, a Hybond-N+ nylon membrane (Amersham Biosciences) to bind free RNA, and gel blot paper (Scleicher and Schuell) to retain any remaining material. The filters were presoaked in binding buffer before loaded onto the apparatus. Binding reactions (50 μL) were set up with 5′-[^{32}P]-end labeled RNA (~1 nM) and eIF3j (0-14 μM) in buffer A, incubated for 1 min at 30°C, and 10 min at room temp. An aliquot (40 μL) was loaded onto the filter apparatus as vacuum was applied to pull the sample through the filters. The filters were washed three times with binding buffer (100 μL), allowed to air dry 30 min, and quantified with phosphorimaging.

ImageQuant (Molecular Dynamics) was utilized to quantify the counts at each position on each of the filters, and the fraction RNA bound at each concentration of eIF3j was calculated using the equation fraction bound = (nitrocellulose)/(Tuffryn + Hybond + nitrocellulose); very little material was found on the Tuffryn or gel blot filters. A negative control was performed with BSA (14 μM), and a positive control was performed with eIF3 (300 nM), which is known to interact non-specifically with RNA with a ~50 nM K_d. Because less than 100% of the RNA bound to the positive control of 300 nM eIF3, the fraction bound was normalized to set the eIF3 positive control to 100% bound.

   Hints: Don't skimp on the volume of binding reactions, if possible. Data quality with at least 40 μL reactions is much higher than with < 20 μL. The vacuum should be just strong enough to pull the samples through within ~2 seconds. Nylon membranes come as charged or uncharged (Hybond-N+ vs. Hybond-N); the charged membrane is needed for short RNAs, but has occasionally been observed to diminish the observed interactions with proteins, if weak. As discussed in Appendix IA, a nitrocellulose membrane with smaller pore size may be necessary with small proteins.

2.2.11 Kinetics experiments

   For kinetics experiments, the 20mer model RNA was labeled with pyrene at its 3′ end for use in 40S subunit binding experiments (Studer et al. 2003; Maag et al. 2005). Pyrene is an environmentally sensitive fluorophore, and its fluorescence emission increases when in hydrophobic environments, such as when mRNA has bound to 40S subunits (Figure 2.2A). This change in pyrene-fluorescence intensity can be monitored over time as pyrene-labeled *mRNA and 40S subunits are mixed to measure the mRNA association rate; the dissociation rate can be determined by mixing pre-formed *mRNA-40S subunit complexes with "infinite" unlabeled 20mer RNA, such that once a labeled mRNA dissociates from 40S subunits, it is out-competed by unlabeled mRNA and essentially never rebinds (Figure 2.2B).

   For kinetics experiments, the fast-mixing device (in Kuriyan lab; estimated to have a mixing time of ~20 ms; Figure 2.2B) attached to a Fluoromax 3 fluorimeter (Jobin Yvon) was thoroughly washed with 0.2N NaOH, followed by extensive washing with DEPC water. Into one syringe was loaded 860 μL of pyrene-labeled RNA (160 nM) in buffer A (20 mM HEPES-KOH, 75 mM KCl, 2 mM Mg(OAc)$_2$, 10% glycerol, 1 mM DTT), and into the other 860 μL of 40S ribosomal subunits (160 nM or 800 nM), either alone or preincubated with eIF3j (2.5 μM), in buffer A, being careful to avoid bubbles. 300 μL binding reactions were conducted (150 μL from each syringe mixed in the cuvette), and the fluorescence intensity of the pyrene dye was monitored (excitation at 345 nm, emission at 378 nm, with both slits at 5 nm, integration and interval times varied between 0.005 and 0.04 sec, so as to obtain ~1000 total points per curve). In general, the first two reactions for any experiment were anomalous, and the 3rd-5th push were used for data evaluation. The fractions out of the stop syringe (containing *mRNA/40S/±eIF3j complex) were recovered and saved on ice. The A_{260} of each fraction was checked using a Nanodrop spectrophotometer, and the peak 1.2 ml were pooled for use in dissociation experiments. This material was reloaded into one of the syringes, and a solution of unlabeled 20mer RNA (3.5 - 11 μM) in buffer A was loaded into the second syringe. Dissociation experiments were conducted with similar parameters as above. Data were fit to a single exponential ($y = A + Be^{-kt}$) to give the observed rate constant, k. The residuals had a small deviation from a random distribution around zero, but were deemed to be acceptable to fit the first-order equation (with the advice of Dr. Jack Kirsch).
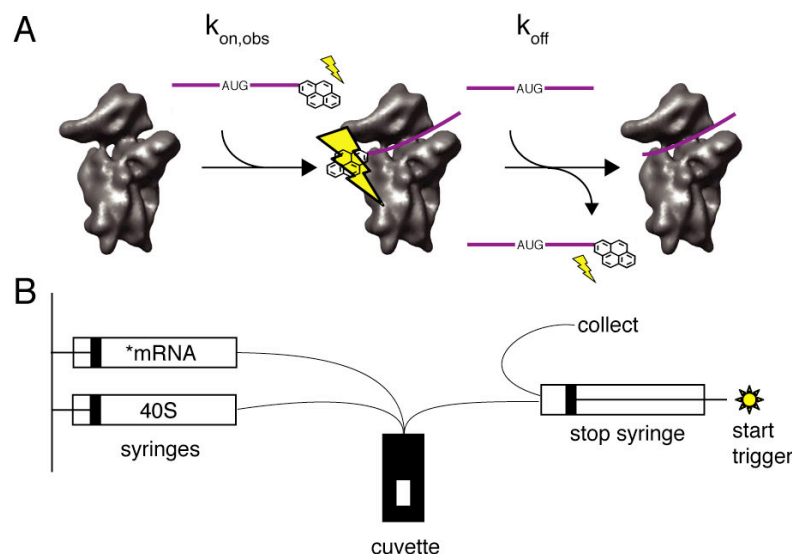
Figure 2.2: Setup of *mRNA-40S kinetics experiments. *(A)* Schematic showing how pyrene-labeled and unlabeled RNA were used to measure the rates of both reactions. *(B)* Schematic showing the construction of the fast mixing device connected to a FluoroMax 3 spectrofluorimeter (Jobin Jvon Horiba; c/o John Kuriyan lab). *mRNA/40S complex from the association reaction was collected from the stop syringe for recycling as the starting material to be mixed with "infinite" unlabeled RNA for the dissociation reaction.

Hints: Fluorescence experiments in the UV range are especially prone to interference by water Raman scattering, and these peaks can be much larger than fluorescence peaks. Always conduct a buffer blank scan to ensure that any peaks you observe in emission or excitation scans are actually due to fluorescence from the dye molecule. If there is any doubt whether you are observing fluorescence or scattering, alter the excitation wavelength by ~5 nm. For true fluorescence, the λ_{max} of emission will remain the same, with a slightly altered intensity. For scattering, the λ_{max} will shift as the λ_{ex} is altered.

For pseudo-first order kinetics, the [40S] should be at least 3x [mRNA], though 10x excess is best. This is because the pseudo-first order approximation assumes an infinite excess of the non-limiting reagent, such that its concentration does not change throughout the course of the association reaction (Frost & Pearson 1961). As the excess of the non-limiting reactant is lowered, the reaction approaches second order, and there is an increasingly large error associated with treating the reaction as first order. Depending on the extent of the reaction monitored, errors on $k_{on,pseudo-first}$ are estimated to be ~20-30% for a 3-fold excess, 10-20% for a 5-fold excess, and only 5-10% for a 10-fold excess (Sicilio & Peterson 1961). In addition, if a reversible association reaction does not go to completion (with concentrations >3-fold above K_d), then the observed association rate constant is the actual pseudo-first order rate constant + the dissociation rate constant (Frost & Pearson 1961). One caveat of going to a higher [40S] is that the pseudo-first order rate constant will be larger, and may be too fast for the instrument (20 ms mixing rate in the Kuriyan lab's fast mixing device). The expected pseudo-first order rate constant at any concentration can be calculated by first calculating the second-order rate constant (dissociation rate constant divided by the K_d), and then multiplying by the new concentration to give the

estimated pseudo-first order rate constant (e.g. 0.9 s^{-1}/19 x10^{-9} M = 4.7 x 10^7 M^{-1}s^{-1}; this predicts the pseudo-first order rate constant at 400 nM [40S] would be 19 s^{-1} and at 100 nM would be 4.7 s^{-1}).

The stopped-flow device in the Marqusee lab has a substantially shorter mixing time, but the syringes require larger volumes of material than the Kuriyan lab fast-mixing device.

2.2.12 Generation of *hcr1Δ* deletion strains

The parent *S. cerevisiae* strains used in the generation of *hcr1Δ* deletion strains were yWG3 and yWG25. yWG3 originally came from Reed Wickner, and is La-, making it useful for generating translation extracts (Searfoss & Wickner 2000). yWG25 (Σ1278b) originally came from Hiten Madhani, and is active for invasive growth.

The *HCR1* ORF was knocked out using *HIS4*, using the method of Longtine and Pringle (Longtine et al. 1998). PCR primers were F: 5′-CTATCCTAACCACCACCTCAAAAAAAA AAAAGTAATAAAACGGATCCCCGGGGTTAATTAA-3′ and R: 5′-AGATGGACAAGTTTA TCATAGCAAAGAAACAATAAGCAGAGAATTCGAGCTCGTTTAAAC-3′. The PCR product from the Longtine plasmid and these primers was cleaned up by phenol/chloroform/IAA extraction and ethanol precipitation. For yeast transformation with the generated PCR products, 5 ml ON cultures of yWG3 and yWG25 were grown in YPD (2% final glucose). Cultures were diluted into 25 ml YPD to OD$_{600}$ = 0.1 and allowed to grow to OD$_{600}$ = 0.5 (~5 h). Cells were pelleted in falcon tubes at 3000xg for 5 min at 4°C. Cells were washed with 25 ml sterile water, then washed in 1 ml sterile water, and 1/5 of these cells were transferred to a fresh Eppendorf tube. Transformation mix (33% PEG3350, 100 mM Li(OAc)$_2$, 300 ng/μL boiled ssDNA carrier, final concentrations) was mixed with the PCR product (at a ratio of 23:1), and this mixture was used to resuspend yeast cells. The cells were vortexed vigorously and incubated at 42°C for 40 min, after which cells were centrifuged for 30 seconds at 16,000 rpm in a room temperature microcentrifuge, and the transformation mix was removed. Cells were gently resuspended in 750 μL sterile water, plated onto three YPD -His plates, and incubated at 30°C for several days. Six colonies from each strain were re-struck on YPD -His plates to ensure that they were in fact His$^+$, genomic DNA was prepped to confirm the insertion of the *HIS4* gene at the *HCR1* locus by PCR with primers 200 nts upstream and downstream of the *HCR1* locus (5′-GGTCGTCCCTGTAT CATATTAGACGAACACACC-3′ and 5′-CGGTGCAAAAACACATTATATTCTCCATCTTG AAAACA-3′) and glycerol stocks were made from the positive isolates. The *hcr1Δ* strain in the yWG3 background was named yKB1, and *hcr1Δ* in yWG25 background was named yKB2.

2.2.13 Growth rate, polysome profiling and invasive growth assays

For growth-rate assays, overnight cultures of WT and *hcr1Δ* strains in both yWG3 and yWG25 backgrounds were grown in YPD, such that they had reached the end of log phase, but had not settled into stationary phase. From these, 5 ml cultures were inoculated at OD$_{600}$ = 0.1, and were grown at 30°C, shaking at 180 rpm. The OD$_{600}$ was checked every 100 minutes and the OD$_{600}$ vs. time was fit to first order exponential equations to yield the growth rate, which was converted to doubling time.

For polysome profiling, 200-400 ml cultures were inoculated from overnight cultures to between OD$_{600}$= 0.002 and 0.02, and allowed to grow at 30°C until the cultures reached either OD$_{600}$ = 0.7 or 1.2. The cultures were poured into chilled 500 ml centrifuge tubes containing cycloheximide (2 or 4 ml 10 μg/ml). (The centrifuge tubes were previously washed with 0.2N NaOH and DEPC water.) Cells were harvested by centrifugation in a Beckman JA-10 rotor at

9,000 rpm for five min at 4°C, washed with 50 ml polysome lysis buffer (20 mM HEPES-KOH, pH 7.4, 2 mM Mg(OAc)$_2$, 100 mM KOAc, 0.1 mg/ml fresh cycloheximide, 3 mM DTT, in DEPC water) and centrifuged as before. Cells were resuspended in 20 ml polysome lysis buffer, transferred to a tared Oakridge tube, pelleted by centrifugation in an Allegra 26R centrifuge (Beckman Coulter) at 3,500 rpm for five minutes at 4°C, and resuspended in 1 ml polysome lysis buffer + RNasin (1 U/µL) / 1 g cell pellet. To the cells were added 5 g cold glass beads / 1 g cell pellet, and cells were lysed by vortexing 6 x 20 seconds, with 30 seconds in between. The crude extract was clarified by centrifugation in an Allegra 26R centrifuge (Beckman Coulter) at 3,500 rpm for five min at 4°C, and the supernatant was further clarified by centrifugation in an Eppendorf microcentrifuge at 10,000 rpm for 10 min at 4°C. The absorbance at 260 nm of the lysate was measured and 10 ODU were loaded onto a 10-50% sucrose gradient in polysome lysis buffer and centrifuged in a Beckman SW41 rotor at 40,000 rpm for 1 hour and 45 min at 4°C. (Sucrose gradients were poured by hand by adding 2.2 ml of 50%, 40%, 30%, 20%, and 10% sucrose in polysome lysis buffer, separated by incubation for 15 minutes at -80°C. Sucrose gradients were unloaded from the bottom at 1.5 ml/min and A$_{260}$ was traced.)

For invasive growth assays, near-saturated cultures (40 µL) were spotted onto YPD plates, which were grown for two nights at 30°C. On the second morning, plates were washed under a gentle, steady stream of water to remove loosely attached cells that had not invaded the media.

2.2.14 Leaky scanning *in vitro* translation assays

The sequences of double AUG leaky scanning reporters are shown in Table 2.1. Inserts for pKB22, pKB23 and pKB29 were ligated into pWG186, digested with StuI and NarI, the sites for which fall between the T7 promoter and luciferase AUG. These constructs were cloned from overlapping oligonucleotides, as described in Appendix II.

Table 2.1: Sequences of leaky scanning reporters. Sequences of regions between T7 promoter and beginning of FF luc ORF**,** in pWG186 backbone. The T7 promoter is colored in red, StuI and NarI restriction sites in orange, upstream AUG in light blue and downstream AUG in royal blue.

Construct Name	Description	Sequence
pKB22	Each AUG in frame	TAATACGACTCACTATAGGCCTAAGCTTggttatccatacgacgttcca gattacgctggtggtaccATGggttatccatacgacgttccagattacg ctggtggtaccATGGAAGACGCCAAAA
pKB23	Second AUG in frame	TAATACGACTCACTATAGGCCTAAGCTTggttatccatacgacgttcca gattacgctggtggtaccATGggttatccatacgacgttccagattacg ctTggtggtaccATGGAAGACGCCAAAAA
pKB29	First AUG in frame	TAATACGACTCACTATAGGCCTAAGCTTggttatccatacgacgttcca gattacgctggtggtaccATGggttatccatacgacgttccagattacg cggtggtaccATGGAAaGACGCCAAAAA

Capped reporter RNAs were *in vitro* transcribed using T7 polymerase from ecl136II-digested plasmids. (This isoschizomer of SacI, ecl136II, was used for plasmid digestion because T7 polymerase prefers a 5′ overhang or blunt end over a 3′ overhang.) A transcription reaction (100 µL) contained digested template plasmid (7 µg), RNasin Plus RNase inhibitor (3 µL; Promega), ATP , UTP and CTP (2 mM), GTP (0.4 mM), m^{7}GpppG cap analog (2 mM; New England Biolabs), T7 polymerase (0.1 mg/ml; purified by Kaihong Zhou), DTT (10 mM), 30 mM Tris-HCl, pH 8.1, 10 mM MgCl$_2$, 2 mM spermidine, and 0.01% TritonX-100. After DNase I

treatment (RNase-free, Promega, 0.09 U/μL for 15 min at 37°C), RNAs were worked up using three iterative ethanol precipitations with ammonium acetate (0.5x vol 7.5M NH$_4$OAc and 2.5x vol EtOH added to aqueous layer) to remove free nucleotides, resuspended in DEPC water, and stored at -20°C until use. The integrity of transcribed RNAs was checked an ethidium bromide-stained agarose TAE gel (1%). Sodium dodecyl sulfate (SDS, 0.025%) was present in the 2x formamide loading dye and was essential for RNAs to run as clean bands in non-denaturing agarose gels.

Translation extracts were prepared as described (Iizuka et al. 1994), and *in vitro* translation experiments were conducted as previously described (Gilbert et al. 2007), with minor modifications. Translation extracts were nuclease treated with 33μg/uL S7 micrococcal nuclease with CaCl$_2$ (0.4 mM) for 5 min at RT. Nuclease treatment was stopped with the addition of EGTA (1.7 mM) and extracts were stored on ice until use. Translation reactions (15 μL) contained reporter RNA (~10 ng) and PMSF (0.3 mM), were incubated for 20 min at RT, and stopped with passive lysis buffer (7.5 μL; Promega).

2.3 Results

2.3.1 eIF3j interacts with eIF1A on the 40S ribosomal subunit

The location of eIF3j in the A site suggested that it may interact with eIF1A on the 40S subunit, which also binds in the A site (Yu et al. 2009). To test this, single-cysteine versions of eIF3j or eIF1A were labeled with fluorescein maleimide and used in anisotropy-based equilibrium binding experiments with purified human 40S ribosomal subunits. *eIF3j binds to 40S subunits on its own with 6 nM affinity and *eIF1A with 19 nM affinity (Figure 2.3); an asterisk refers to the fluorescently-labeled component of a binding reaction. A saturating excess of either factor, however, decreases the affinity of the other by ~12- to 15-fold; Doubling the amount of "saturating" cold eIF3j or eIF3A does not further change the observed affinity of the other (data not shown). Thus, eIF3j and eIF1A bind anticooperatively to the 40S subunit and therefore interact directly or indirectly on the 40S subunit, as predicted by the binding of both factors to the A site. Förster Resonance Energy Transfer (FRET) can be observed between fluorescein-eIF3j and AF568-eIF1A in the presence of 40S subunits, confirming that eIF3j and eIF1A can bind ribosomal subunits simultaneously. The strength of this FRET signal between eIF3j and the C-terminal tail of eIF1A varies based on which amino acid position of eIF3j was conjugated to the donor dye (amino acid 241 > 217 > 235 > 109 = 73 = 179 > 152; preliminary data not shown, collected by Zac Carrico).
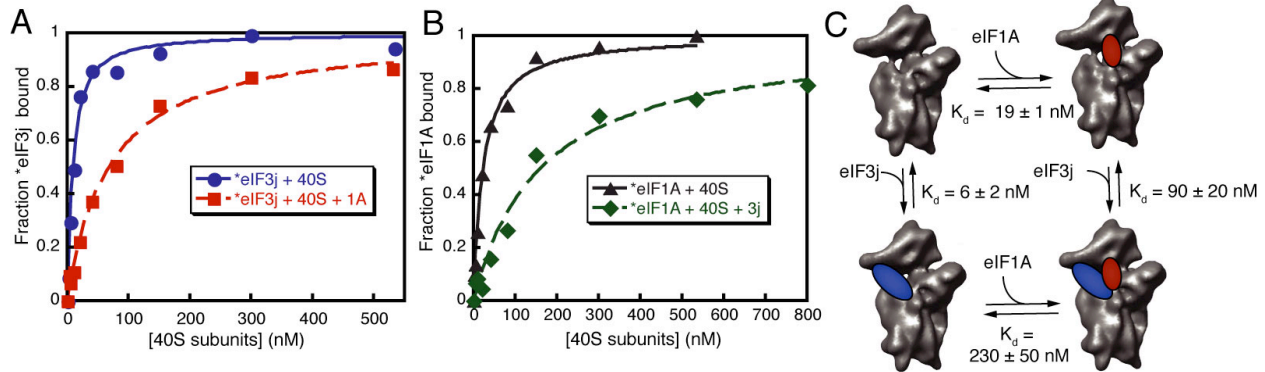
Figure 2.3: Thermodynamics of eIF3j-40S and eIF1A-40S subunit complexes. *(A)* Equilibrium binding of fluorescently labeled eIF3j to the 40S subunit in the absence or presence of a saturating concentration of eIF1A, as measured by anisotropy. *(B)* Equilibrium binding of fluorescently labeled eIF1A to the 40S subunit in the absence or presence of a saturating amount of eIF3j, as measured by anisotropy. *(C)* Diagram of the thermodynamic framework for the binding of eIF3j and eIF1A to the 40S subunit.

2.3.2 eIF3j influences mRNA affinity to the 40S ribosomal subunit

Based on the position of eIF3j in the mRNA binding cleft of the 40S ribosomal subunit, it seemed likely that mRNA bound to the 40S subunit might affect eIF3j's bound conformation, and therefore its affinity to 40S subunits. Indeed, a saturating excess of a model 20mer mRNA (Studer et al. 2003; Maag et al. 2005) decreases *eIF3j's affinity for the 40S ribosomal subunit even more drastically than eIF1A, from a K_d of 6 nM to 160 nM (Figure 2.4A). To examine the effect of eIF3j binding on mRNA affinity for the 40S subunit, this model mRNA was fluorescently labeled at its 5′ end, and used in anisotropy-based equilibrium binding experiments. This affinity of this *model mRNA for 40S ribosomal subunits is ~20 nM, and is reduced substantially in the presence of excess eIF3j (K_d ~270 nM). Doubling the amount of "saturating" cold eIF3j or mRNA does not further change the observed affinity of the other (data not shown), indicating that the observed negative cooperativity in their binding is unlikely to simply be a mass action effect in which eIF3j and mRNA directly compete for the same binding site. If this were the case, further increasing the concentration of one component would further decrease the apparent binding affinity of the other. Rather, it seems eIF3j and the model 20mer RNA can bind simultaneously to the 40S subunit, but when one is bound, the affinity of the other is reduced approximately 20-fold (Figure 2.4C). While there does appear to be a small degree of non-specific interaction between eIF3j and the 20mer model mRNA at 5-10 µM (Figure 2.5), this is not a strong enough interaction to account for the significant decrease in binding affinity of the model mRNA to the 40S subunit under the conditions tested ([eIF3j] ~ 2.5-5 µM). Later experiments revealed however that the strength of eIF3j's nonspecific interaction with RNA is increased as the RNA becomes longer, with gel shifts occurring at 15 µM eIF3j with RNAs longer than, but not shorter than, 100 nts (data not shown).

2.3.3 TC restores mRNA-40S subunit affinity in the presence of eIF3j

Messenger RNA is known to bind poorly to 40S ribosomal subunits when eIF3 is bound without eIF2/Met-tRNA$_i$ (Trachsel et al. 1977). Given that eIF3j is a subunit of eIF3, one of the

first translation factors to bind to the 40S subunit during translation initiation, we hypothesized that eIF3j might function to reduce the affinity of RNA for the mRNA binding cleft in the absence of the eIF2/Met-tRNA$_i$ TC. To explore this possibility, additional factors in the translation initiation pathway were added to mRNA-40S binding experiments in the presence of eIF3j, such as eIF1, eIF1A and TC. The addition of eIF1 and eIF1A only slightly increases mRNA binding affinity (K_d ~ 200 nM), but the addition of TC along with eIF1 and 1A strongly rescues mRNA binding affinity, bringing it back to ~30 nM even in the presence of eIF3j (Figure 2.4B,D). In contrast, in the absence of eIF3j, the addition of eIF1, 1A, and TC have less than a 2-fold effect on binding affinity of mRNA for the 40S subunit (Figure 2.6), indicating that TC's impact on mRNA affinity is eIF3j-dependent.
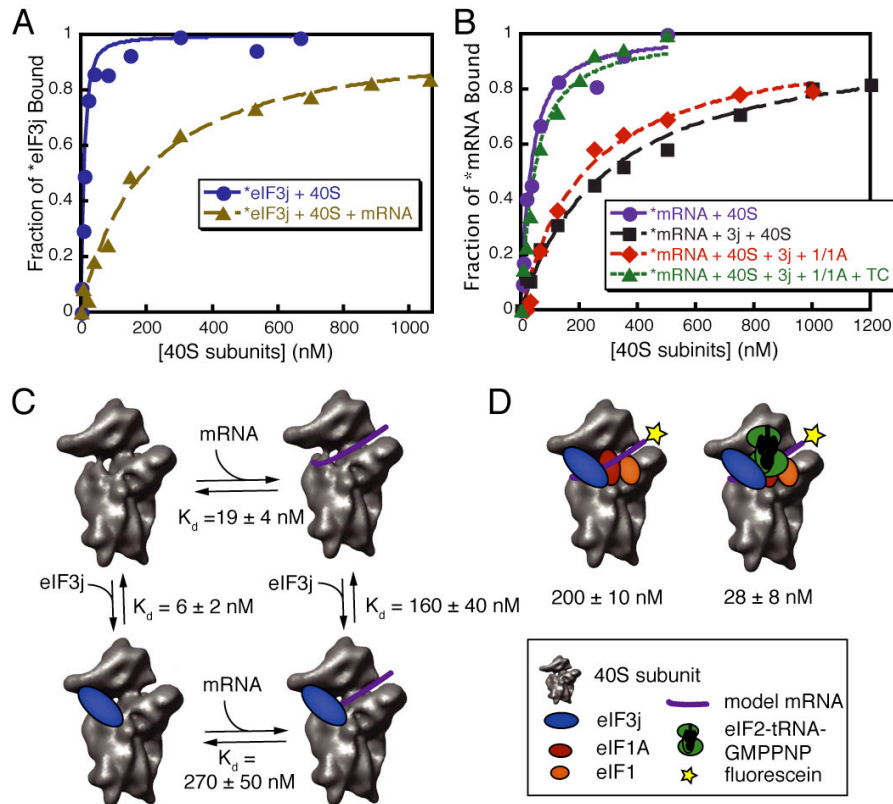
Figure 2.4: Effects of eIF3j and other factors on *mRNA-40S subunit affinity. *(A)* Equilibrium binding of fluorescently labeled eIF3j to the 40S subunit in the absence or presence of a saturating amount of a short unstructured mRNA. *(B)* Equilibrium binding of fluorescently labeled mRNA to the 40S subunit in the absence or presence of saturating concentrations of initiation factors, as measured by anisotropy. *(C)* Thermodynamic framework for binding of eIF3j and mRNA to the 40S subunit. *(D)* Summary of the K_d values for each experiment; the cartoons indicate components added.
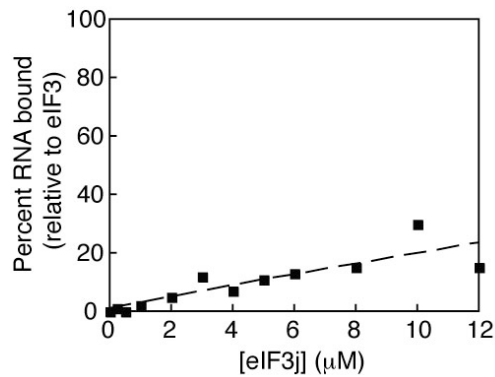
Figure 2.5: Nonspecific interaction between eIF3j and model mRNA. Equilibrium binding of 5′ end labeled model 20mer RNA to eIF3j, as measured by filter binding. The percent of RNA bound to eIF3j at each concentration was normalized to the amount of RNA bound to 300 nM eIF3. The data fit to a linear equation with a y-intercept of 1 and slope of 1.9.
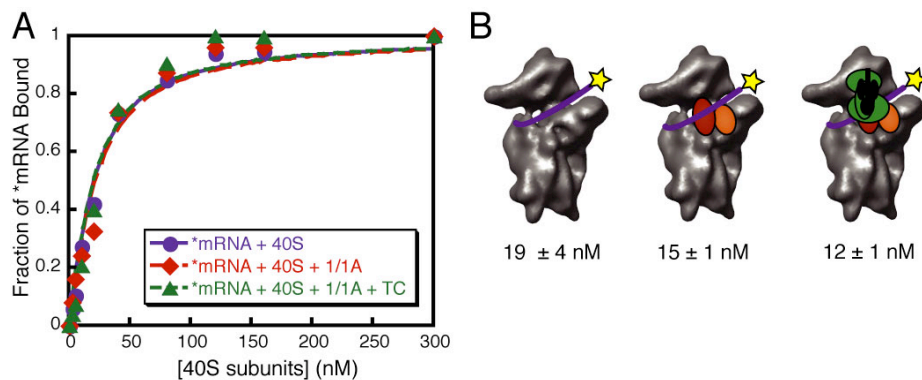
Figure 2.6: mRNA binding to the 40S subunit in the absence of eIF3j. *(A)* Equilibrium binding of fluorescently labeled mRNA to the 40S subunit in the absence or presence of saturating concentrations of initiation factors, as measured by anisotropy. *(B)* Summary of the K_d values for each experiment; the cartoons indicate components added.

2.4 Preliminary Results

With these results in hand, we wished to determine whether this restoration of mRNA affinity by TC depends on base pairing between the mRNA start-codon and anticodon of the tRNA, or whether it occurs independently of base pairing. In the latter case, it might play an important role in defining the association and conformation of mRNA in a pre-scanning initiation complex. In addition, we further explored the interaction of mRNA and eIF3j in the mRNA binding cleft of the ribosome, by studying the mRNA length-dependence of this interaction, as well as the kinetic basis for the decrease in mRNA affinity when eIF3j is present in the binding cleft. We also constructed an *hcr1Δ* strain of *S. cerevisiae* in order to look for conserved physiological defects in scanning and AUG recognition that could result from deletion of eIF3j/HCR1.

35

2.4.1 Mechanistic basis for TC restoration of mRNA-40S subunit affinity

In order to determine how much of TC's effect on mRNA affinity in the presence of eIF3j was independent of codon-anticodon base pairing, we attempted to disrupt the potential for base pairing through both the mRNA and tRNA sides of the interaction, though both strategies proved experimentally challenging. Initial efforts were aimed at mutating the anticodon loop of tRNA$_i$ to eliminate full codon-anticodon base pairing without altering the sequence of the model mRNA, since even small changes in the propensity of this model mRNA to form internal hairpins exert large changes on the observed binding affinity with 40S subunits, as the 40S subunit prefers binding to ssRNA. It had been previously shown that an initiator tRNA with a CCU rather than the WT CAU anticodon does not recognize AUG codons *in vivo*, but supports growth when an AGG start codon is used instead of AUG at an essential gene in yeast (Cigan et al. 1988), suggesting that this mutant tRNA is capable of being charged by the tRNA synthetase, forming ternary complexes, and interacting with pre-initiation complexes. We therefore chose this mutant CCU tRNA as a means of knocking out complete base pairing with the AUG-containing mRNA (Figure 2.7). This mutant tRNA could only be charged *in vitro* to ~18%, in contrast to typical WT tRNA$_i$ charging levels of >50% (data not shown), despite *in vitro* charging assays showing that the middle position of the anticodon is much less important than the outer positions for charging of *E. coli* initiator tRNA (Schulman & Pelka 1984).

Figure 2.7: Mutant CCU initiator tRNA. *Left*: Depiction of complete codon-anticodon base pairs between WT initiator tRNA and an AUG start codon. *Right*: Depiction of incomplete codon-anticodon base pairs between an AUG start codon and CCU mutant tRNA, which does not allow for effective translation *in vivo* in yeast.

Based on eIF2's preference for binding charged over uncharged tRNA in the presence of GTP (Kapp & Lorsch 2004), saturated TC can be achieved by using an excess of partially charged tRNA. Given that previous WT TC experiments had been conducted with a 2-fold excess of partially charged tRNA over eIF2, it was important to ensure that a 7-fold excess of partially charged CCU tRNA would not adversely affect the anisotropy-based binding assay. An examination of the effect on each individual component of TC on the change in the model mRNA's fluorescence polarization (FP) in the absence of 40S subunits showed tRNA on its own had no effect on *mRNA FP (Figure 2.8A). Only eIF2 had a significant influence on the FP of the model mRNA, presumably due to non-specific interactions with RNA, but this was eliminated in the presence of tRNA$_i$, with which eIF2 has a stronger, specific interaction (Figure 2.8A). To ensure that the apparent strengthening of mRNA affinity by TC is indeed an effect of authentic TC and not individual components, we directly compared *mRNA binding to 40S subunits in the presence of TC, excess or "normal" amounts of tRNA, or eIF2 alone in the presence of eIF3j and eIF1/1A (Figure 2.8B). Only the full TC caused a significant change in the

binding curve of *mRNA, and no difference was observed between normal levels and excess WT tRNA (Figure 2.8B). This confirmed that it really is the TC together, and not any individual components, that leads to the stability of *mRNA in the presence of eIF3j.

Figure 2.8: *mRNA FP in absence and presence of 40S subunits. *(A)* Fluorescence anisotropy signal of model 20mer RNA in the presence of increasing concentrations of indicated macromolecules or combinations thereof, in the absence of 40S ribosomal subunits. Note that eIF2 on its own is the only factor to cause a substantial change in anisotropy from the labeled model mRNA, and that this effect is relieved in the presence of tRNA as well. *(B)* Equilibrium binding of fluorescently labeled model 20mer RNA to 40S ribosomal subunits, in the presence of excess amounts of the indicated macromolecules, as measured by anisotropy.

With this reassurance, we wished to determine the equilibrium binding affinity of *mRNA for 40S subunits in the presence of eIF3j, eIF1/1A and mutant CCU TC, to compare it to the 28 nM affinity of *mRNA to 40S subunits in the presence of WT TC. Equilibrium anisotropy-based binding experiments were repeated several times, but significant day-to-day variability was encountered in the ΔFP of *mRNA, the shape of the binding curve, and therefore the measured K_d (Figure 2.9A). The measured K_d values varied dramatically from day-to-day, depending on the estimated saturation point. Estimates of K_d values obtained from these curves are 11±4, 11±3, 20±3, 80±50 and 120±20 nM. Certainly some restoration of mRNA binding affinity occurs with CCU TC, relative to the 200 nM K_d in the presence of eIF3j, eIF1 and 1A alone, but exactly how much was difficult to define. Part of the noisiness and variability of this assay arose because the binding reactions of *mRNA to 40S subunits in presence of CCU TC took much longer to equilibrate compared to reactions with 40S alone (Figure 2.9B,C).
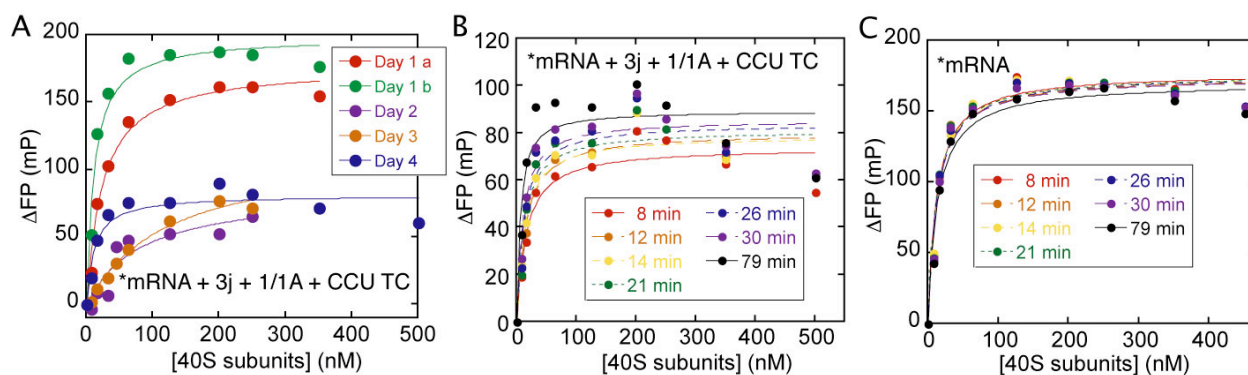
Figure 2.9: Inconsistency of FP data from *mRNA in presence of CCU TC. Attempted equilibrium binding of fluorescently labeled model 20mer RNA to 40S ribosomal subunits in the presence of excess eIF3j, eIF1, eIF1A, and TC containing mutant CCU initiator tRNA, as measured by anisotropy. *(A)* Dramatic variation in the change in fluorescence polarization values from repeated experiments on four separate days. For experiments on day 2 and day 3, the apparent binding affinity was weaker and saturation was not reached. In these cases, the estimated saturation point of the signal has a large effect on the absolute and relative binding affinities derived from the data. From fitting the raw ΔFP data, K_ds obtained ranged from 11±4, 11±3, 20±3, 80±50 and 120±20 nM. *(B)* Time course of change in fluorescence polarization values from Day 4 experiment in *A*. *(C)* Faster equilibration and smaller errors in time course of *mRNA/40S binding experiment, from the same day as *B*, in the absence of additional factors.

Given the experimental challenges encountered using CCU TC and the lack of reproducibility in measuring the binding affinity of *mRNA to 40S subunits in the presence of this mutant TC, the question of whether TC could restore mRNA binding affinity to 40S subunits in the presence of eIF3j was addressed using a mutant "nonAUG" 20mer model mRNA (Figure 2.10A). The nonAUG model RNA was fluorescently labeled at its 5′ end, and anisotropy-based equilibrium binding experiments were conducted with 40S ribosomal subunits. The K_d of *nonAUG to 40S subunits alone was found to be ~ 40 nM and this was reduced to >500 nM in the presence of saturating eIF3j (Figure 2.10B), a roughly similar effect to the original 20mer model mRNAs affinity being reduced from 19 nM to ~270 nM by eIF3j. In the presence of eIF3j, eIF1/1A and TC, there was a substantially smaller change in the FP signal of *nonAUG RNA and, again, the saturation point was difficult to define, but had a large influence on the measured K_d value. Based on our best estimate, the affinity of *nonAUG for 40S subunits in the presence of eIF3j, eIF1/1A and TC is around 140 nM. It was very challenging to quantify the exact amount of TC's effect that is base pairing-independent with these experimental approaches. Still, it appears that TC at least partially increases mRNA-40S subunit binding affinity in the presence of eIF3j even without codon-anticodon base pairing, indicating that this interaction may play an important role in a step of mRNA binding to a pre-scanning 43S complex (where there is not yet an AUG in the P site).

A Sequence of nonAUG 20mer:
5′-GACUCUCGCUCUCCGUCUCU-3′

B

Legend:
- ● - *nonAUG 20mer
- □ - *nonAUG + 3j
- ▲ - *nonAUG + 3j + 1/1A + TC
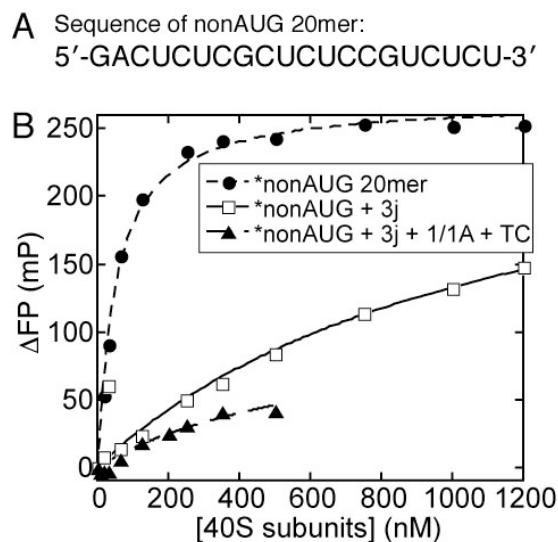
Figure 2.10: Effects of factors on *nonAUG mRNA-40S subunit affinity. *(A)* Sequence of model 20mer without an AUG (nonAUG 20mer) used to eliminate complete codon-anticodon base pairing in the presence of wild type TC. *(B)* Equilibrium binding of fluorescently labeled non-AUG 20mer RNA to 40S ribosomal subunits in the presence of excess indicated macromolecules. Note the small signal change in the presence of TC, related to the other conditions. The K_d derived from fitting these data was very different depending on whether it is assumed that saturation has been reached. From fitting the raw ΔFP data, the K_d obtained was 700±500 nM; from fitting the estimated fraction bound, the K_d obtained was 140±30 nM.

2.4.2 Length-dependence of mRNA/eIF3j interaction

We also wished to characterize the interaction of eIF3j and mRNA in the binding cleft more completely, and wondered whether the model 20mer RNA captured the entire interaction between mRNA and eIF3j in the binding cleft, given that 40S subunits protect ~30-40 nts of mRNA in footprinting experiments (Kozak & Shatkin 1976; Legon & Robertson 1976). We designed model 20mer, 30mer and 40mer mRNAs, which were matched in sequence, but extended on their 3′ ends, and predicted to each have a ΔG for hairpin formation of 0 kcal/mol (Figure 2.11A). These RNAs were fluorescently labeled at their 5′ ends and used in anisotropy-based equilibrium binding experiments with 40S ribosomal subunits. The *30mer and *40mer model RNAs each bind with an increased affinity relative to the 20mer (9 nM vs. 15 nM; Figure 2.11B,C). This suggests that the longer mRNA models can make more contacts through the mRNA binding cleft. In the presence of 2.5 μM eIF3j, the affinities of the *30mer and *40mer are reduced to ~40-120 nM. It appears as though eIF3j does not decrease these affinities as significantly as it decreases the affinity of the 20mer, which is reduced to ~440 nM (Figure 2.11C). This suggests that additional contacts made by the 30mer and 40mer in the binding cleft help them retain binding affinity in the presence of eIF3j. To confirm these effects, it would be important to double or triple the amount of eIF3j present to see that the effect on mRNA affinity has saturated, as it had for the original 20mer model mRNA.

39
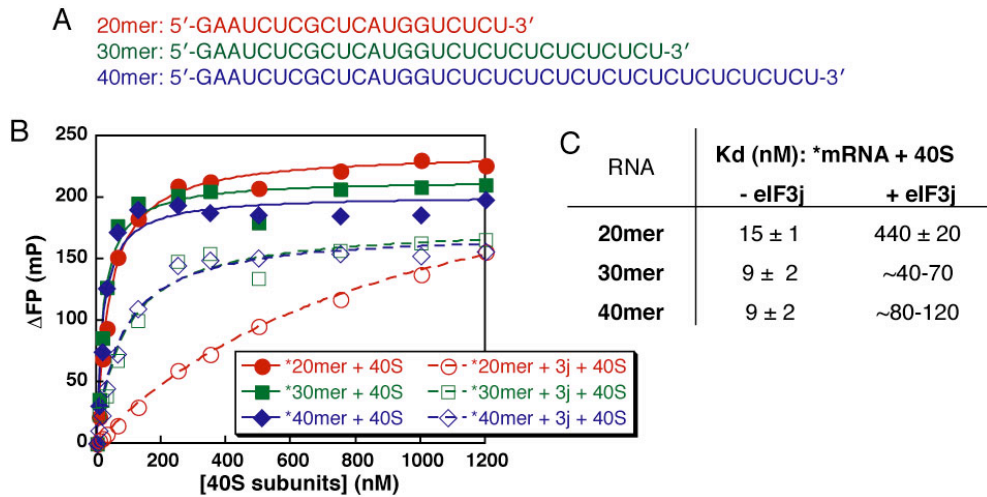
Figure 2.11: Length-dependence of mRNA/eIF3j interaction. *(A)* Sequences of matched 20mer, 30mer, and 40mer model mRNAs, each containing an AUG and predicted by mfold (Zuker 2003) to have a $\Delta G = 0$ for hairpin formation. *(B)* Equilibrium binding of mRNA to 40S subunits with mRNAs of different lengths, in presence and absence of eIF3j, measured using fluorescence anisotropy. *(C)* Summary of the K_d values for each experiment in *B*.

It would be very interesting to measure the affinity of *eIF3j for 40S subunits in the presence of each of these model mRNAs, to help confirm the simultaneous binding of eIF3j and mRNA. Unfortunately, the signal:noise for fluorescence polarization of the fairly large *eIF3j (35 kDa) is much worse than *mRNA, and it also appears that *eIF3j takes much longer to come to equilibrium than *mRNA (Figure 2.12). Thus, the fluorescence polarization signal from *eIF3j was not adequate to reliably make these measurements with precision required to quantitatively distinguish the effects of the mRNAs of different lengths on *eIF3j/40S binding affinity.
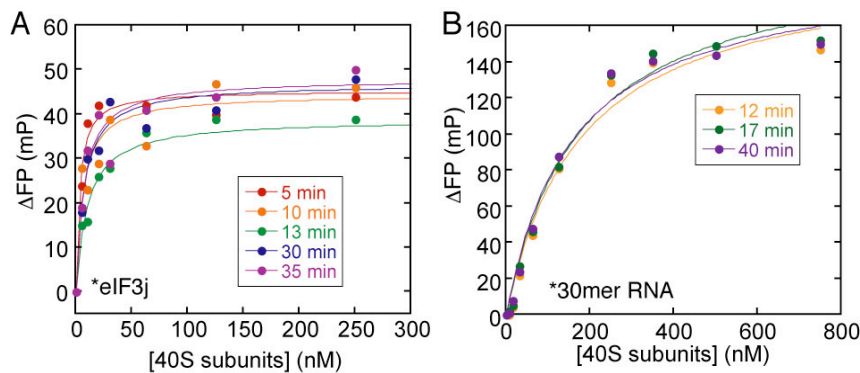
Figure 2.12: Comparison of FP-signal consistency between *eiF3j and *mRNA. Attempted equilibrium binding over time of *(A)* fluorescently-labeled eIF3j to 40S ribosomal subunits in the presence of a large excess of model mRNA or *(B)* fluorescently-labeled model mRNA to 40S subunits in the presence of a large excess of eIF3j, each as measured by anisotropy. Note the much larger noise variance of signal with time in *eIF3j signal than *30mer RNA signal.

40

2.4.3 Kinetic effects of eIF3j on mRNA-40S subunit interactions

eIF3j's function in decreasing the equilibrium binding affinity of mRNA for 40S subunits could arise from a slowing of the mRNA association rate or a speeding of the mRNA dissociation rate, or some combination thereof. How eIF3j exerts its equilibrium binding effect could shed light on the interaction of eIF3j and mRNA in the 40S binding cleft and so we wished to determine the effect of eIF3j on mRNA binding kinetics. For kinetics experiments, the 20mer model RNA, labeled with pyrene was used in 40S subunit binding experiments (Studer et al. 2003; Maag et al. 2005).

