STUDY OF TRANSLATIONAL CONTROL USING CELL-FREE TRANSLATION SYSTEMS AND PRIMER EXTENSION INHIBITION ASSAYS

Cheng Wu

B.S., Fudan University, 1999

A dissertation submitted to the faculty of the

OGI School of Science & Engineering

At Oregon Health & Science University

In partial fulfillment of the

Requirements for the degree

Doctor of Philosophy

In

Biochemistry and Molecular Biology

March 2008

The thesis “Study of translational control using cell-free translation systems and primer extension inhibition assays” by Cheng Wu has been examined and approved by the following Examination Committee:

Matthew S. Sachs, Ph.D., Thesis Advisor
Professor

Michiko Nakano, Ph.D.
Research Associate Professor

Peter Zuber, Ph.D.
Professor

Michael Freitag, Ph.D.
Assistant Professor
Oregon State University

ACKNOWLEDGEMENT

I wish to express my sincere appreciation and gratitude towards my thesis advisor, Dr. Matthew S. Sachs, for his inspiration, enthusiasm, guidance and suggestions, which have made my Ph.D. studies full of excitement and reward. Special thanks go to the dissertation committee members, Drs. Michael Freitag, Michiko Nakano, and Peter Zuber. I am indebted to Michael for his commitment to my studies and for his guidance as I prepare for my thesis. Thanks to Drs. Peter Zuber and Michiko Nakano for their guidance, support, encouragement and insightful opinions throughout my Ph.D. studies, and to their lab members for their support and technical assistance, and generously sharing equipments and reagents.

Sincere thanks to Dr. Peng Fang for his guidance with my research project and for his invaluable friendship and unconditional help. I would like to thank Drs. Anthony Gaba, Christina Spevak, who have kindly assisted me with experiments, and to thank Drs. Meray Baştürkmen, Heather Hood, Vladimir Zeenko, and other former and current lab members Paul ReFalo, Julie Michelle, Ana Luiza Ramos-Crawford, Kanan Lathia, Yanzhen Cui, Seth McEldowney, Julius Mopper, and Richard Feinberg, for their support and discussions. Very special thanks to department moms Nancy Christie and Terrie Hadfield for all their help and kindness throughout my graduate studies.

I would like to give special thanks to my wife Dr. Ying Zhang for her invaluable opinions on my experiments and taking good care of me for all these years.

I am forever indebted to my family especially my mother and father for their continuous love and encouragement, advice and support, which without, this thesis would not have been possible. For all their efforts, patience, and understanding, I dedicate this thesis to my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AAP: arginine attenuator peptide

AdoMet: *S*-adenosyl-L-methionine

AdoMetDC: *S*-adenosylmethionine decarboxylase

ATF4: activating transcription factor 4

CaMV: cauliflower mosaic virus

C/EBPs: CCAAT/enhancer binding proteins

CPS-A: arginine-specific carbamoyl phosphate synthetase

CPE: cytoplasmic polyadenylation element

CPEB: CPE-binding protein

dUNR: upstream of N-ras

FMDV: foot-and-mouth disease virus

FRAP: FKBP-rapamycin-associated protein

GCN2: general control non-derepressible 2

HRI: heme-regulated inhibitor

IRES: internal ribosome entry site

MAPK: mitogen-activated protein kinase

MFC: multifactor complex

Mnk: MAP-kinase-interacting kinase

MSL-2: male specific lethal 2

mTOR: mammalian target of rapamycin

NMD: nonsense-mediated mRNA decay

ORF: open reading frame

OxLDL: oxidized low-density lipoprotein

Pabp: poly(A)-binding protein

PERK: pancreatic endoplasmic reticulum eIF2α kinase

PIC: pre-initiation complex

PI3K: phosphatidylinositol 3-kinase

PKC: proteins kinase C

PKR: protein kinase RNA-activated

PTC: peptidyl transferase center

RACK1: Receptors for Activated C-Kinases

RISC: RNA-induced silencing complex

SAPK: stress-activated protein kinase

SXL: sexual lethal

TC: ternary complex

TDE: translational derepression element

TOP: 5′-terminal oligopyrimidine

uORF: upstream open reading frame

Upf: upframeshift suppressor

UTR: untranslated region;

ABSTRACT

STUDY OF TRANSLATIONAL CONTROL USING CELL-FREE TRANSLATION SYSTEMS AND PRIMER EXTENSION INHIBITION ASSAYS

Cheng Wu

Ph.D., OGI School of Science & Engineering

Oregon Health & Science University

March 2008

Thesis Advisor: Dr. Matthew S. Sachs

The *Neurospora crassa arg-2* gene encodes the small subunit of carbamoyl phosphate synthetase, the first enzyme in fungal arginine (Arg) biosynthesis, which is negatively feed-back regulated by the cellular arginine level. This regulation is mediated by a 24-residue peptide, known as the arginine attenuator peptide (AAP), specified by an evolutionarily conserved upstream open reading frame (uORF) in *arg-2* mRNA. Previous studies with cell-free translation systems from *N. crassa* and *Saccharomyces cerevisiae* have shown that ribosomes synthesizing the *N. crassa arg-2* and the homologous *S. cerevisiae CPA1* uORF-encoded AAPs stall at the uORF termination codon in the presence of high levels of Arg, blocking ribosomes from scanning to the downstream initiation codon.