In order to conserve 40S ribosomal subunits, we set out to compare the $k_{on,obs}$ at a given concentration of 40S and mRNA in the presence and absence of eIF3j, instead of performing a [40S] titration to calculate the association rate constant (k_{on}) under pseudo-first order conditions. Initial experiments were conducted with 80 nM *pyrene-mRNA, in order to have sufficient fluorescence signal, and 80 nM 40S subunits, in the absence or presence of 1.25 μM eIF3j (Figure 2.13A,C). The observed k_{on} of 11.0 s^{-1} for the association of *mRNA to 40S subunits was reduced to 0.48 s^{-1} in the presence of eIF3j (Table 2.2). The bound *mRNA-40S subunit complex was collected from these association experiments, mixed with 3.5 μM excess unlabeled mRNA, and the decrease in fluorescence intensity in time was observed and fit to yield a single exponential rate constant (Figure 2.13B,D,E). The dissociation of *mRNA from 40S subunits yielded a simple curve giving a k_{off} of 0.9 s^{-1} (Figure 2.13B). This observed k_{off} was not significantly affected by doubling the concentration of excess eiF3j (data not shown). For the complex with eIF3j present, the dissociation curve was somewhat more complex, with a very fast phase in the first 10 seconds, which fits to a k_{off} of 0.87 s^{-1}, and a much slower process for which the fluorescence continued to decrease steadily past 2,000 seconds (Figure 2.13D,E). This slower phase of fluorescence-intensity loss reproducibly only occurs in the presence of eIF3j, so is not an effect of simple photobleaching, but may perhaps result from weak non-specific interactions with eIF3j.
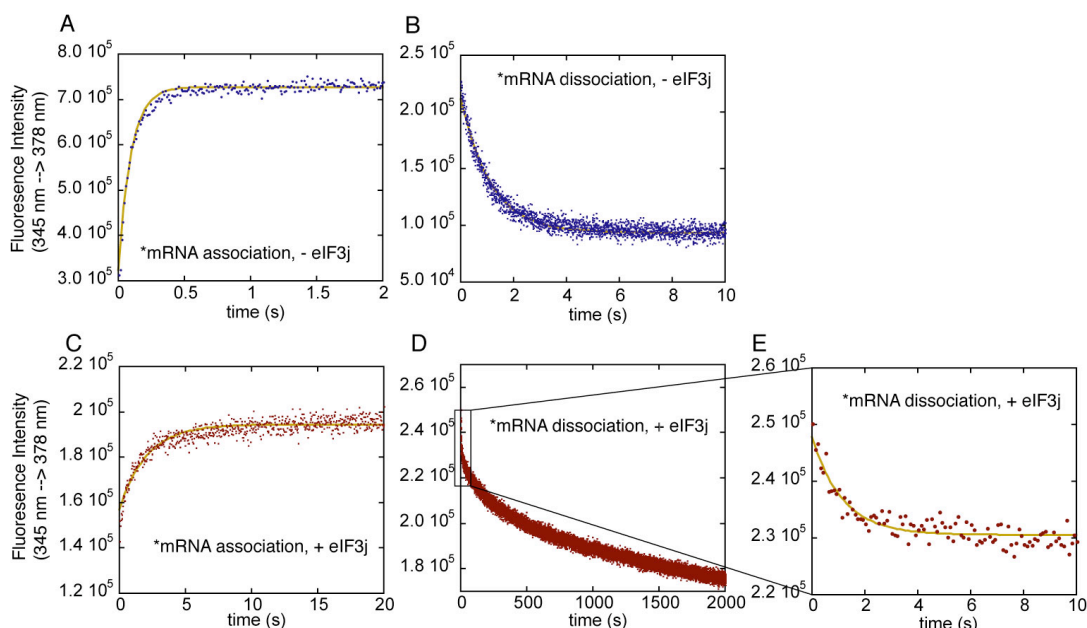
Figure 2.13: **Kinetics of *mRNA-40S subunit association and dissociation.** Preliminary measurements of association *(A,C)* and dissociation *(B,D,E)* rates of pyrene-labeled model 20mer RNA to 40S ribosomal subunits in absence *(A,B)* and presence *(C,D,E)* of excess eIF3j. Data were collected in a fast mixing device connected to a FluoroMax 3 spectrofluorimter (Jobin Jvon Horiba; c/o John Kuriyan lab). For association reactions, 80 nM *mRNA was mixed with 80 nM 40S subunits in absence or presence of 1.25 µM eIF3j (all final concentrations). For dissociation reactions, this *mRNA/40S/(eIF3j) complex was collected and mixed with >3.5 µM unlabeled 20mer RNA.

Table 2.2: Kinetic parameters for *mRNA / 40S subunit interaction. Errors shown are mean deviations between three runs. mRNA was present at 80 nM in all reactions.

	*mRNA + 40S (80 nM)	***mRNA + 40S (80 nM) + eIF3j**	***mRNA + 40S (400 nM)**
$k_{on, obs}$ **(s^{-1})**	11.0 ± 0.7	0.48 ± 0.01	40 ± 17[a]
k_{off} **(s^{-1})**	0.87 ± 0.02[b]	0.9 ± 0.1	---
k_{on} **(s^{-1})** **= $k_{on,, obs}$-k_{off}**	11.0 - 0.9 = 10.1	0.5 - 0.9 = -0.4	---
% mRNA bound to 40S at equilibrium	60% (K_d = 19 nM)	19% (K_d = 270 nM)	95% (K_d = 19 nM)

[a] Error is quite large, but one individual run with a $k_{on, obs}$ of 57 s^{-1} makes clear that at 400 nM 40S, the association reaction is too fast for this fast-mixing instrument (mixing time ~20 ms).

[b] This value was measured with 3.5 µM excess unlabeled RNA. When repeated with 3.5 µM RNA on a separate day, an off rate of 0.73 ± 0.02 s^{-1} was observed; with 5.5 µM RNA, an off rate of 0.85 ± 0.02 s^{-1} was observed, indicating that these concentrations of excess unlabeled RNA are sufficient.

Upon further investigation, it was realized that, if the concentration of binding components are not high enough to drive the binding reaction to completion (*i.e.* a minimum of 3x the K_d, the observed rate constant, $k_{on,obs}$, will be the sum of the association- and dissociation-rate constants, $k_{on} + k_{off}$, and so k_{off} must be subtracted from $k_{on,obs}$ to obtain the true k_{on}. In practice, the association reaction needs to proceed to at least 50% binding in order to be able to reliably subtract the k_{off} without introducing large amounts of error (J. Kirsch, personal communication). The association experiments described above with 80 nM *mRNA and 80 nM 40S would not be expected to go to completion, so the true k_{on} is achieved once the k_{off} of 0.9 s^{-1} is subtracted (Table 2.2). This yields a k_{on} of 10.1 s^{-1} for *mRNA to 40S subunits alone, and of -0.4 s^{-1} in the presence of eIF3j. This negative value arises from the association reaction only going to ~19% completion (Table 2.2). Thus, in order to accurately determine the association rates *mRNA for 40S subunits in the absence and presence of eIF3j, these binding experiments were repeated with 80 nM *mRNA and 400 nM 40S ribosomal subunits. The $k_{on,obs}$ in the absence of eIF3j was measured as 40 ± 17 s^{-1}. The large error in this measurement stems from the rate approaching the mixing rate in the Kuriyan lab instrument of 50 s^{-1}. In the future, it would be possible to measure these faster association rates in a full-fledged stopped flow system, such as that in the Marqusee lab. Despite the preliminary nature of these results (see Discussion), it appears that eIF3j reduces the binding affinity of *mRNA for 40S subunits principally by slowing its on-rate, while leaving the off-rate unaffected.

2.4.4 Testing for leaky scanning in the absence of eIF3j

We hypothesized that eIF3j might help to prevent leaky scanning by discouraging mRNA from tightly associating with the mRNA binding cleft until the eIF2/Met-tRNA$_i$ TC is bound and ready to recognize an initiation codon as scanning occurs. We chose to address this question using a *S. cerevisiae* deletion strain of HCR1, the eIF3j homolog, as the *hcr1Δ* strain was already known to be viable but slow growing, consistent with a kinetic defect in translation (Valasek et al. 2001a). *Hcr1Δ* strains were generated in two backgrounds of *S. cerevisiae*: yWG3, which is La- and useful for making *in vitro* translation extracts, and yWG25, which is active for invasive growth, a developmental transition requiring new protein synthesis, entered into upon glucose starvation and resulting in yeast colonies growing down into the media to scavenge for additional resources (Cullen & Sprague 2000; Gilbert et al. 2007). The doubling times of *hcr1Δ* strains in rich media were measured, and growth defects were observed in both yWG3 and yWG25 backgrounds, as observed previously (Figure 2.14A; Valasek et al. 2001a). Polysome profiling of yWG3 (WT) and yKB1 (*hcr1Δ*) at mid-log and stationary phase showed that deletion of HCR1 leads to shift towards smaller polysomes, demonstrating a translation initiation defect (Figure 2.14B). In addition, yKB2 (*hcr1Δ*) showed an invasive growth defect relative to WT yWG25, though not as severe as the invasive growth-defective yWG3 strain (Figure 2.14C). Invasive growth is known to require translational control of IRES-containing messages (Gilbert et al. 2007), and the defect in a *hcr1Δ* strain may reflect reduced translation rates.
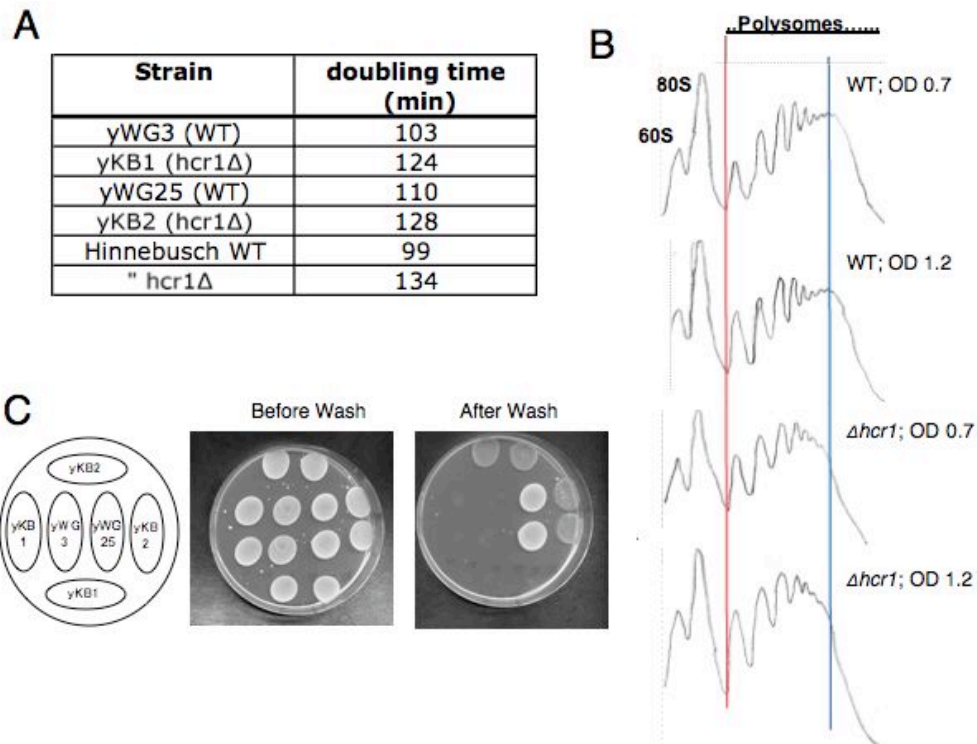
Figure 2.14: Characterization of *hcr1Δ* strains. *(A)* Doubling time of WT and *hcr1Δ* strains in two different strain backgrounds (yWG3 = La-, used for preparation of translation extracts; yWG25 = La+, competent from invasive growth) in YPD at 30°C, and comparison to published doubling times for a *hcr1Δ* strain from the Hinnebusch lab (Valasek et al. 2001a). *(B)* Polysome profiles of WT and *hcr1Δ* strains in yWG3 background, taken at two points within growth curve (OD = 0.7 or 1.2). The red line marks the beginning of polysomes and the blue line indicates the edge of the WT polysome peak. The 60S and 80S peaks are indicated. *(C)* Invasive growth defect of *hcr1Δ* strains in yWG25 background (and yWG3 background as a negative control). *Left*: setup of strains on plate. Each strain is spotted twice on the plate, and two separate isolates of *hcr1Δ* strains are shown. *Center*: yeast growth before wash, showing total growth. *Right*: yeast growth after gentle water wash, showing invasive growth.

With these yeast strains in hand, we wished to test the hypothesis that eIF3j could prevent mRNA from associating in a scanning-competent manner with 40S subunits before TC is present. This model predicts that a leaky scanning defect (*i.e.* failure to recognize the first AUG) would be observed in the absence of eIF3j. A panel of luciferase reporter RNAs was designed to test the degree of leaky scanning using *in vitro* translation reactions in extract from yWG3 (WT) and yKB1 (*hcr1Δ*) strains. These constructs each contained two start codons in identical contexts, and insertions and/or deletions were made such that either the first or the second AUG was out of frame with the luciferase ORF (Figure 2.15A). The reporter RNA with the second AUG in frame with luciferase only yields signal when the first AUG is scanned through without recognition, and so this signal represents leaky scanning. Translation levels of luciferase from *hcr1Δ* extracts were lower than WT, but once normalized to the translation levels of the reporter with both AUGs in frame, no increase was observed in the amount of signal from the leaky-scanning reporter with only the second AUG in frame in the *hcr1Δ* extract (Figure 2.15B). Unfortunately, this experiment was not conducted with a strong positive control (*e.g.* a yeast

44

strain with an established leaky scanning defect). To help validate the *in vitro* translation assay, we tested whether human eIF3j would rescue the translation defects from *hcr1Δ* extract, based on the previous observation that human eIF3j rescues translation defects in *hcr1Δ* yeast strains *in vivo* (Valasek et al. 2001a). Purified human eIF3j was added to *in vitro* translation reactions of the reporter with both AUGs in frame in WT and *hcr1Δ* extracts. No significant increase in translation levels was observed with the addition of exogenous human eIF3j (Figure 2.15C).
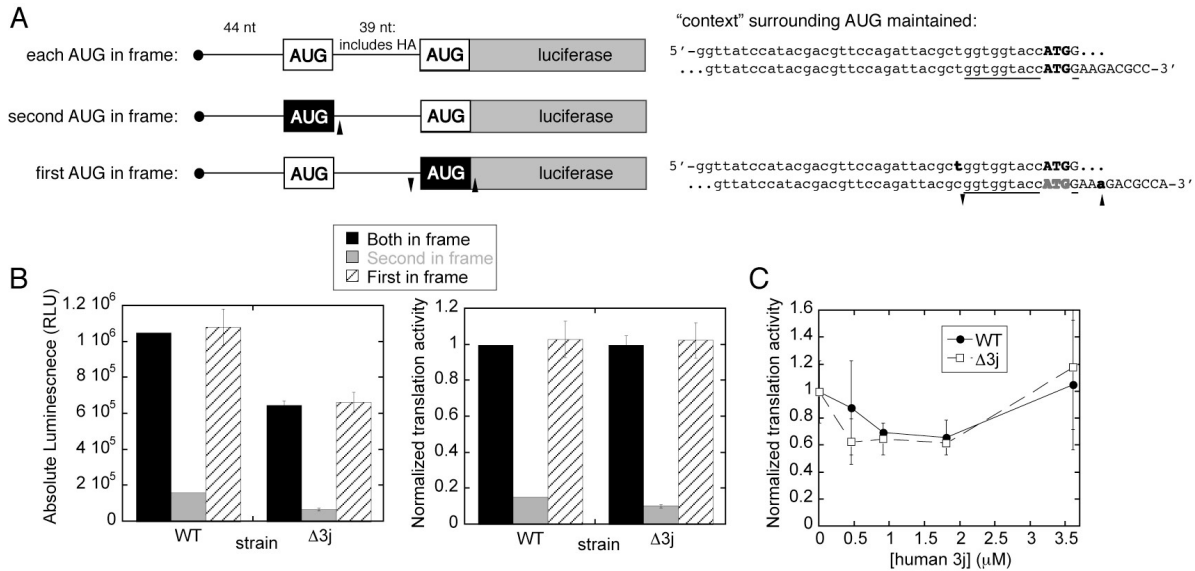
Figure 2.15: *In vitro* translation assay for leaky scanning. *(A)* Design of double AUG luciferase reporters to detect leaky scanning. *Left*: Insertions and/or deletions (indicated with carets) were made to the starting construct with each AUG in frame with the coding sequence of luciferase to selectively shift one or the other out of frame, such that translation from the AUG would not yield functional luciferase enzyme. *Right*: these mutations left the Kozak consensus context (underlined) around each AUG identical. Inserted or deleted nts are shown in bold, as are the AUG start codons. *(B) Left*: Absolute luminescence of reporter RNAs in WT and Δ3j yeast translation extracts, made from strain yWG25. *Right*: Translation activity of reporter RNAs normalized to the activity of the message with both AUGs in frame in WT and Δ3j yeast translation extracts. *(C)* Translation activity of the luciferase reporter with both AUGs in frame in WT and Δ3j yeast translation extracts, in the absence or presence of increasing amounts of purified recombinant human eIF3j, normalized to the translation activity in each extract in the absence of exogenous h3j.

2.5 Discussion

For many years, it was thought that eIF3 bound nearly exclusively to the solvent side of 40S ribosomal subunits, and that the principal role of eIF3j in translation initiation was to aid the rest of the eIF3 complex in binding to the 40S subunit. The C-terminal domain of human eIF3j is now known to be located on the interface surface of the 40S ribosomal subunit, binding in the A site and mRNA binding cleft. Our initial work to biochemically characterize the consequences of eIF3j's presence in the mRNA binding cleft provides insights about its interactions with eIF1A and mRNA. Most significantly, eIF3j decreases mRNA affinity for 40S subunits when TC is not

present. We wished to extend these initial results to give more quantitative insight into the interaction of mRNA and eIF3j on the 40S subunit, the mechanistic basis for TC's effect on mRNA binding affinity in the presence of eIF3j, and the physiological role of eIF3j in translation initiation. These more quantitative questions proved challenging to answer with the current experimental techniques. Here we discuss what has been learned, both in terms of the potential functions of eIF3j and the experimental difficulties encountered.

2.5.1 eIF3j participates in many steps of translation initiation

eIF3j is known to play an important role in eIF3 recruitment to 40S subunits (Fraser et al. 2004; Miyamoto et al. 2005; Nielsen et al. 2006). The j subunit is the most loosely associated protein of the 13 subunits in human eIF3j (Fraser et al. 2004), and binds to the decoding center of the 40S subunit (Fraser et al. 2007). Therefore, it may play roles in translation initiation beyond eIF3 recruitment. It is unclear whether eIF3j ever functions *in vivo* as an independent factor, without the rest of eIF3. For technical reasons, it wasn't possible to include the rest of eIF3 in the *mRNA binding experiments with 40S subunits, but the eIF3j-CTD still binds to the A site and mRNA binding cleft in the presence of the rest of eIF3 (Fraser et al. 2007).

The equilibrium binding affinities of eIF3j and eIF1A for the 40S subunit, alone and together, demonstrate that these factors bind anticooperatively to 40S subunits, and their binding to 40S subunits is thermodynamically coupled. As such, eIF3j likely influences the conformation of eIF1A on the 40S subunit. eIF1A in turn interacts with eIF1 and eIF5 (Figure 1.2) and, together with eIF1, influences the conformation of the 40S mRNA binding cleft (Passmore et al. 2007) and contributes to AUG recognition (Pestova et al. 1998a; Fekete et al. 2005; Fekete et al. 2007). To what extent eIF3j's direct or indirect interaction with eIF1A influences its multiple functions in translation initiation is not yet known.

Equilibrium binding experiments reveal that, in addition to its interaction with eIF1A, eIF3j binds negative cooperatively with mRNA to 40S ribosomal subunits. This quantitatively defines an interaction that has been previously observed only in sucrose gradients (Kolupaeva et al. 2005). eIF3j is a subunit of one of the first translation factors to bind to the 40S subunit during translation initiation and decreases mRNA affinity for the 40S mRNA binding cleft substantially. It is tempting to hypothesize that eIF3j might function to reduce the affinity of RNA for the mRNA binding cleft in the absence of the eIF2/Met-tRNA$_i$ TC. Excitingly, in our *in vitro* system, nearly full affinity of a 20mer model mRNA containing an AUG is restored in the presence of TC, showing that the eIF2/Met-tRNA$_i$ TC indeed overcomes the inhibitory effect of eFI3j and rescues the affinity of mRNA for the 40S ribosomal subunit. Thus, eIF3j appears to regulate access to the mRNA binding cleft, inducing factor-dependent changes in the affinity of mRNA by reducing mRNA affinity until the TC is present. In the absence of eIF3j, TC exerts minimal effects on mRNA/40S subunit binding affinity. Given that eIF3j is still present in binding cleft even when Met-tRNA$_i$ is present (Fraser et al. 2007; ElAntak et al. 2010), it may influence 40S-mRNA interactions during scanning and AUG recognition.

2.5.2 Base pairing-independence of ternary-complex effects

The mechanistic basis of TC restoring mRNA binding affinity in the presence of eIF3j could stem from some combination of one or two effects. Given that the model 20mer mRNA contains an AUG to which the anticodon of tRNA$_i$ could base pair, these base pairs could energetically contribute to the rescue of this model mRNA's affinity. Additionally or alternatively, the TC could influence the conformation of eIF3j in the mRNA binding cleft,

46

exerting an indirect effect on mRNA affinity and conformation. Mechanistically, these contributions are important to distinguish because the base-pairing effect would only occur upon start codon recognition, whereas the indirect effect of TC on mRNA affinity could occur in a pre-scanning conformation, where the 5′ end of the mRNA wouldn't possess an AUG in the P site. In this case, the effect of TC on eIF3j could play a role in dictating an appropriate time to begin scanning (*i.e.* once TC has bound). It is clear that TC at least partially restores mRNA affinity for the 40S subunit in the presence of eIF3j, even without the full codon-anticodon interaction. This question was explored with experiments using either a tRNA with a mutant CCU anticodon or a model mRNA without an AUG start codon. This secondary approach was especially important given that the CCU tRNA still allowed for two of three base pairs with the AUG in the middle of the 20mer RNA sequence (Figure 2.7), and that the model RNA contains AGG sequences near its 5′ end. Unfortunately, both of these approaches yielded very inconsistent and noisy data, perhaps stemming from codon-anticodon base-pair interactions being important for these complex binding reactions to reach equilibrium in a time frame that does not damage protein components. Thus, it was very difficult to quantitatively determine how much of the restored binding energy of mRNA for the 40S subunit comes from codon-anticodon base pairing between TC and mRNA.

What may account for at least a partial restoration of mRNA binding affinity by TC even without codon-anticodon base pairing? Directed hydroxyl radical probing suggests that the conformation of eIF3j on 40S subunits changes in the presence of the ternary complex (Fraser et al. 2009). This conformational change may displace eIF3j from the binding cleft enough to allow mRNA to bind tightly. It would be very interesting to determine whether the 6 nM binding affinity of eIF3j for 40S subunits is altered in the presence of TC, but the FP signal from *eIF3j is prohibitively low for these measurements to be made with the current anisotropy-based system, since TC introduces additional noise to the assays. Still, the partial restoration of mRNA binding affinity by TC even without codon-anticodon base pairing suggests that eIF3j may affect mRNA's interaction with the 40S ribosomal subunit in a pre-scanning conformation, and may therefore play a role in the switch to a scanning-competent conformation.

2.5.3 Additional insights into mRNA-eIF3j interaction

All initial studies of the interaction of eIF3j and mRNA on the 40S subunit made use of a 20mer model mRNA. Equilibrium binding studies of longer model mRNAs indicate that a 20mer does not fully capture the possible interactions of mRNA with the mRNA binding cleft. This is consistent with the 80S ribosome protecting ~30-40 nts in footprinting experiments (Kozak & Shatkin 1976; Legon & Robertson 1976). In addition, it seems that eIF3j has a weaker negative effect on the binding of a 30- or 40-mer than a 20mer, suggesting that eIF3j only displaces a portion of the mRNA from the cleft.

It would be very interesting to measure the affinity of *eIF3j in the presence of each of these model mRNAs, to help confirm the simultaneous binding of eIF3j and mRNA. If mRNA and eIF3j can bind together to the 40S subunit, their equilibrium binding affinities should be thermodynamically coupled, meaning that the affinity of eIF3j should be reduced less by the 40mer than the 20mer. However, fluorescence polarization signal from *eIF3j was not adequate to reliably make these measurements with the precision required to distinguish mutual from competitive binding.

In principal, eIF3j's negative effect on the equilibrium binding affinity of mRNA for the 40S subunit could either arise from a slowing of the mRNA association rate, a speeding of the

mRNA dissociation rate, or some combination thereof. For kinetics experiments, a fluorescence intensity-based assay is preferable to anisotropy-based experiments, so the pyrene-labeled model 20mer RNA was utilized in 40S subunit binding experiments. Pyrene is an environmentally sensitive fluorophore and undergoes a change in fluorescence intensity upon 40S subunit binding. For association reactions, the $k_{on,obs}$ of mRNA for 40S subunits is substantially reduced in the presence of eIF3j. The true k_{on} values, however, have not been precisely measured since, at low concentrations of *mRNA and 40S subunits, the association reactions do not go far enough to completion. With respect to dissociation rates, two phases are seen for *mRNA dissociation from 40S subunits in the presence of eIF3j. The first, fast phase is thought to represent the dissociation of *mRNA from 40S-eIF3j complexes, as the decrease in *mRNA-40S subunit binding affinity by eIF3j would, if anything, be likely to occur from an increase in the mRNA dissociation rate. The second, slow process reproducibly only occurs when eIF3j is present, but the physical process from which it stems is not well understood. With this interpretation, the dissociation rate of mRNA from 40S subunits is essentially unchanged in the presence of eIF3j. There is a slight possibility that 1.25 μM eIF3j did not fully saturate the 80 nM 40S subunits and that the fast phase observed in the presence of eIF3j stems from a small percentage of 40S subunit-*mRNA binary complexes. If this were the case, the dissociation rate of *mRNA from 40S subunits would be drastically reduced by eIF3j, and the association rate would have to be tremendously increased in order for the binding affinity to be reduced by ~15-fold.

Preliminary kinetic measurements therefore suggest that eIF3j decreases the mRNA on-rate, while leaving the off-rate unchanged. To more accurately measure the association rates, a faster mixing stopped-flow device would be required, and additional measurements of *eIF3j's association and dissociation rates would be necessary to fully define the kinetic framework of these interactions. As an additional caveat, it should be noted that this kinetic analysis is built on the assumption that mRNA and eIF3j can indeed bind simultaneously to 40S subunits, yielding ternary 40S-mRNA-eIF3j complexes. If mRNA and eIF3j binding were in fact mutually exclusive, the mRNA off-rate would not be expected to change in the presence of eIF3j because, in either case, the mRNA would dissociate from a 40S-mRNA binary complex, rather than from a 40S-mRNA-eIF3j ternary complex. In addition, the off-rate of eIF3j from the 40S could be rate limiting, such that the $k_{on,obs}$ of mRNA for 40S subunits with eIF3j present could indeed be eIF3j's k_{off}. A more detailed kinetic framework, detailing both *eIF3j and *mRNA association and dissociation rates, will be necessary to distinguish these possibilities with complete certainty, and may be a goal of future efforts.

2.5.4 Physiological role of eIF3j

During translation initiation, mRNA is tethered to the 40S subunit via its 5′ 7-methyl-guanosine cap. This cap-based interaction will contribute to mRNA's overall affinity for the 40S subunit and it is therefore reasonable to question the physiological significance of eIF3j's effect on mRNA binding to the 40S subunit's binding cleft. Scanning is known to take place even in the absence of TC (Hinnebusch 2005), but clearly AUG recognition cannot occur without the codon-anticodon interaction made possible by tRNA. Therefore, it would make sense to couple the beginning of scanning with binding of TC. Based on the observations that eIF3j discourages mRNA from stably binding in the absence of TC, and that TC at least partially restores mRNA affinity without AUG recognition, it seems possible that eIF3j could contribute to a switch between an initial conformation of bound mRNA and a scanning-competent conformation. This hypothesis predicts that, if the function of eIF3j is conserved between yeast and humans, deletion

of HCR1 (the eIF3j homolog) in *S. cerevisiae* would result in a leaky scanning defect or failure of initiating ribosomes to recognize the most 5′ start codon. No evidence for leaky scanning was observed using an *in vitro* translation assay in *hcr1Δ* yeast translation extracts, but no positive control was conducted to establish the effectiveness of this assay. Interestingly, the yKB2 *hcr1Δ* strain did demonstrate a defect in invasive growth, a developmental transition requiring new protein synthesis, entered into upon glucose starvation (Cullen & Sprague 2000; Gilbert et al. 2007). This defect could either reflect a requirement for high translation levels for invasive growth, or might highlight unrealized transcript-specific translation effects of eIF3j.

The yeast translation community has established a powerful *in vivo* assay for looking for leaky scanning using the *GCN4* translational control system (Hinnebusch 2005). This assay has recently been used to show that an *hcr1Δ* strain does indeed exhibit a leaky scanning defect *in vivo* (ElAntak et al. 2010). This result suggests that our *in vitro* leaky scanning assay was poorly designed, but validates a role of eIF3j in AUG recognition and the prevention of leaky scanning in yeast. The leaky scanning defect of deleting HCR1 is partially rescued by eIF1A overexpression, suggesting that these factors may function together on the 40S subunit (ElAntak et al. 2010). In addition, the slow growth phenotype of an *hcr1Δ* strain is partially rescued by overexpression of TC (ElAntak et al. 2010), which is consistent with our proposal of eiF3j helping to change the pre-scanning complex until TC is present.

Unexpectedly, the N-terminal half of HCR1 is sufficient to rescue leaky scanning defect (ElAntak et al. 2010), even though it is the C-terminal portion that has been shown to bind in the mRNA binding cleft in humans (Fraser et al. 2007). The C-terminal domain (CTD) of yeast HCR1 binds to 40S ribosomal proteins that are near the mRNA entry tunnel, but on the solvent side (ElAntak et al. 2010). In yeast, it appears that the function of the HCR1-CTD may partially overlap with that of the CTD of TIF32 (the eIF3a homolog); both of these domains interact independently with eIF3b RRM and bind near the mRNA entry tunnel of the 40S, and mutants in the TIF32-CTD cause the HCR1-CTD to be required for efficient translation and cell growth (Chiu et al. 2010). TIF32 and HCR1 also interact directly with one another (Nielsen et al. 2006), meaning that all three of these subunits – eIF3a, b, and j – interact very intimately in yeast. In contrast, eIF3j in humans seems to principally interact with the eIF3b RRM (ElAntak et al. 2007). It would be very interesting to determine whether eIF3j could be knocked down in human cells and, if so, to conduct *in vivo* experiments on the role of eIF3j in translation initiation in a human system for direct comparison to yeast *in vivo* results.

eIF3j may have other physiological roles in addition to altering the conformation of mRNA bound to 40S subunits once TC is present and preventing leaky scanning. In yeast, HCR1 is known to participate in a late stage of 40S subunit biogenesis, in the conversion of 20S to 18S rRNA (Valasek et al. 2001a). Even after biogenesis is complete, it is possible that eIF3j continues to play a structural role in influencing the conformation of the 40S subunit, perhaps holding the flexible beak in a favorable conformation until the mRNA fills the binding cleft. Given that eIF3j stays associated with 40S subunits during scanning and influences mRNA's interaction with the binding cleft, it could also modulate translational efficiency of messages based on structures near the initiation codon. In addition, eIF3j has been recently shown to play a role in disengaging mRNA from 80S ribosomes, during ribosome recycling after termination (Pisarev et al. 2007).

It is now clear that a subunit of eIF3 joins eIF1, eIF1A, eIF5 and eIF2 as an interface-binding translation initiation factor that contributes to scanning and AUG recognition. There is

much work to be done, both *in vitro* and *in vivo*, to elucidate the specific mechanisms and complete functions of eIF3j in human translation.

Chapter 3

High-throughput screen for Hepatitis C virus translation inhibitors

3.1 Introduction

Hepatitis C virus (HCV) is a significant global health problem with ~180 million people infected worldwide (Suzuki et al. 2007; Webster et al. 2009). Most patients (75-85%) develop chronic disease, leading to cirrhosis and hepatocellular carcinoma in approximately 20% of these cases (Webster et al. 2009). The current treatment for HCV is the combination of pegylated interferon-α and ribavirin, which has limited efficacy and causes significant side effects. New classes of anti-HCV compounds with varied mechanisms of action are needed to improve the current standard of care (Nelson 2009).

The 5′-untranslated region (5′-UTR) of the HCV genome, a positive-sense, single-stranded RNA, is attractive as an anti-viral target due to its importance in the viral life cycle. In particular, nucleotides (nts) 40-372 comprise an internal ribosome entry site (IRES) by forming a specific secondary and tertiary RNA structure essential for viral protein synthesis (Tsukiyama-Kohara et al. 1992; Kieft et al. 1999). Both the structure and mechanism of the IRES have been studied extensively, revealing some of the molecular events that lead to translation initiation (Fraser & Doudna 2007). The IRES RNA binds directly to the human 40S ribosomal subunit and to eukaryotic initiation factor 3 (eIF3; Pestova et al. 1998b; Kieft et al. 2001). The IRES thereby bypasses the need for most initiation factors involved in typical cellular translation initiation, such as the cap-binding eIF4F complex and scanning-associated factors eIFs 1, 1A, and 4A (Pestova et al. 1998b). Specific steps in this pathway can be blocked by mutation or deletion of particular regions of the IRES RNA (Ji et al. 2004; Otto & Puglisi 2004; Locker et al. 2007).

The IRES is one of the most conserved regions of the HCV genome because it must function both in ribosome recruitment to the 5′ end of the plus-strand RNA and in viral replication from the 3′ end of the minus-strand (Suzuki et al. 2007). Despite its functional importance, the IRES is a challenging drug target, as it has no enzymatic activity or explicit active site, but instead uses an extended surface to interact with both the 40S ribosomal subunit and eIF3 (Spahn et al. 2001; Siridechadilok et al. 2005). While inhibition of HCV IRES translation with an antisense oligonucleotide to disrupt the structure of a key region of the IRES has been preliminarily successful (Tallet-Lopez et al. 2003; McHutchison et al. 2006), direct inhibition of IRES binding to 40S subunits or to eIF3 may be difficult to achieve with a small molecule.

Importantly, however, the IRES is more than a molecular scaffold, since mutations that do not affect its affinity for the host translation machinery can nonetheless dramatically reduce translation efficiency (Kieft et al. 2001; Otto & Puglisi 2004). For example, conformational flexibility of the domain II hairpin is required to induce a conformational change in the 40S subunit (Spahn et al. 2001), a necessary precursor to formation of active 80S ribosomes. An attractive possibility is that small molecules can be found that specifically disrupt local structures or conformational dynamics of the RNA required for IRES function.

In an effort to discover such small molecule inhibitors of the HCV IRES, we developed a high-throughput assay based on IRES-dependent *in vitro* translation in rabbit reticulocyte lysate. Conditions for this assay were designed to report on authentic HCV IRES translation relative to a 5′ capped mRNA control. We screened a library of 426,736 small molecules for IRES inhibitors, leading to ~1,700 initial hits of which 59 appeared promising after secondary assays. That a large number of these molecules proved to be off-target luciferase inhibitors underscores the challenges in screens for compounds that recognize RNA targets and in using enzymes as read-outs for activity.

In addition to the functional screen for new HCV IRES inhibitors described here, we also worked with putative small molecule IRES inhibitors discovered by others, with the goal of determining their mechanism of action. Appendix I details these efforts through which cyclic peptides evolved by mRNA display in the Szostak lab were found to have little IRES inhibitory activity, and small molecules discovered by ISIS pharmaceuticals, which the field assumes targets the IRES, displayed non-specific inhibition of translation.

3.2 Methods

3.2.1 Reagents

Restriction endonucleases, calf intestinal alkaline phosphatase and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). [^{35}S]methionine (>1000 Ci/mmol) was obtained from GE Life Sciences (Boston, MA). Plasmid DNA preparation, restriction enzyme digestion, agarose gel electrophoresis of DNA and RNA, DNA ligation and bacterial transformations were carried out using standard methods (Sambrook & Russell 2001).

3.2.2 Plasmid construction and RNA transcription

A DNA fragment encoding the HCV IRES sequence was generated by PCR using a genotype 1a HCV IRES construct (H77 strain, nts 40-372; Yanagi et al. 1997) as the template. The resulting DNA fragment was ligated into the EcoRI and BamHI restriction sites of pUC19 to form the parent plasmid for all subsequent constructs. Derivative plasmids encoding the IRES with domain II deleted (ΔdomII, nts 40–119 deleted) were generated by using QuikChange mutagenesis (Stratagene). The firefly (FF) and *Renilla* (RN) luciferase reporter genes were amplified from pGL3 and pRL-TK, respectively (Promega), and cloned between the BamHI and HindIII restriction sites of pUC19. Capped messages had a template-encoded poly(A) tail of 62 nt. All constructs were verified by DNA sequencing. IRES and 5′ capped RNAs were prepared by *in vitro* transcription (Megascript and mMESSAGE T7 system, respectively, Ambion) according to the manufacturer's instructions. The 5′ capped-messages were generated with a GTP:m^{7}GpppGTP ratio of 1:10 to ensure >90% capping efficiency.

3.2.3 *In vitro* translation reactions

In vitro translations were performed in rabbit reticulocyte lysate and translation activity was measured using either a luciferase reporter assay or radiolabeling with [^{35}S]methionine, according to the manufacturers' instructions (Promega and Life Technologies). Unless otherwise indicated, *in vitro* translations were carried out in 15 µl reactions, containing 2 µL of nuclease-treated RRL (0.5 µL Ambion + 1.5 µL Promega; the composition of RRL was optimized for translation efficiency and accuracy, data not shown), 1 ng/ml and 3 ng/ml of IRES-RN and 5′ cap-FF reporter mRNAs, respectively, and 5 ng/ml polyC RNA as a carrier. Translation reactions were adjusted to final concentrations of 2.6 mM Mg(OAc)$_2$, 45 mM KCl and 90 mM KOAc, and also contained amino acids (1 mM), dimethyl sulfoxide (1%; DMSO) and 0.1% complete protease inhibitor (Roche). Translation reactions were incubated at 30°C for 90 min and terminated by the addition of puromycin to a final concentration of 20 µM.

For high-throughput screening (HTS), reactions were scaled proportionally to 5 µL total volume and the positive controls for IRES-RN and 5′ cap-FF translation inhibition were a DNA oligonucleotide complementary to the IIId loop of the IRES (10 µM) and puromycin (20 µM), respectively. The IIId oligo sequence used was 5′-ACCCAACACTACTCGGC-3′ (Tallet-Lopez

et al. 2003). A combination of DualGlo, BrightGlo, and Dual Luciferase Assay System kits (Promega) was used in preliminary assays, as noted in figure legends, to measure firefly and *Renilla* luciferase activities from the same well. The DualGlo luciferase assay kit (Promega) was selected for HTS due to its long half-life of luminescence, and was used according to the manufacturer's instructions with minor adjustments. Luminescence was measured by TopCount (PerkinElmer) for 96-well plates and by Envision for 384-well plates (PerkinElmer). For the secondary enzyme interference screen, *in vitro* translation reactions were performed as above but in the absence of compounds. After quenching translation reactions with puromycin, compounds were added and incubated with the translated enzymes for 30 min at room temperature before measuring luciferase activity.

For [^{35}S]methionine incorporation experiments, translation reactions were performed as above, except that methionine was omitted from the amino acid mixture, and [^{35}S]methionine (10 µCi) was added. Aliquots from translation mixtures were mixed with SDS sample buffer, boiled for 2 min and resolved on an SDS polyacrylamide gel (10% acrylamide). Gels were fixed in 10% methanol, 7.5% acetic acid, and treated with ENHANCE (New England Nuclear). The level of translation of FF luciferase (61 kDa) and RN luciferase (36 kDa) was quantified by phosphorimaging and densitometry using ImageQuant TL (Molecular Dynamics).

3.2.4 Compound library and liquid handling

The chemical library for high-throughput screening consisted of 426,736 compounds (Gilead Sciences, Inc). Liquid handling protocols were optimized for distribution of (a) the compound library (Bravo, Velocity11), (b) translation mix to minimize formation of air bubbles (Deerac, Tecan), (c) RNA templates to ensure precise dispensing at small volumes (Deerac, Labcyte), and (d) luciferase substrates (microFill, Biotek).

3.2.5 Data analysis

Data analysis was carried out by Graphpad Prizm5.1 and SpotfFire. The Z-factor was calculated as previously described (Zhang et al. 1999b). The normalized percentage of inhibition values (NPI, %) were calculated for each HTS plate by setting the average signal of the negative control (1% DMSO) as 0% inhibition and the average signal of the positive control (IIId oligo or puromycin) as 100% inhibition.

3.3 Results

3.3.1 Translation inhibition screen design

A robust functional assay was established to screen for small molecule inhibitors that block HCV IRES-mediated translation. To enable independent assessment of 5′ cap-dependent and cap-independent translation initiation from the same well, we designed two monocistronic mRNAs encoding distinct luciferase enzymes, one driven by a 5′ cap and the other by the HCV IRES (Figure 3.1A; genotype 1a, H77 strain, nts 40-372). The IRES primary sequence and tertiary structure are highly conserved among all 6 genotypes (Han et al. 1991; Bukh et al. 1992), making it likely that inhibitors found against this genotype would have pan-genotypic activity. For the translation of the reporter constructs, we considered both HeLa cell extract and rabbit reticulocyte lysate (RRL). Although HeLa translation extracts are reported to faithfully recapitulate the translation behavior observed in cells (McCaffrey et al. 2002), low activity prevented obtaining sufficient signal for the statistical analysis required for a high-throughput

screen. In contrast, unmodified RRL has limited 5′ cap-dependence and low sensitivity to mutations in the HCV IRES (this study; Dasso & Jackson 1989; Soto Rifo et al. 2007), but produces high levels of translation. These observations led us to focus on optimization of RRL as the system in which to conduct inhibitor screens.

In light of previous results showing that the salt concentrations in RRL could be adjusted to stimulate faithful 5′ cap-dependent scanning (Kozak 1990), we examined the effect of RRL salt concentration on the fidelity of HCV IRES-mediated translation. To distinguish IRES-dependent from promiscuous translation, we utilized two established methods for inhibiting HCV IRES function: (a) deletion of domain II (nts 40-119, referred to as ΔdomII; Ji et al. 2004) and (b) inhibition with a DNA oligonucleotide complementary to the IIId loop of the IRES, a region required for recruitment of the 40S ribosomal subunit (referred to as 'IIId oligo'; Tallet-Lopez et al. 2003).

Under standard RRL conditions (79 mM KOAc, 0.5 mM $MgCl_2$, Promega), a reporter mRNA containing the ΔdomII IRES was translated with 65% of the activity of an mRNA containing the wild type IRES (WT; Figure 3.1B, black columns). By comparison, the ΔdomII mRNA has been shown to have only 3%, 2%, and 20% activity in mice, HeLa cells and HeLa S10 extracts, respectively (McCaffrey et al. 2002; Otto & Puglisi 2004). Notably, the inhibitory activity of the IIId oligo against IRES-mediated translation was also limited under these conditions (Figure 3.1B). When the salt concentration of the RRL reactions was adjusted to a previously identified condition (2.2 mM Mg^{2+}, 45 mM KCl, 90 mM KOAc; Kozak 1990), the ΔdomII-containing mRNA showed a significant reduction in translation activity (7% of the wild type IRES-containing mRNA). Furthermore, the IIId oligo was more effective at inhibiting IRES-driven translation in the adjusted RRL when compared to standard RRL (Figure 3.1B). The ratio of luciferase activities from translation of WT IRES-containing mRNA versus either ΔdomII IRES or WT IRES in the presence of the IIId oligo defined the "selectivity window" of the assay. This selectivity window can be considered to report both on the authenticity of IRES translation and our ability to detect inhibition of this translation. Further optimization of KCl (Figure 3.1C) and $Mg(OAc)_2$ (Figure 3.1D) showed that the largest selectivity window of 11-fold occurred at 45 mM KCl and 2.6 mM $Mg(OAc)_2$.

Under these adjusted RRL conditions, a 5′ cap-containing mRNA was translated ~5-fold better than the same mRNA containing a non-physiological cap (Figure 3.1E). Thus, the adjusted RRL displays increased fidelity for both IRES- and 5′ cap-dependent translation. Although the overall activity of the salt-adjusted RRL was decreased relative to unaltered conditions (nearly 20-fold reduction for 5′ capped mRNA and ~50-fold for IRES-containing mRNA), these activity levels were significantly higher than those observed with HeLa extracts (which showed at least 20-fold lower activity in the presence of greater than 100-fold more mRNA; data not shown).
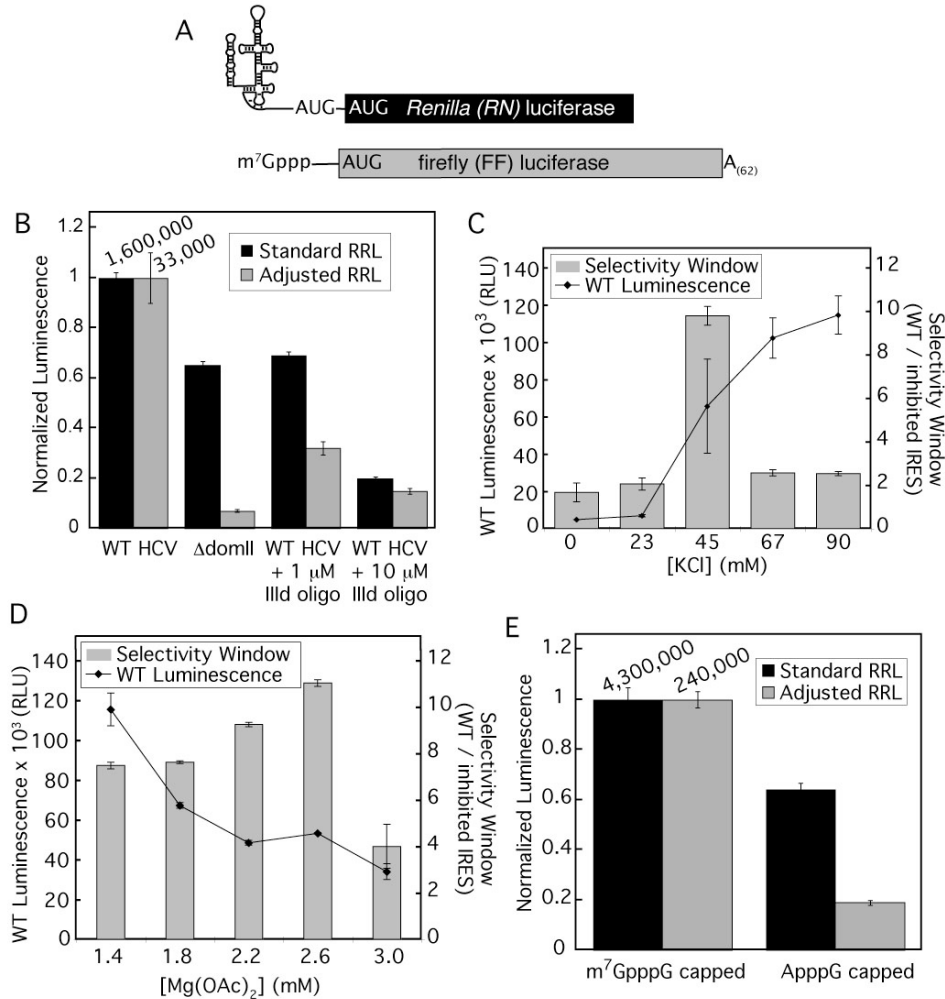
Figure 3.1: Optimization of RRL for authentic translation of HCV IRES. *(A)* Schematic showing IRES-RN and 5′ cap-FF reporter mRNAs used in this study. *(B) In vitro* translation of HCV IRES-FF mRNA in standard or adjusted RRL and the effect of deleting domain II (ΔdomII), or addition of a DNA oligo that hybridizes to the IIId domain of the IRES (IIId oligo). Effects of *(C)* KCl or *(D)* Mg(OAc)$_2$ titration on translation efficiency of WT IRES and the ΔdomII mutant IRES. The absolute luminescence signal of WT IRES-containing mRNA is shown in relative light units (RLU) and selectivity window is defined as the ratio of signal from WT over the ΔdomII IRES. *(E)* Comparison of translation of m^{7}GpppG- and ApppG-capped FF RNAs in standard and adjusted RRL. Luciferase activity was measured using Luciferase Assay System. Note that all of the above *in vitro* translation reactions were carried out in pre-HTS conditions (8.5 μL RRL/15 μL reaction, in the absence of polyC RNA and DMSO, with 100% Promega RRL).

3.3.2 Reporter mRNA choice and assay optimization

To identify IRES-specific inhibitors, we designed a high-throughput screen using two reporter mRNAs encoding distinct luciferase enzymes, *Renilla* (RN) and firefly (FF). To maximize the signal:noise ratio in the primary screen, we wanted each well to contain as much of the two monocistronic mRNAs as possible without introducing competition between them,

which could result from one reporter preferentially monopolizing ribosomes and other translation factors in the extract. Titrations of 5′ capped-FF and IRES-RN mRNAs were performed using the optimized salt conditions established above, measuring the signal from one reporter at a time. As IRES-RN mRNA concentration was increased to moderate to high concentrations, the signal from 5′ capped-FF translation dropped (Figure 3.2A, note the vertical spread), which is indicative of interference between the two messages. Similarly, the highest concentrations of 5′ capped-FF mRNA interfered with the signal from IRES-RN translation (Figure 3.2B), although competition in this direction does not appear to be as strong. Significant interference at high concentrations of either reporter mRNA was also demonstrated for the converse reporter pair (IRES-FF and 5′ capped-RN; data not shown). Based on the results of these two-dimensional titrations of reporter mRNAs, we chose roughly 1 ng/μL IRES-RN RNA and 3 ng/μL 5′ capped-FF RNA for the primary screen; the concentration of templates used was adjusted, within 3-fold, for each new batch of *in vitro* transcribed RNA. Different concentrations of each monocistronic mRNA were ideal to maximize signal while avoiding competition between the mRNAs, underscoring an important advantage of using two monocistronic messages as opposed to a single bicistronic message where the relative concentrations cannot be varied.
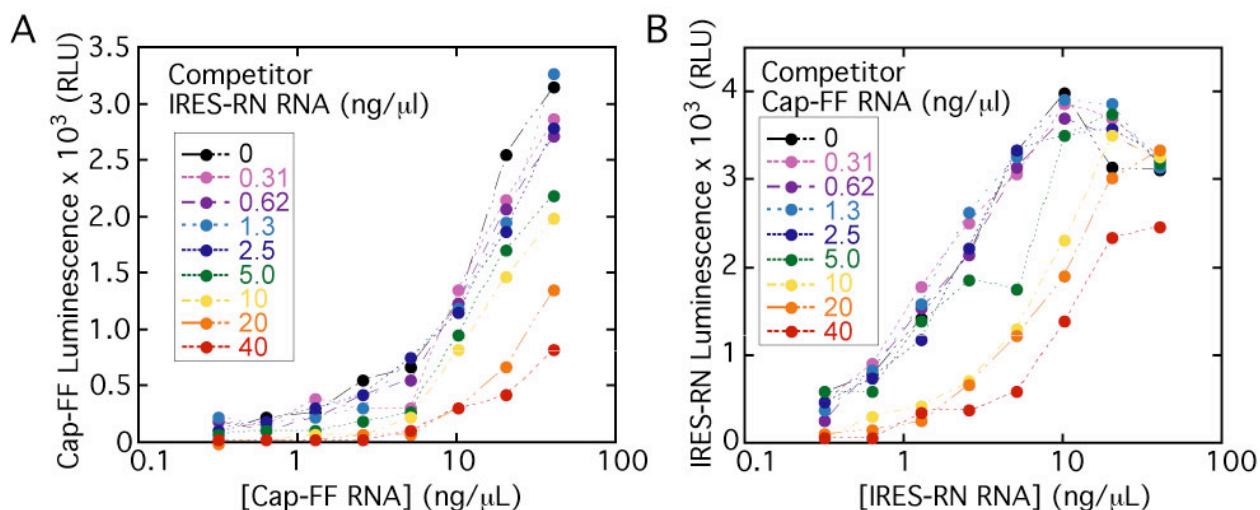
Figure 3.2: Two-dimensional titrations of IRES-RN and 5′ cap-FF mRNAs. The concentration of each of the two mRNAs was varied between 0 ng/ml and 40 ng/ml in *in vitro* translation reactions in adjusted RRL, and the absolute luminescence activities of *(A)* 5′ cap-FF and *(B)* IRES-RN mRNA are plotted. Luciferase activities were measured using the DualGlo system.

The RRL-based translation assay conditions were further adjusted to be suitable for high-throughput screening (HTS) of potential small molecule inhibitors. Because the small molecules to be tested were dissolved in DMSO, we examined the effect of this organic solvent on the activity and selectivity window of the translation reactions. Although the signal from IRES-FF mRNA steadily decreased as DMSO concentration increased, (Figure 3.3A), the selectivity window between WT IRES- and ΔdomII IRES-containing mRNAs, representing the faithfulness

of the extract, remained greater than 6.5 in the presence of up to 2% DMSO. These data show that the final DMSO concentration of 1% was compatible with this assay.

It was also possible to significantly reduce the amount of RRL used in the assay for implementation in the high-throughput screen. The selectivity window and the signal from WT IRES-containing mRNA were measured as RRL concentration was varied from 3.3-57% (vol/vol), while the overall salt concentration was kept constant. The luminescence signal decreased by ~1000-fold over the range of titration, but the selectivity window remained relatively stable from ~57% down to 6.7% (Figure 3.3B). Lastly, it was important to ensure the linearity of reaction kinetics. At 13% RRL, the condition chosen for HTS, the 5′ capped-FF signal remained linear for at least three hours while the IRES-RN signal began to plateau at ~120 min (Figure 3.3C). Thus, an incubation time of 90 min was chosen for the high-throughput screen, as it gave substantial signal and was within the linear range. In the final format for the primary screen, 5 μL translation reactions containing 13% RRL (Ambion:Promega = 1:3) were incubated at 30 °C for 90 min.
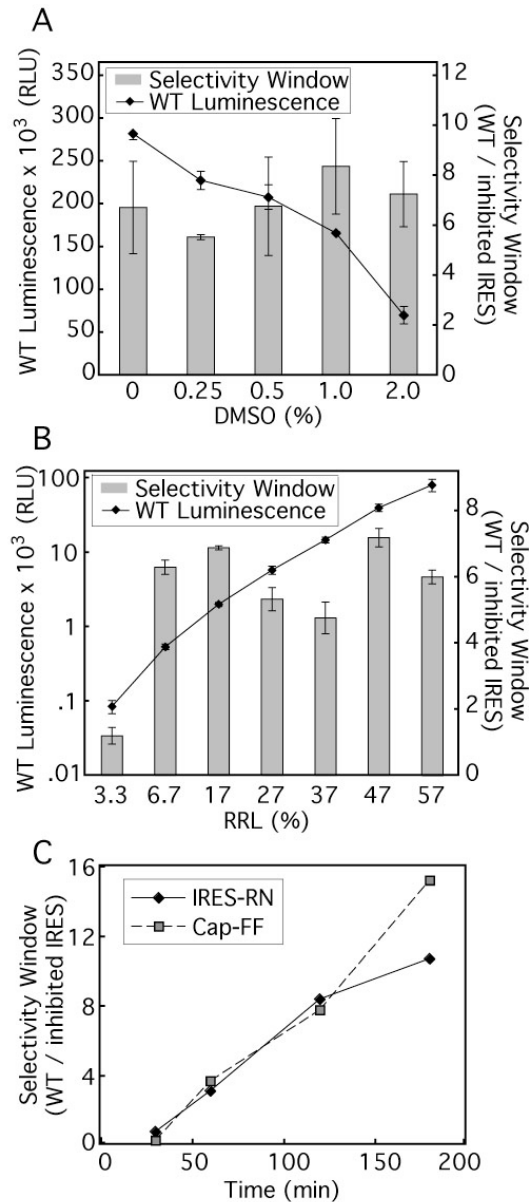
Figure 3.3: Optimization of RRL translation assay for high-throughput screening. The absolute luminescence signal of WT IRES-FF mRNA is shown in relative light units (RLU). *(A)* Effect of DMSO titration on efficiency (luminescence) and authenticity of IRES translation (WT signal over ΔdomII signal defined the selectivity window). These *in vitro* translation reactions were conducted at the pre-HTS RRL concentration (8.5 μL RRL/15 μL reaction, with 100% Promega RRL). *(B)* Effect of RRL concentration on efficiency and authenticity of IRES translation. The selectivity window was measured using the ΔdomII IRES for RRL concentrations greater than and equal to 17% (vol/vol), and addition of the 10 μM IIId oligo for RRL concentrations less than 17% (vol/vol). *(C)* Time course of *in vitro* translation assay. 3 ng/μl 5′ cap-FF and 1 ng/μl IRES-RN mRNA were translated in the optimized RRL conditions for 60-180 minutes. The selectivity window for 5′ cap-FF translation was calculated as the ratio of the signal in the absence of puromycin over the signal at 20 mM puromycin. The selectivity window for IRES-RN was calculated as the ratio of the signal in the absence of IIId oligo over the signal at 10 mM IIId oligo. Luciferase activity in all panels was measured using DualGlo system.

59

3.3.3 Primary Screen

Using the assay conditions outlined above, a total of 426,736 compounds (average MW ~ 550) were tested in a primary screen. The Z-factor was determined for each 384-well plate. The average Z-factors for 5′ cap-FF and IRES-RN reporters were 0.71 and 0.62, respectively (ranging between 0.5 and 0.86. The results from plates with Z-factors which did not meet the cut-off of 0.5 were discarded and the assay plates rerun. Three known translation-inhibiting antibiotics were present in the library and inhibited both the IRES-RN and 5′ cap-FF message in the primary screen (Table 3.1), confirming the robustness of screen. Approximately 1% of the total compounds tested reduced IRES-RN luciferase signal by at least 45% (reported as normalized percentage of inhibition (NPI); Figure 3.4A). Within this group, ~25% were IRES-RN selective as the 5′ cap-FF signal had an NPI less than 40%. In addition to these IRES-RN selective hits, ~650 compounds were identified as highly potent inhibitors of both IRES-RN and 5′ capped-FF reporters (>70% NPI). These inhibitors were included with the selective inhibitors to be tested in a secondary screen since selectivity for the IRES over a 5′ cap might not be observed for a potent inhibitor at the single concentration of 10 mM tested in the primary assay.

Table 3.1: NPI values of known translation inhibitors in HTS. Three known inhibitors of eukaryotic protein synthesis were present in the library. The NPI (%) of these compounds from the primary screen was calculated as described in the Methods section.

	NPI (%)	
Inhibitors	**IRES-RN**	**5′ cap-FF**
Puromycin* (positive control)	100%	100%
emitine	92.2%	95.1%
madumycin derivative	41.7%	39.3%

* Puromycin was used as the positive control for the screen. IC_{50} values were
  determined to be 0.41 μM against IRES-RN and μM against 5′ cap-FF.

3.3.4 Secondary Screen

Of these ~1,700 primary hit compounds, a representative group of 623 compounds was purchased for follow-up dose-response assays. These compounds were chosen based on commercial availability and their drug-like nature, determined by an automated *in silico* triage based on molecular weight (<600 Daltons), lack of reactive groups (e.g. methylating groups), and solubility (calculated partition coefficient, log P < 5). The goals of these assays were (a) to define the selectivity window between IRES- and 5′ cap-dependent inhibition, even for potent hits, and (b) to distinguish authentic IRES inhibitors from compounds that inhibit *Renilla* luciferase preferentially over firefly luciferase. Two formats of secondary screens were used. One was essentially identical to the primary screen except it consisted of the opposite pair of reporters: IRES-FF and 5′ capped-RN mRNA. The second test was a luciferase enzyme interference assay in which the translation of IRES-RN and 5′ capped-FF was performed as in the primary screen but in the absence of any compounds. Once the translation reaction was quenched with puromycin, compounds were added and incubated with the translated enzymes before measuring luciferase activity. In total, these assays yielded six sets of data: IC_{50} values for each compound against IRES-RN, 5′ capped-RN, IRES-FF, 5′ capped-FF (compounds

present during translation reactions), and IC$_{50}$ values against *Renilla* enzyme and firefly enzyme (compounds added after the translation of enzyme). Data for these dose-response follow up assays were plotted as the selectivity index (SI; ratio of IC$_{50}$) calculated for the 5′ cap/IRES and the enzyme (Enz)/IRES comparisons made for each luciferase (Figure 3.4B).

The vast majority of compounds in both the *Renilla* and the firefly assays showed a selectivity index of less than 3 for both Enz/IRES and 5′ cap/IRES, falling into the lower left-hand quadrant in Figure 3.4B. Some of these compounds, colored in green, had very high IC$_{50}$s against the IRES-containing message, suggesting that they were simply false positives from the primary screen. Many other compounds, colored in red or orange, had low IC$_{50}$s against the IRES-containing message, indicating that they were correctly identified by the primary screen as inhibitors of the IRES-containing message. These compounds are therefore most likely luciferase inhibitors that equally inhibit the luciferase signal, whether it is driven by the IRES, 5′ cap, or comes directly from the enzyme.

There were also a large number of compounds present in the upper left-hand quadrants of the plots in Figure 3.4B, showing selectivity for IRES-driven translation over the enzyme alone, but not over 5′ cap-dependent translation. Many of these molecules are likely to be general translation inhibitors capable of blocking translation driven either by the 5′ cap or the IRES. Consistent with this inhibition acting at the level of general translation, these compounds behaved similarly against both RN and FF reporters: of the 21 compounds that were found in this upper left-hand quadrant for FF and the 34 compounds in this quadrant for RN, 18 overlap. It is also possible that some of these compounds are general luciferase inhibitors that were not hits in the enzyme alone assays due to unavoidable differences in the setup of this assay (see Discussion).

We chose compounds that had a selectivity index greater than 3 for both Enz/IRES and 5′ cap/IRES as potential IRES-selective inhibitors. These compounds fell into the upper right quadrant of either plot in Figure 3.4B. As the screening cascade shows (Figure 3.4A), many more compounds met this criterion with the *Renilla* reporter than the firefly reporter. Of these compounds, only one appeared to be a selective inhibitor with both reporter enzymes (GS-036984, Figure 3.5A).
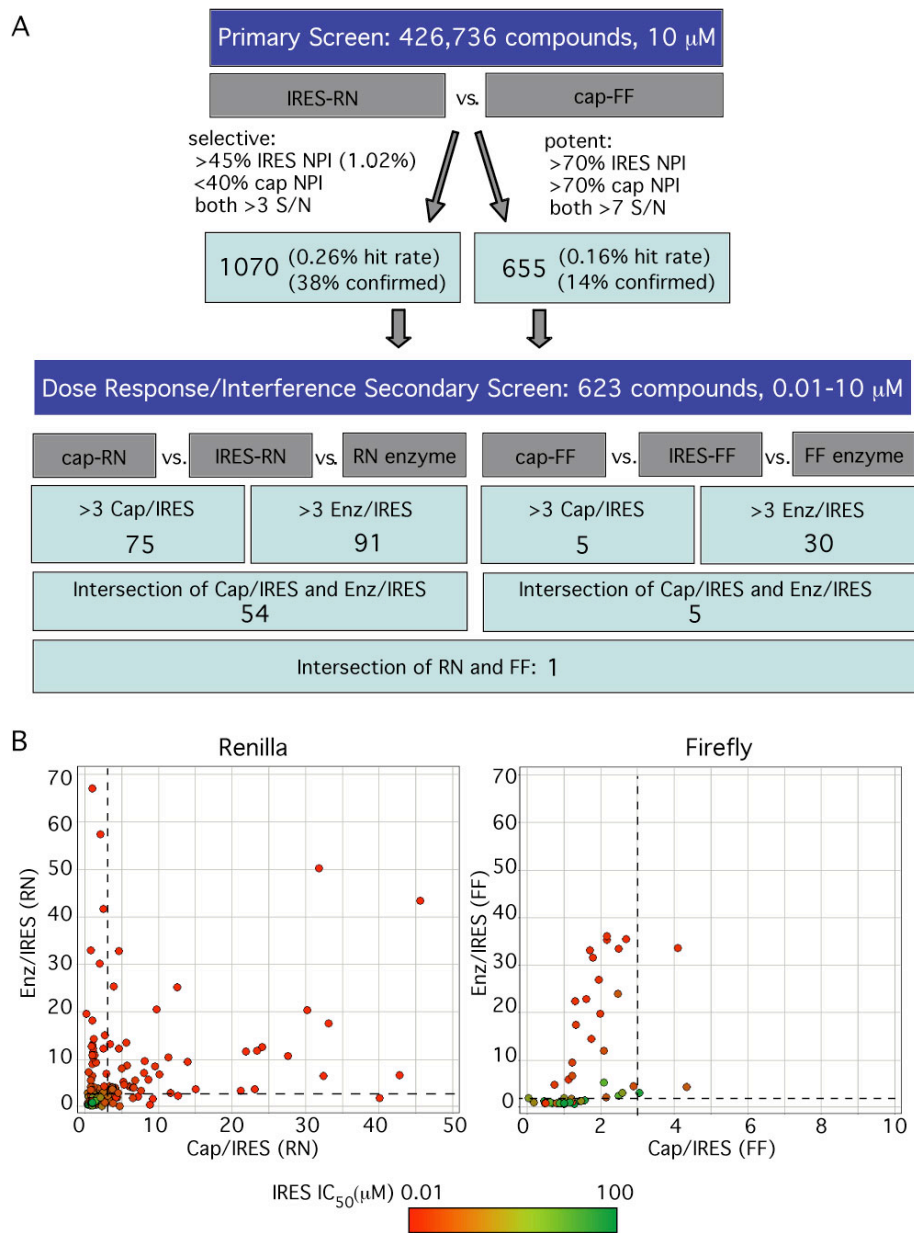
Figure 3.4: Results of high-throughput screen. *(A)* Screening cascade for primary and secondary screens, showing the criteria for hits and the observed hit rates. Normalized percentage of inhibition (NPI) was defined as the relative translation inhibition compared to positive controls (puromycin for 5′ cap-mRNA or IIId oligo for IRES-mRNA). The ~1,700 hits from the primary screen were tested in an identical confirmation assay in triplicates (at least two out of three repeated hits were counted as positive) and the confirmation rates are shown. These confirmation rates are similar to what has been observed for other high-throughput screens (observations at Gilead Sciences, Inc). *(B)* Selectivity index (SI) plots. Enz/IRES SI was defined as the ratio of the IC_{50} of a compound against the luciferase enzyme vs. IRES-driven translation of the enzyme, and the Cap/IRES SI was the ratio of a compound's IC_{50} against 5′ cap-mRNA vs. IC_{50} against IRES-mRNA translation. Data points are shown for the 623 compounds analyzed in the secondary screen, and are colored by their IC_{50} against the IRES reporter, with red being potent inhibitors and green showing no inhibition of the IRES reporter up to 100 mM. Dashed lines indicate the SI>3 cut-off of the secondary assays.
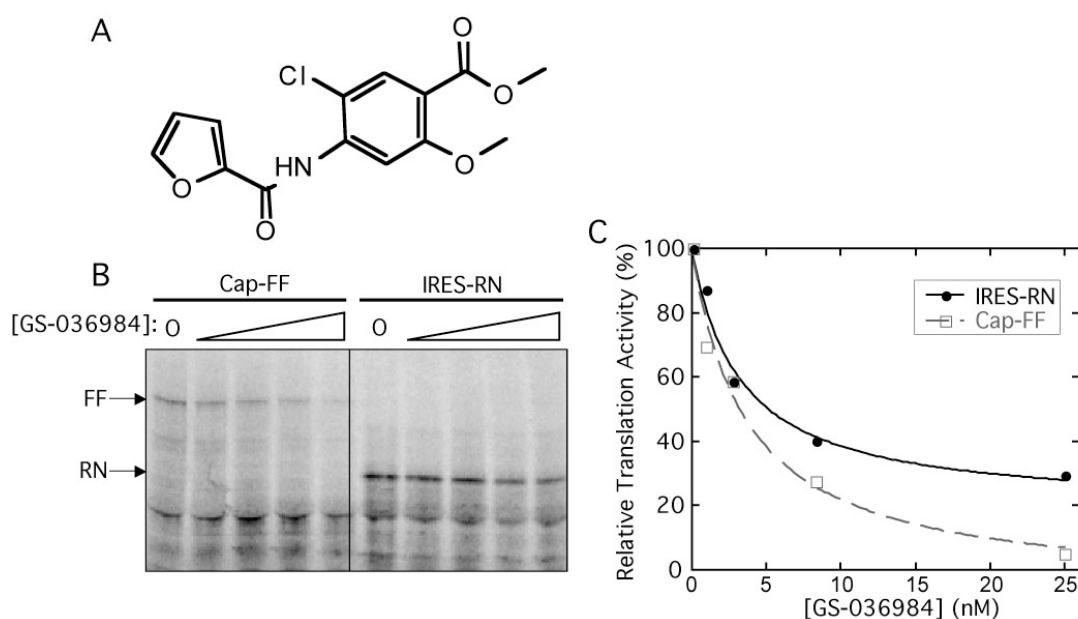
Figure 3.5: Non-selective inhibition of GS-036984 against IRES- and 5′ cap-dependent translation. *(A)* Structure of GS-036984. [^{35}S]methionine incorporation translation assay of IRES-RN and 5′ cap-FF reporter in the presence of increasing concentrations of GS-036984 from 0 to 25 μM. Translation activity was measured by phosphorimaging of the fixed and dried SDS PAGE gel *(B)*, and normalized to levels in the absence of the compound. *(C)* The percentage of inhibition was quantified by phosphorimaging and densitometry as the ratio between the intensity of the band in the presence of absence of compounds, after subtracting the background. Data were fit in GraphPad Prism, yielding IC$_{50}$ values of 3.6 ± 1.0 against 5′ cap-FF and 3.2 ± 0.8 against IRES-RN.

3.3.5 Validation Assay

While GS-036984 appeared to be an IRES-selective inhibitor after the secondary screens using both FF and RN reporters, the selectivity index of 5′ cap/IRES (RN) for this compound was 3.9, barely meeting the cut-off of 3. Therefore we wanted to additionally validate this compound's activity in an assay completely independent of luciferase activity. [^{35}S]methionine incorporation followed by SDS PAGE analysis was used to observe translation of the IRES-RN and 5′ capped-FF messages in the presence of GS-036984 (Figure 3.5B). Contrary to expectations from the screen, this compound was found to inhibit translation from both 5′ cap-FF mRNA and IRES-RN reporters and thus is in fact a general translation inhibitor (Figure 3.5C; see discussion).

We also examined the activity of the 59 compounds that were identified as selective IRES inhibitors with either FF or RN reporters, but not both, using the [^{35}S]methionine-incorporation translation assay. One additional compound related to the scaffold above also inhibited both 5′ cap- and IRES-driven translation. To the best of our knowledge, these compounds represent a new scaffold of general translation inhibitor. The other compounds tested, however, failed to inhibit translation of either message (data not shown).

3.4 Discussion

The HCV IRES RNA is an attractive drug target due to its central role in regulating viral protein synthesis and pan-genotypic sequence conservation. In this study, we screened a small molecule library for specific inhibitors of HCV IRES function. The results of this work highlight both successes in high-throughput assay development and challenges inherent in identifying compounds that recognize RNA targets.

Previous work established that direct IRES binding by an antisense oligonucleotide can block IRES-mediated protein synthesis (Tallet-Lopez et al. 2003). Although this nucleic acid based candidate (ISIS 14803) reached Phase I clinical trials, limited on-target activity and aminotransferase flares led to a halt of clinical development (McHutchison et al. 2006). Additionally, there have been other attempts to identify IRES-specific inhibitors by high-throughput screens with smaller libraries. Fragment-based screening using mass spectrometry has identified small molecule binders to domain II of the IRES which are inhibitors of an HCV replicon assay (Seth et al. 2005). See Appendix IB for further discussion of the activity of these molecules. A dual luciferase-based HTS assay in Krebs extracts with a bicistronic reporter did not reveal any IRES-specific compounds (Novac et al. 2004). We sought to identify new chemical scaffolds that directly target IRES function and could be further optimized to yield potent oral antiviral agents.
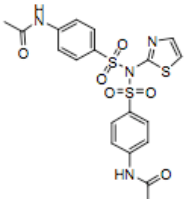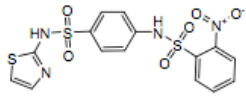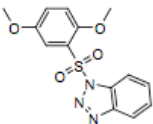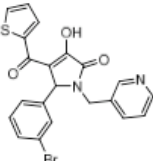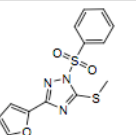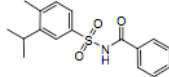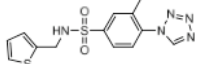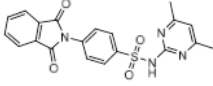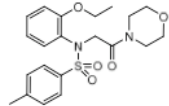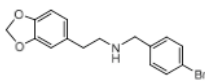
A screening assay was developed using RRL supplemented with additional salts, which improved the fidelity of both cap-dependent and HCV IRES-mediated translation. These results are analogous to what has previously been seen with the encephalomyocarditis virus (EMCV) IRES (although the EMCV IRES requires a different set of translation initiation factors from the HCV IRES) (Jackson 1991). The use of two monocistronic reporters (one IRES-driven and one 5′ cap-driven) enabled each reporter's signal to be maximized independently, while avoiding competition between the two reporter templates, which cannot be achieved using a single bicistronic mRNA. We noted that in general, many more compounds in the secondary screen inhibited the RN enzyme than the FF enzyme, and no compounds inhibited both. This result suggests that the primary screen was effective at excluding general luciferase inhibitors, such as competitive inhibitors of the ATP substrate or protein aggregants. However, ~3% of the primary hits appeared to be general translation inhibitors (in the upper left-hand quadrant of Figure 3.4B) which would have ideally been eliminated in the primary screen.

Due to issues with luciferase interference, an [^{35}S]methionine-incorporation validation assay was essential to verify the inhibition activity of the fifty-nine compounds which appeared to be selective IRES inhibitors in the secondary screen with either the firefly or *Renilla* reporter. Two of these compounds were found to be general translation inhibitors and, to the best of our knowledge, represent a new chemical scaffold of general translation inhibitors. The compound GS-036984 was selected from the primary screen as a potent hit, consistent with it inhibiting both 5′ cap- and IRES-dependent translation. The selectivity indices determined in the secondary screen for 5′ cap/IRES were barely above the cut-off of 3.0 (3.9 with the *Renilla* reporter and 3.3 with the firefly reporter). While this low selectivity suggested a potential preference for inhibiting the IRES, the direct visualization of translation products without interference of luciferase reporters clearly showed that this compound inhibits both 5′ cap-dependent and IRES-dependent translation (Figure 3.5B,C).

In the [^{35}S]methionine-incorporation validation assay, 57 out of 59 of the compounds identified in the secondary luciferase screen showed no inhibition against either IRES- or 5′ cap-dependent translation. This observation suggests that these compounds are likely to be direct

inhibitors of *Renilla* luciferase enzyme. Several scaffolds emerged from the compounds showing selective *Renilla* luciferase inhibition, such as para-amino-sulfonamides, sulfanyltriazoles, and dihydropyrrolones. The ten most potent Renilla-specific inhibitors identified in the screen are shown in Table 3.2. These *Renilla* luciferase inhibitors may not have been eliminated in the secondary screen due to differences in the assay conditions, such as the temperature of incubation and the higher final luciferase protein levels in the enzymatic interference assay due to the lack of DMSO during translation. A subset of compounds that showed no inhibition at the HTS conditions for the enzymatic interference assay (incubation at room temperature for 30 min) demonstrated enhanced inhibitory activity against the luciferase enzymes when the incubation conditions were changed to more closely match the primary assay (30°C for 90 min). Additionally, when the interference assay was performed using recombinant *Renilla* luciferase enzyme rather than *in vitro* translated enzyme, an even larger subset of compounds (~80% of the 54 hits) showed inhibition when incubated at 30°C for 90 min.

Table 3.2: Examples of potent *Renilla* luciferase-specific inhibitors. These compounds showed *Renilla* luciferase-specific inhibition to both *in vitro* translated *Renilla* luciferases (regardless of whether translation was driven by the IRES or a 5′ cap) and recombinant *Renilla* luciferase. These compounds showed no inhibition of firefly luciferase at concentrations up to 100 µM under the same conditions.

	IC$_{50}$ (REN) µM	IC$_{50}$ (FF) µM		IC$_{50}$ (REN) µM	IC$_{50}$ (FF) µM
structure	0.02	> 100	structure	2.02	> 100
structure	0.10	> 100	structure	2.24	> 100
structure	0.73	> 100	structure	2.45	> 100
structure	1.44	> 100	structure	2.98	> 100
			structure	3.46	> 100
			structure	5.12	> 100

The relative abundance of luciferase inhibitors identified in this study reflects the fact that it is much simpler chemically for a compound to block an enzyme active site than to bind to a macromolecular surface and inhibit a conformational change or an intermolecular interaction (note that a large fraction (~40%) of current drug targets are enzymes and many other drug targets also naturally bind to small molecules) (Overington et al. 2006). Our secondary counter screens eliminated most, but not all, of the facile and prevalent direct inhibitors of *Renilla*

luciferase. The high level of *Renilla* luciferase inhibitors selected in this screen suggests there is an inherent problem in using an enzymatic reporter assay as a means to screen for compounds that interact with a particularly challenging target, such as a structured RNA, as the reporter enzyme is itself a classical target of small molecules.

In order to expand our ability to devise new and diverse drugs, the scope of small molecule activities must be expanded to non-traditional targets, such as influencing macromolecular interactions and conformations, and efforts are underway to expand the diversity of small molecules in screening libraries (Nielsen & Schreiber 2008). Based on our experience with enzyme inhibition masking the desired RNA-based inhibition, assays for new functions would do well to avoid traditional small molecule targets as read-outs for activity. Fluorescence, as opposed to luminescence, may be an alternative approach to couple translation activity to a spectroscopic signal. For example, members of the green fluorescence protein family do not have a defined active site for binding small molecules and may therefore be less subject to direct reporter inhibition. A dual reporter screen could be conducted with a pair of red and green fluorescent proteins, though spectroscopic interference from heterocyclic compounds in the library would need to be considered (Inglese et al. 2007). Additionally, in an effort to reduce enzyme interference, the active site of firefly luciferase has been redesigned and evolved to discourage small molecules binding to the luciferin and ATP pockets. This new enzyme, 'Ultra-Glo', shows a marked loss of inhibition from several scaffolds that inhibited the original firefly luciferase (Auld et al. 2009).

Other strategies for identifying RNA-binding small molecules as inhibitors that do not require enzymatic read-outs, such as fragment-based or *in silico* screening, could also be considered as alterative approaches (Hermann & Westhof 2000; Seth et al. 2005; Bodoor et al. 2009). A FRET-based screen for small molecule binders to RNA has previously been used to find compounds that bind to and stabilize the HCV IRES IIId loop (Baugh et al. 2009), but whether these compounds are effective inhibitors of HCV IRES translation has yet to be determined. As a complement to such small molecule binding screens, it is certainly attractive to envision direct functional screens to look for inhibitors of IRES-mediated translation. However, the present study shows that such assays are not as straightforward as they may first appear.

In this study we validated the use of rabbit reticulocyte lysate with increased salt concentrations as a system to study the effects of mutations and inhibitors of the HCV IRES. The current screen has provided insights into how deleterious direct enzyme inhibition can be to an otherwise robust high-throughput assay for a difficult target, such as the HCV IRES. Additionally, a new scaffold has been identified of a general eukaryotic translation inhibitor, the mechanism of which will be interesting to investigate in the future.

Chapter 4

Function and inhibition of the HCV IRES pseudoknot

4.1 Introduction

Hepatitis C virus is a major public health problem, infecting ~180 million people worldwide and causing chronic liver problems including cirrhosis and hepatocellular carcinoma (Ghany et al. 2009). New treatments are needed for HCV infection as the current combination of pegylated interferon-α and ribavirin has adverse side effects as well as limited efficacy with some viral genotypes (Deutsch & Hadziyannis 2008). The 5′ untranslated region (UTR) of the viral genomic RNA contains an internal ribosome entry site (IRES), the structure of which is required for viral protein synthesis (Tsukiyama-Kohara et al. 1992; Wang et al. 1993). Due to its significance in the viral life cycle and its high degree of conservation across genotypes, the IRES is an attractive drug target for new HCV therapeutics. A complete structural and mechanistic understanding of how the HCV IRES functions in translation initiation will facilitate targeted discovery of inhibitors.

The HCV IRES adopts an ion-dependent three-dimensional fold with specific secondary and tertiary structural elements (Brown et al. 1992; Kieft et al. 1999). The RNA is capable of binding directly to 40S ribosomal subunits and eukaryotic initiation factor 3 (eIF3; Pestova et al. 1998b; Sizova et al. 1998; Kieft et al. 2001). Assembly studies suggest that the IRES binds first to 40S subunits and that this complex then recruits eIF3 and the ternary complex (TC) of eIF2-GTP-Met-tRNA$_i^{met}$ (Ji et al. 2004; Otto & Puglisi 2004). While recruitment of 40S subunits occurs before eIF3 in HeLa extracts (Otto & Puglisi 2004), the IRES may bind to 40S subunits that are pre-associated with eIF3 and/or the TC in cells. GTP hydrolysis by the TC, mediated by eIF5, enables 60S ribosomal subunit joining with the help of eIF5B, forming an elongation-competent 80S ribosome (Pestova et al. 1998b; Locker et al. 2007). The IRES may also facilitate an alternative, eIF2-independent pathway of translation initiation under conditions of increased eIF2a phosphorylation during infection (Pestova et al. 2008; Terenin et al. 2008).

Significant progress has been made in mapping various functions of the IRES to specific structural domains (Figure 4.1A). The basal portion of domain (dom) III is required for 40S recruitment (Kieft et al. 2001), while regions in dom IV and dom II are also contacted by the 40S subunit (Kolupaeva et al. 2000; Kieft et al. 2001; Spahn et al. 2001). The apical IIIabc domain contributes to IRES affinity for eIF3 (Sizova et al. 1998; Kieft et al. 2001). Dom II binds in the tRNA exit site, inducing a 40S conformational change (Spahn et al. 2001) and promoting eIF5-mediated GTP hydrolysis (Locker et al. 2007).

The specific tertiary structure that is formed in the basal portion of dom III has been subject to debate. On the basis of predicted base pairing, a pseudoknot structure was proposed in which the IIIf loop nucleotides base pair with nucleotides in a sequence just upstream of the AUG-containing dom IV (Wang et al. 1995). The proposed pseudoknot consists of two stems, SI and SII, as depicted in Figure 4.1A. Base pairing within a stem can be experimentally verified if mutation of either side of a stem inhibits activity and mutation of both sides to introduce compensatory base pairs rescues activity. The importance of base pairing in SI has been experimentally verified for both HCV and the related IRES in classical swine fever virus (CSFV; Wang et al. 1994; Fletcher & Jackson 2002). Similarly, base pairing in SII of the CSFV IRES is critical for activity (Rijnbrand et al. 1997; Fletcher & Jackson 2002). In the HCV IRES, initial studies also showed modest rescue of HCV-mediated translation with compensatory base pairing mutations in SII (Wang et al. 1995). Subsequently, however, HCV IRES mutants containing compensatory base pairing mutations throughout the entirety of SII were inactive (Kieft et al. 2001). This result suggested that the specific sequence within predicted SII, instead of or in addition to secondary structure alone, is needed for IRES function.

To resolve this discrepancy and further expand our understanding of the pseudoknot's function, we constructed a large set of mutants (n=35) to structurally and mechanistically probe how the pseudoknot contributes to the activity of the HCV IRES. Translation activities from two-base pair mutations in SII of the HCV IRES pseudoknot offer strong evidence for functionally important base pairing throughout the six-nucleotide stem, suggesting that the full pseudoknot forms as predicted. SII structure contributes to AUG positioning by the IRES as analyzed by toeprinting analysis using purified human 40S ribosomal subunits. The global structure of the pseudoknot, not just that of SI or SII, is required for robust IRES function. Across a wide panel of mutants, translation activity correlates well with AUG positioning ability. Furthermore, an IRES with a compromised pseudoknot has a more stringent requirement for the proper distance between the AUG and pseudoknot, supporting a model in which the conformation of the pseudoknot positions the mRNA open reading frame (ORF) in the 40S binding cleft. Based on these findings, we validated the pseudoknot as a possible HCV IRES drug target using antisense 2′-OMe oligonucleotides.

4.2 Methods

4.2.1 Cloning of IRES reporter construct and mutants

A wild type (WT) IRES-firefly luciferase (FF luc) reporter was generated from a previous MS2 hairpin-HCV IRES-FF luc construct (Ji et al. 2004) by QuikChange mutagenesis (Stratagene) with the primer 5′-CGGAATTCTAATACGACTCACTATAGCTCCCCTGTGAG GAACTACTG-3′ and its reverse complement. Deletion of three MS2 hairpins from the WT IRES-FF luc construct yielded plasmid pKB84. The IRES sequence in pKB84 is genotype 1b, ultimately derived from the pK1b plasmid (Tsukiyama-Kohara et al. 1992) with the exception of a G357A mutation in the coding sequence of the IRES (downstream of dom IV). All IRES mutations were generated from pKB84 using QuikChange mutagenesis. DNA sequencing was used to check the IRES and FF luciferase sequences of each mutant. The names, sequences, descriptions and preliminary translation activity data for all mutants are presented in Table 4.1, and Table 4.2 presents these mutants in groups, based on the hypotheses they were designed to test.

Hints: Where multiple mutations were being made in the same construct, multiple QC reactions could be conducted simultaneously. As long as primers are fairly close to each other and not overlapping, four primers can be added to a QC reaction for double mutations. From agarose gels, it appears that a superprimer is made *in situ* in such reactions. Primers were designed according to Stratagene's instruction manual, such that primers were between 25 and 50 nts, ended in a G or C, and had T_m>78°C using the following equations: Tm=81.5+ 0.41(%GC)-675/N-% mismatch for point mutations, or Tm=81.5+0.41(%GC)-675/N for insertions/deletions, where N does not include the bases being inserted or deleted. QC PCR reactions (50 μL) contained *Pfu* ultra high fidelity DNA polymerase (1 μL; Agilent), *Pfu* ultra buffer (1x; Agilent), dNTPs (200 μM), DMSO (1%), plasmid template (25 ng), and primers (200 nM). Temperature was cycled ~18 times from 95°C for 30 sec to 55°C for 1 min to 68°C for 16 min (2 min/1kb for *Pfu* polymerase). DpnI (1 μL; NEB) was added to reactions and incubated for > 1 hr at 37°C. An aliquot (~2 μL) of the reaction was used to transform Mach1 cells (20 μL).

Table 4.1: HCV IRES pseudoknot mutants by construct number. All mutants of the HCV IRES pseudoknot that were made in the pKB84 (-FF luc reporter) background are listed in numerical order, with their intralab construct name, publication name, if published (P), translation activity, and mutant sequence. For translation activity values, standard deviations are provided only if the results have been obtained from two independent translation reactions of two independent transcriptions. Otherwise, preliminary data are indicated with a ~. Translation data collected by Shruti Waghray, Naeem Husain and Carl Onak are noted with parenthetical initials. In sequences, SI sequence is shown in green, SII sequence in blue, and mutated nts are shown in red, inserted nts in orange, and nts surrounding the site of a deletion in purple. For SI mutants in which mutations were made on the 5′ and 3′ sides of the stem, these sequences are separated with a semicolon.

Name	Description	P	Trans. Activity (% WT)	Mutant Sequence
pKB84	WT	Y	100	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pSW1	SII mid 5'x	Y	2 ± 1 (SW)	AGGGTGCTTGGCAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pSW2	SII mid 3'x	Y	2 ± 2 (SW)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTGCTAGACCGTGCATCATG
pSW3	SII mid comp	Y	60 ± 8 (SW)	AGGGTGCTTGGCAGTGCCCCGGGAGGTCTGCTAGACCGTGCATCATG
pSW4	SII top 5'x	Y	41 ± 9 (SW)	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pSW5	SII top 3'x	Y	40 ± 9 (SW)	AGGGTGCTTGCGAGTGCCCCGGGAGGTGACGTAGACCGTGCATCATG
pSW6	SII top comp	Y	91 ± 16 (SW)	AGGGTGCTTGCGTCTGCCCCGGGAGGTGACGTAGACCGTGCATCATG
pSW16	SI/J top 5'x	N	21 ± 2	AGGGTCCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pSW17	SI/J top 3'x	N	60 ± 5	AGGGTGCTTGCGAGTGGCCCGGGAGGTCTCGTAGACCGTGCATCATG
pSW18	SI/J top comp	N	71 ± 5	AGGGTCCTTGCGAGTGGCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB143	SII bot 5'x	Y	11 ± 1	AGGGTGCTACCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB144	SII bot 3'x	Y	32 ± 4	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGAUGACCGTGCATCATG
pKB145	SII bot ~comp	Y	3 ± 1	AGGGTGCTACCGAGTGCCCCGGGAGGTCTCGAUGACCGTGCATCATG
pKB146	SII bot comp	Y	71 ± 8	AGGGTGCTACCGAGTGCCCCGGGAGGTCTCGGUGACCGTGCATCATG
pKB147	SII GU to GC	N	67 ± 5	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGCAGACCGTGCATCATG
pKB148	L1/L2 U's to A's	Y	85 ± 1	AGGGTGCATGCGAGAGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB149	Ent SII 3'x	Y	<10%	AGGGTGCTTGCGAGTGCCCCGGGAGGTGAGCATGACCGTGCATCATG
pKB150	Ent SII 5'x	Y	<10%	AGGGTGCAACGCTCAGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB151	Ent SII ~comp	Y	<10%	AGGGTGCAACGCTCAGCCCCGGGAGGTGAGCATGACCGTGCATCATG
pKB152	Ent SII comp	Y	<10%	AGGGTGCAAtGCTCAGCCCCGGGAGGTGAGCATGACCGTGCATCATG
pKB153	SI del 2	Y	3 ± 1	AGGACCCCCCCTCGGGAGAGCCATAGT; AGGGTGCTTGCGAGTGCCCCGAGGTCTCGTAGACCGTGCATCATG
pKB154 (stored as 155)	SI ins 2	Y	87 ± 12	AGGACCCCCCCTCCCCCGGGAGAGCCATAGT; AGGGTGCTTGCGAGTGCCCCGGGGGAGGTCTCGTAGACCGTGCATCATG
pKB155 (stored as 154)	SI/J ins 2	Y	4 ± 2	AGGGTGCGCTTGCGAGTGCGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB156	L2 elim	y	29 ± 7	AGGGTGCTTGCGAGGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB157	L2 ins 2	Y	11 ± 2	AGGGTGCTTGCGAGTTTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB158	L3 elim	Y	102 ± 9	AGGGTGCTTGCGAGTGCCCCGGGAGGCTCGTAGACCGTGCATCATG
pKB159	L3 ins 1	Y	80 ± 7	AGGGTGCTTGCGAGTGCCCCGGGAGGTTCTCGTAGACCGTGCATCATG
pKB160	L3 ins 3	Y	70 ± 11	AGGGTGCTTGCGAGTGCCCCGGGAGGTTTTCTCGTAGACCGTGCATCATG
pKB161	SII del 1	Y	80 ± 11	AGGGTGCTTGCGGTGCCCCGGGAGGTCCGTAGACCGTGCATCATG
pKB162	SII ins 1 (CG)	Y	18 ± 6	AGGGTGCTCTGCGAGTGCCCCGGGAGGTCTCGTAGGACCGTGCATCATG
pKB163	ins 1 U dir AUG			AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCTATG
pKB164	ins 2 U dir AUG	Y	75 ± 17	CTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCTTATG
pKB165	del 1 dir AUG	Y	125 ± 8	CTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATATG

pKB166	ins 1 U top SLIV	N	79 ± 8	CTTGCGAGTGCCCCGGGAGGTCTCGTATGACCGTGCATCATG
pKB167	SIJ/II stack CC/GG → CG/GC –2.9 → –2.0 kcal	N	70 ± 11	AGGGTGCTTGCGACTGCCCCGGGAGGTGTCGTAGACCGTGCATCATG
pKB168	SIJ/II stack CC/GG → GC/CG 2.9 → –3.4	N	16 ± 1	AGGGTGGTTGCGAGTCCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB169	SIJ/II stack CC/GG → AA/UU –2.9 → –0.9	N	50 ± 7	AGGGTGATTGCGATTTCCCCGGGAGGTATCGTAGACCGTGCATCATG
pKB170	SII/dIV stack AG/UC → AU/UA –1.7 → –0.9	N		TGCTTGCGAGTGCCCCGGGAGGTCTCGTATACCGTGCATCATGAGCACAAATACTGAA
pKB171	SII/dIV stack AG/UC → GG/CC –1.7 → –2.9	N	74 ± 8	AGGGTGCTCGCGAGTGCCCCGGGAGGTCTCGTGGACCGTGCATCATG
pKB172	L1 elim	Y	70 ± 11	AGGGTGCTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB173	L1 ins 2	Y	36 ± 7	AGGGTGCTTTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB174	ins 3 U dir AUG	Y	28 ± 6	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCTTTATG
pKB175	del 2 dir AUG	Y	67 ± 10	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCAATG
pKB176	del 3 dir AUG	Y	34 ± 6	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATG
pKB177	ins 2 U top SLIV	N	30 ± 5	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTATTGACCGTGCATCATG
pKB178	ins 3 U top SLIV	N	11 ± 3	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTATTTGACCGTGCATCATG
pKB179	SII: AU to CG	N	~~10 (NH)	AGGGTGCTTGCGGGTGCCCCGGGAGGTCCCGTAGACCGTGCATCAT
pKB181	SII: 147+ GU	N	~~15 (NH)	AGGGTGCTTGTGAGTGCCCCGGGAGGTCTCGCAGACCGTGCATCATG
pKB182	SII: 147+ AU	N	~~15 (NH)	AGGGTGCTTGTGAGTGCCCCGGGAGGTCTCACAGACCGTGCATCATG
pKB183	SII: 162+ GU	N	~2	AGGGTGCTCTGCGAGTGCCCCGGGAGGTCTCGTGGGACCGTGCATCATG
pKB184	SII: 162+ AU	N	~4	AGGGTGCTCTGUGAGTGCCCCGGGAGGTCTCATAGGACCGTGCATCATG
pKB185	SII: 171+ GU	N	~~1 (NH)	AGGGTGCTCGTGAGTGCCCCGGGAGGTCTCGTGGACCGTGCATCATG
pKB186	SII: 171+ AU	N	~~10 (NH)	AGGGTGCTCGTGAGTGCCCCGGGAGGTCTCATGGACCGTGCATCATG
pKB187	SII ins 1(AU)	N	~39	AGGGTGCTATGCGAGTGCCCCGGGAGGTCTCGTATGACCGTGCATCATG
pKB188	SI end 5'x	Y	37 ± 3	AGGACCCCCGGTCCCGGGAGAGCCATAGT AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB189	SI end 3'x	Y	32 ± 5	AGGGTGCTTGCGAGTGCCCCGGGACCTCTCGTAGACCGTGCATCATG
pKB190	SI end comp	Y	50 ± 1	AGGACCCCCGGTCCCGGGAGAGCCATAGT AGGGTGCTTGCGAGTGCCCCGGGACCTCTCGTAGACCGTGCATCATG
pKB191	II/III del 2	Y	73 ± 4	AGGACCCCCTCCCGGGAGAGCCATAGT
pKB192	II/III ins 2	Y	82 ± 5	AGGACCCCCCCCCTCCCGGGAGAGCCATAGT
pKB193	SII: GU to AU	N	~110	AGGGTGCTTACGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB194	SII: AU to GC	N	~78	AGGGTGCTGGCGAGTGCCCCGGGAGGTCTCGTCGACCGTGCATCATG
pKB197	SI/J del 1	Y	55 ± 8	AGGGTGTTGCGAGTCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB198	SI/J elim	Y	1.3 ± 0.5	AGGGTTTTGCGAGTCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB199	SI ins 1	Y	95 ± 18	AGGACCCCCCCTCCCGGGAGAGCCATAGT; AGGGTGCTTGCGAGTGCCCCGGGGAGGTCTCGTAGACCGTGCATCATG
pKB200	SI del 1	Y	90 ± 4	AGGACCCCCCCTCCGGGAGAGCCATAGT; AGGGTGCTTGCGAGTGCCCCGGAGGTCTCGTAGACCGTGCATCATG
pKB207	SI/II stack GC/GC → GG/CC –3.4 → –2.9	N	104 ± 24	AGGACCCCCGCTCCCGGGAGAGCCATAGT; AGGGTGCTTGCGAGTGCCCCGGGAGCTCTCGTAGACCGTGCATCATG
pKB208	SI/II stack GC/GC → CG/CG –3.4 → –2.0	N	76 ± 5	AGGACCCCCGCTCCCGGGAGAGCCATAGT; AGGGTGCTTGCGACTGCCCCGGGAGCCTGTCGTAGACCGTGCATCATG
pKB209	SI/II stack GC/GC → AA/UU –3.4 → –0.9	N	85 ± 5	AGGACCCCCACTCCCGGGAGAGCCATAGT; AGGGTGCTTGCGAATGCCCCGGGAGTTTTCGTAGACCGTGCATCATG
pKB210	pSW4; ins 2 U dir AUG	Y	17 ± 4	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCATCTTATG
pKB211	pSW4; ins 3 U dir AUG	Y	5 ± 0.8	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCATCTTTATG

pKB212	pSW4; ins 1 U top SLIV	Y	5 ± 0.9	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTATGACCGTGCATCATG
pKB213	pSW4; ins 2 U top SLIV	Y	4 ± 2	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTATTGACCGTGCATCATG
pKB214	pSW4; ins 3 U top SLIV	Y	2 ± 1	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTATTTGACCGTGCATCATG
pKB215	del 6 dir AUG	Y	27 ± 7	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGATG
pKB216	del 9 dir AUG	Y	49 ± 5	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGAATG
pKB217	pSW4; del 6 dir AUG	Y	3 ± 1	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGATG
pKB218	pSW4; del 9 dir AUG	Y	2 ± 1	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGAATG
pKB219	176 (delete 3) + ACA to UCA	N	17 ± 4	GCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATGAGCTCAAATCCTGAA
pKB220	ACA to UCA mutation in WT background	N	80 ± 3	GCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATGAGCTCAAATCCTGAA
pKB222	pSW4; del 1 dir AUG	Y	22 ± 3	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCATATG
pKB223	pSW4; del 2 dir AUG	Y	16 ± 2	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCAATG
pKB224	pSW4; del 3 dir AUG	Y	9 ± 2	AGGGTGCTTGCGTCTGCCCCGGGAGGTCTCGTAGACCGTGCATG
pKB225	SI/J bot 5'x	N	39 ± 3	AGGGTGGTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB226	SI/J bot 3'x	N	80 ± 9	AGGGTGCTTGCGAGTCCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB227	SI/J bot comp	N	16 ± 1	AGGGTGGTTGCGAGTCCCCCGGGAGGTCTCGTAGACCGTGCATCATG
pKB228	pKB84 code single point mut from G to A	N		GACCGTGCATCATGAGCACAAATCCTAAACCTCAAAGAAAAAGGATCCTCATG
pKB229	SLIV: C334U	N	~95	CCCGGGAGGTCTCGTAGATCGTGCATCATGAGCACAAATCCTGAA
pKB230	SLIV: C334A	N		CCCGGGAGGTCTCGTAGAACGTGCATCATGAGCACAAATCCTGAA
pKB231	SLIV: C335U	N		CCCGGGAGGTCTCGTAGACTGTGCATCATGAGCACAAATCCTGAA
pKB232	SLIV: C335A	N	~100	CCCGGGAGGTCTCGTAGACAGTGCATCATGAGCACAAATCCTGAA
pKB233	SLIV: G336A	N	~105	CCCGGGAGGTCTCGTAGACCATGCATCATGAGCACAAATCCTGAA
pKB234	SLIV: G336U	N		CCCGGGAGGTCTCGTAGACCTTGCATCATGAGCACAAATCCTGAA
pKB235	SLIV: U337C	N	~75	CCCGGGAGGTCTCGTAGACCGCGCATCATGAGCACAAATCCTGAA
pKB236	SLIV: U337A	N	~105	CCCGGGAGGTCTCGTAGACCGAGCATCATGAGCACAAATCCTGAA
pKB237	SLIV: G338A	N	~100	CCCGGGAGGTCTCGTAGACCGTACATCATGAGCACAAATCCTGAA
pKB238	SLIV: G338U	N	~120	CCCGGGAGGTCTCGTAGACCGTTCATCATGAGCACAAATCCTGAA
pKB239	SLIV: C339U	N	~80	CCCGGGAGGTCTCGTAGACCGTGTATCATGAGCACAAATCCTGAA
pKB240	SLIV: C339G	N	~100	CCCGGGAGGTCTCGTAGACCGTGGATCATGAGCACAAATCCTGAA
pCO1	Vary pos'n of UU ins	N	21 ± 3 (CO)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGTTACCGTGCATCATG
pCO3	Vary pos'n of UU ins	N	20 ± 4 (CO)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACTTCGTGCATCATG
pCO5	Vary pos'n of UU ins	N	42 ± 8 (CO)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTTGCATCATG
pCO7	Vary pos'n of UU ins	N	63 ± 16 (CO)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGTTCATCATG
pCO9	Vary pos'n of UU ins	N	29 ± 9 (CO)	AGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATTTCATG

Table 4.2: HCV IRES pseudoknot mutants by groups. All mutants of the HCV IRES pseudoknot that were made in the pKB84 (-FF luc reporter) background are listed in groups according to the types of mutations of made. For sequence details, translation activity values, and whether or not the mutants have been published, please refer to *Table 4.1*.