The use of cell-free translation systems from fungi and primer extension inhibition assays has provided key insights into the mechanism underlying the AAP-mediated ribosome stalling. Here, I present results suggesting that (i) the nascent AAP can transiently stall ribosomes when placed either at the N-terminus or internally

within a polypeptide; (ii) ribosomes appear to resume translation after release from the stalling; and (iii) the nascent AAP is in its peptidyl-tRNA form while protruding into the ribosome exit tunnel when the ribosome stalls after synthesizing the AAP.

The 18-residue 2A peptide of foot-and-mouth disease virus (FMDV) mediates a co-translational cleavage at its C-terminus resulting in separated N- and C-terminal products. The 2A reaction was recapitulated in the *N. crassa* cell-free translation systems. The majority of N-terminal products are released peptides when the ribosome pause occurs instead of being linked to the tRNA. Primer extension inhibition assays revealed a ribosome pause at the C-terminus of 2A. These results suggest that the nascent 2A peptide adopts a confirmation that stalls the translating ribosome and promotes the hydrolysis of the peptidyl-tRNA bond.

Salmonella enterica alternatively expresses two distinct types of flagellin proteins, FljB and FliC. FljA is co-expressed with FljB to inhibit FliC expression. Mutations that allow synthesis of FliC in the presence of FljA are located in the 5′-UTR of the *fliC* mRNA, which suggests that FljA binds to this region to inhibit translation. A modified primer extension inhibition assay was used to map the position of bound FljA on *fliC* mRNA, revealing that the binding site overlaps the ribosome binding site. When FljA binds to the *fliC* mRNA, it blocks ribosomes access to the translation initiation site, thereby inhibiting the synthesis of FliC.

The successful use of cell-free translation systems and primer extension inhibition assays in studying these three cases proves they are powerful tools that enable biochemical determination of factors and mechanisms contributing to translational processes.

CHAPTER 1 INTRODUCTION

Proteins constitute most of the dry mass of a cell, and play crucial roles in virtually all biological processes. They catalyze nearly all chemical reactions in biological systems, act as transducers of motion, signal integrators, material transporters, and provide mechanical support, immune protection, and other functions essential for life. Thus, the protein synthesis is essential to cell maintenance, growth and development. The process of translation synthesizes proteins according to instructions given by messenger RNA (mRNA) templates. This process takes place in three stages: initiation, elongation, and termination; it is regulated at multiple levels to finely control gene expression. This chapter summarizes the mechanism of eukaryotic translation and the mechanisms of regulating translation. It also summarizes the mechanisms by which upstream open reading frames (uORFs) and nascent peptides regulate translation. uORFs are found in ~10% of eukaryotic mRNA transcripts, and are very common in proto-oncogenes that cause cancers. This makes the study of uORFs crucial. Two examples of genes containing uORFs that will be discussed in the following chapters are the *Neurospora crassa arg-2* gene and the homologous *Saccharomyces cerevisiae CPA1* gene. They encode an evolutionarily conserved arginine attenuator peptide (AAP) that regulates translation in response to arginine. The nascent AAP can also function as an internal domain within a long polypeptide, being able to arrest the translating ribosome near the AAP's C-terminus through a novel mechanism that resembles other instances of nascent peptide regulated control in eukaryotic and prokaryotic genes.

1.1 The Molecular Mechanism of Eukaryotic Translation

Eukaryotic mRNA contains multiple structural elements that influence translation (Figure 1.1) (Mignone *et al.*, 2002). The most distinctive elements are the $m^7G(5')ppp(5')N$ cap structure at the 5′-end and the polyadenylate [poly(A)] tail at the

3′-end. The cap structure is essential for efficient translation (Hershey & Merrick, 2000; Kozak, 1999; Sachs, 2000) and maintaining mRNA stability (Decker and Parker, 1994; Jacobson and Peltz, 1996). It also functions in mRNA transport and localization processes (Cougot *et al.*, 2004; Lewis and Izaurralde, 1997; Lewis and Tollervey, 2000). The poly(A) tails, with a typical length of approximately 70 nucleotides in yeast or ~200 – 250 nucleotides in mammals, are found in virtually all eukaryotic mRNAs (Jacobson, 1996). The majority of eukaryotic translation appears to be cap-dependent (Kozak, 1989b; Kozak, 1999). The cap structure and poly(A) tail act as translational enhancers by interacting with each other through specific binding proteins and initiation factors. This results in the functional circularization of mRNA and the synergistic stimulation of translation (Gallie, 1991; Iizuka *et al.*, 1994; Munroe and Jacobson, 1990; Preiss and Hentze, 1998; Tarun and Sachs, 1995; Tarun and Sachs, 1996). Other regulating elements can be found in the 5′-untranslated region (5′-UTR), the open reading frame (ORF) region(s), and the 3′-untranslated region (3′-UTR) (Hentze *et al.*, 2007).

1.1.1 Initiation

The initiation of translation is a multi-step process that includes the recruitment of ribosomal subunits to the mRNA to form the initiation complex (Figure 1.2) (Asano and Sachs, 2007). First, eIF2, GTP, and Met-tRNA$_i^{Met}$ are assembled into a ternary complex (TC). Second, the 40S ribosomal subunit is recruited to this complex to form a 43S pre-initiation complex (PIC). This 43S PIC is then loaded onto the mRNA cap structure and starts to scan the mRNA until it locates the initiation codon. Finally, the large (60S) ribosomal subunit joins the 40S subunit to form the 80S elongating ribosome, thus concluding the initiation stage (Kapp and Lorsch, 2004; Pestova *et al.*, 2007). The process of initiation requires more than 10 initiation factors (Asano and Sachs, 2007; Kapp and Lorsch, 2004; Pestova *et al.*, 2007). The molecular mechanism is discussed further below.

1.1.1.1 Formation of the TC

Initiation factor eIF2 consists of three subunits: α, β, and γ. The α subunit plays

an important role in translational regulation as discussed further in section *1.2.1.1*. The β subunit may be involved in the binding of Met-tRNA$_i^{Met}$ to eIF2 (Huang *et al.*, 1997). Three lysine-repeat tracts in the N-terminal region of the eIF2β subunit are predicted to interact with eIF5, eIF2B, and RNA (Asano *et al.*, 1999; Das and Maitra, 2000; Laurino *et al.*, 1999). The γ subunit, which acts as a GTPase, directly binds to both GTP and Met-tRNA$_i^{Met}$ (Dorris *et al.*, 1995; Erickson and Hannig, 1996; Gaspar *et al.*, 1994; Harashima and Hinnebusch, 1986; Huang *et al.*, 1997; Naranda *et al.*, 1995). eIF2 specifically binds to initiator tRNA by recognizing the methionyl residue and an A1:U72 pair at the end of its acceptor stem (Hershey and Merrick, 2000). GTP is required for the formation of the active eIF2·Met-tRNA$_i^{Met}$·GTP complex. After the initiation codon is recognized, eIF2 will be released from the complex as an inactive eIF2·GDP form, which cannot bind Met-tRNA$_i^{Met}$ until GDP is replaced with GTP by eIF2B (Hershey and Merrick, 2000; Hinnebusch, 2000; Kimball, 1999). The phosphorylation of a conserved serine residue at position 51 on the eIF2α subunit will inhibit the exchange of bound GDP to GTP catalyzed by eIF2B, thus reducing eIF2 capacity to function in subsequent initiation events (Dever *et al.*, 1992; Hershey, 1991).

1.1.1.2 Formation of the 43S PIC

The TC binds to the 40S ribosomal subunit with the assistance of eIF1, eIF1A and eIF3. Both eIF1A and eIF3 can promote and enhance the 43S PIC formation (Algire *et al.*, 2002; Chaudhuri *et al.*, 1999; Kolupaeva *et al.*, 2005; Maag *et al.*, 2005; Majumdar *et al.*, 2003). A stable multifactor complex (MFC) consisting of TC, eIF1, eIF3, and eIF5 has been isolated in yeast (Asano *et al.*, 2000). The MFC may work as a preformed unit to facilitate the cooperative binding of all components to the 40S subunit or may form after each component independently binds to the 40S subunit to stabilize the 43S PIC assembly (Hinnebusch, 2006). The MFC can also compete with the 60S subunit to prevent premature association (Spahn *et al.*, 2001). As the largest and most complex initiation factor, eIF3 is involved in many steps of the initiation pathway and is important for PIC binding to mRNA and scanning the mRNA for AUG recognition (Hinnebusch, 2006; Jivotovskaya *et al.*, 2006). With the aminoacylated Met-tRNA$_i^{Met}$ positioned at the P (peptidyl) site of the 40S subunit (Benne *et al.*, 1978), the PIC is ready for the entry

of mRNA and the recruitment of large 60S ribosomal subunit.