Type of Mutant	Mutant name	Description
SII base pairs	pSW1	SII mid 5'x
	pSW2	SII mid 3'x
	pSW3	SII mid comp
	pSW4	SII top 5'x
	pSW5	SII top 3'x
	pSW6	SII top comp
	pKB143	SII bot 5'x
	pKB144	SII bot 3'x
	pKB145	SII bot ~comp
	pKB146	SII bot comp
	pKB149	Ent SII 3'x
	pKB150	Ent SII 5'x
	pKB151	Ent SII ~comp
	pKB152	Ent SII comp
SII stability length	pKB147	SII GU to GC
	pKB179	SII: AU to CG
	pKB181	SII: 147+ GU
	pKB182	SII: 147+ AU
	pKB183	SII: 162+ GU
	pKB184	SII: 162+ AU
	pKB185	SII: 171+ GU
	pKB186	SII: 171+ AU
	pKB187	SII ins 1(AU)
	pKB193	SII: GU to AU
	pKB194	SII: AU to GC
	pKB161	SII del 1
	pKB162	SII ins 1 (CG)
SI length bp	pKB153	SI del 2
	pKB154 (155 in freezer)	SI ins 2
	pKB188	SI end 5'x
	pKB189	SI end 3'x
	pKB190	SI end comp
	pKB199	SI ins 1
	pKB200	SI del 1
SI/J length bp	pKB155 (154 in freezer)	SI/J ins 2
	pKB197	SI/J del 1
	pKB198	SI/J elim
	pKB225	SI/J bot 5'x
	pKB226	SI/J bot 3'x
	pKB227	SI/J bot comp
	pSW16	SI/J top 5'x
	pSW17	SI/J top 3'x

	pSW18	SI/J top comp
Loops	pKB148	L1/L2 U's to A's
	pKB156	L2 elim
	pKB157	L2 ins 2
	pKB158	L3 elim
	pKB159	L3 ins 1
	pKB160	L3 ins 3
	pKB172	L1 elim
	pKB173	L1 ins 2
	pKB191	II/III del 2
	pKB192	II/III ins 2
PK AUG distance	pKB163	ins 1 U dir AUG
	pKB164	ins 2 U dir AUG
	pKB165	del 1 dir AUG
	pKB166	ins 1 U top SLIV
	pKB174	ins 3 U dir AUG
	pKB175	del 2 dir AUG
	pKB176	del 3 dir AUG
	pKB177	ins 2 U top SLIV
	pKB178	ins 3 U top SLIV
	pKB215	del 6 dir AUG
	pKB216	del 9 dir AUG
	pKB210	pSW4; ins 2 U dir AUG
	pKB211	pSW4; ins 3 U dir AUG
	pKB212	pSW4; ins 1 U top SLIV
	pKB213	pSW4; ins 2 U top SLIV
	pKB214	pSW4; ins 3 U top SLIV
	pKB222	pSW4; del 1 dir AUG
	pKB223	pSW4; del 2 dir AUG
	pKB224	pSW4; del 3 dir AUG
	pKB217	pSW4; del 6 dir AUG
	pKB218	pSW4; del 9 dir AUG
Move UU between SLIV and AUG	pCO1	Vary pos'n of UU ins
	pCO3	Vary pos'n of UU ins
	pCO5	Vary pos'n of UU ins
	pCO7	Vary pos'n of UU ins
	pCO9	Vary pos'n of UU ins
Single point mutations between SLIV and AUG	pKB229	SLIV: C334U
	pKB230	SLIV: C334A
	pKB231	SLIV: C335U
	pKB232	SLIV: C335A
	pKB233	SLIV: G336A
	pKB234	SLIV: G336U
	pKB235	SLIV: U337C
	pKB236	SLIV: U337A
	pKB237	SLIV: G338A
	pKB238	SLIV: G338U
	pKB239	SLIV: C339U
	pKB240	SLIV: C339G
Coaxial	pKB167	SIJ/II stack CC/GG → CG/GC –2.9 →

Stacking		-2.0 kcal
	pKB168	SIJ/II stack CC/GG → GC/CG -2.9 → -3.4
	pKB169	SIJ/II stack CC/GG → AA/UU -2.9 → -0.9
	pKB170	SII/dIV stack AG/UC → AU/UA -1.7 → -0.9
	pKB171	SII/dIV stack AG/UC → GG/CC -1.7 → -2.9
	pKB207	SI/II stack GC/GC → GG/CC -3.4 → -2.9
	pKB208	SI/II stack GC/GC → CG/CG -3.4 → -2.0
	pKB209	SI/II stack GC/GC → AA/UU -3.4 → -0.9

4.2.2 Transcription and quantification of reporter mRNA

Uncapped reporter RNAs were *in vitro* transcribed using T7 polymerase from HindIII-digested plasmids. Transcription reactions (50 μL) contained digested template plasmid (5 μg), RNasin Plus RNase inhibitor (1.5 μL; Promega), NTPs (5 mM), T7 polymerase (0.1 mg/ml; purified by Kaihong Zhou), pyrophosphatase (50 ng; Roche), DTT (10 mM), 30 mM Tris HCl, pH 8.1, 25 mM $MgCl_2$, 2 mM spermidine, and 0.01% Triton X-100. After DNase I treatment (RNase-free, Promega, 0.09 U/μL for 15 min at 37°C), RNA was purified from free nucleotides and enzyme using RNA Clean&Concentrator-25 columns (Zymo Research). RNA was eluted in HEPES-KOH pH 7.5 (20 mM), distributed into small, single-experiment aliquots and stored at -80°C until use. Initial quantification of RNA concentration was determined by absorbance at 260 nm. Concentrations were adjusted based on densitometry measurements of the full-length RNA, run on an ethidium bromide-stained 1% agarose TAE gel. Sodium dodecyl sulfate (SDS, 0.025%) was present in the 2x formamide loading dye and was essential for RNAs to run as clean bands in non-denaturing agarose gels. The intensities of full-length RNA bands were normalized to a lane-by-lane loading control of 40 ng linearized dsDNA plasmid. All six RNAs for a given translation experiment were run in duplicate on the same 1% agarose gel immediately before the translation experiment, and requantified as above. In rare cases, RNAs that showed significant smearing or degradation were discarded and transcribed again.

Hints: Plasmid DNA was prepped with HiSpeed Plasmid Midi Preps (Qiagen). The manufacturer's protocol was followed, with modifications in the final steps: DNA was eluted from Midi-concentrators with warm EB buffer (600 μL). This buffer was passed through the concentrator a second time, and was followed with additional warm EB buffer (200 μL). If needed, plasmid DNA was concentrated by EtOH precipitation before restriction digestion.

The majority of the error in these *in vitro* translation reactions arises from quantification of RNA concentrations, since different mutant RNAs were being compared. I only transcribed and translated six RNAs at once: WT and five mutants. I ran the RNAs on an agarose gel in duplicate after transcription and before translation reactions. It was essential to recheck concentrations of dilutions immediately before each translation reaction. Huge lane-to-lane differences were observed in RNA band strength, coming from parameters such as the depth of the gel, curvature of the ethidium front, etc. Thus, the lane-by-lane DNA loading control was

essential to normalize RNA signal and get reasonable standard deviations between duplicate samples.

4.2.3 In vitro translation reactions

Rabbit reticulocyte lysate (RRL) *in vitro* translations were performed using Promega's standard nuclease treated Rabbit Reticulocyte Lysate System. Each reaction (15 μL) contained nuclease treated RRL (56% (v/v)), amino acids (20 mM), RNasin Plus RNase inhibitor (1.3 U/ml; Promega), Complete Protease Inhibitor Cocktail (Roche: 1 tablet in 1 ml used as 50x), DTT (2 mM), and 1.8 mM $MgCl_2$, 45 mM KCl, and 26 mM KOAc, to obtain final salt concentrations of 2.2 mM Mg^{2+}, 45 mM KCl, and 90 mM KOAc (Kozak 1990; Berry et al. 2010). Translation reactions were initiated with the addition of 10 ng reporter mRNA (quantified as described above), incubated at 30°C for 30 minutes, and stopped with the addition of puromycin (7.5 μL 60 mM stock). These conditions were in the linear range for signal with respect to both RNA concentration and translation time. Luciferase activity was measured in a Veritas microplate luminometer (Turner BioSystems) after addition of Luciferase Assay Reagent (50 μL; Promega) to each well, according to the manufacturers' instructions. A mutant's translation activity is reported as the mean of its activity normalized to wild type from each of four independent translation experiments (from two independent transcriptions), and is reported with the standard deviation between these four values.

For *in vitro* translation experiments containing 2′-OMe oligonucleotides, serial dilutions of 10x oligonucleotides were made, and the oligonucleotides were mixed with translation extracts prior to the addition of WT IRES-FF luc reporter RNA. The means for duplicate translation reactions at each oligonucleotide concentration were plotted and fit to the inhibition curve %Translation Activity = 100 / (1 + 10^(log(IC_{50}) - log[inhibitor]) using Kaleidagraph (Synergy Software). Sequences for the 2′-OMe oligonucleotides were IIId: 5′-ACCCAACACU ACUCGGC-3′, SII / dom IV: 5′-UGCACGGUCUACGAGA-3′, SI / SII: 5′-CUACGAGAC CUCCCGG-3′, SII: 5′-CGGUCUACGAGA-3′, and dom IV: 5′-GAUGCACGGUC-3′.

4.2.4 Primer extension inhibition (toeprinting) assays

40S ribosomal subunits were purified from HeLa cytoplasmic lysate (a gift from R. Tjian's laboratory) as described previously (Fraser et al. 2007) and in Chapter 2. Reverse transcription was conducted from a 5′ end-labeled DNA primer that hybridizes to nts 18-37 of the FF luciferase ORF (5′-GCGCCGGGCCTTTCTTTATG-3′; Fraser et al. 2009). Toeprinting reactions were performed based on previous descriptions (Pestova et al. 1998b; Fraser et al. 2009) with minor modifications. Briefly, toeprinting reactions (10 μL) contained IRES-FF luc RNA (~50 nM), 5′ end-labeled primer (50 nM) and 40S subunits (150 nM) in toeprinting buffer (50 mM HEPES-KOH pH 7.5, 50 mM KCl, 2 mM $Mg(OAc)_2$, and 1 mM DTT). Binding reactions were heated to 37°C for 10 min, then incubated for 5 min at RT and 5 min on ice. Reverse transcription was initiated with the addition of extension buffer (1 μL) containing $Mg(OAc)_2$ (80 mM), DTT (10 mM), dNTPs (4 mM) and Superscript III reverse transcriptase (12U/μL; Invitrogen). Reactions were incubated at 30°C for 45 min and stopped on ice with the addition of an equal volume of formamide loading dye. cDNAs were resolved without further purification on denaturing acrylamide sequeincing gels (10%) alongside dideoxy sequencing reactions and detected by phosporimaging. While primer extension gels are cropped above the IIIe loop (at approximately nt 300) in figures, gels were inspected to ensure the presence of

significant stops for each IRES mutant at higher positions on the gel and corresponding to full length RNA, indicating that RNA was not degraded in these reactions.

To quantify toeprint strength, densitometry was performed in ImageQuant TL (Molecular Dynamics) by measuring the intensity within a fixed area around the P-site toeprint for each mutant and normalizing to the toeprint intensity from the WT IRES within the same experiment. Toeprint strength is reported as the mean observation and standard deviation from at least two independent reverse transcription experiments from two independent RNA transcriptions, to control for variability within the assay as well as variations in lane loading. For each mutant, translation activity was plotted against toeprint strength and ordinary least squares linear regression analysis was performed in Kaleidagraph (Synergy Software) to yield an estimated slope and a Pearson correlation coefficient (r).

Hints: *Pouring and running sequencing gels:* First of all, wash plates very well. Siliconize one plate extensively (~800 μL Sigmacote) and full plate lightly (~300 μL Sigmacote). On this second plate, do not resiliconize the top two inches, where the comb will sit. Oversiliconization around the comb will reduce the friction too much, and the wells will be pulled up and become curvy or come out of the plates altogether; undersiliconization will cause the friction to be too high and wells will rip. There is an ideal depth for the comb to be placed: deep enough that the wells will be retained within the glass plates even if some slippage occurs as the comb is pulled out, but shallow enough that the gel loading tips will reach easily into the well so samples can be deposited directly within them. Pull out the comb very slowly and apply pressure as evenly as possible. Don't be discouraged if the wells are disrupted; tedious work with a spacer will help to push them back into place. Run sequencing gels at 45-60W.

Toeprinting reaction workups: Ethanol precipitation did not improve quality of toeprinting sequencing gels, as the salt front runs far below the lengths of cDNAs resolved. Phenol/chloroform extractions also did not drastically change the appearance of the toeprinting gels. Some days, on some gels, a fair amount of radioactivity would remain in the wells, which phenol/chloroform extraction would presumably remove. However, it was determined that more consistent results would be obtained if the number of post-toeprinting steps was reduced. Any error introduced by variable amounts of counts in the wells for different RNAs in the absence of phenol/chloroform extraction should be evident in the averaging of multiple toeprints from different days.

Dideoxy sequencing conditions: Sequencing reactions (10 μL) contained IRES-FF luc RNA (50 nM) and 5′ end-labeled primer (50 nM) in toeprinting buffer. Binding reactions were heated to 65°C for 5 min, then incubated for 5 min at 37°C, 5 min at RT, and 5 min on ice. Reverse transcription was initiated with the addition of the appropriate ddNTP (1 μL 3 mM stock), as well as extension buffer (1 μL) containing Mg(OAc)$_2$ (80 mM), DTT (10 mM), dNTPs (4 mM), AMV reverse transcriptase (4 U/μL; Amersham) and Superscript III reverse transcriptase (37 U/μL; Invitrogen). Reactions were incubated at 30°C for 45 min and stopped on ice with the addition of an equal volume formamide loading dye.

4.3 Results

4.3.1 Stem II forms within the HCV IRES pseudoknot
The proposed pseudoknot structure in the HCV IRES is adjacent to dom IV, a stem loop that contains the AUG start codon (Figure 4.1A). The pseudoknot is predicted to form two stems, SI and SII. These are separated by three single uridine loops, L1-L3 (Figure 4.1A). Despite

77

conservation of a predicted pseudoknot structure across related hepacivirus/pestivirus (HP) IRESs (Hellen & de Breyne 2007), there have been differing results for the HCV and CSFV IRESs regarding the formation of stem II of the pseudoknot (Wang et al. 1995; Rijnbrand et al. 1997; Kieft et al. 2001; Fletcher & Jackson 2002). Mutation of either side of the predicted stem II severely inhibited activity in both systems, whereas compensatory mutations to restore base pairing showed no restoration of activity in the HCV pseudoknot when the entire SII sequence was mutated at once (Figure 4.1B, "Ent SII"; Kieft et al. 2001). When mutations were made two base pairs at a time in the CSFV IRES, however, substantial rescue of activity was observed for compensatory base pairing mutations (Fletcher & Jackson 2002).

To reconcile this discrepancy, mutations analogous to those made in CSFV were introduced into the HCV IRES, changing two nucleotides at a time throughout SII to their complements, as shown in Figure 4.1C. Translation activity of an IRES-firefly luciferase (FF luc) reporter lacking dom I (Figure 4.1A) was measured in salt-adjusted rabbit reticulocyte lysate (RRL) previously optimized for authentic IRES translation (Berry et al. 2010). RNA concentration and time of the translation reaction were adjusted to fall within the linear range of translation signal (data not shown). Disruption of two base pairs at a time on either side of SII in the HCV pseudoknot led to a reduction to 2-40% activity relative to wild type (WT, Figure 4.1C, Table 4.3). Mutations in the middle base pairs were the most deleterious, and mutations in the top base pairs were the least detrimental.

A significant increase in activity (to 60-91% of WT levels) was observed for all three sets of base pairs when compensatory mutations were made on both sides of SII (Figure 4.1C, Table 4.3). Thus, SII base pairing is important for activity in the HCV IRES as previously shown for the related CSFV IRES and initially suggested for the HCV IRES (Wang et al. 1995; Rijnbrand et al. 1997; Fletcher & Jackson 2002). After confirming the lack of activity for the Ent SII comp mutant (data not shown; Figure 4.1B; (Kieft et al. 2001)), we wished to determine if this was due to the uridines (U's) in L1 and L2 being changed to their complementary adenosines (A's). To test this possibility, we made only these loop mutations in the WT IRES background. The U's to A's mutant retained 85% activity (Figure 4.1C, Table 4.3), indicating that the sequence of L1 and L2 is not required for high translation activity.
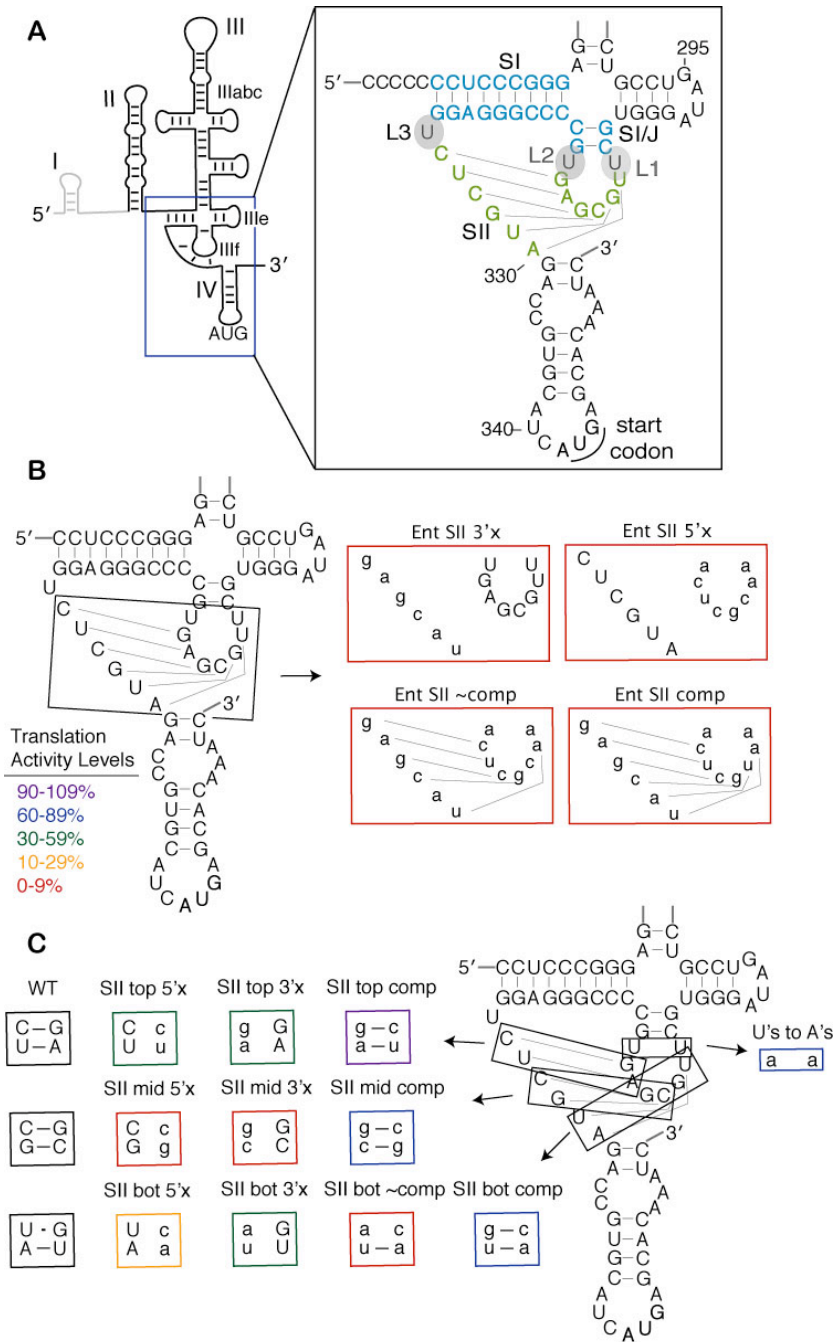
Figure 4.1: Stem II base pairs form in HCV IRES pseudoknot and contribute to IRES translation efficiency. (*A*) Line diagram of predicted HCV IRES secondary structure with major domains and AUG start codon labeled. Dom I is shown in grey and is not included in IRES-luciferase reporter constructs. The inset shows the sequence and predicted secondary structure of the pseudoknot with stem and loop nomenclature indicated. SI is shown in light blue, and SII is shown in light green. (*B*) Mutations of the entire (Ent) SII sequence, as previously reported in Kieft et al. 2001 and (*C*) mutations of SII made two base pairs at a time. Mutated nucleotides are shown in lowercase, and mutants are boxed in colors corresponding to their translation activity levels as defined in *B*. Disruptive mutations are named according to whether the 5′ or 3′ side of SII (in the primary sequence) was mutated to its complement.

Table 4.3: Translation activities of IRES mutants.

Stem II base pair mutations		Global stem / loop mutations	
Mutant	Translation Activity[a]	Mutant	Translation Activity[a]
WT	100 %	WT	100 %
Ent SII 3'x	< 10 %[b]	SII del 1	80 ± 11 %
Ent SII 5'x	< 10 %[b]	SII ins 1	18 ± 6 %
Ent SII ~comp	< 10 %[b]	SI/J elim	1.3 ± 0.5 %
Ent SII comp	< 10 %[b]	SI/J del 1	55 ± 8 %
SII top 5'x	41 ± 9 %	SI/J ins 2	4 ± 2 %
SII top 3'x	40 ± 9 %	SI del 1	90 ± 4 %
SII top comp	91 ± 16 %	SI del 2	3 ± 1 %
SII mid 5'x	2 ± 1 %	SI ins 2	87 ± 12 %
SII mid 3'x	2 ± 2 %	II/III del 2	73 ± 4 %
SII mid comp	60 ± 8 %	II/III ins 2	82 ± 5 %
SII bot 5'x	11 ± 1 %	L2 elim	29 ± 7 %
SII bot 3'x	32 ± 4 %	L2 ins 2	11 ± 1 %
SII bot ~comp	3 ± 1 %	L1 elim	70 ± 11 %
SII bot comp	71 ± 8 %	L1 ins 2	36 ± 7 %
U's to A's	85 ± 1 %	L3 elim	102 ± 9 %
		L3 ins 1	80 ± 7 %
		L3 ins 3	70 ± 11 %
		SI end 5'x	37 ± 3 %
		SI end 3'x	32 ± 5 %
		SI end comp	50 ± 1 %

[a]reported as the mean and standard deviation between four translation reactions from two independent transcriptions.
[b]reported in Kieft et al. 2001.

4.3.2 Stem II base pairing contributes to AUG positioning

Previous reports have suggested that SII base pairing is important for AUG positioning in the CSFV IRES (Pestova et al. 1998b). To test this possibility in the HCV IRES, ribosomal toeprinting was utilized to analyze the effect of the HCV IRES pseudoknot stem II on AUG positioning. Primer extension was conducted on IRES-FF luc mRNA under native conditions using a primer that hybridizes to nucleotides (nts) 18-37 of the FF luc ORF. Primer extension in the absence of 40S subunits yielded prominent stops at U239, A321/G318, and U302 (Figure 4.2A). These stops correspond to the 3′ end of SII, the middle of SI, and the 3′ end of IIIe, respectively (Figure 4.2B), and likely result from secondary and/or tertiary structure within the IRES. In the presence of 40S subunits, new strong stops appeared at nts 343-346, corresponding to the nucleotides at and immediately downstream of the AUG start codon (Figure 4.2A). During translation initiation, this AUG binds to the P site of the 40S ribosomal subunit. No P-site (+2-5, relative to the AUG) toeprint was observed for a mutant in which dom II was deleted (ΔdomII; Figure 4.2A), consistent with previous reports that this mutant does not stably position the adjacent ORF in the mRNA binding cleft (Pestova et al. 1998b; Kolupaeva et al. 2000; Otto & Puglisi 2004).
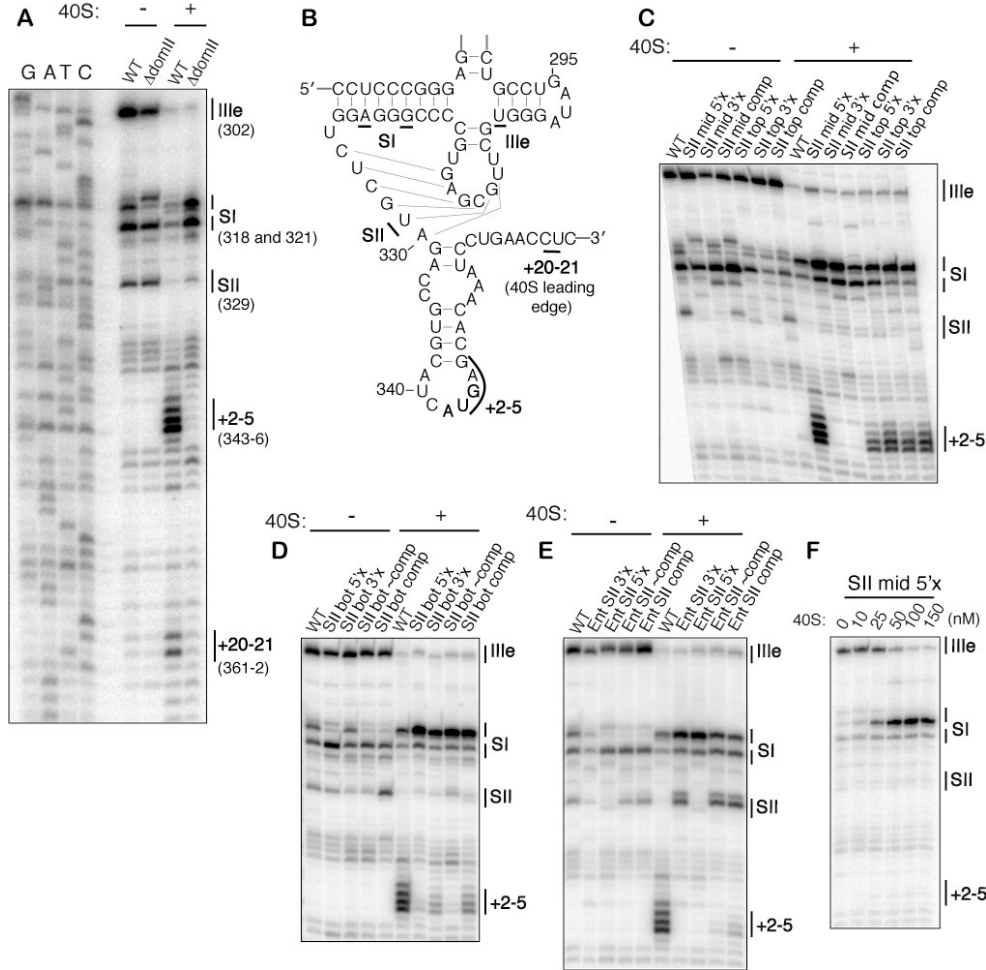
Figure 4.2: Stem II base pairing contributes to AUG positioning by the IRES. (*A*) Establishment of primer extension and toeprinting stops for WT IRES in absence and presence of 40S subunits and comparison to ΔdomII IRES. Primer extension stops were mapped based on dideoxy sequencing reactions. (*B*) Location of major primer extension stops mapped onto pseudoknot predicted secondary structure. Primer extension inhibition and toeprinting of (*C*) middle and top SII base pair mutants, (*D*) bottom SII base pair mutants, and (*E*) entire SII base pair mutants. (*F*) Primer extension reactions of SII mid 5′x IRES-FF luc mRNA (50 nM) in the presence of increasing concentrations of 40S ribosomal subunits. The positions of major stops are indicated at the right of each gel and are defined in *A* and *B*.

While predominant stops at the leading edge of the 40S subunit have been reported previously using either purified rabbit 40S subunits or rabbit reticulocyte lysate (Pestova et al. 1998b; Kolupaeva et al. 2000; Otto & Puglisi 2004; Locker et al. 2007), the P-site (+2-5) stop in this system was previously found to be as strong or stronger than the leading edge stop (Fraser et al. 2009). The +2-5 stop in the P site can be interpreted as an intermediate in the translation initiation pathway in which the AUG is positioned in the P site, but the downstream ORF is not locked into the mRNA entry tunnel to give rise to a leading edge stop. The leading edge (+20/21) stop was weak under the current toeprinting conditions, but when present, its intensity correlated well with the +2-5 stop across mutants, and no mutant ever showed a stronger leading edge toeprint than wild type. Thus, we utilized the stronger +2-5 toeprint in this study to characterize

81

the ability of pseudoknot mutants to place the AUG in the mRNA binding cleft. An additional difference between the toeprints observed here and in several previous reports is that the leading edge stop falls at +20/21 relative to the AUG, rather than +14-15 or +16-18 (Pestova et al. 1998b; Kolupaeva et al. 2000; Otto & Puglisi 2004; Locker et al. 2007). As with the +2-5 P-site stop, the leading edge stop at +20/21 has been consistently observed with the present HeLa 40S / FF luciferase reporter system (Fraser et al. 2009). The appearance of a strong stop at +2-5 and the leading edge stop being at +20/21 rather than the previously observed +16-18 may be due to the use of HeLa rather than rabbit 40S subunits, the FF luc reporter construct, the particular reverse transcription primer, or reverse transcription conditions.

The primer extension patterns of IRESs with mutated SII base pairs showed a weakening of the SII stop in the absence of 40S subunits, though this effect was less pronounced with the bottom base pairs, as compared with the middle and top base pairs (Figure 4.2C,D). This SII stop was significantly strengthened in the compensatory mutants for all sets of base pairs, suggesting that the base pairing mutations had the expected structural effects. The P-site toeprint was clearly reduced in the presence of 40S subunits for the disrupted base pairs of SII that resulted in the strongest translation defects (middle and bottom). Toeprinting was restored by compensatory base pairing mutations (Figure 4.2C,D), as was translation activity (Figure 4.1C, Table 4.3). The P-site toeprint with 40S subunits had near WT strength for the disrupted top base pairs of SII that caused the smallest defect in translation activity, despite destabilization of SII in the absence of 40S subunits (Figure 4.2C).

Primer extension on the previously published Ent SII mutants ((Kieft et al. 2001); Figure 4.1B) showed that SII was indeed destabilized by disruptive mutations and restored by compensatory mutations (Figure 4.2E, - 40S). Nevertheless, very little P-site toeprint was seen for the compensatory mutant, consistent with the reported <10% translation activity (Kieft et al. 2001). Closer examination of the primer extension patterns in the absence of 40S subunits revealed that despite SII formation in the compensatory mutants, there were other structural problems with these mutants as evidenced by the lack of a top SI stop. These structural defects may account for the lack of toeprints and translation activity by these IRESs. Mutation of the bottom SII base pairs on the 5′ side (with respect to primary sequence) led to a similar loss of the top SI stop in the absence of 40S subunits that was not restored by compensatory mutations (Figure 4.2D). This UG sequence may therefore participate in a tertiary interaction with SI.

Based on previous analysis of mutations in SII (Kieft et al. 2001), we hypothesized that the present mutants would be defective in translation initiation downstream of 40S binding. Strong changes in the primer extension pattern were observed for all mutants in the presence of 150 nM 40S subunits (Figure 4.2C-E), suggesting that mutant IRESs were still capable of binding to 40S subunits, even if their toeprints were weak. Accordingly, a dose-response toeprinting experiment of the mutant SII mid 5′x RNA showed a gradual increase in the top SI stop as a function of 40S subunit concentration (Figure 4.2F). This top SI stop is strengthened for nearly all mutants examined in this study, and likely represents a direct impediment posed to reverse transcriptase elongation by the binding of the 40S subunit. While a concurrent loss in the IIIe stop is observed, this is likely a direct consequence of the strengthening of the downstream top SI stop, as there are no additional strong stops observed upstream of the IIIe stop in the presence of 40S subunits.

4.3.3 Global pseudoknot structure contributes to translation and toeprint activity

To further probe the overall structure of the pseudoknot, we systematically shortened or lengthened each of the predicted stems and loops of the pseudoknot. Based on their positions within the tertiary structure of the pseudoknot, we envisioned that some elements would be more tolerant of such changes than others. Stem II was more tolerant of shortening than lengthening by one base pair (80% vs. 16% translation activity) and SI/J was very sensitive to lengthening and shortening, especially by two base pairs (Figure 4.3A, Table 4.3). SI was more tolerant of changes than the other stems, with the IRES showing 87% activity when SI was lengthened by two base pairs. Shortening of SI by one base pair was highly tolerated (90% activity) whereas shortening by two base pairs led to an inactive IRES (3% activity). This translation defect from the shortened SI was not due to insufficient space between domains II and III, as deletion of 2 nts from the poly(C) linker between these domains still yielded 83% activity. Rather, reverse transcription of the SI del 2 mutant (Figure 4.3B) showed that the entire pseudoknot structure was highly destabilized in this mutant, suggesting that the global structure cannot tolerate deletion of two base pairs from SI, at least not deletion of the two GC base pairs specifically tested here. Notably, most mutants examined in this study maintained strong aspects of the WT primer extension pattern, suggesting that the vast majority of mutants did not compromise the global pseudoknot structure (Figures 4.2, 4.3, and data not shown. Note: Please see Figure 4.10 in section 4.5 for gels of primer extension inhibition and toeprinting reactions for loop-length and dom II/III spacer mutants which were not included in the original publication of this work.)

With respect to the loops of the pseudoknot, L2 had the most stringent length requirement with deletion of the single uridine or lengthening by 2 uridines leading to 29% and 11% translation activity, respectively (Figure 4.3C, Table 4.3). L1 had an intermediate tolerance, with elimination of the loop preferred over its lengthening by 2 nt (70% vs. 36%, respectively). L3, on the other hand, was highly tolerant of either insertions or deletions, with the elimination of L3 yielding a surprising 102% activity. The observation that two of the three predicted pseudoknot loops can be eliminated without deleterious effects suggests that if the pseudoknot stems are to resemble coaxially stacked helices, which require spanning loop sequences, the termini of predicted stems might actually melt to allow for longer loops.

To test the ability of the terminus of stem I to melt in the pseudoknot, each side of the first two dom II-proximal base pairs was mutated to its complement, or both to their complements to allow for compensatory base pairing. While disruption of the base pairs either from the 5′ or 3′ side in primary sequence did lead to a reduced translation activity (37% and 32%, respectively), the compensatory base pair mutation showed minimal restoration of activity (50%, Figure 4.3C, Table 4.3). This suggests that the sequence of the dom II-proximal terminus of SI is more important than base pairing within the stem. Mutation of the 3′ side of the SI terminus had a deleterious effect on the SI stops in primer extension reactions, even in the compensatory mutant (Figure 4.3D, -40S), suggesting that this sequence may be involved in a tertiary interaction with other pseudoknot elements.
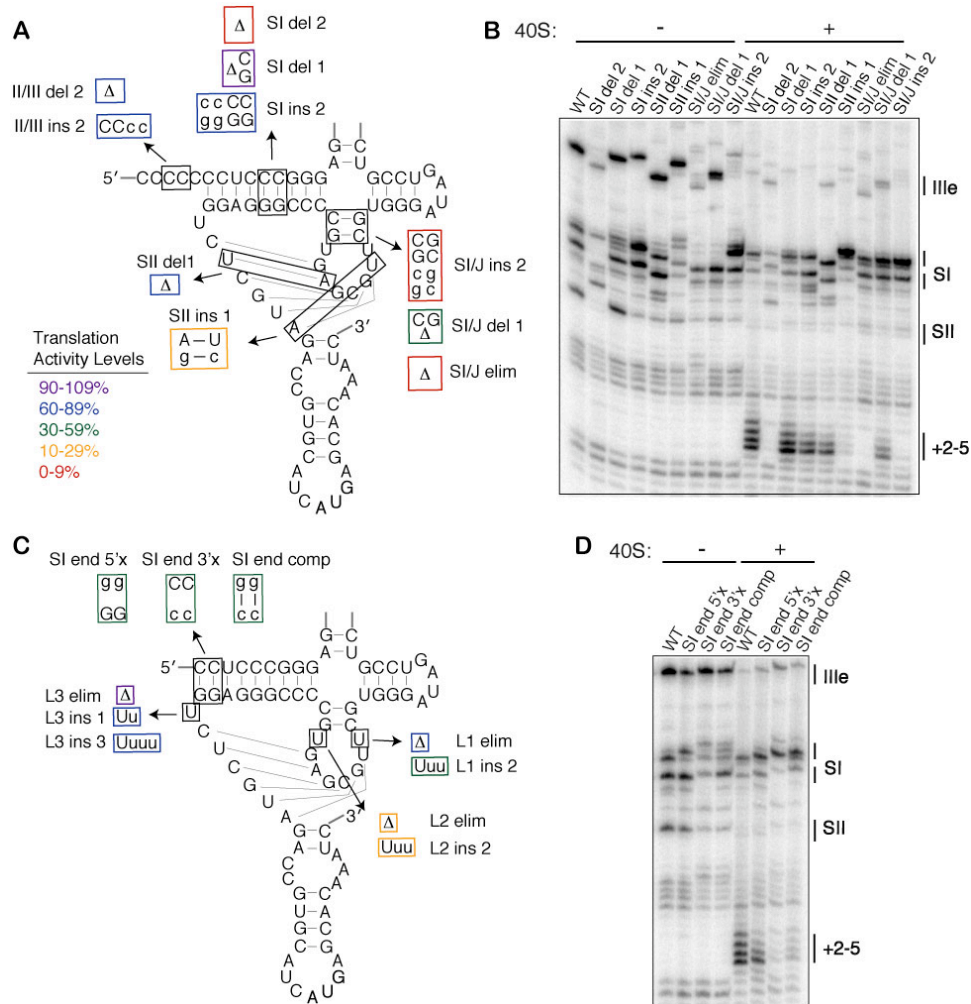
Figure 4.3: Stem- and loop- length tolerance of pseudoknot structure. *(A)* Mutations of pseudoknot stem lengths. Mutated nucleotides are shown in lowercase and mutants boxed in colors according to their translation activity, as defined in the legend. Deletions of nucleotides are shown as Δ. *(B)* Primer extension inhibition and toeprinting of stem length mutants. The positions of major stops are indicated at the right of each gel and are defined in *Figure 4.2A,B*. *(C)* Mutants of pseudoknot loop lengths and base pairing at the terminus of SI, with mutations and activities represented as in *A*. *(D)* Primer extension inhibition and toeprinting of SI terminus base pair mutants, labeled as in *B*.

4.3.4 Strong correlation between toeprint strength and translation activity

Given this diverse set of mutations of the pseudoknot sequence, we wished to determine how well 40S toeprint strength would correlate with IRES-mediated translation activity. Reproducible changes in the primer extension pattern were observed for all mutants in the presence of 150 nM 40S subunits (Figures 4.2, 4.3 and 4.10), suggesting that mutant IRESs were all capable of binding to 40S subunits. The P-site (+2-5) toeprint stop from each mutant was quantified and normalized to the WT IRES from the same experiment. Each mutant's translation activity was then plotted against this normalized toeprint strength (Figure 4.4). Regression analysis of translation activity on toeprint strength produced a slope coefficient of 0.99 ± 0.11 and a correlation (r) of 0.84, which is significant at the .0001 level using a Student's t-test. The

slope coefficient near 1.0 shows that translation activity is proportional to toeprint strength, and is consistent with the IRES pseudoknot structure principally functioning to position the initiation codon. Although the correlation between translation activity and toeprint strength is strong, the relationship is not perfect, indicating that the pseudoknot structure may affect other aspects of translation initiation beyond positioning the start codon. The dispersion of observations from the regression line may also be due in part to the large amount of error inherent in a dynamic assay such as toeprinting, seen in the horizontal error bars in Figure 4.4.
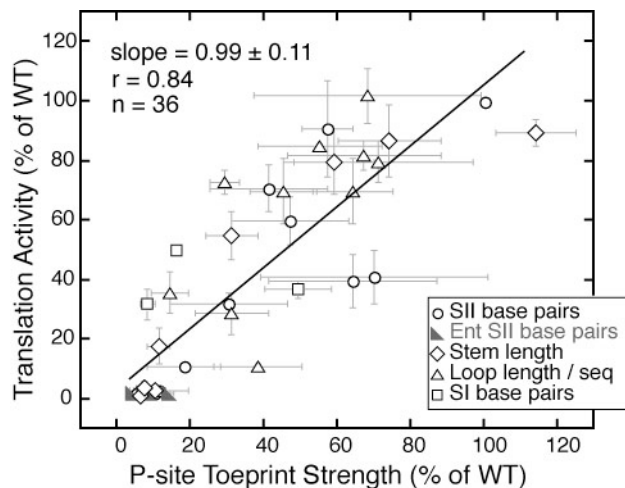
Figure 4.4: Correlation between toeprint strength and translation activity. Translation activity (as quantified in *Table 4.3*) plotted against toeprint stop intensity, quantified by densitometry and normalized to the toeprint of WT IRES. Plotted points reflect the mean value of activity across four *in vitro* translations from two independent transcriptions, and the mean value of toeprint strength from two independent transcriptions. The error bars around points extend one standard deviation above and below the mean. Mutants are grouped into categories based on what structural features they disrupt, as indicated by the legend, but linear regression was conducted on the full set of 36 mutants together. The estimated slope with standard error is 0.99 ± 0.11; the correlation is 0.84, which is significantly different from zero at the .0001 level.