1.1.1.3 Association of the 43S PIC with mRNA

The 43S PIC binds mRNA through the 5′-cap structure. Several initiation factors are involved in this process. eIF4F is a heterotrimeric complex consisting of eIF4A, eIF4E, and eIF4G (Grifo *et al.*, 1983). eIF4G serves as a scaffold protein, allowing multiple initiation factors to assemble at the 5′-cap of mRNA and promote its recruitment by the 43S PIC (Hinton *et al.*, 2007; Prevot *et al.*, 2003). eIF4G has multiple binding domains for poly(A) binding protein (Pab1p), and eIF4E (Imataka *et al.*, 1998; Lamphear *et al.*, 1995; Mader *et al.*, 1995). eIF4E recognizes and binds the cap structure (Sonenberg *et al.*, 1978). The association between the cap structure, eIF4E, eIF4G, Pab1p, and poly(A) tail leads to the circularization of mRNA and facilitates the association of the 43S PIC with mRNA (Preiss and Hentze, 1998; Sachs, 2000). eIF4G also has binding sites for eIF3 and eIF4A (Korneeva *et al.*, 2001; Lamphear *et al.*, 1995; Marcotrigiano *et al.*, 2001). eIF4A is an ATP-dependent RNA helicase that unwinds the RNA duplex and removes any secondary structure that may obstruct the ribosome loading onto the cap structure (Preiss and Hentze, 2003; Rozen *et al.*, 1990). The helicase activity is increased when eIF4A is present in the eIF4F complex (Pestova and Kolupaeva, 2002; Rozen *et al.*, 1990), especially when it binds to the eIF4G (Korneeva *et al.*, 2005), with the additional assistance of RNA-binding proteins eIF4B and eIF4H (Rogers *et al.*, 2001). The 43S PIC is recruited to mRNA *via* the direct interaction between eIF3 of the 43S PIC and the eIF4G (Lamphear *et al.*, 1995). mRNA lacking its 3′-end will not be translated efficiently. This “closed-loop” model may provide a guard against the synthesis of truncated proteins that could be harmful to the cell (Kahvejian *et al.*, 2001).

1.1.1.4 Recognition of AUG Codon and Assembly of the 80S Ribosome

Cap-dependent scanning is the most common mechanism for translation initiation on capped and polyadenylated mRNAs (Hershey and Merrick, 2000; Kozak, 1999). After the 43S PIC loads onto the 5′-cap structure of mRNA, it starts to move in a 5′ – 3′ direction until it finds an initiation codon (typically an AUG triplet). Multiple ATP

molecules are hydrolyzed during the scanning process to enable eIF4A to remove any secondary structures (Kozak, 1980). Typically, the AUG codon closest to 5′-cap structure will be selected as the primary initiation site. However, the nucleotide sequence flanking the AUG codon may highly influence this selection (Kozak, 1989a; Kozak, 2002). An AUG codon with highly conserved purines, usually an A at the -3 position and a G at the +4 position of its context is favored for efficient initiation. However, if a pyrimidine occupies that -3 position, the scanning 43S PIC usually bypasses the AUG codon and selects a downstream AUG codon as the initiation site (Kozak, 1989b). The recognition of the AUG codon is through the anticodon-codon interaction between the Met-$tRNA_i^{Met}$ and mRNA, and requires the presence of several initiation factors: eIF1, eIF1A, eIF2, and eIF5 (Asano *et al.*, 2000; Donahue, 2000; Fekete *et al.*, 2005; Hershey and Merrick, 2000; Pestova *et al.*, 1998a; Pestova and Hellen, 2000). Once the AUG codon is selected for initiation, eIF5 triggers the hydrolysis of the GTP associated with eIF2γ (Chakravarti and Maitra, 1993; Huang *et al.*, 1997; Merrick and Hershey, 1996). Recent evidence suggests that phosphate (P_i) release is more important for AUG recognition than the hydrolysis event. Hydrolysis can occur during scanning but eIF1 prevents P_i release from eIF2·GDP-P_i at non-AUG codons (Algire *et al.*, 2005). AUG recognition triggers a conformational change in the PIC which eventually releases eIF1 and P_i (Cheung *et al.*, 2007). eIF2·GDP and other initiation factors are also released from the complex at this point (Singh *et al.*, 2006). A second GTPase, eIF5B, with GTP bound to it, stimulates the joining of the 60S ribosomal subunit to the 40S subunit to form the 80S ribosome (Pestova *et al.*, 2000). Once the assembly is completed, eIF5B hydrolyzes the GTP and releases from the ribosome (Lee *et al.*, 2002; Pestova *et al.*, 2000; Shin *et al.*, 2002).

1.1.1.5 Internal Initiation

Cap-dependent scanning is the predominant pathway for eukaryotic translation initiation. But there are alternative mechanisms to accomplish this. In some cases, the ribosome can be recruited to a specific region in the 5′-UTR of mRNA, known as internal ribosome entry site (IRES), to initiate translation (Prats and Prats, 2002). Internal initiation was first discovered in picornaviral mRNA (Jang *et al.*, 1988; Pelletier and

Sonenberg, 1988). A substantial number of genes are IRES-controlled, including those specifying viral proteins, oncogenes, growth factors, transporters and activators of apoptosis (Doudna and Sarnow, 2007; Elroy-Stein and Merrick, 2007; Hellen and Sarnow, 2001; Holcik, 2004; Holcik and Sonenberg, 2005; Kozak, 2003; Lewis and Holcik, 2005; Martinez-Salas *et al.*, 2002; Stoneley and Willis, 2004; Vagner *et al.*, 2001).

1.1.2 Elongation

The elongation cycle starts with the insertion of an aminoacyl-tRNA into the empty A (aminoacyl) site on the ribosome, followed by the formation of a peptide bond (Moore and Steitz, 2003). Finally, a GTP-driven process translocates peptidyl-tRNA from the A to the P site, leaving the A site vacant to accept the next aminoacyl-tRNA (Wintermeyer *et al.*, 2001). Deacylated tRNA moves to the E (exit) site and leaves the ribosome from there. Most elongation factors and the ribosome core are highly conserved in all organisms (Doudna and Rath, 2002; Spahn *et al.*, 2001), which suggests that the mechanism of elongation is essentially the same in prokaryotes, archaea, and eukaryotes (Ramakrishnan, 2002; Spahn *et al.*, 2001).

1.1.2.1 EF-Tu/eEF1A

Cognate aminoacyl-tRNA is delivered to the empty A site by elongation factor Tu (EF-Tu) in prokaryotes or eEF1A in eukaryotes (Moldave, 1985). The binding of EF-Tu/eEF1A to aminoacyl-tRNA is responsible for the protection of the activated ester bond from hydrolysis by water. Like eIF2, EF-Tu/eEF1A has GTPase activity and contains a bound GTP. Once the aminoacyl-tRNA is positioned in the A site, correct anticodon-codon recognition alters the conformation of the small ribosomal subunit, which in turn triggers the hydrolysis of GTP (Ogle *et al.*, 2001; Pape *et al.*, 1998; Rodnina and Wintermeyer, 2001). The GDP form of EF-Tu/eEF1A dissociates from the ribosome and the aminoacyl-tRNA. A second elongation factor, EF-Ts/eEF1B recycles EF-Tu/eEF1A back to its active form.

Translation is a complicated process with high speed and high fidelity. The error frequency is about 10^{-4} (Kurland, 1992). Since there is no way to replace the incorrect

amino acid residue once the peptide bond is formed, the incoming aminoacyl-tRNA must be carefully selected. Both cognate and non-cognate aminoacyl-tRNA can be delivered to the A site by EF-Tu/eEF1A·GTP. Proofreading occurs both before and after hydrolysis of GTP bound to EF-Tu/eEF1A. Only cognate aminoacyl-tRNA interacts with mRNA strongly enough in both states to stay in the complex, whereas the non-cognate charged-tRNAs leave in one of these two intervals (Thompson *et al.*, 1986). The allosteric linkage between the E and A sites also improves the proper selection of aminoacyl-tRNAs in the decoding process, as discussed in section 1.1.2.2.

1.1.2.2 Peptidyl Transferase and Translocation

A peptide bond is formed between the peptidyl-tRNA in the P site and the aminoacyl-tRNA in the A site. This reaction is catalyzed by the peptidyl transferase center (PTC) within the large ribosomal subunit (Moore and Steitz, 2003). Upon the release of EF-Tu/eEF1A·GDP, the CCA end of the aminoacyl-tRNA interacts with the PTC. The activated peptidyl unit is transferred to the amino group of the aminoacyl-tRNA. Peptide-bond formation changes the interaction of both tRNAs with the large ribosomal subunit but not the small subunit. The deacylated tRNA now occupies the E site of the large subunit with its acceptor end, while its anticodon stays in the P site of the small subunit (Green and Noller, 1997). A GTP-driven process now translocates the deacylated tRNA to the E site and peptidyl-tRNA to the P site (Moazed and Noller, 1989; Rodnina *et al.*, 1997; Wintermeyer *et al.*, 2001). mRNA moves a distance of three nucleotides so that the next codon is ready to interact with the incoming aminoacyl-tRNA. There is evidence indicating that when a deacylated tRNA occupies the E site, only cognate aminoacyl-tRNA inserted into the A site is able to induce a conformational change in the ribosome to release that deacylated tRNA (Nierhaus, 1990; Pape *et al.*, 1999).