4.3.5 Inhibition of translation by varying pseudoknot-AUG distance

Since the pseudoknot structure affects AUG and ORF positioning in the mRNA binding cleft of the 40S subunit, we tested the dependence of translation activity on the distance between the pseudoknot and the AUG start codon. The WT spacing is 11 nt for the HCV IRES and is well conserved between most related IRESs at 9-13 nts between the end of SII and the AUG (Hellen & de Breyne 2007). Uridines were inserted or nucleotides were deleted directly before the AUG to vary the SII-AUG distance within the IRES (Figure 4.5A). Either deletion or insertion of nucleotides between the pseudoknot and start codon impaired translation activity, with the exception of deletion of a single nucleotide (Figure 4.5B, black line). Thus, the WT IRES sequence has nearly the ideal distance between the pseudoknot and start codon. Primer extension inhibition analysis showed that these mutants retained a strong P-site toeprint with 40S subunits, even as the sequence present in the P site was varied (data not shown).

A

Insert 3 U's	5'-CUCGUAGACCGUGCAUCuuuAUGAGCACAAA-3'
Insert 2 U's	5'-CUCGUAGACCGUGCAUCuuAUGAGCACAAA-3'
Wild Type	5'-CUCGUAGACCGUGCAUC**AUG**AGCACAAA-3'
Delete 1 nt	5'-CUCGUAGACCGUGCAUAUGAGCACAAA-3'
Delete 2 nt	5'-CUCGUAGACCGUGCAAUGAGCACAAA-3'
Delete 3 nt	5'-CUCGUAGACCGUGCAUGAGCACAAA-3'
Delete 6 nt	5'-CUCGUAGACCGAUGAGCACAAA-3'
Delete 9 nt	5'-CUCGUAGAAUGAGCACAAA-3'
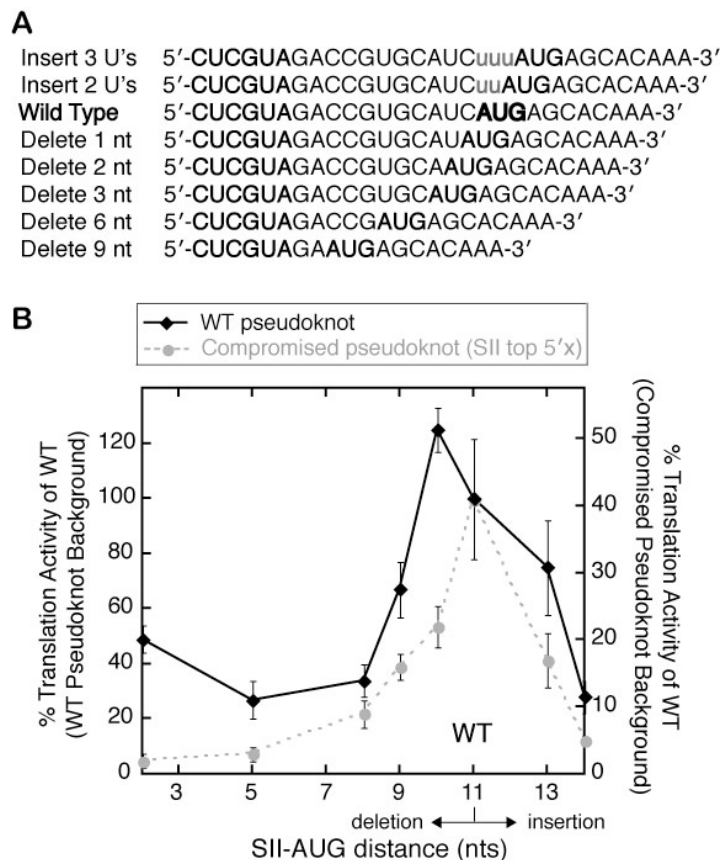
Figure 4.5: Inhibition of IRES translation activity by varying pseudoknot-AUG distance. (*A*) Sequences of mutations made to either lengthen or shorten the pseudoknot-AUG distance. Inserted nucleotides are shown in lowercase letters, and the start codon and SII are in bold. (*B*) Translation activity plotted against pseudoknot-AUG distance with deletions plotted to the left of the WT AUG position and insertions to the right. Translation activities from WT IRES background are plotted on the left vertical axis and those from a compromised pseudoknot (SII top 5′x mutant) on the right vertical axis, with 100% activity on the left axis scaled to 40% starting activity on the right. Plotted points represent the mean activities of four translation reactions from two independent transcriptions and the error bars extend one standard deviation above and below the mean.

Electron cryomicroscopy (cryo-EM) reconstructions of the HCV IRES bound to the 40S ribosomal subunit show that the pseudoknot is thought to bind to the back of the 40S subunit platform, ~50-70 Å away from the P site (Spahn et al. 2001). One model for pseudoknot function is that the global structure of the pseudoknot orients dom IV towards the P site. This hypothesis predicts that a compromised pseudoknot would have a smaller window of tolerance for pseudoknot-AUG distance changes since the orientation of the ORF would be imperfect. To test this possibility, nucleotides were inserted or deleted between the pseudoknot and AUG of the SII top 5′x mutant IRES. This mutant possesses a compromised pseudoknot but still retains 40% activity and a P-site toeprint. The compromised pseudoknot mutant indeed showed a more stringent requirement for the correct pseudoknot-AUG distance when compared to a WT pseudoknot (Figure 4.5B, grey vs. black lines).

4.3.6 Translation inhibition with pseudoknot-targeted 2′-OMe oligonucleotides

The role of the HCV IRES pseudoknot in orienting the ORF toward the 40S mRNA binding cleft suggests that the pseudoknot may be an attractive drug target. It is much easier to imagine a small molecule perturbing the conformation of the RNA structure in the pseudoknot region than disrupting the binding affinity between the 40S subunit or eIF3, given the large surfaces of these interaction areas. As a proof-of-principle demonstration that the pseudoknot structure can be disrupted in *trans*, several 2′-OMe oligonucleotides were designed with complementarity to different regions of the pseudoknot and dom IV (Figure 4.6A). These oligonucleotides were titrated into *in vitro* translations of a WT IRES-FF luc reporter in salt-adjusted RRL. As a positive control, a previously established inhibitory 2′-OMe oligonucleotide which hybridizes to the IIId loop of the IRES ('IIId,' Tallet-Lopez et al. 2003) was tested and had an IC_{50} in our system of ~30 nM (Figure 4.6B, grey). The reverse complement of this sequence served as a negative control and showed no inhibition of IRES-FF luc signal through concentrations of 5 μM (data not shown). Three of the four 2′-OMe oligonucleotides targeted against the pseudoknot and dom IV showed potent inhibition, with IC_{50}s around ~140 nM (Figure 4.6B). The 2′-OMe oligonucleotide that in part targeted SI of the pseudoknot had a considerably weaker IC_{50} of ~400 nM, likely due to a lack of access to this kinetically stable stem. These oligonucleotides introduced strong new stops in primer extension inhibition reactions, but the P-site toeprint of IRES-40S subunit complexes was downstream of these new stops (Figure 4.6C). This toeprint was reduced by the pseudoknot-targeted oligonucleotides in a manner proportional to their inhibitory activity (Figure 4.6C).
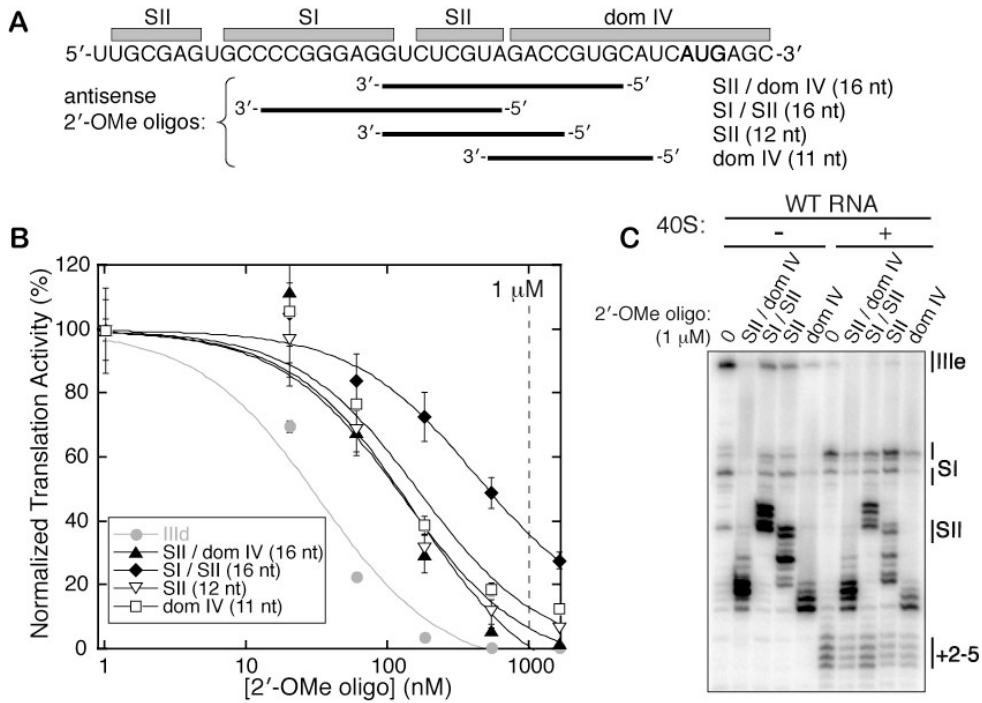
Figure 4.6: Inhibition of HCV IRES translation with pseudoknot-targeted 2′-OMe oligonucleotides. (*A*) Diagram showing to which regions of the IRES each of the four 2′-OMe oligonucleotides are complementary, with the start site AUG shown in bold. (*B*) Titrations of the 2′-OMe oligonucleotides in WT IRES-FF luciferase *in vitro* translation reactions, fit to sigmoidal inhibition curves. Plotted points represent the mean activities of four translation reactions from two independent transcriptions and the error bars extend one standard deviation above and below the mean. The 1 mM concentration utilized in toeprinting experiments in *C* is indicated with a vertical dashed line. (*C*) Primer extension inhibition and toeprinting on WT IRES-FF luc mRNA in the presence of 1 μM 2′-OMe oligonucleotides.

4.4 Discussion

We set out to establish whether the complete pseudoknot structure forms in the HCV IRES and to probe its structural tolerance and functional contributions to translation initiation. Our results show that full pseudoknot structure, including SII base pairs, does form in the HCV IRES and that pseudoknot formation leads to correct positioning of the AUG start codon in the downstream mRNA on the 40S ribosomal subunit. We also find that disruption of the pseudoknot structure using antisense 2′-OMe oligonucleotides blocks IRES-mediated translation, providing evidence that the pseudoknot may be an appealing target for therapeutic intervention.

4.4.1 SII base pairing contributes to IRES function

Conservative disruptions and compensatory base changes in the proposed SII base pairing interactions of the HCV IRES show that SII structure is required for efficient IRES-mediated translation initiation. The bottom and middle base pairs contribute more to translation activity than the top base pair (Figure 4.1C, Table 4.3), perhaps due to proximity to the AUG start codon or to preferential 40S subunit contacts with this portion of the stem. Although compensatory base pairs throughout the stem restore translation activity, the middle and bottom pairs only rescue activity to 60% and 71% of WT levels. This observation likely reflects some

contribution of sequence identity to activity. Mutation of the entire SII stem to its complementary sequence results in an inactive IRES (Kieft et al. 2001). Here we conclude that this inactivity is not due to L1 and L2 sequences being changed from uridines to adenosines. Instead, numerous simultaneous mutations cause global misfolding of the pseudoknot, as evidenced by the lack of an SI stop in compensatory mutants (Figure 4.2E).

An intriguing possibility is that SII may be a dynamic part of the pseudoknot structure, as suggested from RNase probing experiments showing sensitivity to both single- and double-stranded RNases (Wang et al. 1995; Kolupaeva et al. 2000; Fletcher & Jackson 2002). Comparison of the primer extension inhibition patterns of IRES mutants in the absence and presence of 40S ribosomal subunits shows that most mutants (with the exception of the Ent SII series) lose their stable SII stop in the presence of 40S subunits. Given that stops upstream of SII are still strong, this may represent a weakening of SII base pairing upon productive 40S binding. It is still an open question whether the dynamic nature of this stem has a functional role in translation initiation. Mutations in SII that strengthen or weaken the thermodynamic stability of this stem did not show a clear correlation with activity level (data not shown).

4.4.2 The pseudoknot positions the start codon on the ribosome

Comparison of 40S toeprint strength to translation activity among the 35 mutants analyzed in this study reveals a strong linear correlation (Figure 4.4); mutations that perturb ORF positioning on the 40S subunit also result in defective translation initiation. This suggests that the primary function of the pseudoknot is to position the initiation codon, and that the entire pseudoknot structure is required for this function. This global structure may extend further into the main stem of dom III, as it has recently been shown that there is a Watson-Crick interaction between a IIIe loop pyrimidine and a conserved purine/purine mismatch in the stem of dom III (Easton et al. 2009). This interaction serves to stabilize SI of the pseudoknot, and may or may not contribute to the P-site toeprint in 48S pre-initiation complexes (Otto & Puglisi 2004; Easton et al. 2009).

Cryo-EM analysis maps the position of the pseudoknot to the mRNA exit channel at the back of the platform of the 40S ribosomal subunit (Spahn et al. 2001). This location on the small ribosomal subunit is similar to that of the 16S rRNA segment that binds to the Shine-Dalgarno sequence of bacterial mRNAs. In bacteria, this interaction establishes the path of mRNA through the exit channel and into the mRNA binding cleft (Yusupova et al. 2001; Boehringer et al. 2005). Toeprinting analysis shows that the pseudoknots in HCV and CSFV IRESs contribute to AUG positioning (this study; Pestova et al. 1998b; Otto & Puglisi 2004). Notably, many more contacts between the pseudoknot and the platform of the 40S ribosomal subunit are observed by cryo-EM in a structure of the IRES with only the 40S subunit than in a structure with the 80S ribosome (Spahn et al. 2001; Boehringer et al. 2005). There are at least two distinct, but not mutually exclusive, mechanisms by which the pseudoknot could help to position mRNA stably in the binding cleft. One possibility is that the structure sets the angle at which dom IV is presented to the interface region of the 40S subunit, orienting it appropriately through the mRNA exit tunnel. A second possibility is that the pseudoknot structure allosterically manipulates the mRNA binding cleft by influencing the conformation of the 40S subunit.

Supporting the allosteric manipulation model is the previous observation that toeprint strength in 48S complexes may not actually correlate with codon-anticodon stability (Locker et al. 2007). This suggests that the lack of a stable toeprint from a dom II deletion mutant may be partially due to the absence of a conformational change that affects the shape of the mRNA

binding cleft (Spahn et al. 2001). Such conformational changes could clearly have direct effects on how strongly the mRNA binding cleft blocks an elongating reverse transcriptase molecule. The pseudoknot is positioned at or near the intersection between the 40S body and head, a flexible region of the small ribosomal subunit. It is possible that, similarly to dom II, the pseudoknot may allosterically alter the shape of the mRNA binding cleft and thus influence toeprint strength and the initiation pathway.

Modulation of pseudoknot-AUG distance in the WT IRES background shows that the IRES maintains a near-optimal distance between the pseudoknot and the initiation codon. This distance range resembles that between the Shine-Dalgarno sequence and the AUG in bacterial mRNAs, which is most commonly 4-12 nts in *E. coli* (Ringquist et al. 1992; Ma et al. 2002). The stimulation of translation activity observed upon reducing the pseudoknot-AUG distance by 1 nt may be an effect of destabilizing the stem-loop of dom IV (Honda et al. 1996). The deleterious effect of larger deletions in this spacing likely results from the AUG not being able to reach the P site. It should be noted that this genotype 1b sequence contains an A at position 350 and does not possess the downstream ACG codon that was proposed to serve as an alternate translation start site in other HCV IRES sequences (Reynolds et al. 1995). The observation that insertions between the pseudoknot and AUG hurt translation activity considerably is consistent with previous conclusions that scanning does not occur in CSFV or HCV IRES-mediated translation initiation (Reynolds et al. 1996; Rijnbrand et al. 1997). It also suggests that AUG is not a "ball on a string" suspended from the pseudoknot but is more actively guided to the binding cleft, quite possibly through the same path as that observed in bacteria (Yusupova et al. 2001; Boehringer et al. 2005).

We predicted that if the pseudoknot positions the ORF in the mRNA binding cleft, a compromised pseudoknot would tolerate a smaller range of SII-AUG distances than the WT pseudoknot for effective translation initiation. Indeed, this is observed for the SII top 5′x mutant relative to the WT IRES (Figure 4.5B), supporting a model in which the pseudoknot structure sets the angle at which dom IV is presented to the interface region of the 40S subunit, orienting it appropriately through the mRNA exit tunnel. It is also possible that the pseudoknot structure allosterically manipulates the mRNA binding cleft by influencing the conformation of the 40S subunit and thus influences the P-site primer extension stops.

4.4.3 Towards the global structure of the pseudoknot

In the absence of other tertiary structural constraints, the most thermodynamically stable way for two helices of a pseudoknot to interact is to coaxially stack to form a continuous helix in three-dimensional space (Pleij et al. 1985; Wyatt et al. 1989; Brierley et al. 2007). The fact that all three predicted loops in the HCV IRES pseudoknot are single nucleotides, and especially the observation that L1 and L3 can be eliminated without deleterious effect (Figure 4.3C, Table 4.3), suggests that this pseudoknot may not assume a classic, coaxially stacked structure. A lack of base pairing at the SI terminus could partially compensate for the short loop sequences. The possibility that SII might coaxially stack with either SI/J or SI was considered, but mutations made to test these possibilities gave no direct evidence for these arrangements (data not shown). The pseudoknot may adopt a complex globular structure that evades modeling based on mutational studies alone. It will be critical in the future to determine the high-resolution structure of this region of the IRES. Current efforts are directed at solving a crystal structure of the HCV IRES pseudoknot (see appendix II).

The present mutational study provides new insights into the nature of the pseudoknot structure, including the length tolerance of loops and stems within the pseudoknot. In addition, activity assays with mutant IRESs suggest a potential sequence-specific requirement at the dom II-proximal terminus of SI, as compensatory base pairing mutations in the final two base pairs of this stem do not show a significant rescue in activity over mutations of either individual side of the stem (Figure 4.3C, Table 4.3). Primer extension inhibition analysis suggests that mutation of the 3′ side (in primary sequence) of this stem preferentially disrupts SI stability (Figure 4.3D), indicating a potential role of these nucleotides in a tertiary structure. In contrast, previous mutational analysis shows that compensatory mutations fully rescue activity when the 2nd-5th bases in this stem are changed from CUCC to AGAA (Wang et al. 1994), and similar results have been seen for internal positions in SIa of the CSFV IRES (Fletcher & Jackson 2002). The simplest explanation for this difference would be that a sequence-specific requirement occurs in the terminal base pair of SI, rather than in the second internal base pair. Other sequence-specific requirements are observed in this region of the pseudoknot in the related CSFV IRES, where mutating the A-rich L3 to contain more uridines leads to a drop in activity (Fletcher & Jackson 2002).

4.4.4 The pseudoknot as a drug target

The HCV IRES is more than a molecular scaffold for recruiting eIF3 and the 40S ribosomal subunit, as mutations can dramatically reduce translation activity without strongly affecting the affinity of the IRES for these translation factors (Kieft et al. 2001; Otto & Puglisi 2004). The places within the IRES where conformations of the RNA contribute to activity are vulnerable regions at which small molecules may be able to interfere. The pseudoknot is likely to be such a structure. Previous binding affinities measured for severe pseudoknot (Ent SII) mutants suggest that proper pseudoknot structure does not strongly contribute to binding affinity for 40S subunits (Kieft et al. 2001). In addition, for all mutations made throughout the global structure of the pseudoknot, consistent changes are observed between primer extension inhibition patterns for the IRES in the absence and presence of 40S subunits (Figures 4.2, 4.3 and 4.10). This suggests that any differences in the binding affinities of mutant IRESs for the 40S subunit are relatively small and not significant enough to cause the observed lack of translation activity in rabbit reticulocyte lysate reactions (which we estimate to contain ~600 nM 40S subunits (Meyer et al. 1982)).

2′-OMe oligonucleotides that can base pair with sequences within SII of the pseudoknot show potent inhibition of IRES activity (Figure 4.6). A 2′-OMe oligonucleotide targeted to the IIId loop of the IRES (used here as a positive control for inhibition) was dropped after a Phase I trial due to a lack of on-target activity and aminotransferase flares (McHutchison et al. 2006). Nonetheless, inhibition from these pseudoknot-targeted 2′-OMe oligonucleotides serves as a proof of principle that disrupting the pseudoknot structure would be a viable target for novel HCV therapeutics. Indeed, the fact that point mutations are capable of drastically reducing translation activity suggests that a small molecule might be able to accomplish this goal by perturbing the specific pseudoknot conformation required for AUG positioning. To fully realize the goals of understanding the molecular mechanism of the HCV IRES pseudoknot's function and to incapacitate this function with a small molecule, a critical objective will be the determination of an atomic resolution structure of this region of the IRES (see appendix II).

4.5 Addendum: unpublished Results and Discussion

Preliminary results for additional HCV IRES pseudoknot mutants, which were not included in the publication encompassing the rest of the chapter, are described here. The names, sequences, descriptions and preliminary translation activity data for all mutants are presented in Table 4.1, and Table 4.2 presents these mutants in groups, based on the hypotheses they were designed to test.

4.5.1 SII stability

One of the six base pairs predicted in SII of the HCV IRES pseudoknot is a GU wobble base (Figure 4.1A). Mutation of this GU wobble to a full GC base pair resulted in a drop to 67 ± 5 % of the WT IRES translation activity (Table 4.1). This moderate drop in activity could either be due to a sequence-specific requirement for the uridine base, or potentially to the GC base pair overstabilizing the helical nature of SII. The latter possibility was especially intriguing given the prior evidence for dynamic structure of this stem (Wang et al. 1995; Kolupaeva et al. 2000; Fletcher et al. 2002). To further test whether the dynamics or flexibility of SII play a functional role in HCV IRES translation, we tested whether there was a general correlation between the predicted thermodynamic stability of SII and IRES translation efficiency, assuming that the predicted thermodynamic stability of the stem would correlate overall with its kinetic stability and dynamics. A panel of mutants (Table 4.2) was generated which varied a single base pair identity within SII (GC/AU/GU) or multiple base pairs at once to compensatorily adjust the stability of the stem. When translation activity of each of these mutants was plotted against the mutant's predicted SII stability (Figure 4.7), essentially no correlation was observed (R=0.29). Thus, there is no evidence to support the notion that overstabilizing SII base pairing is detrimental to IRES activity; instead, defects seems more likely to stem from sequence-specific effects. As a caveat, however, many of the data points plotted in Figure 4.7 (those without error bars) are rough estimates of activities from preliminary measurements, but could be revised upon further study.
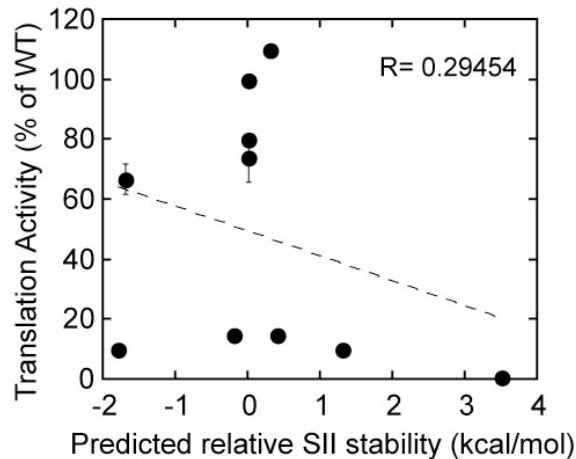
Figure 4.7. Lack of correlation between SII stability and translation activity. The translation activity (relative to WT HCV IRES) for all pseudoknot mutants which varied the sequence of base pairs (between GC, AU, and GU) is plotted against the relative predicted thermodynamic stability of the stem (calculated by mfold (Zuker 2003)), by generating mock hairpins with the 6 base pair stems with a GUAA tetraloop in between; this should accurately predict the relative stabilities of the various mutants). The WT SII sequence is predicted by this algorithm to have a stability of -8.5 kcal/mol, and all of the energies are shown relative to this value. Linear regression shows the correlation has an R-value of 0.29, strongly suggesting that translation activity is not correlated with thermodynamic stability of SII. Error bars represent standard deviations between average values for four separate experiments (two independent translation reactions from two independent transcriptions) and are shown for the two mutants for which such extensive measurements have been collected. Data points without error bars should be considered preliminary.

4.5.2 Coaxial stacking

Another question we hoped to be able to address through mutational analysis was what the overall topology of the pseudoknot's fold is in the context of the IRES. A classic H-type pseudoknot folds with two stems coaxially stacking and spanned by loop sequences (Pleij et al. 1985; Wyatt et al. 1989; Brierley et al. 2007). Although the loop sequences of the HCV IRES pseudoknot are predicted to be extremely short (each single U bases, Figure 4.1), there is an unusual four-way junction in the middle of the pseudoknot domain, so it is difficult to predict what the three-dimensional arrangement of the two pseudoknot stems would be. We envisioned three potential ways the pseudoknot stems could coaxially stack: SI/J stacking with SII, SII stacking with SLIV, or SI stacking with SII (Figure 4.8A). The latter interaction would almost certainly require melting of at least one of the predicted SI/J base pairs. The base-pair identities of the terminal base pairs of each of these predicted stems were mutated to vary the energy of potential coaxial stacking between the stems, using nearest neighbor base stacking energetic parameters (Freier et al. 1986). If two of the stems were indeed coaxially stacking, the expectation would be that there would be a strong correlation between the predicted energy arising from coaxial stacking of the stems and the translation activity of the mutant. No striking correlation was observed for any of the three potential stacking interactions (Figure 4.8B). This suggests that either no two stems are coaxial stacked in three dimensions or that there are other sequence-dependent factors that contribute more to the translation activity of the IRES than the energy derived from coaxial stacking.

93
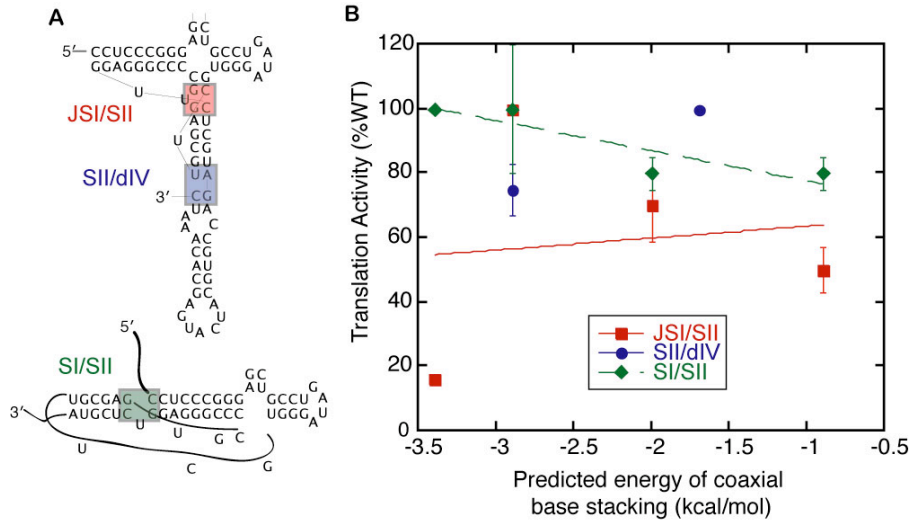
Figure 4.8: Mutational analysis of potential coaxial stacking in pseudoknot domain. *(A)* Two potential tertiary structures of HCV IRES pseudoknot, with SII coaxially stacking on SI/J and dom IV hairpin (top) or SII stacking on SI, requiring the SI/J region to not be base paired (bottom). Colored, shaded regions represent the base pairs on either helix that would be coaxially stacking in the proposed arrangements. These base pairs were mutated from their WT sequences, and *(B)* the translation activity of the mutants were plotted against the predicted energy of the resulting coaxial base stacking interaction, (nearest neighbor energies taken from (Freier et al. 1986)).

4.5.3 SI/J base pairing

In the course of considering an alternate arrangement of SI and SII that would require melting of SI/J base pairs (Figure 4.8A, bottom), we wondered how much these base pairs contributed to the structure and activity of the IRES. The top and bottom base pairs were independently disrupted by mutating one or the other nucleotide to its complement, or compensatorily restored by mutating both nucleotides to their complements at once (Figure 4.9A), and these mutants were tested for their *in vitro* translation activity. For the top base pair, the compensatory mutant has 71% activity vs. 21% or 60% for the disruption of either side of the base pair (Figure 4.9B), suggesting that base pairing perhaps plays a modest role in the predicted top base pair of SI/J. The mutations of the predicted bottom base pair of SI/J strongly suggest that sequence is more important than base pairing at this position. Mutating either side to its complement yields 39% and 80% activity, whereas the compensatory mutant has only 16% translation activity (Figure 4.9B). A caveat to this type of mutational analysis is that changing the sequence of one portion of the pseudoknot structure may have unintended consequences in the folding of other regions of the structure.
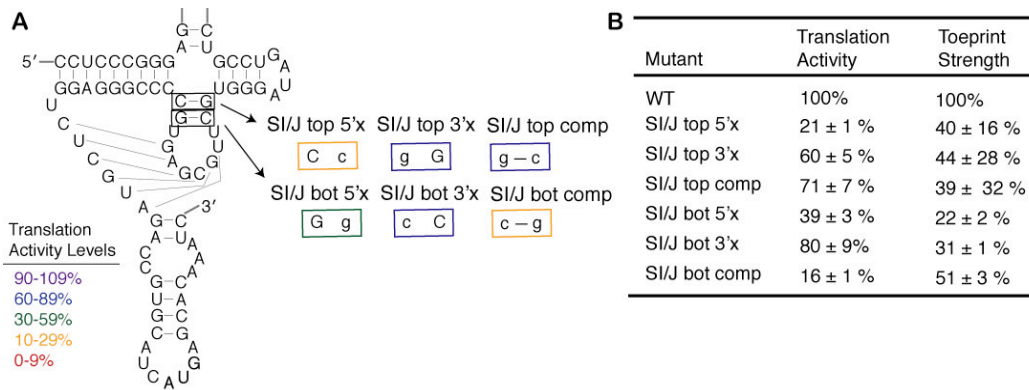
Figure 4.9: Mutational analysis of SI/J base pairs. (*A*) Mutations of base pairs within the predicted SI/J region. Mutated nucleotides are shown in lowercase, and mutants are boxed in colors corresponding to their translation activity levels as defined in the legend. *(B)* Quantitative translation activity and toeprint strength (relative to wild type HCV IRES) for each mutant. Toeprint strength was quantified by densitometry and normalized to the toeprint of WT IRES. Reported values reflect the mean value of activity across four *in vitro* translations from two independent transcriptions, and the mean value of toeprint strength from two independent transcriptions, and the standard deviation around mean is reported.

The results of primer extension inhibition and toeprinting reactions of these SI/J base pairing mutants are intriguing, and seem potentially at odds with the *in vitro* translation results, as certain structural stops are weakened by disrupting base pairs and are restored by compensatory base pairing mutants (Figure 4.10C). These observations suggest that SI/J base pairs do contribute structurally to the pseudoknot domain, and paint a different picture than translation activity values alone. Primer extension analysis shows that both the top SI stop and the SII stop are reduced when the predicted SI/J base pairs are weakened and both stops are strengthened for the top bp compensatory mutant, while only the SII stop is restored for the bottom bp compensatory mutant. In addition, the correlation between translation activity and toeprint strength for these SI/J base pairing mutations is not as strong as the overall correlation for global pseudoknot mutations (Figures 4.9B and 4.4). Together, these observations may suggest that the sequence within SI/J base pairs is important for another aspect of translation initiation by the HCV IRES beyond AUG positioning, as read out by the toeprinting assay.
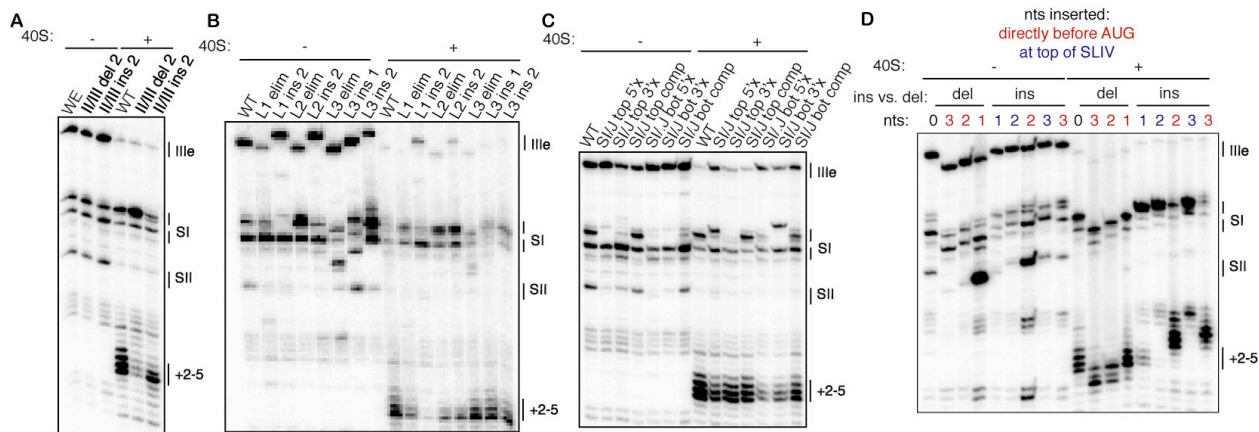
Figure 4.10: Unpublished toeprinting reactions of pseudoknot-mutant HCV IRESs. Representative primer extension inhibition and toeprinting reactions of *(A)* insertion/deletion mutants of the loop between dom II / SI, *(B)* loop-length mutants, *(C)* SI/J base pair mutants, and *(D)* mutants with shortened or lengthened pseudoknot-AUG distances. Quantification of the toeprints in *(A-B)* has been published, but not the gel images themselves. The positions of major stops are indicated at the right of each gel and are defined in *Figure 4.2B*.

4.5.4 Positional effects of domain IV insertions

The effects of inserting or deleting nucleotides directly before the AUG in order to vary the pseudoknot-AUG distance were previously described (section 4.3.4). It is additionally interesting to note that when insertions were made at the very top of SLIV, which would still affect the overall SII-AUG distance in the same way, the inhibitor effect was larger than when the insertions were made directly before the AUG (Figure 4.11A). This difference was especially clear for the UU insertions. Notably, the AUG toeprints for the mutants with insertions at the top of SLIV were much weaker than those directly before the AUG (Figure 4.10D), suggesting that the insertions at the top of SLIV interfere with AUG positioning. One possibility for this would be that the nucleotide sequence at some point within SLIV is important for an interaction with the 40S ribosomal subunit, allowing the RNA to lock in toward the mRNA binding cleft. Insertions at the top of SLIV would push this putative sequence element out of register, whereas insertions directly before the AUG would leave the dom IV sequence up to the AUG in register. Alternatively, the hairpin structure that dom IV forms before melting may be important for AUG positioning.
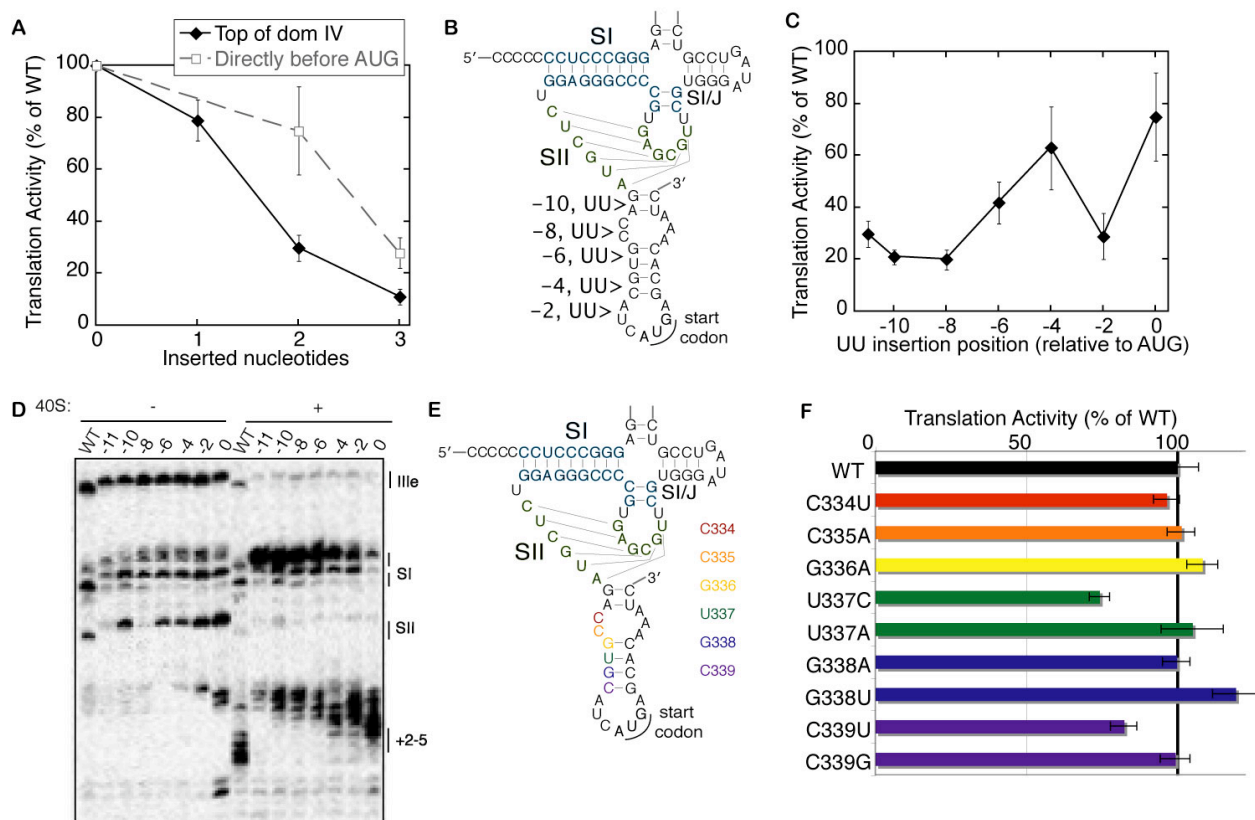
Figure 4.11: Positional effects of insertions between SII and the AUG. Mutations were made to lengthen the pseudoknot-AUG distance by inserting uridine nucleotides either directly before the AUG or at the top of SLIV and *(A)* the translation activity of each mutant is plotted against pseudoknot-AUG distance. (Note: only the mutants with insertions made directly before the AUG have been published.) *(B)* Depiction of additional insertion mutants made throughout the length of SLIV, cloned by Carl Onak (CO). *(C)* Translation activity of UU insertion mutants plotted against the position of the insertion in SLIV, as depicted in *(B)*, collected by CO. *(D)* Primer extension inhibition and toeprinting of UU insertion mutants, conducted by CO. The positions of major stops are indicated at the right of each gel and are defined in *Figure 4.2B*. *(E)* Depiction and *(F)* translation activity of SLIV sequence mutations.

To test this possibility, the position of the UU insertion was varied throughout the dom IV sequence (Figure 4.11B). While there is not a clear partition between "low activity" and "high activity" insertion positions in the translation activity values (Figure 4.11C), activity improves as the insertion is moved from -8 to -6 or -4, relative to the AUG, and this corresponds to a strengthening of the AUG toeprints (Figure 4.11D). These data preliminarily suggest that nucleotide(s) in the middle of dom IV could be important for translation activity and AUG positioning, and so point mutations were made within this region to look for a dependency on translation activity from a specific sequence within dom IV (Figure 4.11E). No single point mutation reduced IRES translation activity more than 30% (Figure 4.11F). For U337 and C339, which did show a slight drop in activity when mutated to a C and U, respectively, no loss of activity was seen when mutated to an A or G, respectively. Thus, the WT U337 and C339 nucleotides are not required for full translation activity.

Chapter 5

Summary

The overarching goal of this work has been to elucidate the molecular mechanisms by which messenger RNA is recruited to human 40S ribosomal subunits during translation initiation and positioned correctly in the mRNA binding cleft. The general mechanisms of mRNA recruitment that would affect nearly every cellular message in a human cell have been studied, as well as the specialized strategy utilized by the Hepatitis C virus (HCV) internal ribosome entry site (IRES), which serves both as a drug target and as a minimal model of translation initiation.

When this work, it had recently been shown using directed hydroxyl radical probing that that, while the majority of eukaryotic initiation factor 3 (eIF3) binds to the solvent-exposed side of the 40S ribosomal subunit, the C-terminal portion of the eIF3j subunit binds to the interface surface of the 40S, near the mRNA entry channel and aminoacyl (A) site. This location of eIF3j suggested new functions for this subunit of eIF3 in mRNA recruitment, by potentially influencing the way mRNA interacts with the mRNA binding cleft during translation initiation. A fluorescence anisotropy assay was developed to examine the equilibrium association of mRNA with purified human 40S ribosomal subunits. A model unstructured 20mer RNA was found to bind to the 40S subunit with high affinity, while the presence of eIF3j in the mRNA binding cleft decreased its affinity by more than 20-fold. In this system, a high affinity mRNA-40S complex was only observed in the presence of the eIF2-tRNA$_i^{met}$-GTP ternary complex (TC). The presence of eIF3j in the mRNA binding cleft caused mRNA affinity to the 40S to be dependent on the presence of TC, which we hypothesized might help to prevent mRNA from associating with the ribosome in a scanning-competent manner until tRNA is properly in place. Indeed, although the human and yeast systems may not be completely equivalent, it has since been shown by the Valasek lab that the deletion of eIF3j (HCR1) from *S. cerevisiae* confers a leaky-scanning phenotype. Preliminary studies have been aimed at further dissecting the molecular consequences of eIF3j/mRNA/TC interactions on 40S subunits during translation initation.

Given that the HCV IRES does not require many of the cap-binding and scanning factors used by a typical cellular message, we were interested in how the specialized mRNA of the HCV IRES positions its initiation codon in the mRNA binding cleft, and turned our attention to the structure most proximal to the AUG in the IRES: a proposed pseudoknot. Site-directed mutational analysis and functional assays were utilized to comprehensively study how elements of the pseudoknot structure contribute to translation initiation. Discrepancies in the literature were resolved to show that Watson-Crick base pairing in the predicted second stem of the pseudoknot does indeed contribute to translation activity. In addition, IRESs with mutant pseudoknots could still bind to 40S subunits, but were deficient at AUG positioning, as analyzed by 'toeprinting' or reverse transcription inhibition assays with purified 40S subunits. This result suggested a model in which the conformation of the pseudoknot positions the mRNA open reading frame (ORF) in the 40S binding cleft. Consistent with this, an IRES with a compromised pseudoknot had a more stringent requirement for the proper distance between the AUG and the pseudoknot. A strong correlation between the toeprinting activity and translation activity of across all pseudoknot mutants suggested there is no other factor in a cellular extract that can rescue AUG positioning for the IRES if the pseudoknot is defective at this task in a simple binary complex with 40S subunits, underscoring the importance of RNA structure and conformation for HCV translation. Current efforts are geared determining a high-resolution

structure of the HCV IRES pseudoknot to better understand how the globular fold of this RNA directs the ORF toward the mRNA binding cleft, and to pave the way toward targeted design of small molecule HCV IRES inhibitors.

In parallel to the work on the HCV IRES pseudoknot, and in collaboration with Gilead Sciences, we conducted a high-throughput screen for inhibitors of the HCV IRES. Small molecule inhibitors of the HCV IRES would be powerful tools, both as starting points for new therapeutics and as chemical biological probes to reveal new mechanistic steps in translation initiation by the IRES. The HCV IRES is a challenging drug target, and several reported inhibitors have lacked translation inhibition activity or failed to show IRES-specific inhibition in our hands. A robust functional assay was established using IRES-luciferase mRNA reporters in rabbit reticulocyte lysate. We determined ideal salt concentrations for authentic translation by the IRES, which was critical for the high-throughput screen as well as the mutational analysis described above. ~400,000 compounds were screened against a dual luciferase reaction with IRES-*Renilla* and 5′ cap-firefly luciferase mRNA reporters. Molecules that specifically inhibited IRES-*Renilla* activity were taken on to a secondary screen with the swapped reporter pair (IRES-firefly and 5′ cap-*Renilla*) with the aim of eliminating *Renilla* enzyme inhibitors. Unfortunately, even after extensive counter screening, the most promising hits proved to be *Renilla* luciferase inhibitors, and no specific IRES inhibitors were identified. We learned an important lesson through this experience: even with a robust assay, it is unwise to use a classical drug target, such as an enzyme, as an intermediate to look for inhibitors of a more challenging drug target, such as the HCV IRES. Enzymatic inhibition will be so much more facile than inhibition of a critical conformation of the IRES that no counter screen is likely to eliminate all of the false hits arising from luciferase interference. In the future, HCV IRES inhibitors may be more efficiently identified using fragment-based approaches to look for small molecules that bind to regions of the IRES for which we know RNA conformations are important, such as the pseudoknot described above. The pseudoknot has been validated as a possible HCV IRES drug target using antisense 2′-OMe oligonucleotides that disrupt this RNA structure.

The last decade and a half has witnessed significant progress in the field's understanding of the molecular events that underlie translation initiation in eukaryotes. In particular, structural biological and quantitative, biophysical methods have played critical roles in this progress. The work presented here contributes to our understanding of the interaction between mRNA and 40S ribosomal subunits during translation initiation. Our findings should stimulate future investigations into the *in vivo* and *in vitro* consequences of the interface-binding factor eIF3j on scanning and AUG recognition. In addition, we have characterized a critical step in translation initiation by the IRES in the HCV genomic RNA. Once the IRES RNA has globally bound to the solvent-side of 40S ribosomal subunits, a secondary binding event of the start codon to the mRNA binding cleft is facilitated by an RNA pseudoknot structure upstream of the AUG, underscoring the idea that the HCV IRES functions as more than a mere scaffold for capturing 40S ribosomal subunits. Based on the likely dynamic nature of the HCV IRES pseudoknot and the critical role of its precise conformation, this region holds great promise as a highly conserved RNA drug target for novel Hepatitis C virus therapeutics.

Appendix I

Attempts to inhibit HCV IRES translation by published small molecules

IA. Lack of HCV IRES inhibition by an mRNA display-selected cyclic peptide

- Description of my extensive efforts to reproduce published *in vitro* translation results from the Szostak lab. These were meant to set the stage for mechanistic studies of the inhibition by the cyclic peptide, but unfortunately the compound's inhibition of and binding to the HCV IRES could not be reproduced, leading to the retraction of the paper: Litovchick A, Szostak JW. (2008) Selection of cyclic peptide aptamers to HCV IRES RNA using mRNA display. *Proc Natl Acad Sci,* **105,** 15293-8.