1.1.2.3 EF-G/eEF2 and eEF3

Translocation is a GTP-consuming process and requires prokaryotic elongation factor G (EF-G) or eukaryotic eEF2. EF-G/eEF2, like eIF2 and EF-Tu/eEF1A, is a

GTPase cycling between GTP and GDP forms (Kaziro, 1978). Hydrolysis of the GTP bound to EF-G provides driving force not only for translocation but also for the release of EF-G from the ribosome (Rodnina *et al.*, 1997). A two-step model has been proposed recently in which the eEF2/EF-G binding induces the ratcheting motion of the 40S subunit, followedy by GTP hydrolysis that uncouples the mRNA-tRNA complex from the PTC so translocation of the mRNA-tRNA moiety may be completed by a head rotation of the 40S subunit (Taylor *et al.*, 2007).

A distinct eukaryotic elongation factor, eEF3, is found exclusively in fungi and is essential for yeast viability (Qin *et al.*, 1990; Skogerson and Wakatama, 1976). Evidence shows that eEF3 is a ribosome-dependent ATPase (Dasmahapatra and Chakraburtty, 1981) and is responsible for the release of deacylated tRNA from E site (Andersen *et al.*, 2006). eEF3 is mostly associated with polysomes and interacts with eEF1A (Anand *et al.*, 2003; Kovalchuke *et al.*, 1998).

1.1.3 Termination

When the elongating ribosome encounters a stop codon (UAA, UAG and UGA) in its A site, the elongation phase ends and the termination process is triggered. Normal cells do not have tRNAs with anticodons complementary to stop codons with the exception of selenocysteinyl-tRNA (Stortchevoi, 2006). Eukaryotic class I release factor eRF1 is responsible for recognizing the stop codon and promoting the cleavage of the ester bond between the peptidyl and tRNA, thus releasing the nascent peptide (Arkov *et al.*, 1998; Arkov *et al.*, 2002; Caskey *et al.*, 1971; Seit-Nebi *et al.*, 2001; Zavialov *et al.*, 2002). The sequence surrounding the termination codon also has an effect on termination efficiency (Bonetti *et al.*, 1995; Major *et al.*, 1996; McCaughan *et al.*, 1995). Multiple factors involved in the nonsense-mediated mRNA decay (NMD) pathway may also affect termination efficiency (Leeds *et al.*, 1992; Weng *et al.*, 1996a; Weng *et al.*, 1996b).

1.1.3.1 eRF1 and eRF3

In prokaryotes, there are two class I release factors (RFs) that each recognizes

different stop codons. RF1 recognizes UAA and UAG; and RF2 recognizes UAA and UGA (Scolnick *et al.*, 1968). Eukaryotes only have a single class I release factor, eRF1, which can recognize all three stop codons and promote hydrolysis of peptidyl-tRNA (Dontsova *et al.*, 2000; Frolova *et al.*, 1994; Konecki *et al.*, 1977). Studies on crystal structures reveal no significant similarity between eRF1 and RF2 (Song *et al.*, 2000; Vestergaard *et al.*, 2001), but all three release factors resemble tRNA. This may allow the release factors to act as tRNA to recognize the stop codon (Song *et al.*, 2000). A well conserved GGQ motif is found in all three release factors (Frolova *et al.*, 1999). This motif mimics the CCA-end at aminoacylated tRNA and is required for the peptidyl-tRNA hydrolysis (Song *et al.*, 2000). The activity of class I release factors requires class II release factor (Frolova *et al.*, 1996; Mikuni *et al.*, 1994; Stansfield *et al.*, 1995; Zhouravleva *et al.*, 1995). eRF3 is the only class II release factor in eukaryotes and functions as a GTPase. The hydrolysis of GTP bound to eRF3 releases eRF1 and eRF3 from the ribosome following the peptidyl-tRNA hydrolysis.

The insight to eRF3 function was first provided by studies on eRF3's prokaryotic counterpart RF3 (Zavialov *et al.*, 2001). RF3 binds to the posttermination ribosome in the form of RF3·GDP. The RF1/RF2-ribosome complex serves as guanine nucleotide exchange factor (GEF) for RF3·GDP. Once RFs 1 and 2 trigger peptidyl-tRNA hydrolysis resulting in the peptide release, there is a conformational change that allows RF3·GDP to convert to RF3·GTP. This change makes RF3 now have higher affinity for the ribosome than RFs 1 and 2, causing RFs 1 and 2 to dissociate from the ribosome. GTP is hydrolyzed by RF3. RF3·GTP converts back to RF3·GDP and dissociates from the ribosome for another cycle. Recent evidence shows that GTP hydrolysis by eRF3 increases the fidelity of eukaryotic termination by eRF1, and is required to permit subsequent hydrolysis of peptidyl-tRNA and efficient release of the nascent peptide (Alkalaeva *et al.*, 2006; Salas-Marco and Bedwell, 2004).