IB. Lack of specific HCV IRES inhibition by ISIS heteropolycycles

- Description of *in vitro* translation experiments to examine the inhibition of benzimidazoles found to bind to the HCV IRES in the paper: Seth PP, Miyaji A, Jefferson EA, Sannes-Lowery KA, Osgood SA, Propp SS, Ranken R,Massire C, Sampath R, Ecker DJ, Swayze EE, Griffey RH. (2005). SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J Med Chem* **48,** 7099-7102.

Appendix IA. Lack of HCV IRES-inhibition by an mRNA display-selected cyclic peptide

IA.1 Introduction

A recent report from Jack Szostak's lab described an mRNA display-based system to select and evolve cyclic peptides that could bind to HCV IRES RNA (Litovchick & Szostak 2008). The winning peptide, when cyclized or linear, was shown to inhibit HCV IRES-dependent translation in HeLa or rabbit reticulocyte lysate, but not to inhibit translation from a capped reporter RNA (Figure IA.1; Litovchick & Szostak 2008). In addition, published equilibrium binding experiments showed binding affinities for the evolved peptides to the HCV IRES around 4 nM, whereas nonspecific interactions with RNAs other than the IRES were 10-500 times weaker (Litovchick & Szostak 2008). The authors expressed interest in the Doudna lab studying the mechanistic and structural basis of translation inhibition by these peptides and sent us peptide samples for our studies. The extensive unsuccessful efforts to reproduce published *in vitro* translation results from the Szostak lab in our laboratory are described below.
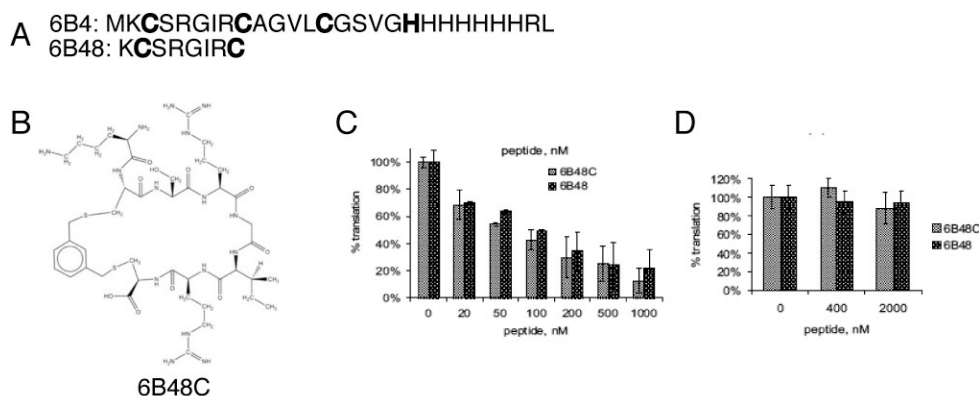
Figure IA.1: Published HCV IRES-binding and -inhibiting peptides. *(A)* Sequence of original winning peptide (6B4) evolved to bind to IRES RNA by Szostak lab, and of shortened, minimal peptide (6B48). Residues that could cyclize with dibromoxylene are shown in bold. Note that 6B4C will be a mixture of regioisomers of doubly cyclized peptides. *(B)* Chemical structure of 6B48 peptide once cyclized with dibromoxylene. Representative *in vitro* translation results from HeLa extracts using *(C)* IRES-Gluc reporter and (D) capped leader-Gluc reporter. Panels B-D are reproduced from the original publication (Litovchick & Szostak 2008).

IA.2 Methods

1A.2.1 Peptides

The peptide sequences used in these studies were MKCSRGIRCAGVLCGSV GHHHHHHHRL (6B4) and KCSRGIRC (6B48; Litovchick & Szostak 2008). 6B4 was used in linear form, whereas 6B48 was cyclized with dibromoxylene (6B48C). The 6B4 peptide was synthesized by GenScript, according to the same procedure as was used for the original order from primary author of the Szostak publication, Sasha Litovchick. Two samples of 6B48C peptide were utilized in these studies: initial experiments were conducted with 6B48C peptide synthesized Litovchick, and later experiments utilized a fresh stock of 6B48C peptide, synthesized by Yollete Guillen in the Szostak lab. MALDI mass spectrometry after several

months of storage in the Doudna lab at -20°C verified the Litovchick 6B48C peptide had not degraded or oxidized.

For initial experiments, peptide dilutions were made in DEPC water alone. For later experiments, peptide stocks were 1 mM 6B48C (in water with 2 mM TCEP; Litovchick stated 2 equivalents of TCEP present with his peptide), and 825 μM 6B4 (in water with 1.7 mM TCEP). Eventually, 10x serial dilutions were made in 20 mM HEPES-KOH pH 7.4, 1 mM TCEP to ensure the peptide was not damaged as dilutions sat on ice before translation experiments. Early experiments in RRL showed that TCEP itself had no effect on translation levels of any reporter mRNAs.

IA.2.2 RNA reporter constructs and preparation

The original IRES-luciferase construct used in the Doudna lab (pKB2) did not have dom I present, and had three MS2 hairpins at the 5′ end (Figure IA.2). A reporter without the MS2 hairpins and with dom I (pKB101) was cloned, as well as other constructs (pKB84 and pKB86) with or without dom I or MS2 hairpins (Figure IA.2). Ultimately, Litovchick shared his own IRES-*Gaussia* luciferase construct, in the form of a PCR product.
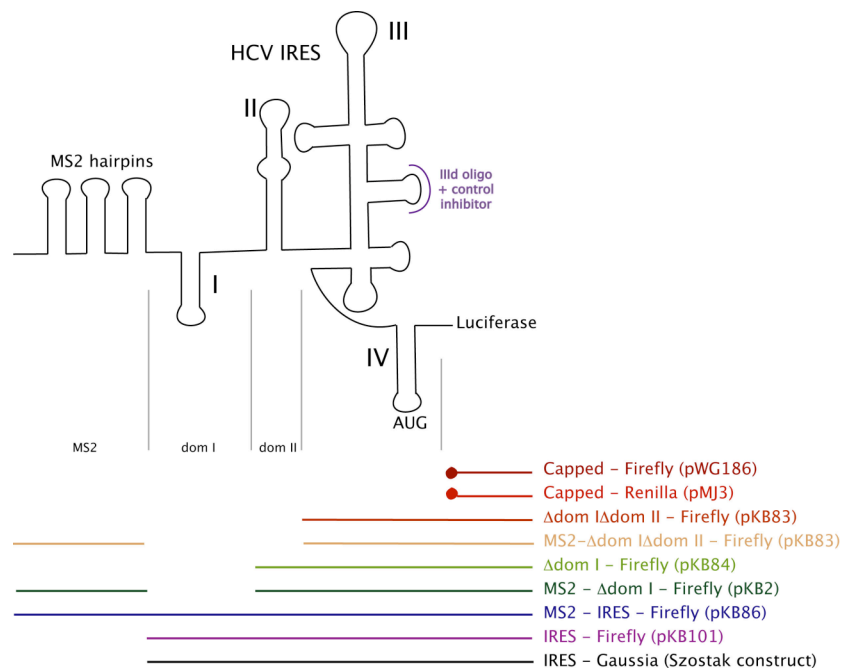
Figure IA.2: Diagram of RNA reporter constructs. The names of each reporter construct is shown and it is depicted whether each possesses three MS2 hairpins, domain I and/or domain II of the HCV IRES, and whether the reporter open reading frame is a firefly, *Renilla*, or *Gaussia* luciferase.

Initial RNA preparation utilized RNA Clean&Concentrator-25 columns (Zymo Research) to remove free nucleotides and enzyme (see section 4.2.2). Experiments were eventually performed with RNA transcribed from the Litovchick IRES-Gluc PCR product, with RNA worked up according to a protocol from Litovchick (personal communication), given potential concerns of the RNA workup affecting the IRES structure. This protocol utilizes LiCl precipitation and ethanol precipitation to remove proteins and concentrate RNA. After

transcription reactions, an equal volume of LiCl (10M) was added, and precipitations were put through 2 freeze/thaw cycles at -80°C. Precipitations were spun at 4°C for 15 minutes at high speed in a microcentrifuge, the supernatant was removed and RNA pellets were resuspended in 1.5M NaOAc, pH5.2, 0.05 mM Mg(OAc)$_2$ and two volumes of EtOH (100%) were added. EtOH precipitations were placed at -80°C for >20 minutes, spun at 4°C for 15 minutes at high speed in a microcentrifuge. The supernatant was removed and RNA pellets were resuspended in "RNA buffer" (20 mM HEPES-KOH 7.4, 100 mM KCl, 1 mM MgCl$_2$) and diluted for experiments in this buffer.

IA.2.3 In vitro translation reactions

Several different translation extract systems were used during the course of these experiments. 1) HeLa extracts were prepared according to the protocol of Otto and Puglisi, 2004, as referenced by Litovchick (Otto & Puglisi 2004). 2) Novagen RRL, with either KCl or KOAc added, was used according to the manufacturer's instructions. RRL is known to vary in activity significantly between the preparations of different companies, and this was the RRL Litovchick used. 3) Promega RRL was supplemented with additional salts to improve the fidelity of the extracts. "Kozak" conditions (2.2 mM Mg^{2+}, 45 mM KCl, 90 mM KOAc (Kozak 1990)) resulted from the addition of 1.8 mM MgCl$_2$, 45 mM KCl, and 26 mM KOAc to Promega RRL translation reactions; "Zhang" conditions simply supplemented the translation reactions with an additional 120 mM KCl (Zhang et al. 1999a). Optimal RNA concentrations for the various extract conditions were determined by RNA titrations, to maximize signal while remaining well within the linear range.

Throughout these experiments, a DNA oligo that hybridizes to the IIId loop of the HCV IRES was used as a positive control for inhibition in RRL and HeLa extracts (IIId oligo; Tallet-Lopez et al. 2003). The IC$_{50}$ of the IIId oligo was found to be ~400-500 nM in both RRL with 4 nM reporter RNA and in HeLa extract with 100 nM reporter RNA. After several full titrations of the IIId oligo inhibitor in each extract, later experiments only included one concentration of the IIId oligo (5 µM; ~10x the IC$_{50}$).

IA.2.4 Filter binding

For filter binding experiments, RNA was 5′ end labeled using γ-[^{32}P]-ATP, annealed in binding buffer (20 mM Tris-HCl pH 7.4, 100 mM KOAc, 200 mM KCl, 2.5 mM MgCl$_2$ and 1 mM DTT) by heating at 65°C for 1 minute then cooling slowly to room temperature in the heat block ("slow cool"). For each experiment, four filters were utilized. From top to bottom, these were: a Tuffryn filter with pores sized 0.2 or 0.45 µm (Pall Corporation) to collect aggregates, a nitrocellulose membrane with pores sized 0.1, 0.2 or 0.45 µm (Protan; Whatman) to bind protein-RNA complexes, a Hybond-N+ nylon membrane (Amersham Biosciences) to bind free RNA, and gel blot paper (Scleicher and Schuell) to retain any remaining material. The filters were presoaked in binding buffer before loaded onto the apparatus. Binding reactions (50 µL) were set up in binding buffer with 5′-^{32}P-end labeled RNA (1,000 cpm) and peptide (0-25 µM), and incubated for 30 min at room temp; aliquots (45 µL) were loaded onto the filter apparatus as a vacuum was applied to pull the sample through the filters. The filters were washed twice with binding buffer (45 µL), allowed to air dry 30 min, and quantified with phosphorimaging. ImageQuant (Molecular Dynamics) was utilized to quantify the counts at each position on each of the filters. The fraction RNA bound at each concentration of peptide was calculated using the equation fraction bound = (nitrocellulose)/(Tuffryn + Hybond + nitrocellulose), and very little

material was found on the Tuffryn or gel blot filters. A negative control was performed with 5′-^{32}P-end labeled human initiator tRNA to examine non-specific binding of the peptides with structured RNA, and a positive control was performed with 40S ribosomal subunits (100 nM), which is known to interact specifically with HCV IRES RNA with a ~2 nM K_d (Kieft et al. 1999).

Hints: As discussed in the results section, care must be taken to choose the appropriate pore size of filters for filter binding. Specifically, the Protran nitrocellulose membranes (Whatman) is made with 0.1-, 0.2- or 0.45-micron pores. The product literature for Protran nitrocellulose membranes suggests that 0.1 μm should retain peptides or proteins <8 kDa, 0.2 μm for 8-20 kDa and 0.45 μm for >20 kDa. While the tuffryn membrane (which collects aggregates) would ideally have the same pore size as the nitrocellulose membrane, this membrane is only available in 0.2- or 0.45-micron sizes.

IA.2.5 RNase footprinting

HCV IRES RNA was transcribed from a double-ribozyme construct (hammerhead ribozyme at 5′ end and HDV ribozyme at the 3′ end), gel purified and 5′-^{32}P-end labeled as previously described (Kieft et al. 1999). Radiolabeled IRES RNA (20,000 cpm) was used in each RNase cleavage reaction (10 μL). RNA was annealed in annealing buffer (20 mM Tris, pH 7.5, 200 mM KCl and 2.5 mM MgCl$_2$) by heating at 65°C for 1 minute, then cooling slowly to room temperature in the heat block. Annealed RNA was then mixed with RNase-free tRNA (1 mg/ml final concentration) and peptide dilutions, made in annealing buffer (final peptide concentrations: 25 nM-25 μM). RNase cleavage was initiated with the addition of RNase A (1 μg/μL; Ambion/Fermentas), Rnase T1 (1 U/μL; Ambion/Fermentas), or RNase V1 (0.1 U/μL; Ambion). The RNase concentrations had been previously titrated to give ~1% cleavage to maintain the overall integrity of the IRES RNA (Clarke 1999). Cleavage reactions were incubated at 37°C for 10 minutes and quenched by transfer to ice and with the addition of an equal volume formamide loading dye. A hydrolysis ladder was generated with 5′-^{32}P-end labeled HCV IRES RNA (40,000 cpm) in hydrolysis buffer (20 μL reaction; 5 mM NaHCO$_3$/Na$_2$CO$_3$, pH 9.2, 0.1 mM EDTA), incubated at 95°C for 5 minutes, then quenched on ice and with an equal volume of formamide dye. A T1 ladder was generated with 5′-^{32}P-end labeled HCV IRES RNA (40,000 cpm) in a reaction (20 μL) in annealing buffer in the presence of tRNA (0.2 mg/ml). RNA was heated to 55°C for 1 min in the complete buffer, and T1 RNase (0.1 U/μL, final; Ambion/Fermentas) was then added and allowed to incubate with RNA for 2 min at 55°C, before reactions were quenched on ice and with an equal volume of formamide dye. RNase cleavage reactions were resolved on a denaturing acrylamide sequencing gel (0.5x TBE, 8M urea, 10% (29:1)) by loading 3 μL next to T1 and hydrolysis ladders. Gels were dried for at 80°C for 1-2 hours under vacuum and visualized using phosphorimaging.

Hints: RNase A cleaves at ssUpN or ssCpN positions, leaves a 3′ phosphate, and retains activity in EDTA. RNase T1 cleaves ssGpN positions, leaves a 3′ phosphate and retains activity in EDTA. RNase V1 cleaves dsRNA, leaves a 3′OH and requires divalent metal ions for activity. To prepare calf liver tRNA for RNase cleavage experiments, care was taken to ensure that no RNases were contaminating the tRNA stock. The tRNA stock was protease K-treated then phenol/chloroform extracted. For this procedure, lyophilized protease K (Sigma) was resuspended to 5 mg/ml in 50 mM Tris-HCl, pH 8.0 and 10 mM CaCl$_2$. Calf liver tRNA (400 μL 25 mg/ml) was treated with protease K (50 μg/ml) in 450 μL at 50 mM Tris-HCl, pH 8.0, 5 mM

CaCl₂, and 2 mM DTT for 2 hours at 37°C. The reactions were extracted with phenol/chloroform, then with chloroform alone, and RNA was recovered by EtOH precipitation.

IA.3 Preliminary Results and Discussion

IA.3.1 Determination of ideal *in vitro* translation conditions

As was discussed extensively in Chapter 3, the salt concentrations present in rabbit reticulocyte lysate have a large effect on the activity and fidelity of HCV IRES-mediated *in vitro* translation. For the inhibition experiments with Szostak peptides presented here, two RRL translation conditions were used. These salt conditions are referred to as Kozak and Zhang conditions, and are described in details in the methods section of this appendix. Kozak and Zhang conditions show increased fidelity of HCV IRES-mediated translation, relative to standard Promega conditions (Figure IA.3A). Despite Zhang and Kozak conditions yielding 5- and 12-fold less firefly luciferase activity from WT HCV reporters, respectively, inhibition of IRES translation by either deletion of dom II or the addition of IIId oligo was much more efficient (Figure IA.3A).

For each translation extract condition, it was important to determine appropriate RNA concentrations and translation times to fall within the assay's linear range. RNA titrations and time courses suggested that up to 40 minute time points were acceptable for Zhang and Kozak conditions and that Kozak conditions only maintained linearity through ~10 ng RNA whereas Zhang conditions were linear through 200 ng RNA (Figure IA.3B,C). HeLa translation extracts required an order of magnitude still more reporter RNA; very little signal was observed with <500 ng reporter RNA and signal was linear through at least 1800 ng (Figure IA.3D).

IA.3.2 Representative in vitro translation inhibition results

In vitro translation reactions from WT HCV IRES or capped reporter-RNA controls were conducted in the presence of increasing concentrations of either the 6B48C cyclic peptide or IIId oligo as a positive control. Over the course of many repeated experiments, the cyclic peptide showed minimal inhibition of HCV IRES translation through ~5 μM (Figure IA.4). While strong inhibition of HCV IRES translation was observed at concentrations of peptide above 10 μM, equal inhibition was seen against a control, capped reporter (Figure IA.4A), indicating that this was non-specific inhibition of translation in general. In contrast, the IIId oligo positive control specifically inhibited the HCV IRES-containing reporter, actually causing a slight stimulation of cap-dependent translation at a concentration of 10 μM. Similar results were observed with pKB86, pKB84 and pKB101 reporter RNAs, indicating that the lack of HCV IRES-specific inhibition was not due to the presence of MS2 hairpins or the lack of dom I in the reporter RNAs (data not shown; see Methods for construct descriptions).

After initial reactions did not show inhibition of translation by the Szostak peptides, it was proposed that these denaturing conditions could have damaged the structure of the HCV IRES RNA. Attempts to anneal reporter RNAs prior to translation reactions in RRL showed no stimulatory effect (data not shown), suggesting that the IRES structure was properly folded. When additional salts are added to RRL, the WT IRES translates well, whereas a dom II deletion mutant shows almost no translation (Figure IA.3A), indicating that signal from RRL is not coming from promiscuous use of the start codon in the presence of denatured IRES RNA.
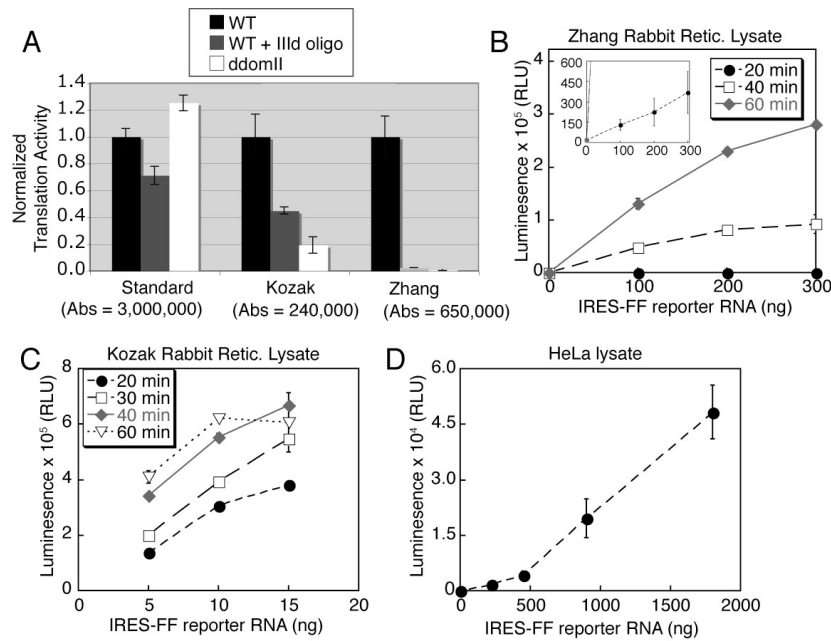
Figure IA.3: HCV IRES translation in different translation extracts. *(A)* Activity and fidelity of Promega RRL under Standard, Kozak or Zhang conditions (see methods). pKB2 or pKB3 template RNA (100 ng) was used in 60 min translation reactions, in the absence or presence of IIId oligo (5 μM). Time course and titration of pKB2 reporter RNA in *(B)* Zhang conditions or *(C)* Kozak conditions of Promega RRL. *(D)* Titration of pKB2 reporter RNA in HeLa extracts in 90 min reactions.
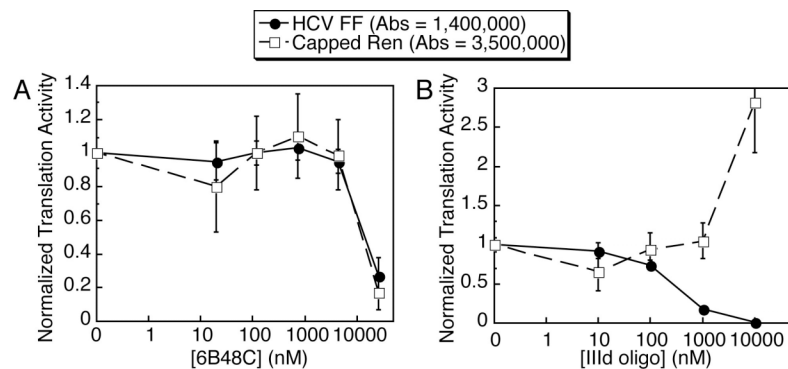
Figure IA.4: In vitro translations of HCV IRES- and capped-reporters in the presence of putative inhibitors. Translation of pKB2 (2 nM; HCV-FF luc) and pMJ3 (capped Ren) in Zhang conditions of Promega RRL, titrated with *(A)* 6B48C cyclic peptide (Litovchick preparation, diluted in water) or *(B)* IIId oligo (positive control).

107
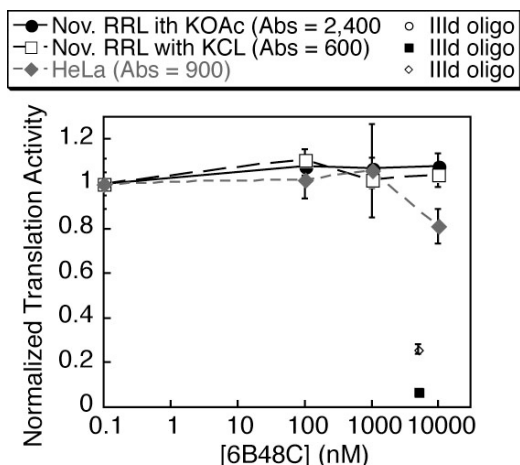
Figure IA.5: In vitro translation of Szostak IRES-Gluc reporter RNA in the presence of putative inhibitors. Translation of IRES-Gluc RNA (330 nM) in HeLa extracts or IRES-Gluc RNA (43 nM) in Novagen RRL, supplemented with KOAc or KCl. Reporter RNAs were prepared according to the Litovchick protocol (always in buffer with HPES/KCl/TCEP and only frozen once). In vitro translation reactions were either titrated with 6B48C cyclic peptide (Guillen preparation, diluted in buffer with HEPES/TCEP) or conducted in the presence of IIId oligo (5 μM; positive control).

To more closely repeat the translation conditions from the Szostak publication (Litovchick & Szostak 2008), HeLa translation extracts and Novagen RRL were used for inhibition experiments with the Litovchick IRES-Gluc reporter RNA. While Litovchick transcribed his reporter RNA from this propagated PCR product (personal communication), sequencing of individual clones from the PCR product showed that many point mutations had been introduced after many rounds of PCR with *Taq* polymerase, and about 1/3 of the clones (5 out of 13) had insertions or deletions that put luciferase out of frame with the HCV AUG. Very minimal inhibition of HCV IRES translation was observed in any of these extracts through 10 μM 6B48C peptide, while substantial inhibition was seen with 5 μM IIId oligo (Figure IA.5). This experiment was repeated with independent batches of peptide (Genscript-synthesized 6B4 and Guillen-synthesized 6B48C, in addition to the original Litovchick-synthesized 6B48C), but little to no inhibition by the peptides was observed (data not shown).

IA.3.3 RNase footprinting of HCV IRES with peptide

In the course of attempting to repeat the translation inhibition of the cyclic peptide, we also wanted to determine the domain(s) of the HCV IRES to which it bound. As a first step to determine the region of RNA that mediates this interaction, RNase footprinting was attempted (Clarke 1999). A combination of single stranded (ss) and double stranded (ds) ribonucleases was used to cleave HCV IRES RNA. The RNase cleavage pattern was compared in the presence of increasing concentrations of the Szostak peptides, to observe specific changes in cleavage pattern when the peptide binds. Fairly good separation and coverage was obtained for nts 87-211 (Figure IA.6A), nts 49-117 (Figure IA.6B) and nts 180-278 (data not shown) of the HCV IRES by running gels of cleaved 5′ end labeled RNAs for varying lengths of times. No changes in the cleavage patterns were observed in the presence of up to 25 μM peptide, despite a published binding affinity of the peptide for IRES RNA of 4 nM (Litovchick & Szostak 2008). This lack of

a peptide footprint could result from either a lack of binding between the IRES and peptide, or an inability of RNase footprinting to detect the binding. This could occur if the peptide binds in a region of the RNA where there are no strong cleavages by any of the RNases, or if it binds near the very 3′ end of the IRES RNA where it is difficult to resolve fragments on sequencing gels.
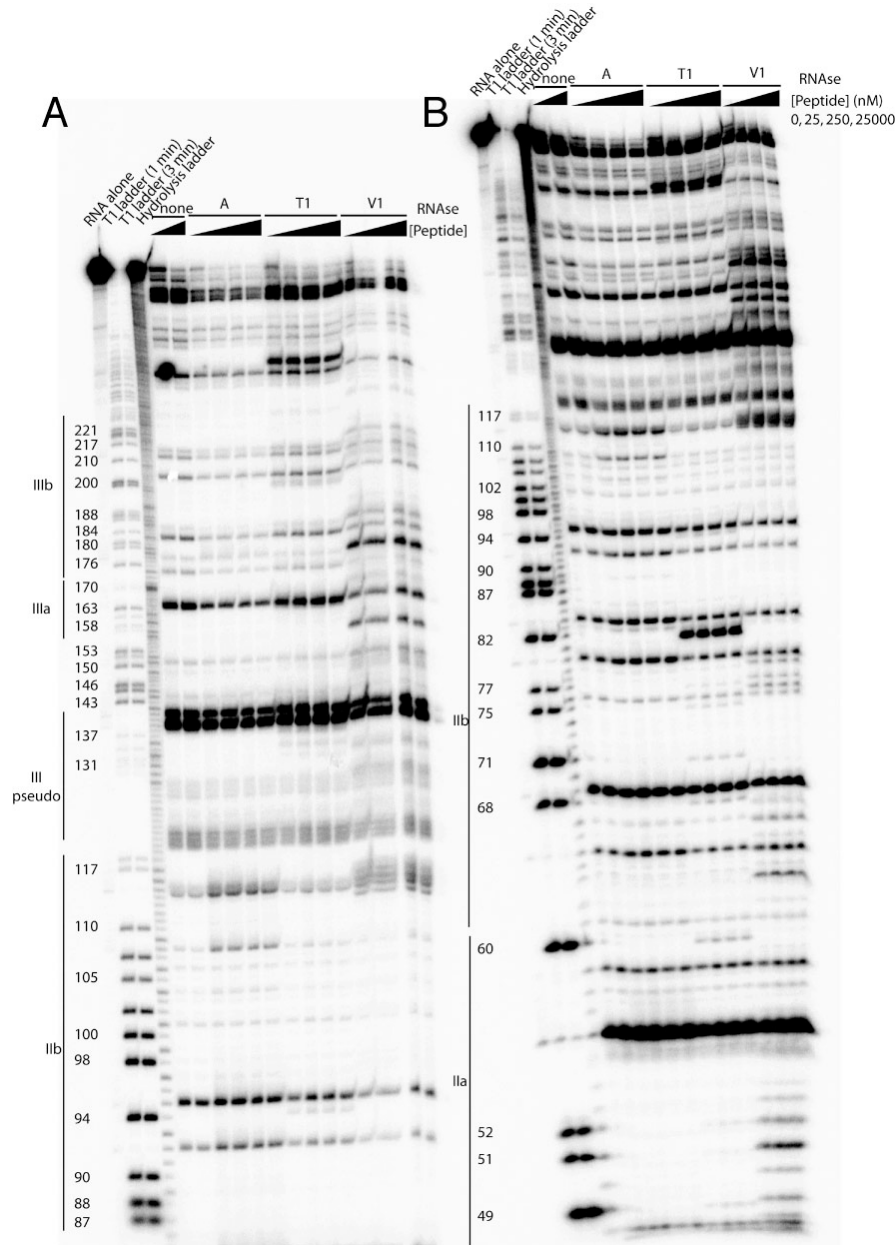
Figure IA.6: RNase footprinting with Szostak peptides and HCV IRES RNA. RNase cleavage reactions of HCV IRES RNA, resolved on a 10% sequencing gel for *(A)* 6 hours and *(B)* 4 hours. 5′ end-labeled HCV IRES RNA was incubated with increasing amounts of 6B48C peptide (Litovchick preparation; 0, 25 nM, 250 nM, 25 μM), and then cleaved with limiting concentrations of RNases A, T1, or V1. Cleaved RNAs were run alongside a hydrolysis ladder and T1 digestion ladder.

IA.3.4 Filter binding between peptide and IRES

In order to determine whether there was binding of the cyclic peptide to HCV IRES RNA that was not detectable by RNase footprinting, we used filter-binding experiments to look for a specific interaction between the peptide and HCV IRES RNA. Initially, it was unclear whether we would be able to detect binding of these very small peptides using a filter-binding assay. (The 6B48C peptide is an 8mer with a mass of 1025 Da and the 6B4 peptide is a 27mer of mass 3026 Da.) The critical parameter seemed to be the pore size of nitrocellulose membranes, as the product literature for Protran nitrocellulose membranes (Whatman) suggests that 0.1 µm should retain peptides or proteins <8 kDa, 0.2 µm for 8-20 kDa and 0.45 µm for >20 kDa. As expected, increased interaction of the 6B48C peptide was observed when using a nitrocellulose membrane with a 0.1-micron pore size, rather than 0.2- or 0.45-micron pores (Figure IA.7A). Thus, the binding of 6B48C or 6B4 was examined to 5′ end-labeled IRES or a nonspecific tRNA control using 0.1-micron nitrocellulose membrane. No binding was observed until > 250 nM peptide was present. There does appear to be slightly increased binding to HCV IRES RNA compared to tRNA at 2.5 µM, but no difference in binding to the negative control is seen at 25 µM peptide (Figure IA.7B). These data suggest that it is highly unlikely that any observed binding of HCV IRES RNA to the peptides represents specific, biologically relevant, interaction.

After this result was obtained, Yollete Guillen (Szostak lab) repeated equilibrium ultrafiltration experiments with ^{35}S-labeled peptide, which was the original assay used to show a high-affinity, specific interaction between peptide and HCV IRES (Litovchick & Szostak 2008). In agreement with the above results, she did not observe any binding (personal communication).
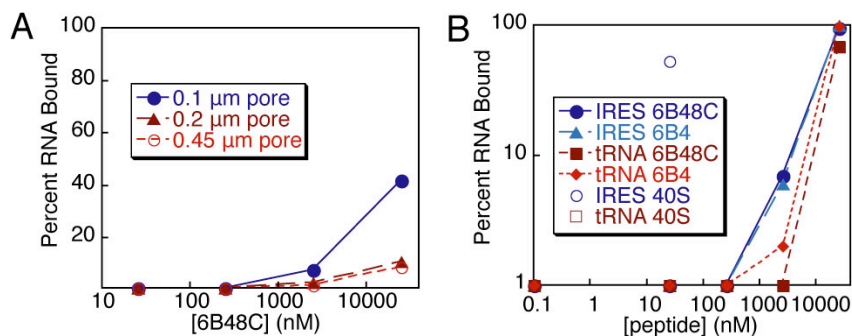
Figure IA.7: Filter binding of HCV IRES RNA with Szostak peptides. *(A)* Filter-binding assay of HCV IRES RNA with cyclic 6B48C peptide (Guillin preparation) to determine the effect of nitrocellulose membrane-pore size on retention of peptide-RNA complexes. Percent RNA bound to peptide is plotted, as determined by the ratio of counts on the nitrocellulose membrane to the total counts on the nitrocellulose and nylon membranes. No significant counts were seen on the Tuffryn membrane (0.2 or 0.45 µM). *(B)* Filter binding of HCV IRES RNA or human initiator tRNA (negative control) with either cyclic 6B48C peptide (Guillen preparation) or linear 6B4 peptide (GenScript preparation), using a 0.1 micron nitrocellulose membrane. Only at the highest peptide concentration were significant counts seen on the Tuffryn membrane, in which case ~10% of the RNA was aggregated in all four titrations. Also shown is a positive control of 40S ribosomal subunits (100 nM), which binds HCV IRES RNA, but not tRNA.

IA.4 Conclusions

      The reproduction of translation inhibition and IRES binding assays with the Szostak peptides were meant to set the stage for mechanistic studies of the inhibition by the peptides. Unfortunately, the compounds' inhibition of and binding to the HCV IRES could not be reproduced, despite extensive efforts documented here and also by Yollete Guillen in the Szostak lab. It is not at all clear what lead to the strikingly different results originally published. Due to the inability of the results to be reproduced, the original publication has since been retracted (Litovchick & Szostak 2009).

Appendix IB. Lack of specific HCV IRES inhibition by ISIS heteropolycycles

IB.1 Introduction

The conformation of dom IIa is thought to be critical for IRES activity. A mass spectrometry-based screen was conducted by ISIS pharmaceuticals for small molecules that bind to this domain of the HCV IRES (Seth et al. 2005). This approach identified a class of benzimidazoles that showed good inhibitory activity against an HCV IRES replicon assay without showing any significant cellular toxicity using one particular assay (an MTT assay). However, no published data have directly validated that these compounds specifically inhibit translation from the HCV IRES as opposed to translation more generally or another process to which the replicon assay could be highly sensitive. Despite the lack of validation that the activity of this class of molecules stems from a direct and specific effect on HCV IRES-driven translation, biophysical characterization using FRET and NMR has shown that these benzimidazoles do indeed bind to a particular bulge in dom II RNA and that their binding disrupts the overall conformation of the IRES RNA (Parsons et al. 2009; Paulsen et al. 2010).
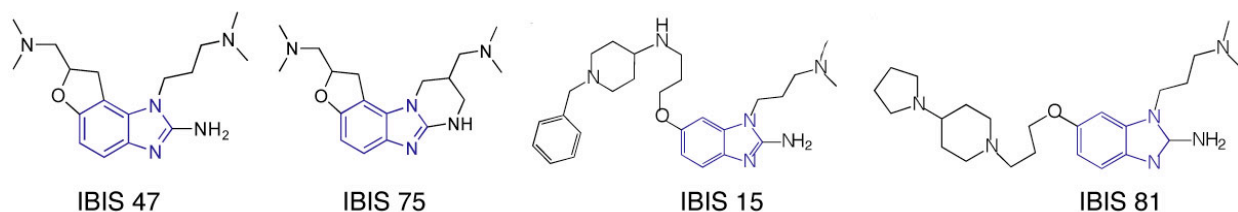
Figure IB.1: Structures of benzimidazole compounds (putative HCV IRES inhibitors) given as a gift by ISIS pharmaceuticals. Compounds were resuspended in water at 10-20 mM and stored at -20°C. ESI FT ICR mass spectrometry confirmed that all compounds were >95% pure after prolonged storage in frozen aqueous solution.

Four benzimidazole compounds were given to us by ISIS pharmaceuticals to examine their mechanism of action against HCV IRES (Figure IB.1). Based on the pKa's of these compounds' ionizable groups, each compound would have multiple positive charges in physiological solutions. Before undertaking mechanistic studies, we wished to examine the activity of these compounds directly on translation using *in vitro* translation assays with HCV IRES or control luciferase RNAs. Based on the proposal that these benzimidazole compounds bind to dom II of the HCV IRES, it was expected that these compounds would inhibit the translation of WT IRES preferentially over ΔdomII constructs and over translation from a reporter RNA whose translation was not IRES-dependent.

IB.2 Methods

IB.2.1 RNA constructs and preparation

RNA reporters used were WT (pKB2) and ΔdomII (pKB83), and uncapped luciferase RNA (Promega; Figure IA.2). Note that IRES reporter constructs contained three MS2 hairpins at their 5′ ends. pKB2 and pKB83 were *in vitro* transcribed as described in section 4.2.2, but RNAs were worked up using three iterative ammonium acetate ethanol precipitations to remove

free nucleotides, rather than with spin columns. Uncapped firefly luciferase RNA, supplied with RRL from Promega, was used as a control for nonspecific inhibition.

IB.2.2 *In vitro* translation reactions

In vitro translation reactions (15 μL) were set up with nuclease-treated RRL (10.5 μL; Promega), amino acids (1 mM), RNasin Plus RNase inhibitor (0.3 μL; Promega), protease inhibitors (2 mM PMSF and Roche protease inhibitor cocktail, with 1 tablet/1 ml H_2O = 50x), and firefly luciferase reporter RNA (300 ng), and were incubated at 30°C for 1.5 hours, then quenched on ice, and luciferase activity was measured in a 20/20[n] luminometer (Turner Biosystems) after addition of LARII reagent (Promega). Four benzimidazole compounds, which are putative HCV IRES inhibitors, were given to us as a gift by ISIS pharmaceuticals. These compounds were resuspended in water at 10-20 mM and stored at -20°C. ESI FT ICR mass spectrometry confirmed that all compounds were >95% pure after prolonged storage in frozen aqueous solution. Serial dilutions of benzimidazole compounds were prepared in water and compounds were added to the extract prior to initializing reactions with the addition of reporter RNA. Luminescence values were normalized to the activity level in the absence of any compounds and IC_{50} values were calculated using the equation Relative Translation Activity = 1 / $(1 + 10^{\log(IC50) - \log[inhibitor]})$ in Kaleidagraph.

IB.3 Preliminary Results and Discussion

Titrations of all four benzimidazole compounds (Figure IB.1) were conducted against *in vitro* translation reactions from WT or ΔdomII HCV IRES-FF luciferase or from a control, uncapped FF-luc RNA. Representative titrations for IBIS 47 and IBIS 75 (Figure IB.2) show that, while WT HCV IRES-FF luc activity is indeed inhibited, the translation from a ΔdomII IRES reporter or from uncapped luciferase RNA is inhibited to a similar extent. IC_{50} values from all titrations are shown in Table IB.1, and almost no specificity for WT IRES inhibition is seen for any of the benzimidazole inhibitors in RRL.

These compounds were also tested in other translation extract systems for inhibition of HCV IRES translation and general translation. All four compounds significantly inhibit the translation of a control luciferase reporter in HeLa extracts at 10 μM (data not shown). In addition, the translation in RRL of both EMCV- and CrPV-IRES reporters is inhibited by these compounds at 10 μM (data not shown). The CrPV IRES inhibition is especially striking since this IRES requires no translation initiation factors beyond the ribosome itself for initiation. Interestingly, the compounds do not inhibit translation of luciferase reporter in *S. cerevisiae* translation extracts at 10 μM (data not shown), but since the HCV IRES is not believed to be active in yeast extracts, it could not be confirmed that the compounds still inhibit HCV IRES-driven translation. An explanation of this observation is that the factor/ribosomal binding site with which the benzimidazoles interact is not conserved between yeast and mammalian translation machinery.
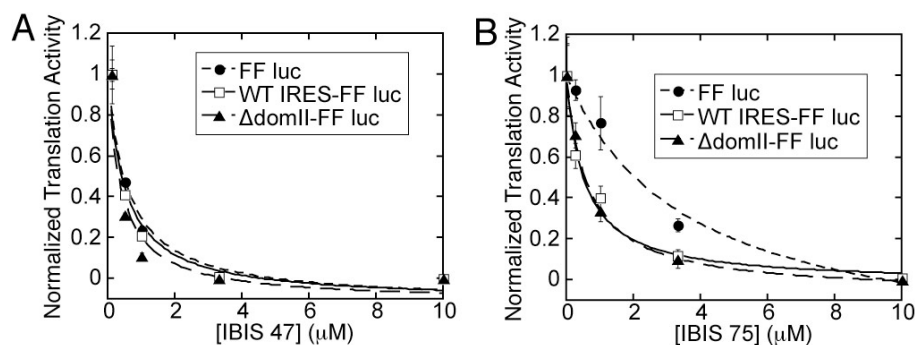
Figure IB.2: **Representative inhibition data with benzimidazole compounds**. Translation reactions were conducted in standard Promega RRL for 90 minutes using of WT IRES-FF luc, ΔdomII IRES-FF luc, or FF luc RNA (300 ng), and varying concentrations of *(A)* IBIS 47 and *(B)* IBIS 75. Luminescence values were normalized to the activity level in the absence of any compounds and IC_{50} values were calculated using the equation Rel. Trans. Act. = $1 / (1 + 10^{log(IC50) - log[inhibitor]}$ to yield calculated IC_{50} values, which are displayed in Table IB.1.

Table IB.1: **Summary of inhibition of benzimidzole compounds**. IC_{50} values for four IBIS compounds against the translation of three reporter RNAs in standard RRL.

	IC_{50} (nM)*			
	IBIS 15	**IBIS 47**	**IBIS 81**	**IBIS 75**
FF luc	3.3 ± 0.6	0.6 ± 0.2	3.6 ± 1.2	3.6 ± 1.1
WT HCV IRES-FF luc	1.0 ± 0.3	0.5 ± 0.2	1.0 ± 0.3	0.5 ± 0.1
ΔdomII HCV IRES-FF luc	1.6 ± 0.8	0.4 ± 0.2	14 ± 30$^\perp$	0.64 ± 0.03

*   Calculated by fitting to: Rel. Trans. Act. = $1 / (1 + 10^{\wedge log(IC50) - log[inhibitor]}$.
      Representative data are shown in *Figure IB.2*.

$\perp$ This value is derived from particularly noisy data and can only be interpreted as showing that significant inhibition indeed occurs at IBIS81 concentrations under 50 nM.

IB.4 Conclusions and Caveats

These data suggest that, in addition to binding to dom II of the HCV IRES, the benzimidazoles may have a second binding site on mammalian ribosomes which leads to general translation inhibition. While the compounds were reported to have good activity against an HCV IRES replicon assay without the cellular toxicity one would expect for a general translation inhibitor (Seth et al. 2005), the *in vitro* translation data presented here suggest that more effort should be invested in establishing the *in vivo* mechanism of action of these inhibitors.

As shown and discussed in Chapters 3 and 4, Promega RRL is a highly promiscuous system, which may not represent regulated translation initiation mechanisms. This promiscuity of standard RRL is underscored by the fact that uncapped control FF-luc RNA (from Promega) and ΔdomII HCV IRES-FF luc RNA were efficiently translated in these reactions. The data in this appendix were collected before this promiscuity was recognized and have not been repeated

with salt-adjusted RRL, in which the salt concentrations have been adjusted for authentic and cap- and HCV IRES-dependent translation. It would be worth repeating these inhibition assays in a more faithful *in vitro* translation system to confirm the lack of specificity of translation inhibition.

Appendix II

Towards the structure of the HCV IRES pseudoknot

II.1 Introduction

As described in Chapter 4, the pseudoknot domain of the HCV IRES plays a critical role in allowing the IRES to position its ORF in the mRNA binding cleft, once bound to the 40S ribosomal subunit. Although the structures of most domains of the HCV IRES have been solved by X-ray crystallography and NMR spectroscopy, the pseudoknot domain has eluded high-resolution structure determination to date (Figure 1.6). Such a structure would provide insight into how the HCV ORF is oriented towards the mRNA binding cleft, and could open doors to search for small molecules that could inhibit necessary conformations of this RNA structure, which is critical for the virus.

II.1.1 Approach

In RNA crystallography, the limiting step is frequently obtaining well-ordered crystals that diffract to (near-)atomic resolution; even for a well-behaved molecule, the surface of an RNA molecule presents fewer unique chemical features than a typical protein, and the electrostatics of packing many negatively charged phosphates in close proximity can be unfavorable. Obtaining a chemically and conformationally homogenous RNA sample is a necessary first step for crystallization. After this, many potential variables can be adjusted in the search for well-ordered crystals, such as RNA sequence, concentration, crystallization conditions, temperature, and cryoprotectant. In protein crystallography, it is typical to screen more than 12x96 crystallization conditions per protein target to obtain initial crystal hits. We chose to adopt a more focused screening approach, using only one 96-well crystallization screen per RNA construct, aiming to screen many distinct constructs rather than a large number of crystallization conditions for a smaller number of constructs. This approach is based on the hypothesis that the sequence of an RNA will have a larger effect on its crystallizability than crystallization conditions per se, and that an RNA with a propensity to crystallize will do so within the first couple hundred of conditions, especially if those conditions are tailored for nucleic acids (Ferre-D'Amare & Doudna 2001).

General sequence considerations for RNA crystallization constructs have been reviewed (Holbrook et al. 2001; Golden 2007). Because the pseudoknot domain of the IRES is known to be conformationally flexible (Wang et al. 1995; Kolupaeva et al. 2000; Fletcher et al. 2002), we wanted to engineer RNA sequences along the periphery of the pseudoknot domain that would favorably assist in crystallization. We utilized GNRA tetraloop/tetraloop receptor interactions (Ferre-D'Amare et al. 1998; Ferre-D'Amare & Doudna 2001), as well as U1A hairpins for cocrystallization with the RNA binding domain (RBD) of the U1A splicesomal protein (Ferre-D'Amare & Doudna 2000; Rupert & Ferre-D'Amare 2004). Another very useful strategy to systematize the phasing of RNA structures is to include GU wobble base pair-containing motifs which have a high propensity to bind to compounds such as cobalt or iridium hexamine (Keel et al. 2007).

II.1.2 Models for HCV IRES pseudoknot structure

Although no structure has been solved for the HCV IRES pseudoknot, this domain has attracted attention from computational modelers in their efforts to predict RNA structures based on energy minimization and prior biochemical observations. Eric Westhof (University of Strasbourg) shared his unpublished computational prediction for the HCV IRES pseudoknot structure (Figure II.1B). This model predicts that SI and SII run mostly parallel to one another, perhaps interacting with one another's minor grooves. The overall topology of the model predicts

that the IIIe loop runs alongside dom III, consistent with a recently identified Watson-Crick interaction between a IIIe loop pyrimidine and a conserved purine/purine mismatch in the stem of dom III (Easton et al. 2009). The Westhof model is consistent with several interesting mutational observations (Chapter 4). First of all, L1 and L3, which can be deleted without detrimental effects on IRES activity, are predicted to be solvent exposed, whereas L2, which cannot be deleted or lengthened, is predicted to be stacking in between SII and SI/J. In addition, a GG dinucleotide sequence in SI and a UG dinucleotide sequence in SII, which have been implicated biochemically in tertiary contacts, are in rough positions where they could participate in tertiary interactions with neighboring stems, perhaps even with each other (Figure II.1B, red and purple).