1.1.3.2 Hydrolysis of Peptidyl-tRNA

The hydrolysis of peptidyl-tRNA takes place within the large subunit at the peptidyl transferase center (PTC) of the ribosome. Water is required for hydrolysis but is excluded during peptide bond formation in elongation phase. The terminal glutamine

residue in the RF1 GGQ motif may be responsible for coordinating a water molecule and may mediate a nucleophilic attack on the ester bond of peptidyl-tRNA in the P site that results in hydrolysis (Frolova *et al.*, 1999; Seit-Nebi *et al.*, 2001; Song *et al.*, 2000). Never the less, changing the glutamine to a glycine does not affect the efficiency of peptidyl-tRNA hydrolysis (Seit-Nebi *et al.*, 2001; Song *et al.*, 2000). Deacylated tRNA or a CCA trinucleotide in the A site can induce peptidyl-tRNA hydrolysis as efficiently as *E. coli* RF1/RF2 in the presence of 20% ethanol (Zavialov *et al.*, 2002). These results suggest that the GGQ motif may conduct peptidyl-tRNA hydrolysis not only by providing a channel for water but by activating the PTC similarly to tRNA activating the PTC during elongation.

1.1.4 Recycling of Ribosomal Subunits

After one round of translation, the 80S ribosome must be dissociated into free 40S and 60S subunits so that the next round of translation can be initiated. Since the physiological conditions favor the association of the subunits, help from some initiation factors is required for dissociation (Hershey and Merrick, 2000). The mechanism of ribosome recycling in eukaryotes is still poorly understood. eIF1A and eIF3 are both involved in the formation of the 43S PIC (Algire *et al.*, 2002; Chaudhuri *et al.*, 1999; Kolupaeva *et al.*, 2005; Maag *et al.*, 2005; Majumdar *et al.*, 2003). Their binding to the 40S subunit will introduce a conformational change thus promoting the dissociation of the 80S ribosome and preventing re-association of 40S and 60S subunits (Srivastava *et al.*, 1992). eIF6 is thought to bind to the 60S subunit, preventing it from re-associating with the 40S subunit (Raychaudhuri *et al.*, 1984; Russell and Spremulli, 1979). There is evidence indicating that eIF6 is involved in ribosome biogenesis; its depletion impairs the biogenesis of the 60S subunit (Basu *et al.*, 2001; Si and Maitra, 1999). Phosphorylation of eIF6 through a RACK1-PKC pathway releases eIF6 from the 60S subunit, thereby allowing the 60S subunit to join to the 40S subunit (Ceci *et al.*, 2003).

It has been shown that eRF3 is connected to the poly(A) tail by interacting with Pab1p (Hoshino *et al.*, 1999). Disrupting this interaction inhibits translation (Uchida *et al.*, 2002). Based on the closed-loop model, it is proposed that the 40S subunit is shuttled back to the 5′-end of mRNA instead of being released to the cytoplasm during

the termination and recycling steps to facilitate reinitiation of translation.

1.2 Translational Control

Gene expression is controlled in all organisms. In eukaryotes, some regulation is critical for processes including cell growth, proliferation, and development. The control of gene expression can be accomplished at many levels, including the transcriptional and posttranscriptional level. Translational control is one major posttranscriptional regulatory pathway, providing flexible and reversible regulation over protein synthesis in the absence of transcriptional control (Mathews *et al.*, 2007). Regulation can be global or local. Global control regulates the translation of most mRNAs in the cell, and is typically achieved through the modification, especially phosphorylation, of components in the translation machinery. mRNA-specific control only regulates the translation of mRNAs in a specific group, while general protein synthesis is not affected. Regulatory functions are provided by *cis*-acting elements in the mRNA. Since initiation is the first step of translation and is rate-limiting (Mathews *et al.*, 2007), control of initiation is the predominant pathway to regulate translation.

1.2.1 Global Control of Translation

The overall rate of protein biosynthesis in cells is tightly controlled in response to a variety of signals including nutrient availability, energy levels and viral infection. Global control is achieved by regulating the activity of key components in the translation machinery. Most translation factors are phosphoproteins whose activity can be turned on and off through phosphorylation and dephosphorylation (Hershey, 1989; Mathews *et al.*, 2007). This turns out to be the most common pathway of regulating translation in eukaryotic cells. There are several other mechanisms used to regulate translation, e.g. during viral infections.

1.2.1.1 Phosphorylation of eIF2

The role of eIF2 in initiation has been discussed in section *1.1.1.1*. eIF2 is part

of the TC, selects the initiator tRNA and delivers it to the small ribosomal subunit. eIF2 cycles between the GTP and GDP form. Once translation is initiated, hydrolysis of GTP will release eIF2 from the ribosome in the inactive form. This has to be converted to the active form to participate the next round of initiation. The rate of this conversion determines the rate of initiation (Hinnebusch, 2000).