The precise tertiary interactions between the two stems of the pseudoknot could conceivably be sufficient for the IRES to orient its ORF correctly towards the mRNA binding cleft. Since SII connects to dom IV and SI connects to dom II, which has a very specific geometry of interaction with the 40S subunit's E site, the precise orientation of SI and SII, potentially locked down by tertiary interactions just as the stems end, may be enough to cast the open reading frame in the correct direction relative to dom II and the rest of dom III.

The HCV IRES pseudoknot domain has also been computationally modeled by the lab of Kevin Weeks (Lavender et al. 2010). Their modeling algorithm utilized pseudoatoms, such that three mock atoms represent each nucleotide (one atom for the nucleobase, the sugar, and the phosphate). The positions of these pseudoatoms were then optimized using molecular dynamics simulations with constraints generated from previous biochemical data; base pairs were not physically modeled, but rather treated as constraints. The overall topology predicted by this model is similar to the Westhof model, in terms of the orientation of stems at the four-way junction. However, this model predicts less direct interaction between SI and SII of the pseudoknot, and the SII geometry seems to be distorted from a perfect A-form helix (Figure II.1C,D). Because of the nature of the pseudoatoms, it is more difficult to determine the predicted roles of the three predicted single-uridine loops, but the nucleobases all seem to be predicted to fall within the axis of the helix, rather than flipping outside of it. The GG and UG dinucleotide sequences that we have biochemically implicated in tertiary interactions are still generally predicted to be generally close to one another in the Weeks model, but do not seem to be quite as appropriately placed to mediate such interactions as they are in the Westhof model.

One of the more intriguing observations from the Weeks model is that dom IV is predicted to point perpendicularly from the rest of the IRES, as defined by the dom II / dom III axis (Fig II.1C, SLIV coming out of the plane of the page). The authors suggest that this geometry is reminiscent of a tRNA structure, and point to the 7 nt loop at the base of SL IV as another parallel to the structure of a tRNA. Given the proximity of the E site to where to the pseudoknot binds, the authors suggest that this structural similarity to a tRNA may allow the IRES to use a form of tRNA mimicry to bind the 40S subunit. It is important to remember, however, that SLIV is thought to ultimately unfold in order for the AUG to reach the P site. Thus, any such tRNA mimicry would presumably function during an initial "encounter" complex with the 40S subunit, with the SLIV helix eventually melting before translation initiation can be complete.
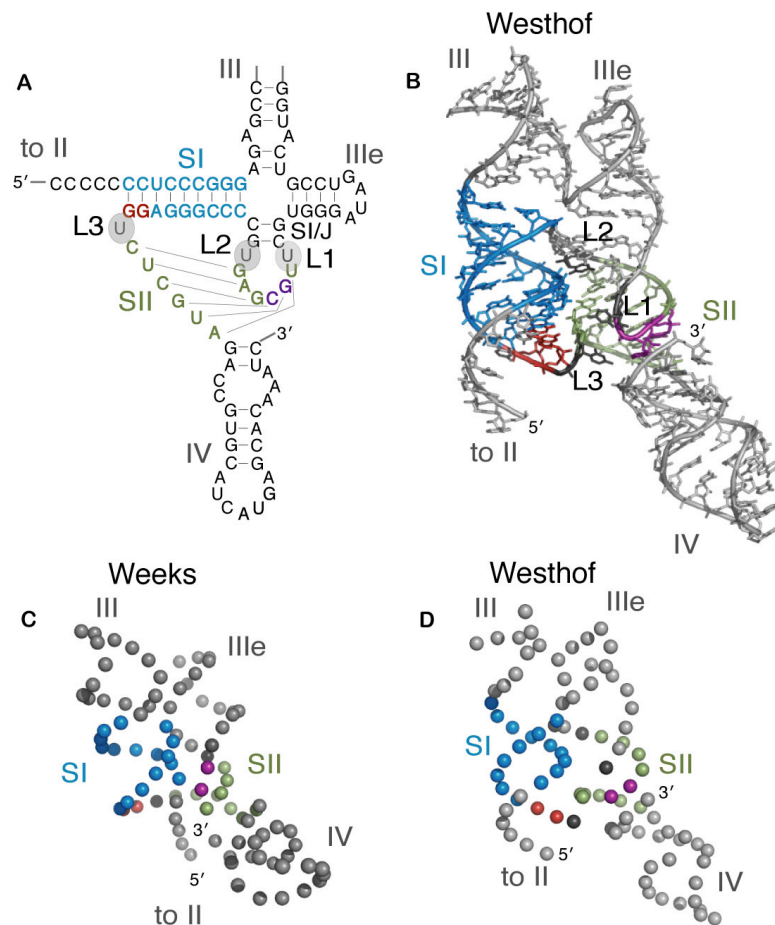
Figure II.1: Computational models of HCV IRES pseudoknot structure. *(A)* Secondary structure of the pseudoknot domain, with SI nts shown in blue, SII nts in green, a SI GG dinucleotide in red and a SII UG dinucleotide in purple. *(B)* Computational model of this domain shared as a personal communication by Eric Westhof of the University of Strasbourg. *(C)* Computational model of the HCV pseudoknot derived from pseudoatoms and known biochemical results (Lavender et al. 2010). Only the sugar pseudoatoms are shown as spheres. *(D)* Westhof computational model with only phosphates shown as spheres, for comparison to Weeks model. In all structures, SI and SII are colored as in *A* and the positions of loops 1, 2 and 3 are labeled. A GG dinucleotide sequence at the terminus of SI, which we have biochemically implicated as participating in tertiary contacts, is colored in red. Structures were rendered using Pymol.

II.2 Methods

II.2.1 Design and cloning of crystallization constructs

In order to reduce any 3′ heterogeneity introduced by T7 RNA polymerase, and to avoid the necessity for guanosines at the 5′ end of the RNA product, crystallization constructs were cloned in between two self-cleaving ribozymes: a hammerhead (HH) ribozyme on the 5′ end and a hepatitis delta virus (HDV) ribozyme on the 3′ end (Figure II.2) (Ke & Doudna 2004). The

sequences of all constructs tested to date are shown in Table II.1; Table II.2 shows the sequences of additional U1A constructs that have not yet been tested.

All crystallography constructs were cloned into a pUC19 vector backbone, between the EcoRI and BamHI restriction sites. For cloning, the desired sequence was broken up into overlapping oligonucleotide (oligo) pairs (Figure II.2) and these DNA oligos were synthesized by IDT. Oligos were phosphorylated in reactions (10 μL) containing the oligo (1 μM), T4 polynucleotide kinase (PNK; 1 μL; NEB), 10x T4 PNK buffer (1 μL; NEB), and ATP (10 mM). Kinase reactions were incubated at 37°C for 1 hour. Oligo pairs were annealed at 150 nM in T4 DNA ligase buffer (1x; Invitrogen) by heating at 70°C for 1 min, then slow cooling to room temperature. Ligation reactions (30 μL) contained pUC19 plasmid DNA (~180 ng; 1.6 nM), digested with EcoRI and BamHI, oligo pairs (10 nM), T4 ligase (1 μL; Invitrogen), T4 DNA ligase buffer (1x; Invitrogen), DTT (10 mM), and ATP (1.3 mM). Ligation reactions were allowed to incubate at 16°C overnight, and ligation product (3 μL) was used to transform mach1 competent cells (20 μL).

Figure II.2: Cloning strategy for crystallography constructs. The construct design for pSW25 is shown, as well as the breakdown of the sequence and reverse complement into individual oligonucleotides. Color key: green = first 11 nts of construct and the 11 nts changed within HH to hybridize, grey highlight = regions of overlap between oligo pairs, red = nucleotides in restriction sites.

Table II.1: Sequences of pseudoknot crystallography constructs tested to date. For schematic representations of constructs, refer to Figures II.3, II.6 and II.9.

	Length (nt)	Sequence
pSW19	110	CCTCCCGGGAGAGCCATAGTGGGAAACTGCTAGCCGAGTAGTGTTGGGTCGCGAAAGGCC TTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pSW20	140	CCTCCCGGGAGGCCATgctaagTAGGAAACTAtatgcAGCCGAGTAGTGTTGGGTCGCGA AAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACC GTGCATCATGAGCACGAATC
pSW21	116	CCTCCCGGGAGGCCATgctaagTAGGAAACTAtatgcAGCCGAGTAGTGTTGGGTCGCGA AAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pSW22	145	CCTCCCGGGAGGCCATAGTGGGAAACTGCTAGCCGAGTAGTGTTGGGTCGCGAAAGGCCT TGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGgctaag TGCATCATGAGGCAtatgcCGAATC
pSW23	133	CCTCCCGGGAGGCCATAGTGGGAAACTGCTAGCCGAGTAGTGTTGGGTCGCGAAAGGCCT TGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATC ATGAGCACGAATC
pSW24	63	CCTCCCGGGAGAGCCGAAAGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTC TCGTA
pSW25	82	CCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTA
pSW26	88	CCTCCCGGGAGAGCGAAAGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCT CGTAGACCGTGCATCATGAGCACGAATC
pSW27	101	CCTCCCGGGAGAGCCGAAAGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTC TCGTAGACCGgctaagTGCATCATGAGGCAtatgcCGAATC
pSW28	106	CCTCCCGGGAGAGCCgctaagTAGGAAACTAuaugcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pSW29	120	CCTCCCGGGAGAGCCATggctgccattgcactccggcagccAGCCGAGTAGTGTTGGGTC GCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pSW30	144	CCTCCCGGGAGAGCCATggctgccattgcactccggcagccAGCCGAGTAGTGTTGGGTC GCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA GACCGTGCATCATGAGCACGAATC
pSW31	85	CCTCCCGGGAGAGCCggctgccattgcactccggcagccGGTACTGCCTGATAGGGTGCT TGCGAGTGCCCCGGGAGGTCTCGTA
pSW32	109	CCTCCCGGGAGAGCCggctgccattgcactccggcagccGGTACTGCCTGATAGGGTGCT TGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pSW40	80	CCTCCCGGGAGAGCCgctaagTGGAAACAtatgcGGTACTGCCTGATAGGGTGCTTGCGA GTGCCCCGGGAGGTCTCGTA
pSW47	84	CCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAAA
pSW68	82	CCTCCCGGGAGAGCCgctaagCTAGGAAACTAGtatgcGGTACTGCCTGATAGGGTGCTT GCGAGTGCCCCGGGAGGTCTCGTA
pSW69	83	CCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAA
pSW70	81	CCCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTG CGAGTGCCCCGGGAGGTCTCGTA
pKB241	86	CCTCCCGGGAGAGCCgctaagTGCTAGGAAACTAGCAtatgcGGTACTGCCTGATAGGGT GCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB242	84	CCTCCCGGGAGAGCCgctaagGCTAGGAAACTAGCtatgcGGTACTGCCTGATAGGGTGC TTGCGAGTGCCCCGGGAGGTCTCGTA
pKB243	83	CCTCCCGGGAGAGCCgctaagCTAGGAAACTAGtatgcGGTACTGCCTGATAGGGTGCTT GCGAGTGCCCCGGGAGGTCTCGTAG
pKB244	84	CCTCCCGGGAGAGCCgctaagCTAGGAAACTAGtatgcGGTACTGCCTGATAGGGTGCTT GCGAGTGCCCCGGGAGGTCTCGTAGA
pKB245	85	CCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAGAC
pKB246	86	CCTCCCGGGAGAGCCgctaagTAGGAAACTAtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAGACC
pKB247	82	CCTCCCGGGAGAGCCgctaagGGGGAAACTCtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTA
pKB248	83	CCTCCCGGGAGAGCCgctaagGGGGAAACTCtatgcGGTACTGCCTGATAGGGTGCTTGC GAGTGCCCCGGGAGGTCTCGTAG

pKB249	82	CCTCCCGGGAGAGCCgctaagGGCGAAAGTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB250	83	CCTCCCGGGAGAGCCgctaagGGCGAAAGTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAG
pKB251	84	CCTCCCGGGAGAGCCgctaagGGGGGAAACTTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB252	85	CCTCCCGGGAGAGCCgctaagGGGGGAAACTTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAG
pKB253	83	CCTCCCGGGAGAGCCgctaagGGGGAAAGCCTtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB254	84	CCTCCCGGGAGAGCCgctaagGGGGAAAGCCTtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAG
pKB267	84	CCTCCCGGGAGAGCCgctaagGGGGAAACTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGA
pKB268	84	CCTCCCGGGAGAGCCgctaagGGCGAAAGTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGA
pKB269	86	CCTCCCGGGAGAGCCgctaagGGGGGAAACTTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGA
pKB270	85	CCTCCCGGGAGAGCCgctaagGGGGAAAGCCTtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGA
pKB271	85	CCTCCCGGGAGAGCCgctaagGGGGAAACTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGAC
pKB272	85	CCTCCCGGGAGAGCCgctaagGGCGAAAGTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGAC
pKB273	87	CCTCCCGGGAGAGCCgctaagGGGGGAAACTTCtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGAC
pKB274	86	CCTCCCGGGAGAGCCgctaagGGGGAAAGCCTtatgcGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGAC

Table II.2: Additional cloned U1A hairpin-containing crystallography constructs. These constructs were designed to move the U1A hairpin into different positions around the pseudoknot domain and to introduce varied linker lengths between the hairpin and the HCV IRES sequence. The four constructs tested to date (pSW29-32) placed the U1A hairpin where the IIIabc(d) domain would be, with a 5 bp linker, which may not have been a favorable arrangement for crystallization. The constructs below have been cloned, but not tested.

	Linker length	U1A pos	Length	Sequence
pSW29–32	5 bp	III	85–144	
pKB255	replace loop	IIId	118	CCTCCCGGGAGAGCCATAGTGGGAAACTGCTAGCCGAGTAGTGccattgcact ccggCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCG GGAGGTCTCGTA
pKB256	replace loop	IIId	142	CCTCCCGGGAGAGCCATAGTGGGAAACTGCTAGCCGAGTAGTGccattgcact ccggCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCG GGAGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pKB257	3 bp	III	116	CCTCCCGGGAGAGCCATggcccattgcactccgggccAGCCGAGTAGTGTTGG GTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGG AGGTCTCGTA
pKB258	3 bp	III	140	CCTCCCGGGAGAGCCATggcccattgcactccgggccAGCCGAGTAGTGTTGG GTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGG AGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pKB259	3 bp	III	81	CCTCCCGGGAGAGCCggcccattgcactccgggccGGTACTGCCTGATAGGGT GCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB260	3 bp	III	105	CCTCCCGGGAGAGCCggcccattgcactccgggccGGTACTGCCTGATAGGGT GCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pKB262	7 bp	III	148	CCTCCCGGGAGAGCCATggctgcgccattgcactccggcgcagccAGCCGAGT AGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGT GCCCCGGGAGGTCTCGTAGACCGTGCATCATGAGCACGAATC
pKB263	7 bp	III	89	CCTCCCGGGAGAGCCggctgcgccattgcactccggcgcagccGGTACTGCCT GATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTA
pKB264	7 bp	III	91	CCTCCCGGGAGAGCCggctgcgccattgcactccggcgcagccGGTACTGCCT GATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGA
pKB265	replace loop	IV	141	CCTCCCGGGAGAGCCATAGTGGGAAACTGCTAGCCGAGTAGTGTTGGGTCGCG AAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCT CGTAGACCGTGCccattgcactccggGCACGAATC
pKB266	replace loop	IV	73	CCTCCCGGGAGAGCCGAAAGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCG GGAGGTCTCGTAGACCGTGC

II.2.2 DNA plasmid preparation and digestion

DNA was initially prepped using HiSpeed Plasmid Maxi Kits (Qiagen). The manufacturer's protocol was followed, with modifications in the final steps: DNA was eluted from Maxi-concentrators with warm EB buffer (600 μL). This buffer was passed through the concentrator a second time, and was followed with additional warm EB buffer (200 μL). Plasmid DNA was concentrated by EtOH precipitation before restriction digestion. Two maxi preps typically yielded ~300-700 μg plasmid / 750 flask of LB. For constructs that proved useful, larger plasmid preps were carried out using Qiafilter Plasmid Mega Kits (Qiagen), typically yielding 1-1.5 mg plasmid / 2 750 mL flasks of LB. The manufacturer's instructions were followed, except that, at the end of the protocol, iPrOH-precipitation pellets were resuspended and transferred to Eppendorf tubes from Oakridge tubes as they were washed with cold EtOH (70%), facilitating the final resuspension of the plasmid pellet in a small volume.

Overnight digestion of plasmid DNA was carried out with BamHI or HindIII (NEB), according to manufacturer's instructions at 37°C. BamHI was initially used, yielding a HDV ribozyme product of 73 nts. For RNA constructs of approximately this length, it was useful to cleave pUC19 DNA at a downstream HindIII restriction site, in order to lengthen the HDV ribozyme product to 103 nts so that it would be easily separable from the product RNA. The HH ribozyme product, for comparison, is 58 nts.

A typical large-scale digestion reaction (~400 μl) might contain DNA (500 μg) with restriction enzyme (~800 units; NEB). After complete digestion of the plasmid was confirmed using a ethidium bromide-stained agarose gel (1% agarose, 1x TAE), plasmid digestions were worked up by extraction with an equal volume of phenol/chloroform/ isoamyl alcohol (25:24:1; Sigma). The aqueous layer was subsequently ethanol precipitated with the addition of NaOAc, pH 5.2 (3M; 0.1x vol) and EtOH (100%; 2.5x vol). Precipitations were incubated at -80°C for >20 min, and spun at high speed in a microcentrifuge at 4°C for 20 min. Pellets were washed with cold EtOH (70%), dried, and resuspended in DEPC water.

II.2.3 RNA transcription, purification, and folding

Transcription reactions (5 mL) contained digested template plasmid (500 μg), RNasin Plus RNase inhibitor (~10 μL; Promega), NTPs (x4, 5 mM of each), T7 polymerase (0.1 mg/mL; purified by Kaihong Zhou), pyrophosphatase (5 μg; Roche), DTT (10 mM) in 30 mM Tris HCl, pH 8.1, 25 mM $MgCl_2$, 2 mM spermidine, and 0.01% Triton X-100 (Ke & Doudna 2004). Transcription reactions were incubated at 37°C for 2-3 hours, after which cycles of ribozyme cleavage were performed. For these cycles, additional $MgCl_2$ (35 mM), with or without cleavage oligo (3.5 mM), was added to reactions. DNA cleavage oligos were designed to anneal to the IRES RNA to free the 5′ and 3′ ends for ribozyme cleavage (oSW85 = 5′-TACGAGACCTCC CGGGGCACTCG-3′ for constructs w/o dom IV, oSW86 = 5′-GCACGGTCTACGAGACCT CCCGGGGCACTCG-3′ for constructs with dom IV, and oSW87 = 5′-GCACTTAGCCGGTCT ACGAGAGACCTCCCGGGGCACTCG-3′ for constructs with a tetraloop receptor in dom IV). Reactions were heated at 75°C for 5 min, placed at room temperature for 8 min, and then heated at 42°C for 12 min. Analytical cleavage reactions confirmed the appropriate number of cycles for each construct to fully cleave the RNA while reducing nonspecific degradation. Three rounds of ribozyme cleavage cycles were typical for these RNA constructs, although ribozymes surrounding a less structured RNA will typically display extensive cotranscriptional cleavage, especially if transcription reactions are transferred to 42°C for an additional hour. After ribozyme cleavage, NaOAc, pH 5.2 (3M; 0.1x vol), and EtOH (100%; 2.5x vol) were added to precipitate

the RNAs. Precipitations were incubated at -80°C overnight in 50 mL falcon tubes, then spun at 3,900xg for 20 min at 4°C. Pellets were washed with cold EtOH (70%), dried, and resuspended in the minimal volume of DEPC water and EDTA (~150 mM; ~2 mL). An equal volume of formamide was added and samples were heated at 75°C for 2 min before loading onto denaturing acrylamide gels (0.5x TBE, 8 M urea, 10% 29:1). Bromophenol blue and xylene cyanol were loaded alongside RNAs and gels run until the xylene cyanol (top dye; corresponding to ~60 nts in a 10% denaturing gel) ran off the bottom. The product RNA band was identified in preparative gels by UV shadowing with fluorescent TLC plates, cut out, crushed and passively eluted into DEPC water overnight at 4°C.

Once eluted, acrylamide was filtered away from RNAs and washed using 50 mL 0.22 μM cellulose acetate tubetop filters (Corning). The RNA solutions were brought to 50 mM KCl, and washed three times with KCl (50 mM) by concentration to < 1 mL and dilution to 15 mL, using 10,000 MWCO centrifugal filter units (Amicon Ultra). After the final wash, RNA was annealed by dilution to ~30 ng/μL in 20mM HEPES-KOH pH 7.5, 50mM KCl, and 2.5mM MgCl$_2$, heating at 65°C for 5 min, and cooling at room temperature (medium cool). Low concentrations of RNA (< 50 ng/μL) were essential to prevent dimerization of RNA constructs. A low amount of dimerization was confirmed for each RNA preparation by running an analytical native acrylamide gel (1X THE, pH 7.5, 2.5 mM MgCl2, 10% 29:1 acrylamide) at 4°C, typically for 3 hours at 12 W (10X THE buffer: 330 mM Tris, 660 mM HEPES, 1 mM EDTA, pH 7.5; to make, dissolve 40 g Tris base and 157 HEPES in 1 L DEPC water; pH will come to ~7.5 without any adjustement). RNAs were visualized by staining with SYBR Gold (Invitrogen; 10,000x stock diluted in water) for 5 min. After annealing, RNA was concentrated first to ~500 μL using 10,000 MWCO centrifugal filter units (Amicon Ultra), and then to a final concentration of 4-7 mg/mL (often ~50-100 μL) using Microcon Ultracel YM-10 concentrators, which were first washed with ~250 μL DEPC water. Once annealed, RNAs were stored at 4°C until used for crystallography.

Note: The addition of a 5 mL HiTrap DEAE FF ion exchange column was briefly introduced to the RNA purification protocol, after filtering off acrylamide and before annealing, in order to remove any low MW acrylamide species that might inhibit crystallization. RNAs were loaded onto this column and washed with 5 column volumes (CVs) of 50 mM KCl, and eluted between 400 and 600 mM KCl. Purification of pSW47 RNA was compared side-by-side with and without this ion exchange step. While the ion exchange step seemed to slightly reduce the amount of aggregation in crystal drops, it did not drastically change the conditions in which crystals were observed and, most significantly, did not lead to crystals with improved diffraction.

Hints: Midway through this project, the lab switched to using Spin=X UF concentrators (Corning; PES) rather than Amicon Ultra (cellulose) centrifugal filter units. These Corning concentrators work well for protein samples, but tend to leak significant quantities of RNA, even at 10,000 MWCO.

II.2.4 U1A RBD purification and RNA binding

An expression plasmid containing the U1A RBD (Oubridge et al. 1994) was freshly transformed into BL21/DE3 cells. Starter cultures were grown in LB + 100 μg/mL ampicillin, and nine 750 mL flasks of 2xYT + 100 μg/mL ampicillin were inoculated. Expression was conducted overnight at 37°C without any induction. Cells were gently pelleted and resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 100 mM KCl, 0.5 mM EDTA. 0.5 mM fresh PMSF, 0.01% Triton X-100, 100 units/mL DNase I, and protease inhibitor cocktail tablets (1/50 mL;

Roche), using 50 mL lysis buffer for the pellet from 3 liters of culture. Resuspended cells were transferred to 50 mL falcon tubes and freeze-thawed four times by cycling between liquid nitrogen and room temperature water. Samples can be stored overnight at -80°C before the final thaw. After the final thaw, samples were centrifuged at 18,000xg for 30 min at 4°C. The sample was then distributed equally into three Oakridge tubes to which PEI (5%, pH 7.9; 0.1x vol) was added. Samples were mixed by inversion and allowed to incubate on ice for 30 min. After additional mixing by inversion, samples were centrifuged at 30,000xg for 30 min at 4°C. The supernatant was decanted, measured, and transferred to a beaker on ice with slow stirring. To the supernatant was slowly added saturated ammonium sulfate, in 5 mL amounts, to a final concentration of 35%. The sample was slowly stirred on ice for an additional 10 min, then centrifuged at 18,000xg for 30 min at 4°C. The supernatant was decanted, measured, and transferred to a beaker at room temperature with slow stirring. To this solution, solid ammonium sulfate was slowly added to a final concentration of 75% (w/v). An additional 5-10 min of stirring allowed the ammonium sulfate to fully dissolve and the sample was centrifuged at 18,000xg for 30 min at 4°C. The pellet was resuspended in ~ 35 mL dialysis buffer (20 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.5 mM EDTA, 0.5 mM fresh PMSF), and dialyzed overnight into 2 L of the same buffer at 4°C with gentle stirring, using Snakeskin pleated dialysis tubing (3500 MWCO; Pierce).

In the morning, the sample (~45 mL) was centrifuged at 30,000xg for 30 min at 4°C, and applied to a SP Sepharose FF column (homepacked, column volume ~33 mL) equilibrated in buffer M (20 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.5 mM EDTA, 0.5 mM fresh PMSF). Proteins were eluted using an 8 CV linear gradient up to 100% buffer N (20 mM HEPES-KOH pH 7.5, 1 M KCl, 0.5 mM EDTA, 0.5 mM fresh PMSF). The column could be run at 4 mL/min with a pressure limit of 0.3 MPa, and 3.5 mL fractions were collected. U1A eluted at ~50% buffer N, whereas an earlier peak (~35% buffer N) contained a 16 kDa contaminating protein. The fractions containing U1A (11 kDa) were pooled, concentrated to < 3 mL and applied to a Superdex 75 (16/60; 120 mL) column (GE healthcare), which had been washed with NaOH (0.5M) and water, and then equilibrated in buffer M. The column had a pressure limit of 0.5MPa and could be run at ~1 mL/min. U1A eluted at ~85 mL and very little was in the void volume. The fractions containing U1A (11 kDa) were pooled and applied to a CHT 10-I (10 mL) column (Biorad), which had been washed with NaOH (0.5M) and water, and then equilibrated in buffer O (10 mM potassium phosphate, pH 7.5, 50 mM KCl). U1A was eluted using a 10 CV linear gradient from 0 to 100% buffer P (10 mM potassium phosphate, pH 7.5, 50 mM KCl, 0.5 M ammonium sulfate). The column could be run at 0.4-0.7 mL/min with a pressure limit of 5.5 MPa, and 2 mL fractions were collected. U1A eluted at ~40% buffer P. The fractions containing U1A were pooled, dialyzed overnight into 1L of U1A storage buffer (10 mM HEPES KOH, pH 7.5, 0.1 EDTA), concentrated using an 3,500 MWCO centrifugal filter units (Amicon Ultra) to a final concentration of ~17 mg/mL, frozen in liquid nitrogen and stored at -80°C.

Since exact determination of protein or RNA concentration by UV spectroscopy is challenging, the ideal ratio of U1A:RNA for cocrystallography was determined by conducting gel shifts in 0.5x TBE (Rupert & Ferre-D'Amare 2004). Typical observed ratios at which complete shifting of U1A hairpin-IRES RNA was seen were 0.7-0.8:1 U1A:RNA. (This presumably corresponds to a true ratio of ~1:1, with the error resulting from error in protein- and RNA-concentration determination.)

Hints: This protein preparation is notoriously difficult. Kaihong Zhou has magic hands that make it work, and the critical step seems to be during or before the ammonium sulfate

precipitations. When things go wrong, one typically purifies a 16 kDa protein, rather than the 11 kDa U1A RBD. Stefanie Mortimer directly compared U1A expression in 2xYT vs. LB media, in the presence of 50 or 100 µg/mL ampicillin, as these were two variables where Kaihong's protocol differs from standard lab practice. Indeed, expression levels of U1A were much higher in 2xYT than LB media, while the lower ampicillin concentration (50 µg/mL) yielded slightly more U1A expression than 100 µg/mL. In order to distinguish the 11 kDa U1A RBD from the 16 kDa impurity, 15% SDS-PAGE gels are essential.

II.2.5 RNA crystallization

Initial crystallization screens were carried out in 96-well plates using a Mosquito crystallization robot (TTP Labtech). The principal screen utilized was the Qiagen Nucleix Suite, with conditions geared toward nucleic acid and protein/nucleic acid crystallography. 96-well hanging drop trays were set up with 3-6 mg/mL RNA, and a "buffer only" drop to help identify salt crystals. RNA and buffer solutions were deposited in 8-well 5 µL micro-reservoir strips (TTP Labtech). The program (Doudna/Blake/2 drop 2x aspiration sample then ML) set up 200-300 nL hanging drops by drawing in 100-150 nL of RNA, followed by an equal volume of mother liquor from Microtest 96 tissue culture plates (Falcon), and then depositing them together onto ViewDropII 96-well plate seals (TTP Labtech). Given enough material, initial Nucleix trays were set up both at 18°C and 4°C. For most constructs, more promising hits were observed at 18°C than 4°C. If significantly greater or less than 50% of the wells resulted in precipitation after 1 week, the RNA concentration was decreased or increased, respectively.

Preliminary hits were repeated and optimized in larger format trays, which were set by hand. Crystallization was performed at 18°C using the hanging drop vapor diffusion method, using equal volumes (1 µL + 1 µL) of mother liquor and concentrated RNA solution. Easy Xtal trays (Qiagen) were used with 500 µL mother liquor. Hanging drops were set up with RNA first, then mother liquor, followed by 1 pipette stroke of mixing up and down. For seeding, drops were set as above and then an additional 0.2 µL of seeding mix (crushed or vortexed crystals in mother liquor) was gently deposited into the drop. For additive screens (with Hampton Research Additive Screen), drops were set as above and then an additional 0.2 µL of additive mix was gently deposited into the drop.

Hints: The Ferre-D'Amare lab routinely uses a dehydration strategy as a standard stage of crystal screening. After each crystal is screened for diffraction, it is transferred from the loop directly to dehydration solution(s), allowed to dehydrate for varying degrees of times, and shot again to see whether diffraction for that particular crystal improved. Methods and approaches for this type of post-crystallization treatments have been reviewed (Heras & Martin 2005).

II.3 Preliminary Results and Discussion

Initial crystallography constructs were designed to include the pseudoknot domain, with or without dom IV, with or without IIId, and with or without U1A hairpins or tetraloop receptors. The initial panel of crystallography constructs is depicted in Figure II.3, and the sequences for each construct are listed in Table II.1. The crystallization-trial results with these constructs are described in Figure II.4. Most of the RNA constructs ran as clean, discrete bands on native acrylamide gels, indicating that they were well-folded after annealing at < 40 ng/µL RNA and subsequent concentration. Constructs pSW22 and pSW27 were the exception to this trend, as these RNAs ran as smeary comet tails on native gels. This behavior likely resulted from insertion of the tetraloop receptor in SLIV destabilizing the stem and resulting in multiple conformations

127

exchanging during electrophoresis. The 4 U1A-hairpin constructs all folded nicely, and gave clean gel shifts with U1A RBD, but no hits beyond clear precipitate were observed. More U1A-hairpin constructs have been cloned, varying the position of the U1A hairpin around the pseudoknot domain and the linker lengths between the hairpin and the HCV IRES sequence (Figure II.5; Table II.2). pSW26 yielded needles which could be reproduced and optimized into boxy plates, but which never yielded better than 20Å diffraction. The best hits from this initial panel came from pSW25, as needles from 96-well screens diffracted initially to 16Å, and these could be reproduced and optimized to eventually give diffraction to 5.5Å.

Once pSW25 showed promise, the sequence of this construct was systematically varied in parallel to efforts to optimize crystallization conditions for the original construct. Two matrices of follow-up constructs were designed. The first matrix varied 5′ and 3′ overhang length vs. the length of the linker between the tetraloop and tetraloop receptor. Schematic depictions of these constructs are shown in Figure II.6, and the sequences listed in Table II.1. Crystallization trials with this matrix of constructs are described in Figure II.7, and photographs of representative crystals are shown in Figure II.8. Without any overhang, varying the linker length still produces crystals, but that diffract more poorly than crystals with a 3 bp linker. 4 and 5 bp linkers produce crystals that diffract to 9-10Å, whereas 2 or 6 bp linkers give very small seeds or globules, but no single crystals. Maintaining a 3 bp linker, the addition of even a single nucleotide as a 5′ overhang leads to very small crystals (Figures II.7 and II.8). In contrast, addition of a 3′ overhang (between 1-4 nts) still gives crystals that diffract to at least 6Å, with the best diffraction to date coming from pSW47, which possesses a 3′ 2 nt overhang. The overhang lengths tested so far have been taken from WT HCV IRES sequence, but whether the sequence of this overhang can be varied to improve crystallization is an important parameter to explore.

The second matrix of follow-up constructs around pSW25 varied 3′ overhang length vs. wobble base pair motifs; these constructs are shown in Figure II.9. Crystallization of a construct with the linker sequence swapped out for a wobble base pair motif could allow for streamlined phasing since these motifs tend to bind well to cobalt and iridium hexamine (Keel et al. 2007). Crystallization trials with this matrix of constructs are described in Figure II.10, and photographs of representative crystals are shown in Figure II.11. Of the 4 GU motifs tested so far with no overhang, no substitution leads to enhanced diffraction relative to pSW25 (5.5Å). PM13 leads to the most comparable diffraction (6Å), whereas PM14 and PM5 lead to 8 and 11Å diffraction, respectively. It will be interesting to test pKB267 which contains the most benign GU motif, PM13, in the context of the favorable 2 nt 3′ overhang, as this construct may be very useful for acquiring phase information after soaking in iridium hexamine.

In addition to varying the RNA construct, crystallization conditions have been extensively varied, altering parameters such as: the concentration of lithium sulfate, ammonium sulfate, spermine, magnesium chloride, cobalt hexamine, potassium chloride, cobalt chloride, nickel chloride, and zinc chloride, as well as crystallization temperatures of 4, 18 or 30 degrees, macro- or micro-seeding, additive screens, and inclusion of a weak ion exchange column in the RNA purification. Dehydration in higher concentrations of lithium sulfate has also been attempted without success (Heras & Martin 2005).

Nickel chloride was identified as a favorable additive in an additive screen, and cobalt and zinc chloride also proved to be useful additives. The best data sets collected so far come from pSW47 RNA in the presence of $NiCl_2$. Crystallization conditions and statistics are shown in Figure II.12, along with statistics for the best native and anomalous data sets collected to date. Molecular replacement has been attempted with the 4.0Å native dataset, using the Westhof

computational model as a search model, but has not yet returned any interpretable electron density maps. Anomalous signal is present in crystals grown in the presence of $NiCl_2$ and $ZnCl_2$, but the signal is only above noise through ~5.5 Å. To date, no heavy atom substructure has been solved. Future directions will focus on soaking in heavier atoms than Ni^{2+} and Zn^{2+}, in order to obtain stronger anomalous signal, even at low resolution (Golden 2000; Garman & Murray 2003; Golden 2007). As a last resort, brute force molecular replacement could be attempted with simple secondary structural elements as search models (Robertson & Scott 2008).
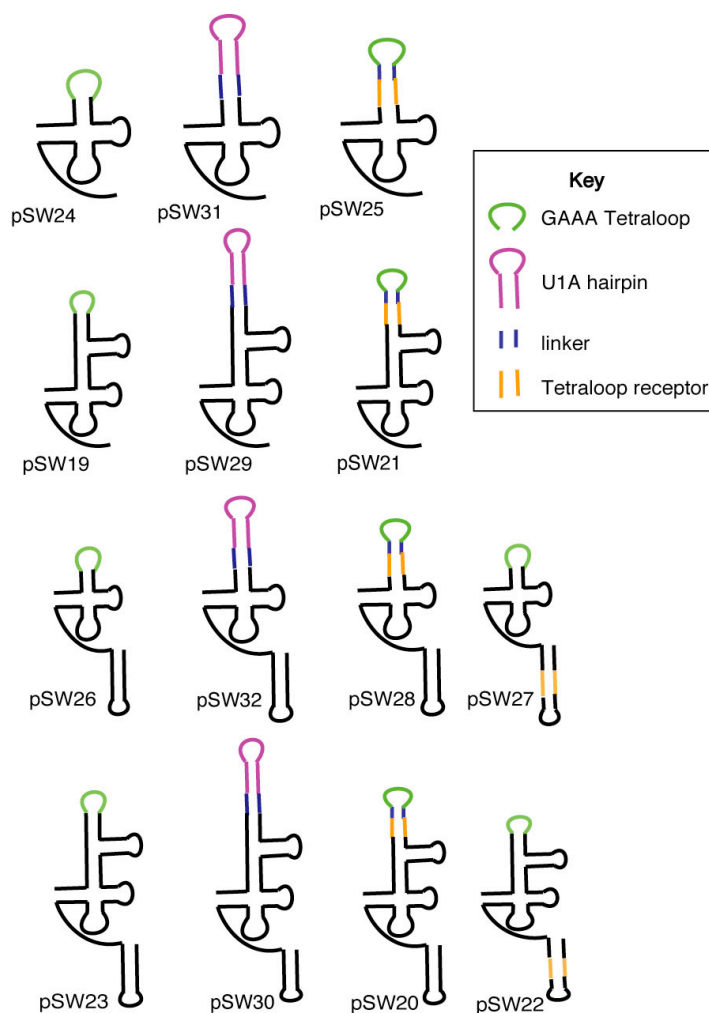

Figure II.3: Schematic depictions of initial panel of crystallography constructs. Constructs were designed to include the pseudoknot domain, with or without dom IV, with or without IIId, and with or without U1A hairpins or tetraloop receptors. For the sequences of constructs, please refer to *Table II.1*.

129

Construct:	**pSW24**	**pSW31**	**pSW25**	
Cloned?:	y	y	y	
Well-folded:	y	y	y	
Trays set:	nucleix(4/)	nucleix (4/18), MPD	nucleix(4/18) +optimization	
crytals?:	clear P	clear P	needles/rods	
Best diffraction:			16A from mosq.	
Tested by:	SW	TC	SW/KB/YB	

Construct:	**pSW19**	**pSW29**	**pSW21**	
Cloned?:	y	y	no	
Well-folded:	y	y		
Trays set:	nucleix (4/18)	nucleix(4/18), MPD		
crytals?:	clear P	clear P		
Best diffraction:				
Tested by:	NH	TC	SW	

Construct:	**pSW26**	**pSW32**	**pSW28**	**pSW27**
Cloned?:	y	y	y	y
Well-folded:	y	y	no, ppts.	no
Trays set:	nucleix (4/18) + optimization	nucleix (4/18), MPD, PEG		
crytals?:	needles/plates			
Best diffraction:	20 A			
Tested by:	SW	TC	SW	NH

Construct	**pSW23**	**pSW30**	**pSW20**	**pSW22**
Cloned?:	y	y	y	y
Well-folded:	y	y	y	no
Trays set:	nucleix(4/18)	nucleix (4/18)	nucleix(4/18)	
crytals?:	clear P	clear P	clear P	
Best diffraction:				
Tested by:	SW	TC	SW	NH

Figure II.4: Details of trials with initial panel of crystallography constructs. Information for each construct is provided about its progress and promise through the crystallographic pipeline: cloning, well-foldedness post-annealing, which 96-well screens (Qiagen) were set at which temperatures, any crystals or clear precipitate observed in these screens, and the best diffraction observed to date, where applicable. Key: SW = Shruti Waghray, KB= Katie Berry, TC = Tony Chen, NH = Naeem Husain, YB Yun Bai; clear P = clear precipitate, which may or may not be microcrystals; 4/18 = trays set at both 4 and 18 degrees. If not noted, trays set at 18 degrees; ppts: precipitation observed during attempted RNA purification.
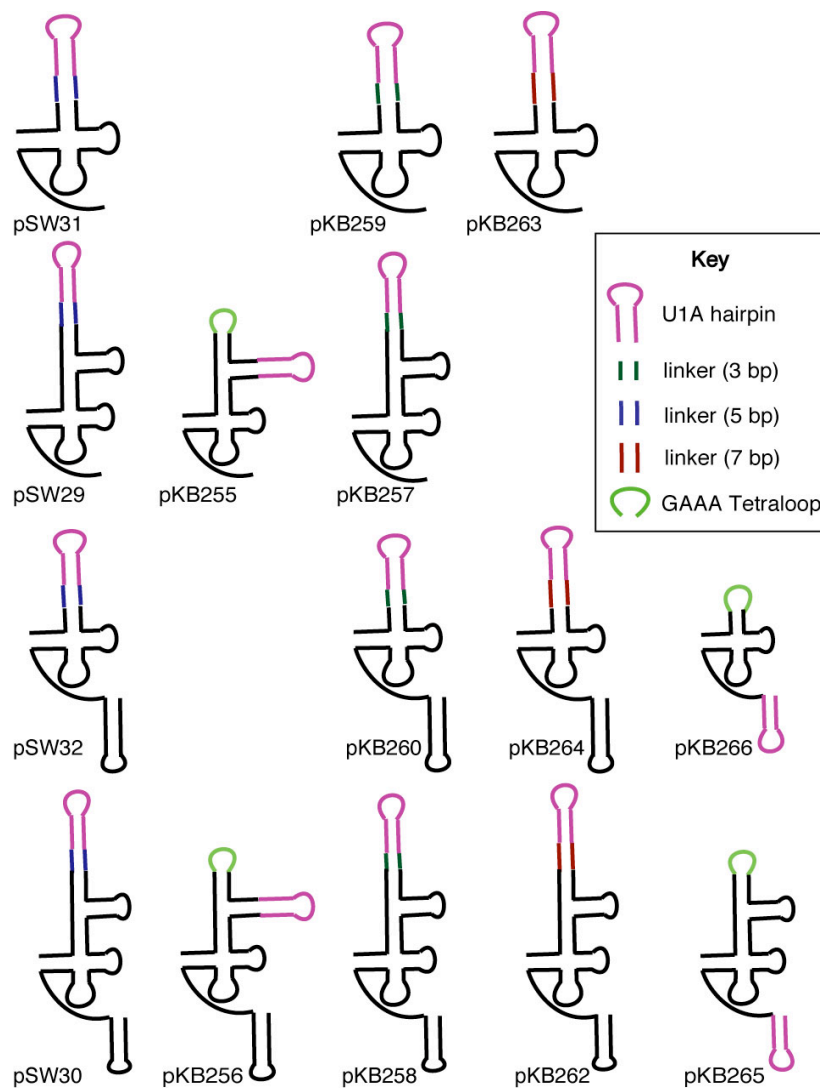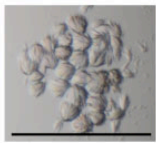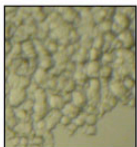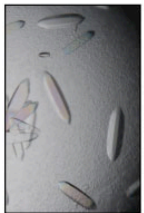
Figure II.5: Schematic depiction of additional cloned U1A-hairpin constructs. These constructs were designed to move the U1A hairpin into different positions around the pseudoknot domain and to introduce varied linker lengths between the hairpin and the HCV IRES sequence. The four constructs tested to date (pSW29-32) placed the U1A hairpin where the IIIabc(d) domain would be, with a 5 bp linker, which may not have been a favorable arrangement for crystallization. Beyond pSW29-32, these constructs have been cloned, but not tested. For construct sequences, please refer to *Table II.2*.
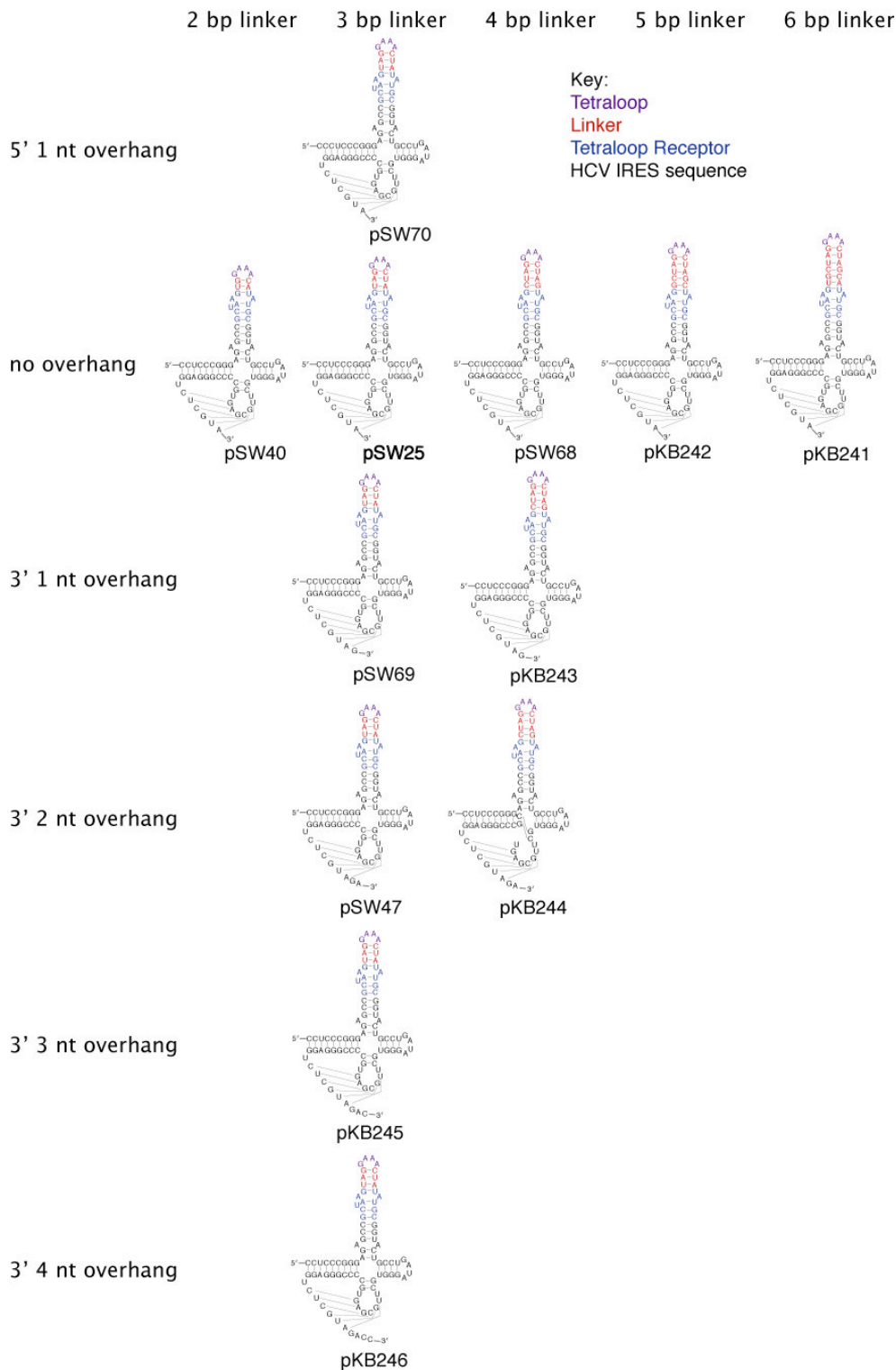
Figure II.6: Crystallographic construct panel (overhang- vs. linker-length). Constructs were designed around pSW25, which contained neither a 5′ or 3′ overhang and had a linker length of 3 base pairs (bps) between tetraloop and receptor. Structures are shown for only those constructs within the grid that have been both cloned and tested (see *Figures II.7* and *II.8*). For the sequences of all constructs, please refer to *Table II.1*.

Overhang:		2 bp linker	3 bp linker	4 bp linker	5 bp linker	6 bp linker
5' 1 nt	Construct:		pSW70			
	Cloned?:		y			
	Well-folded:		y			
	Trays set:		6: l, m, a, p, sp			
	crytals?:		small.			
	Best diffraction:					
	Tested by:		KB			
none	Construct:	pSW40	pSW25	pSW68	pKB242	pKB241
	Cloned?:	y	y	y	y	y
	Well-folded:	y	y	y	y	y
	Trays set:	5: n, s, l, sp, p, m, c, ad	>10	3: l, m, p, sp, s	5: n, l, m, p, s	5: l, m, a, p sp
	crytals?:	seeds.	rods.	rods.	blocky rods.	globules.
	Best diffraction:		5.5 A	9 A	10 A	
	Tested by:	KB	SW	KB	KB	KB
3' 1 nt	Construct:		pSW69	pKB43		
	Cloned?:		y	y		
	Well-folded:		y	y		
	Trays set:		10: l, m, a	2: m, l, a		
	crytals?:		rods with edges.	rods.		
	Best diffraction:		5 A	11 A		
	Tested by:		KB	KB		
3' 2 nt	Construct		pSW47	pKB244		
	Cloned?:		y	y		
	Well-folded:		y	y		
	Trays set:		20: l, m, a, mn, ch, d, 4, 30, ad, c, n, z	2: m, l, a		
	crytals?:		leaves.	rods		
	Best diffraction:		3.5 A	10 A		
	Tested by:		KB	KB		
3' 3 nt	Construct:		pKB245			
	Cloned?:		y			
	Well-folded:		y			
	Trays set:		3: m, l. a, 4			
	crytals?:		leaves.			
	Best diffraction:		5.5 A			
	Tested by:		KB			
3' 4 nt	Construct		pKB246			
	Cloned?:		y			
	Well-folded:		y			
	Trays set:		3: m, l. a, 30			
	crytals?:		rods with edges.			
	Best diffraction		6 A			
	Tested by:		KB			

Figure II.7: Details of trials with overhang- vs. linker-length panel. Information for each construct is provided about its progress and promise through the crystallographic pipeline: cloning, well-foldedness post-annealing, how many trays (typically 15- or 24-well screens, set by hand) have been set and varying which parameters, description of crystals observed, and the best diffraction observed to date. Constructs are shown in the same grid pattern from *Figure II.6*. Key: n= Nucleix; p = vary pH; mn = included Mn^{2+}; L = varied $[Li_2(SO_4)_2]$, a= varied $[(NH_4)_2(SO_4)_2]$, s=seeding, sp= varied [spermine], m=varied $[Mg^{2+}]$, ch = varied $[Co(NH_3)_6]$ c= varied $[CoCl_2]$, ad = additive screen, n = varied $[NiCl_2]$, z = varied $[ZnCl_2]$, k = varied [KCl], d = put over DEAE column in purification, 4 = set at 4 degrees, 30 = set at 30 degrees, initial number = # of trays set with construct.

133

Figure II.8: Photographs of crystal hits with overhang- vs. linker-length panel. Representative crystals are shown for each construct, in the same grid pattern as *Figures II.6* and *II.7*, demonstrating the variation in crystal morphology that is observed as linker- and overhang-lengths are varied. Note: while scales are not consistent throughout, the size crystals appear does reflect their observed size in the trays.
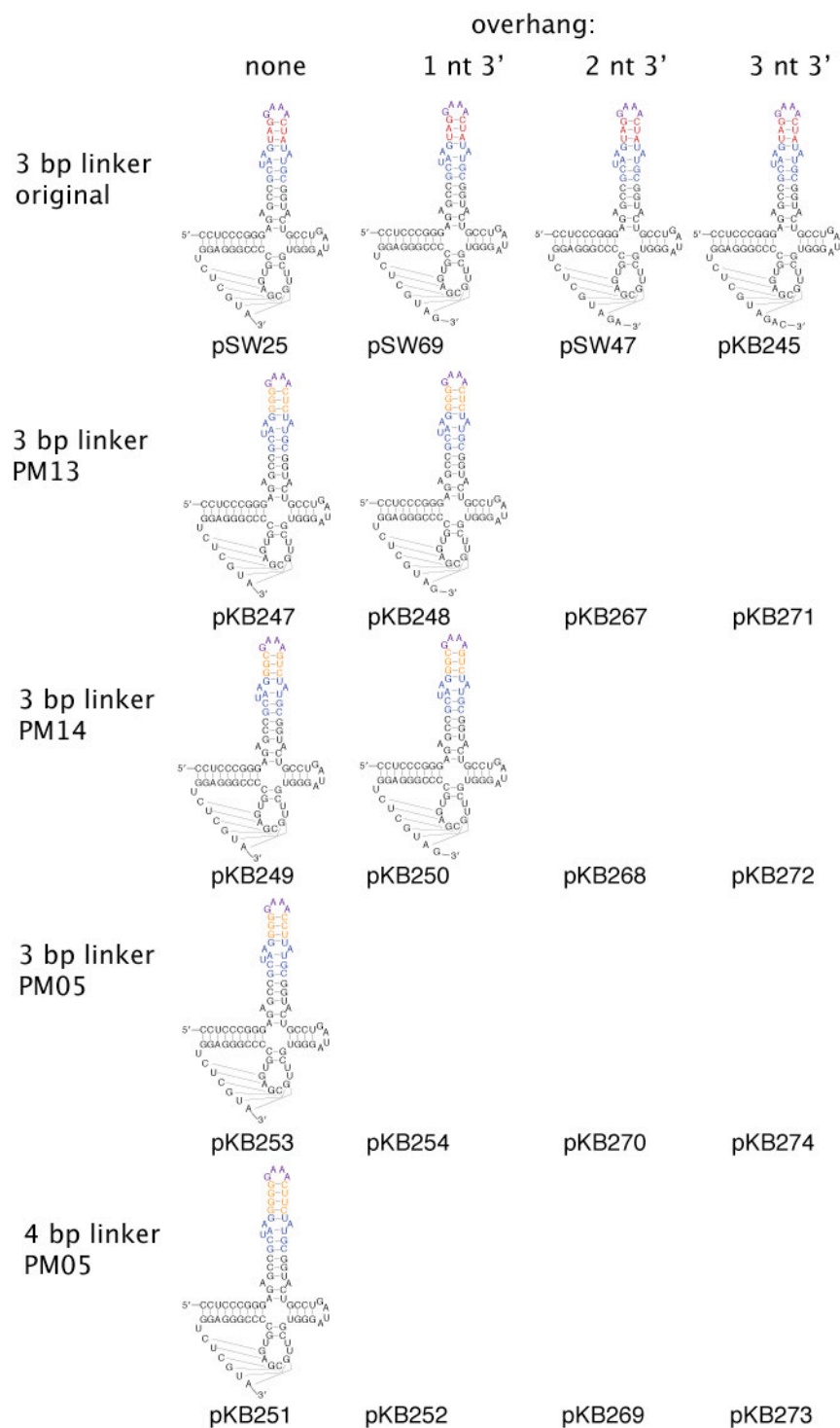
Figure II.9 Crystallographic construct panel (overhang length vs. GU motif). Constructs were designed around pSW25, which contained no 3′ overhang and linker between tetraloop and receptor of 3 base pairs (bps) without a GU wobble motif. These GU wobble motif sequences were taken from (Keel et al. 2007). Structures are shown for only those constructs within the grid that have been both cloned and tested (see *Figures II.10* and *II.11*). For the sequences of all constructs, please refer to *Table II.1*.

135

	overhang:			
	none	1 nt 3'	2 nt 3'	3 nt 3'
3 bp linker original	Construct: **pSW25** Cloned?: y Well-folded: y Trays set: >10 crytals?: rods. Best diffraction: 5.5 A Tested by: SW,.YB	**pSW69** y y 10: l, m, a rods with edges. 5 A KB	**pSW47** y y 20: l, m, a, mn, ch, d, 4, 30, ad, c, n, z leaves. 3.5 A KB	**pKB245** y y 3: m, l. a, 4 leaves. 5.5 A KB
3 bp linker PM13	Construct: **pSW247** Cloned?: y Well-folded: y Trays set: 2: m, l, a, ch crytals?: Best diffraction: 6 A Tested by: KB	**pKB248** y y 2: l, m, 30 leaves/ 6 A KB	**pKB267** y	**pKB271** y
3 bp linker PM14	Construct: **pSW249** Cloned?: y Well-folded: y Trays set: 2: m, l, a, ch crytals?: Best diffraction: 8 A Tested by: KB	**pKB250** y y 2: l, m, 30 rods. 8 A KB	**pKB268** no	**pKB272** y
3 bp linker PM05	Construct **pKB253** Cloned?: y Well-folded: y Trays set: 2: l, m, 30 crytals?: no, but > dimer. Best diffraction: Tested by: KB	**pKB254** y	**pKB270** no	**pKB274** no
4 bp linker PM05	Construct: **pKB251** Cloned?: y Well-folded: y Trays set: 2: l, m, 30 crytals?: leaves. Best diffraction: 11 A Tested by: KB	**pKB252** y	**pKB269** y	**pKB273** y

Figure II.10 Details of trials with overhang length vs. GU motif panel. Nomenclature for GU wobble motifs is taken from (Keel et al. 2007). Information for each construct is provided about its progress and promise through the crystallographic pipeline: cloning, well-foldedness post-annealing, how many trays (typically 15- or 24-well screens, set by hand) have been set and varying which parameters, description of crystals observed, and the best diffraction observed to date. Constructs are shown in the same grid pattern as in *Figure II.9*. Key: n= Nucleix; p = vary pH; mn = included Mn^{2+}; L = varied $[Li_2(SO_4)_2]$, a= varied $[(NH_4)_2(SO_4)_2]$, s=seeding, sp= varied [spermine], m=varied $[Mg^{2+}]$, ch = varied $[Co(NH_3)_6]$ c= varied $[CoCl_2]$, ad = additive screen, n = varied $[NiCl_2]$, z = varied $[ZnCl_2]$, k = varied [KCl], d = put over DEAE column in purification, 4 = set at 4 degrees, 30 = set at 30 degrees, initial number = # of trays set with construct.
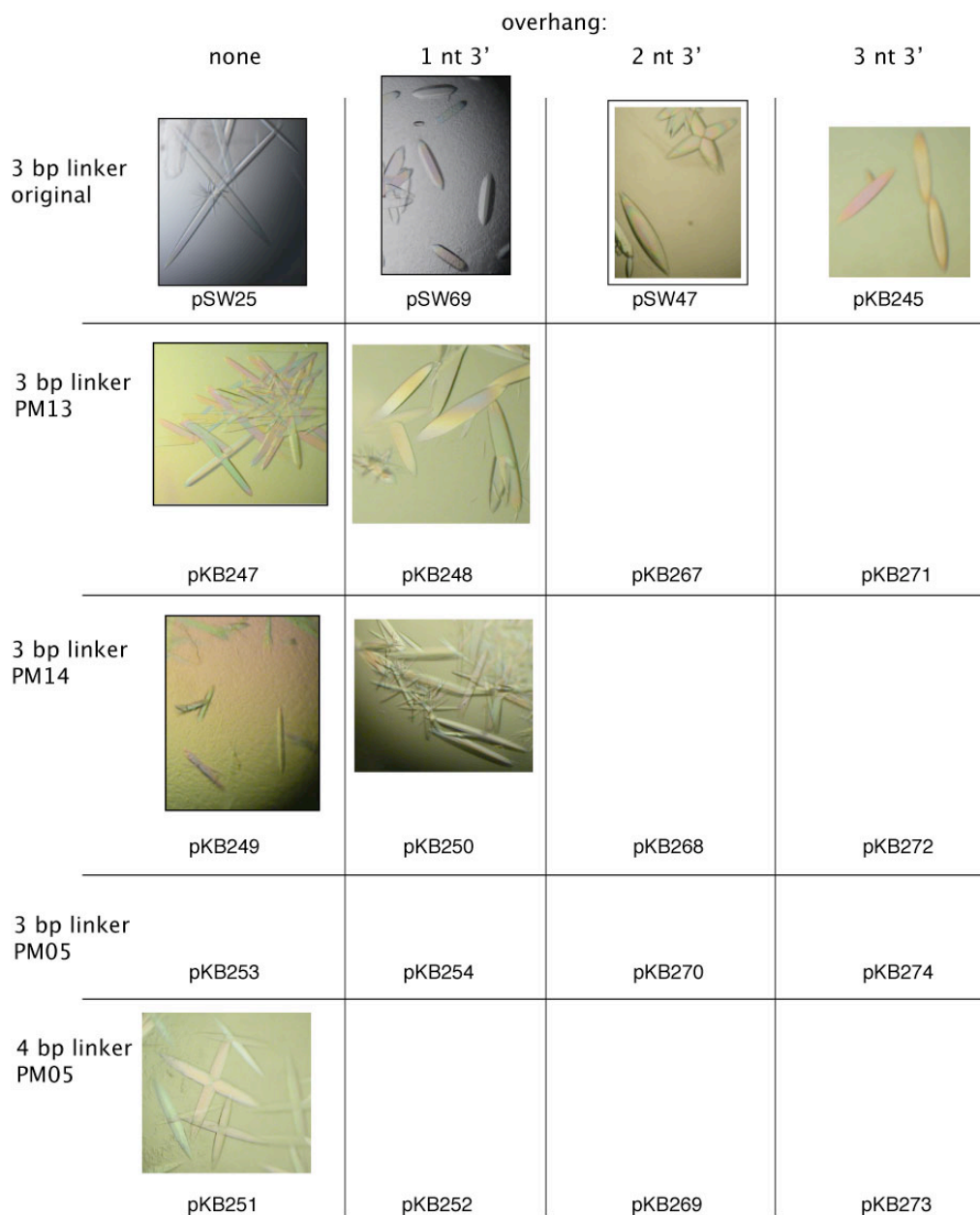
Figure II.11 Photographs of crystal hits with overhang length vs. GU motif panel. Representative crystals are shown for each construct, in the same grid pattern as *Figures II.9* and *II.10*, demonstrating the variation in crystal morphology that is observed as linker- and overhang-lengths are varied. Note: while scales are not consistent throughout, the size crystals appear does reflect their observed size in the trays.

A **Crystallization Conditions:**

50 mM MES 6.5, 0.55 M $Li_2(SO_4)_2$, 45 mM $MgCl_2$, 1 mM spermine, ~10 mM $NiCl_2$, 4 mg/mL RNA, 18 deg, 2 (1+1) μL hanging drop, 2-5 days, cryo-protected in presence of 35% ethylene glycol.

Crystallographic Statistics

Space Group	$P4_{3(1)}2_12_1$
Unit Cell (Å)	a=109.5 b=109.5 c=84

Data Collection	Native	$NiCl_2$ (SAD)
Wavelength (Å)	1.1159	*1.488*
Resolution (Å)	100-4.0 (4.3-4.0)	100-5.4 (5.7-5.4)
No. of reflections		
Total	66,778	28,474
Unique	7,320	3,558
Completeness (%)	99.5 (100.0)	99.6 (100)
I/σ	14.5 (3.4)	33.3 (12.8)
R_{sym} (%)	10.4 (74)	3.6 (15.1)
SigAno		2.53 (1.08)

B ~100 μm
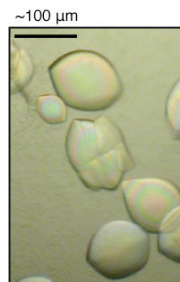
Figure II.12: Crystallographic information for best datasets. *(A)* Crystallization conditions, statistics and data collection information for the best datasets collected to date, using construct pSW47. *(B)* Photograph of representative crystal containing $NiCl_2$, which lead to superior diffraction.

BIBLIOGRAPHY

Acker MG, Shin BS, Dever TE, Lorsch JR. 2006. Interaction between eukaryotic initiation factors 1A and 5B is required for efficient ribosomal subunit joining. *J Biol Chem* **281:** 8469-8475.

Algire MA, Lorsch JR. 2006. Where to begin? The mechanism of translation initiation codon selection in eukaryotes. *Curr Opin Chem Biol* **10:** 480-486.

Algire MA, Maag D, Lorsch JR. 2005. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol Cell* **20:** 251-262.

Alotte C, Martin A, Caldarelli SA, Di Giorgio A, Condom R, Zoulim F, Durantel D, Hantz O. 2008. Short peptide nucleic acids (PNA) inhibit hepatitis C virus internal ribosome entry site (IRES) dependent translation in vitro. *Antiviral Res* **80:** 280-287.

Alvarez-Salas LM. 2008. Nucleic acids as therapeutic agents. *Curr Top Med Chem* **8:** 1379-1404.

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

Auld DS, Zhang YQ, Southall NT, Rai G, Landsman M, Maclure J, Langevin D, Thomas CJ, Austin CP, Inglese J. 2009. A Basis for Reduced Chemical Library Inhibition of Firefly Luciferase Obtained from Directed Evolution. *J Med Chem* **52:** 1450-1458.

Baird SD, Turcotte M, Korneluk RG, Holcik M. 2006. Searching for IRES. *Rna* **12:** 1755-1785.

Balvay L, Soto Rifo R, Ricci EP, Decimo D, Ohlmann T. 2009. Structural and functional diversity of viral IRESes. *Biochim Biophys Acta* **1789:** 542-557.

Bartenschlager R, Cosset FL, Lohmann V. 2010. Hepatitis C virus replication cycle. *J Hepatol* **53:** 583-585.

Batey RT, Sagar MB, Doudna JA. 2001. Structural and energetic analysis of RNA recognition by a universally conserved protein from the signal recognition particle. *J Mol Biol* **307:** 229-246.

Battiste JL, Pestova TV, Hellen CU, Wagner G. 2000. The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. *Mol Cell* **5:** 109-119.

Baugh C, Wang S, Li B, Appleman JR, Thompson PA. 2009. SCAN--a high-throughput assay for detecting small molecule binding to RNA targets. *J Biomol Screen* **14:** 219-229.

Berry KE, Peng B, Koditek D, Pagratis N, Perry J, Parish J, Zhong W, Doudna JA, Shih I-h. 2010. Optimized high-throughput screen for Hepatitis C virus translation inhibitors. *J Biomol Screen,* in press.

Bodoor K, Boyapati V, Gopu V, Boisdore M, Allam K, Miller J, Treleaven WD, Weldeghiorghis T, Aboul-ela F. 2009. Design and implementation of an ribonucleic acid (RNA) directed fragment library. *J Med Chem* **52:** 3753-3761.

Boehringer D, Thermann R, Ostareck-Lederer A, Lewis JD, Stark H. 2005. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13:** 1695-1706.

Brierley I, Pennell S, Gilbert RJ. 2007. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol* **5:** 598-610.

Brown EA, Zhang H, Ping LH, Lemon SM. 1992. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* **20:** 5041-5045.

Bukh J, Purcell RH, Miller RH. 1992. Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci U S A* **89:** 4942-4946.

Bushell M, Wood W, Clemens MJ, Morley SJ. 2000. Changes in integrity and association of eukaryotic protein synthesis initiation factors during apoptosis. *Eur J Biochem* **267:** 1083-1091.

Cheung YN, Maag D, Mitchell SF, Fekete CA, Algire MA, Takacs JE, Shirokikh N, Pestova T, Lorsch JR, Hinnebusch AG. 2007. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes Dev* **21:** 1217-1230.

Chiu WL, Wagner S, Herrmannova A, Burela L, Zhang F, Saini AK, Valasek L, Hinnebusch AG. 2010. The C-terminal region of eukaryotic translation initiation factor 3a (eIF3a) promotes mRNA recruitment, scanning, and, together with eIF3j and the eIF3b RNA recognition motif, selection of AUG start codons. *Mol Cell Biol* **30:** 4415-4434.

Cigan AM, Feng L, Donahue TF. 1988. tRNAi(met) functions in directing the scanning ribosome to the start site of translation. *Science* **242:** 93-97.

Clarke PA. 1999. RNA footprinting and modification interference analysis. *Methods Mol Biol* **118:** 73-91.

Collier AJ, Gallego J, Klinck R, Cole PT, Harris SJ, Harrison GP, Aboul-Ela F, Varani G, Walker S. 2002. A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Biol* **9:** 375-380.

Conte MR, Kelly G, Babon J, Sanfelice D, Youell J, Smerdon SJ, Proud CG. 2006. Structure of the eukaryotic initiation factor (eIF) 5 reveals a fold common to several translation factors. *Biochemistry* **45:** 4550-4558.

Costa-Mattioli M, Sossin WS, Klann E, Sonenberg N. 2009. Translational control of long-lasting synaptic plasticity and memory. *Neuron* **61:** 10-26.

Costantino DA, Pfingsten JS, Rambo RP, Kieft JS. 2008. tRNA-mRNA mimicry drives translation initiation from a viral IRES. *Nat Struct Mol Biol* **15:** 57-64.

Cullen PJ, Sprague GF, Jr. 2000. Glucose depletion causes haploid invasive growth in yeast. *Proc Natl Acad Sci U S A* **97:** 13619-13624.

Das S, Ghosh R, Maitra U. 2001. Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J Biol Chem* **276:** 6720-6726.

Dasso MC, Jackson RJ. 1989. On the fidelity of mRNA translation in the nuclease-treated rabbit reticulocyte lysate system. *Nucleic Acids Res* **17:** 3129-3144.

Deutsch M, Hadziyannis SJ. 2008. Old and emerging therapies in chronic hepatitis C: an update. *J Viral Hepat* **15:** 2-11.

Dorner AJ, Semler BL, Jackson RJ, Hanecak R, Duprey E, Wimmer E. 1984. In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate. *J Virol* **50:** 507-514.

Easton LE, Locker N, Lukavsky PJ. 2009. Conserved functional domains and a novel tertiary interaction near the pseudoknot drive translational activity of hepatitis C virus and hepatitis C virus-like internal ribosome entry sites. *Nucleic Acids Res* **37:** 5537-5549.

ElAntak L, Tzakos AG, Locker N, Lukavsky PJ. 2007. Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit. *J Biol Chem* **282:** 8165-8174.

ElAntak L, Wagner S, Herrmannova A, Karaskova M, Rutkai E, Lukavsky PJ, Valasek L. 2010. The indispensable N-terminal half of eIF3j/HCR1 cooperates with its structurally conserved binding partner eIF3b/PRT1-RRM and with eIF1A in stringent AUG selection. *J Mol Biol* **396:** 1097-1116.

Fekete CA, Applefield DJ, Blakely SA, Shirokikh N, Pestova T, Lorsch JR, Hinnebusch AG. 2005. The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection in vivo. *Embo J* **24:** 3588-3601.

Fekete CA, Mitchell SF, Cherkasova VA, Applefield D, Algire MA, Maag D, Saini AK, Lorsch JR, Hinnebusch AG. 2007. N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. *EMBO J* **26:** 1602-1614.

Ferre-D'Amare AR, Doudna JA. 2000. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol* **295:** 541-556.

Ferre-D'Amare AR, Doudna JA. 2001. Methods to Crystallize RNA. *Current Protocol in Nucleic Acid Chemistry*: John Wiley & Sons. pp 7.6.1-7.6.13.

Ferre-D'Amare AR, Zhou K, Doudna JA. 1998. A general module for RNA crystallization. *J Mol Biol* **279:** 621-631.

Fletcher CM, Pestova TV, Hellen CU, Wagner G. 1999. Structure and interactions of the translation initiation factor eIF1. *Embo J* **18:** 2631-2637.

Fletcher SP, Ali IK, Kaminski A, Digard P, Jackson RJ. 2002. The influence of viral coding sequences on pestivirus IRES activity reveals further parallels with translation initiation in prokaryotes. *RNA* **8:** 1558-1571.

Fletcher SP, Jackson RJ. 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.

Fontanes V, Raychaudhuri S, Dasgupta A. 2009. A cell-permeable peptide inhibits hepatitis C virus replication by sequestering IRES transacting factors. *Virology* **394:** 82-90.

Fraser CS, Berry KE, Hershey JW, Doudna JA. 2007. eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol Cell* **26:** 811-819.

Fraser CS, Doudna JA. 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol* **5:** 29-38.

Fraser CS, Hershey JW, Doudna JA. 2009. The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat Struct Mol Biol* **16:** 397-404.

Fraser CS, Lee JY, Mayeur GL, Bushell M, Doudna JA, Hershey JW. 2004. The j-subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40 S ribosomal subunits in vitro. *J Biol Chem* **279:** 8946-8956.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci U S A* **83:** 9373-9377.

Fringer JM, Acker MG, Fekete CA, Lorsch JR, Dever TE. 2007. Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining. *Mol Cell Biol* **27:** 2384-2397.

Frost AF, Pearson RG. 1961. *Kinetics and Mechanism*. New York, NY: John Wiley and Sons, Inc.

Gallego J, Varani G. 2002. The hepatitis C virus internal ribosome-entry site: a new target for antiviral research. *Biochem Soc Trans* **30:** 140-145.

Garman E, Murray JW. 2003. Heavy-atom derivatization. *Acta Crystallogr D Biol Crystallogr* **59:** 1903-1913.

Gebauer F, Hentze MW. 2004. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5:** 827-835.

Geissmann T, Marzi S, Romby P. 2009. The role of mRNA structure in translational control in bacteria. *RNA Biol* **6:** 153-160.

Ghany MG, Strader DB, Thomas DL, Seeff LB. 2009. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* **49:** 1335-1374.

Gilbert WV. 2010. Alternative ways to think about cellular internal ribosome entry. *J Biol Chem* **285:** 29033-29038.

Gilbert WV, Zhou K, Butler TK, Doudna JA. 2007. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* **317:** 1224-1227.

Golden BL. 2000. Heavy atom derivatives of RNA. *Methods Enzymol* **317:** 124-132.

Golden BL. 2007. Preparation and Crystallization of RNA. In: Doublie S, ed. *Methods in Molecular Biology*. Totowa, NJ: Humana Press. pp 239-257.

Gomez E, Mohammad SS, Pavitt GD. 2002. Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation. *Embo J* **21:** 5292-5301.

Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, Tekamp-Olson P, Kuo G, Choo QL, Houghton M. 1991. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. *Proc Natl Acad Sci U S A* **88:** 1711-1715.

Hellen CU, de Breyne S. 2007. A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: evidence for modular exchange of functional noncoding RNA elements by recombination. *J Virol* **81:** 5850-5863.

Henke JI, Goergen D, Zheng J, Song Y, Schuttler CG, Fehr C, Junemann C, Niepmann M. 2008. microRNA-122 stimulates translation of hepatitis C virus RNA. *Embo J* **27:** 3300-3310.

Heras B, Martin JL. 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* **61:** 1173-1180.

Hermann T, Westhof E. 2000. Rational drug design and high-throughput techniques for RNA targets. *Comb Chem High Throughput Screen* **3:** 219-234.

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59:** 407-450.

Hinnebusch AG. 2006. eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci* **31:** 553-562.

Hinnebusch AG, Dever TE, Asano K. 2007. Mechanism of Translation Initiation in the Yeast *Saccharomyces cerevisiae*. In: Matthews MB, Sonenberg N, Hershey JWB, eds. *Translational Control in Biology and Medicine*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 225-268.

Holbrook SR, Holbrook EL, Walukiewicz HE. 2001. Crystallization of RNA. *Cell Mol Life Sci* **58:** 234-243.

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

Honda M, Brown EA, Lemon SM. 1996. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2:** 955-968.

Huang HK, Yoon H, Hannig EM, Donahue TF. 1997. GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in Saccharomyces cerevisiae. *Genes Dev* **11:** 2396-2413.

Iizuka N, Najita L, Franzusoff A, 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.

Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, Auld DS. 2007. High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* **3:** 466-479.

Izumi RE, Das S, Barat B, Raychaudhuri S, Dasgupta A. 2004. A peptide from autoantigen La blocks poliovirus and hepatitis C virus cap-independent translation and reveals a single tyrosine critical for La RNA binding and translation stimulation. *J Virol* **78:** 3763-3776.

Jackson RJ. 1991. The ATP requirement for initiation of eukaryotic translation varies according to the mRNA species. *Eur J Biochem* **200:** 285-294.

Jackson RJ, Hellen CU, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11:** 113-127.

Jan E, Sarnow P. 2002. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J Mol Biol* **324:** 889-902.

Jang SK. 2006. Internal initiation: IRES elements of picornaviruses and hepatitis c virus. *Virus Res* **119:** 2-15.

Jarvis LM. 2010. The Waiting Game. *Chem Eng News* **88:** 12-17.

Jefferson EA, Seth PP, Robinson DE, Winter DK, Miyaji A, Osgood SA, Swayze EE, Risen LM. 2004. Biaryl guanidine inhibitors of in vitro HCV-IRES activity. *Bioorg Med Chem Lett* **14:** 5139-5143.

Ji H, Fraser CS, Yu Y, Leary J, Doudna JA. 2004. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc Natl Acad Sci* **101:** 16990-16995.

Jopling CL, Schutz S, Sarnow P. 2008. Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome. *Cell Host Microbe* **4:** 77-85.

Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309:** 1577-1581.

Kapp LD, Lorsch JR. 2004. GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J Mol Biol* **335:** 923-936.

Ke A, Doudna JA. 2004. Crystallization of RNA and RNA-protein complexes. *Methods* **34:** 408-414.

Keel AY, Rambo RP, Batey RT, Kieft JS. 2007. A general strategy to solve the phase problem in RNA crystallography. *Structure* **15:** 761-772.

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

Kieft JS, Zhou K, Jubin R, Doudna JA. 2001. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7:** 194-206.

Kieft JS, Zhou K, Jubin R, Murray MG, Lau JY, Doudna JA. 1999. The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *J Mol Biol* **292:** 513-529.

Kikuchi K, Umehara T, Fukuda K, Kuno A, Hasegawa T, Nishikawa S. 2005. A hepatitis C virus (HCV) internal ribosome entry site (IRES) domain III-IV-targeted aptamer inhibits translation by binding to an apical loop of domain IIId. *Nucleic Acids Res* **33:** 683-692.

Kolupaeva VG, Pestova TV, Hellen CU. 2000. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J Virol* **74:** 6242-6250.

Kolupaeva VG, Unbehaun A, Lomakin IB, Hellen CU, Pestova TV. 2005. Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA* **11:** 470-486.

Kozak M. 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15:** 1109-1123.

Kozak M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44:** 283-292.

Kozak M. 1990. Evaluation of the fidelity of initiation of translation in reticulocyte lysates from commercial sources. *Nucleic Acids Res* **18:** 2828.

Kozak M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* **266:** 19867-19870.

Kozak M. 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299:** 1-34.

Kozak M. 2005. A second look at cellular mRNA sequences said to function as internal ribosome entry sites. *Nucleic Acids Res* **33:** 6593-6602.

Kozak M, Shatkin AJ. 1976. Characterization of ribosome-protected fragments from reovirus messenger RNA. *J Biol Chem* **251:** 4259-4266.

Lanford RE, Hildebrandt-Eriksen ES, Petri A, Persson R, Lindow M, Munk ME, Kauppinen S, Orum H. 2010. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* **327:** 198-201.

Lavender CA, Ding F, Dokholyan NV, Weeks KM. 2010. Robust and generic RNA modeling using inferred constraints: a structure for the hepatitis C virus IRES pseudoknot domain. *Biochemistry* **49:** 4931-4933.

Legon S, Robertson HD. 1976. The binding of 125I-labelled rabbit globin messenger RNA to reticulocyte ribosomes. *J Mol Biol* **106:** 23-36.

Linder P. 2003. Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol Cell* **95:** 157-167.

Litovchick A, Szostak JW. 2008. Selection of cyclic peptide aptamers to HCV IRES RNA using mRNA display. *Proc Natl Acad Sci U S A* **105:** 15293-15298.

Litovchick A, Szostak JW. 2009. Retraction for Litovchick and Szostak. "Selection of cyclic peptide aptamers to HCV IRES RNA using mRNA display". *Proc Natl Acad Sci U S A* **106:** 7263.

Lloyd RE. 2006. Translational control by viral proteinases. *Virus Res* **119:** 76-88.

Locker N, Easton LE, Lukavsky PJ. 2007. HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J* **26:** 795-805.

Lomakin IB, Kolupaeva VG, Marintchev A, Wagner G, Pestova TV. 2003. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev* **17:** 2786-2797.

Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14:** 953-961.

Lorsch JR, Dever TE. 2010. Molecular view of 43 S complex formation and start site selection in eukaryotic translation initiation. *J Biol Chem* **285:** 21203-21207.

Lukavsky PJ, Kim I, Otto GA, Puglisi JD. 2003. Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* **10:** 1033-1038.

Lukavsky PJ, Otto GA, Lancaster AM, Sarnow P, Puglisi JD. 2000. Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nat Struct Biol* **7:** 1105-1110.

Ma J, Campbell A, Karlin S. 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J Bacteriol* **184:** 5733-5745.

Maag D, Algire MA, Lorsch JR. 2006. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *J Mol Biol* **356:** 724-737.

Maag D, Fekete CA, Gryczynski Z, Lorsch JR. 2005. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell* **17:** 265-275.

Maag D, Lorsch JR. 2003. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J Mol Biol* **330:** 917-924.

Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. 1997. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89:** 951-961.

Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. 1999. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol Cell* **3:** 707-716.

Marintchev A, Kolupaeva VG, Pestova TV, Wagner G. 2003. Mapping the binding interface between human eukaryotic initiation factors 1A and 5B: a new interaction between old partners. *Proc Natl Acad Sci U S A* **100:** 1535-1540.

Matthews MB, Sonenberg N, Hershey JWB. 2007. Origins and Principles of Translational Control. In: Matthews MB, Sonenberg N, Hershey JWB, eds. *Translational Control in Biology and Medicine*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 1-40.

McCaffrey AP, Ohashi K, Meuse L, Shen S, Lancaster AM, Lukavsky PJ, Sarnow P, Kay MA. 2002. Determinants of hepatitis C translational initiation in vitro, in cultured cells and mice. *Mol Ther* **5:** 676-684.

McHutchinson JG, Bartenschlager R, Patel K, Pawlotsky J-M. 2006. The face of future hepatitis C antiviral drug development: Recent biological and virologic advances and their translation to drug development and clinical practice. *J Hepatol* **44:** 411-421.

McHutchison JG, Patel K, Pockros P, Nyberg L, Pianko S, Yu RZ, Dorr FA, Kwoh TJ. 2006. A phase I trial of an antisense inhibitor of hepatitis C virus (ISIS 14803), administered to chronic hepatitis C patients. *J Hepatol* **44:** 88-96.

Meerovitch K, Pelletier J, Sonenberg N. 1989. A cellular protein that binds to the 5'-noncoding region of poliovirus RNA: implications for internal translation initiation. *Genes Dev* **3:** 1026-1034.

Meyer LJ, Milburn SC, Hershey JW. 1982. Immunochemical characterization of mammalian protein synthesis initiation factors. *Biochemistry* **21:** 4206-4212.

Miyamoto S, Patel P, Hershey JW. 2005. Changes in ribosomal binding activity of eIF3 correlate with increased translation rates during activation of T lymphocytes. *J Biol Chem* **280:** 28251-28264.

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

Nanda JS, Cheung YN, Takacs JE, Martin-Marcos P, Saini AK, Hinnebusch AG, Lorsch JR. 2009. eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *J Mol Biol* **394:** 268-285.

Nelson DR. 2009. Hepatitis C drug development at a crossroads. *Hepatology* **50:** 997-999.

Nielsen KH, Valasek L, Sykes C, Jivotovskaya A, Hinnebusch AG. 2006. Interaction of the RNP1 motif in PRT1 with HCR1 promotes 40S binding of eukaryotic initiation factor 3 in yeast. *Mol Cell Biol* **26:** 2984-2998.

Nielsen TE, Schreiber SL. 2008. Towards the optimal screening collection: a synthesis strategy. *Angew Chem Int Ed Engl* **47:** 48-56.

Nimjee SM, Rusconi CP, Sullenger BA. 2005. Aptamers: an emerging class of therapeutics. *Annu Rev Med* **56:** 555-583.

Novac O, Guenier AS, Pelletier J. 2004. Inhibitors of protein synthesis identified by a high throughput multiplexed translation screen. *Nucleic Acids Res* **32:** 902-915.

Ogle JM, Brodersen DE, Clemons WM, Jr., Tarry MJ, Carter AP, Ramakrishnan V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292:** 897-902.

Olsen DS, Savner EM, Mathew A, Zhang F, Krishnamoorthy T, Phan L, Hinnebusch AG. 2003. Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. *EMBO J* **22:** 193-204.

Otto GA, Puglisi JD. 2004. The pathway of HCV IRES-mediated translation initiation. *Cell* **119:** 369-380.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372:** 432-438.

Overington JP, Al-Lazikani B, Hopkins AL. 2006. How many drug targets are there? *Nat Rev Drug Discov* **5:** 993-996.

Parsons J, Castaldi MP, Dutta S, Dibrov SM, Wyles DL, Hermann T. 2009. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat Chem Biol* **5:** 823-825.

Passmore LA, Schmeing TM, Maag D, Applefield DJ, Acker MG, Algire MA, Lorsch JR, Ramakrishnan V. 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* **26:** 41-50.

Paulin FE, Campbell LE, O'Brien K, Loughlin J, Proud CG. 2001. Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr Biol* **11:** 55-59.

Paulsen RB, Seth PP, Swayze EE, Griffey RH, Skalicky JJ, Cheatham TE, 3rd, Davis DR. 2010. Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc Natl Acad Sci U S A* **107:** 7263-7268.

Pelletier J, Kaplan G, Racaniello VR, Sonenberg N. 1988. Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Mol Cell Biol* **8:** 1103-1112.

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

Pestova TV, Borukhov SI, Hellen CU. 1998a. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394:** 854-859.

Pestova TV, de Breyne S, Pisarev AV, Abaeva IS, Hellen CU. 2008. eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J* **27:** 1060-1072.

Pestova TV, Kolupaeva VG. 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* **16:** 2906-2922.

Pestova TV, Lomakin IB, Lee JH, Choi SK, Dever TE, Hellen CU. 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403:** 332-335.

Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU. 1998b. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev* **12:** 67-83.

Pilipenko EV, Pestova TV, Kolupaeva VG, Khitrina EV, Poperechnaya AN, Agol VI, Hellen CU. 2000. A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev* **14:** 2028-2045.

Pisarev AV, Hellen CU, Pestova TV. 2007. Recycling of eukaryotic posttermination ribosomal complexes. *Cell* **131:** 286-299.

Pisarev AV, Kolupaeva VG, Yusupov MM, Hellen CU, Pestova TV. 2008. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* **27:** 1609-1621.

Pleij CW, Rietveld K, Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* **13:** 1717-1731.

Prabhu R, Garry RF, Dash S. 2006. Small interfering RNA targeted to stem-loop II of the 5' untranslated region effectively inhibits expression of six HCV genotypes. *Virol J* **3:** 100.

Pudi R, Ramamurthy SS, Das S. 2005. A peptide derived from RNA recognition motif 2 of human la protein binds to hepatitis C virus internal ribosome entry site, prevents ribosomal assembly, and inhibits internal initiation of translation. *J Virol* **79:** 9842-9853.

Reynolds JE, Kaminski A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ. 1996. Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA* **2:** 867-878.

Reynolds JE, Kaminski A, Kettinen HJ, Grace K, Clarke BE, Carroll AR, Rowlands DJ, Jackson RJ. 1995. Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J* **14:** 6010-6020.

Rijnbrand R, Thiviyanathan V, Kaluarachchi K, Lemon SM, Gorenstein DG. 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.

Rijnbrand R, van der Straaten T, van Rijn PA, Spaan WJ, Bredenbeek PJ. 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.

Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L. 1992. Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol Microbiol* **6:** 1219-1229.

Robertson MP, Scott WG. 2008. A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallogr D Biol Crystallogr* **D64:** 738-744.

Rogers GW, Jr., Richter NJ, Lima WF, Merrick WC. 2001. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J Biol Chem* **276:** 30914-30922.

Romero-Lopez C, Diaz-Gonzalez R, Barroso-delJesus A, Berzal-Herranz A. 2009. Inhibition of hepatitis C virus replication and internal ribosome entry site-dependent translation by an RNA molecule. *J Gen Virol* **90:** 1659-1669.

Romero-Lopez C, Diaz-Gonzalez R, Berzal-Herranz A. 2007. Inhibition of hepatitis C virus internal ribosome entry site-mediated translation by an RNA targeting the conserved IIIf domain. *Cell Mol Life Sci* **64:** 2994-3006.

Ron D, Harding HP. 2007. eIF2α Phosphorylation in Cellular Stress Responses and Disease. In: Matthews MB, Sonenberg N, Hershey JWB, eds. *Translational Control in Biology and Medicine*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 345-368.

Roy S, Gupta N, Subramanian N, Mondal T, Banerjea AC, Das S. 2008. Sequence-specific cleavage of hepatitis C virus RNA by DNAzymes: inhibition of viral RNA translation and replication. *J Gen Virol* **89:** 1579-1586.

Rupert PB, Ferre-D'Amare AR. 2004. Crystallization of the hairpin ribozyme: illustrative protocols. *Methods Mol Biol* **252:** 303-311.

Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Schneider RJ, Sonenberg N. 2007. Translational Control in Cancer Development and Progression. In: Matthews MB, Sonenberg N, Hershey JWB, eds. *Translational Control in Biology and Medicine*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 401-432.

Schulman LH, Pelka H. 1984. Recognition of tRNAs by aminoacyl-tRNA synthetases: Escherichia coli tRNAMet and E. coli methionyl-tRNA synthetase. *Fed Proc* **43:** 2977-2980.

Searfoss AM, Wickner RB. 2000. 3' poly(A) is dispensable for translation. *Proc Natl Acad Sci U S A* **97:** 9133-9137.

Seth PP, Miyaji A, Jefferson EA, Sannes-Lowery KA, Osgood SA, Propp SS, Ranken R, Massire C, Sampath R, Ecker DJ, Swayze EE, Griffey RH. 2005. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J Med Chem* **48:** 7099-7102.

Shin BS, Maag D, Roll-Mecak A, Arefin MS, Burley SK, Lorsch JR, Dever TE. 2002. Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. *Cell* **111:** 1015-1025.

Shine J, Dalgarno L. 1974. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71:** 1342-1346.

Sicilio F, Peterson MD. 1961. Ratio errors in pseudo first order reactions. *J Chem Ed* **38:** 576-577.

Silvera D, Formenti SC, Schneider RJ. 2010. Translational control in cancer. *Nat Rev Cancer* **10:** 254-266.

Siridechadilok B, Fraser CS, Hall RJ, Doudna JA, Nogales E. 2005. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310:** 1513-1515.

Sivan G, Elroy-Stein O. 2008. Regulation of mRNA Translation during cellular division. *Cell Cycle* **7:** 741-744.

Sizova DV, Kolupaeva VG, Pestova TV, Shatsky IN, Hellen CU. 1998. Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J Virol* **72:** 4775-4782.

Sonenberg N, Hinnebusch AG. 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136:** 731-745.

Soto Rifo R, Ricci EP, Decimo D, Moncorge O, Ohlmann T. 2007. Back to basics: the untreated rabbit reticulocyte lysate as a competitive system to recapitulate cap/poly(A) synergy and the selective advantage of IRES-driven translation. *Nucleic Acids Res* **35:** e121.

Spahn CM, Kieft JS, Grassucci RA, Penczek PA, Zhou K, Doudna JA, Frank J. 2001. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291:** 1959-1962.

Studer SM, Feinberg JS, Joseph S. 2003. Rapid kinetic analysis of EF-G-dependent mRNA translocation in the ribosome. *J Mol Biol* **327:** 369-381.

Suzuki T, Aizaki H, Murakami K, Shoji I, Wakita T. 2007. Molecular biology of hepatitis C virus. *J Gastroenterol* **42:** 411-423.

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

Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ, Sonenberg N. 2001. The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7:** 382-394.

Tallet-Lopez B, Aldaz-Carroll L, Chabas S, Dausse E, Staedel C, Toulmé J. 2003. Antisense oligonucleotides targeted to the domain IIId of the hepatitis C virus IRES compete with 40S ribosomal subunit binding and prevent in vitro translation. *Nucleic Acids Res* **31:** 734-742.

Tarun SZ, Jr., Sachs AB. 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J* **15:** 7168-7177.

Terenin IM, Dmitriev SE, Andreev DE, Shatsky IN. 2008. Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat Struct Mol Biol* **15:** 836-841.

Thumma SC, Kratzke RA. 2007. Translational control: a target for cancer therapy. *Cancer Lett* **258:** 1-8.

Trachsel H, Erni B, Schreier MH, Staehelin T. 1977. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J Mol Biol* **116:** 755-767.

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

Unbehaun A, Borukhov SI, Hellen CU, Pestova TV. 2004. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of

codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* **18:** 3078-3093.

Unbehaun A, Marintchev A, Lomakin IB, Didenko T, Wagner G, Hellen CU, Pestova TV. 2007. Position of eukaryotic initiation factor eIF5B on the 80S ribosome mapped by directed hydroxyl radical probing. *EMBO J* **26:** 3109-3123.

Valasek L, Hasek J, Nielsen KH, Hinnebusch AG. 2001a. Dual function of eIF3j/Hcr1p in processing 20 S pre-rRNA and translation initiation. *J Biol Chem* **276:** 43351-43360.

Valasek L, Nielsen KH, Zhang F, Fekete CA, Hinnebusch AG. 2004. Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol Cell Biol* **24:** 9437-9455.

Valasek L, Phan L, Schoenfeld LW, Valaskova V, Hinnebusch AG. 2001b. Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *EMBO J* **20:** 891-904.

Vattem KM, Wek RC. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A* **101:** 11269-11274.

Wang C, Le SY, Ali N, 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 C, Sarnow P, Siddiqui A. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J Virol* **67:** 3338-3344.

Wang C, Sarnow P, Siddiqui A. 1994. A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *J Virol* **68:** 7301-7307.

Webster DP, Klenerman P, Collier J, Jeffery KJ. 2009. Development of novel treatments for hepatitis C. *Lancet Infect Dis* **9:** 108-117.

Williams DD, Price NT, Loughlin AJ, Proud CG. 2001. Characterization of the mammalian initiation factor eIF2B complex as a GDP dissociation stimulator protein. *J Biol Chem* **276:** 24697-24703.

Wilson JE, Powell MJ, Hoover SE, Sarnow P. 2000. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol Cell Biol* **20:** 4990-4999.

Wyatt JR, Puglisi JD, Tinoco I, Jr. 1989. RNA folding: pseudoknots, loops and bulges. *Bioessays* **11:** 100-106.

Yamasaki S, Anderson P. 2008. Reprogramming mRNA translation during stress. *Curr Opin Cell Biol* **20:** 222-226.

Yanagi M, Purcell RH, Emerson SU, Bukh J. 1997. Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc Natl Acad Sci U S A* **94:** 8738-8743.

Yu Y, Marintchev A, Kolupaeva VG, Unbehaun A, Veryasova T, Lai SC, Hong P, Wagner G, Hellen CU, Pestova TV. 2009. Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res* **37:** 5167-5182.

Yusupova GZ, Yusupov MM, Cate JH, Noller HF. 2001. The path of messenger RNA through the ribosome. *Cell* **106:** 233-241.

Zhang J, Yamada O, Ito T, Akiyama M, Hashimoto Y, Yoshida H, Makino R, Masago A, Uemura H, Araki H. 1999a. A single nucleotide insertion in the 5'-untranslated region of hepatitis C virus leads to enhanced cap-independent translation. *Virology* **261:** 263-270.

Zhang JH, Chung TD, Oldenburg KR. 1999b. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4:** 67-73.

Zhao Q, Han Q, Kissinger CR, Hermann T, Thompson PA. 2008. Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* **64:** 436-443.

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