The conversion of eIF2·GDP to eIF2·GTP is catalyzed by the exchange factor eIF2B. Phosphorylation of Ser51 on the eIF2α subunit converts eIF2·GDP from a substrate to an inhibitor of eIF2B (Hinnebusch, 2000). Numerous eIF2α kinases have been identified (Dever, 2002; Dever *et al.*, 2007), and each is activated by different factors. The mammalian eIF2α kinase PKR is activated by double-stranded RNA, and is essential for the antiviral response (Kaufman, 2000). The mammalian endoplasmic reticulum (ER) resident kinase PERK is activated to relieve ER stress (Ron and Harding, 2000; Ron and Harding, 2007). The heme-regulated inhibitor (HRI) responds to iron deficiency and heme depletion (Chen, 2000). GCN2 (general control non-derepressible 2) is activated by amino acid starvation (Hinnebusch, 2000).

Studies in budding yeast have shown that the activation of GCN2 will activate the expression of GCN4, an activator of amino acid biosynthesis, in a mechanism dependent on the presence of upstream open reading frames (uORFs) (Geballe and Sachs, 2000; Hinnebusch, 1996; Morris and Geballe, 2000), as discussed further below.

1.2.1.2 eIF4E and 4E-BPs

eIF4E is a component of the heterotrimeric eIF4F complex. Its interaction with eIF3 and the 5′-cap structure recruits the 43S PIC to the mRNA. The activity of eIF4E is controlled at several levels. eIF4E mRNA levels are increased in response to growth factors. The transcription of eIF4E mRNA is activated by Myc (Schmidt, 2004). Overexpression of eIF4E leads to enhanced translation of many highly structured mRNAs involved in cell growth, proliferation and survival (Mamane *et al.*, 2004). eIF4E has been defined as a *bona fide* oncogene(Lazaris-Karatzas and Sonenberg, 1992), a potential target for anticancer therapies (Avdulov *et al.*, 2004; Oridate *et al.*, 2005).

Phosphorylation of eIF4E stimulated by hormones, growth factors, cytokines, and mitogens leads to increases in translation (Gingras *et al.*, 1999; Wang *et al.*, 2003), but is

not required for translation (McKendrick *et al.*, 2001; Morley and Naegele, 2002). Evidence shows that the interaction between eIF4E and the 5′-cap structure is enhanced when eIF4E is phosphorylated (Minich *et al.*, 1994). eIF4E is phosphorylated through the p38-mitogen-activated-protein kinase (MAPK) pathway (Kimball and Jefferson, 2000). The MAP-kinase-interacting kinase, Mnk1, contains a region that binds to eIF4G. Interaction of Mnk1 with eIF4G is important to stimulate the phosphorylation of eIF4E. Mutations reducing this interaction decrease the ability of Mnk1 to phosphorylate eIF4E (Parra-Palau *et al.*, 2003).

The activity of eIF4E is negatively regulated by translational repressors, e.g. eIF4E-binding proteins (4E-BPs). 4E-BPs are a family of three proteins that mimic the eIF4E binding site in eIF4G (Marcotrigiano *et al.*, 1999), competing with eIF4G for eIF4E binding. When bound to eIF4E, 4E-BPs prevent the assembly of eIF4F, thereby repressing the cap-dependent translation (Haghighat *et al.*, 1995). The block to translation caused by the 4E-BPs is reversible by their phosphorylation at certain key residues (Gingras *et al.*, 1999; Gingras *et al.*, 2001). Regulation of phosphorylation of 4E-BPs is achieved by a step-wise pathway, involving FRAP/mTOR and PI3K/Akt kinases (Gingras *et al.*, 1999; Gingras *et al.*, 2001; Mamane *et al.*, 2006; Miron *et al.*, 2003).

In apoptotic cells, the caspase-dependent cleavage of 4E-BP1 leads to insufficient phosphorylation and inhibits cap-dependent translation (Proud, 2005; Tee and Proud, 2002). Enhancement of the eIF4E/4E-BP interaction, through the expression of a constitutively active 4E-BP mutant (all phosphorylation sites mutated to Ala), leads to inhibition of cell cycle progression in a manner that mimics the treatment of rapamycin, a mTOR inhibitor (Fingar *et al.*, 2002; Fingar *et al.*, 2004). This 4E-BP mutant also slows G1 progression, blocks *c-myc*-induced transformation and prevents tumor growth (Avdulov *et al.*, 2004; Lynch *et al.*, 2004). These findings suggest that the ability of rapamycin to inhibit cancer cell growth is in part mediated through the dephosphorylation of 4E-BPs.

1.2.1.3 eIF4G

eIF4G is another component of heterotrimeric complex eIF4F. It plays an

important role in recruiting the 43S PIC to the cap structure in the mRNA (section *1.1.1.3*). eIF4G contains a binding domain for the eIF4E kinase Mnk1, whose function has been discussed in section *1.2.1.2*. eIF4G itself can be phosphorylated to inhibit cap-dependent translation by the cytostatic p21-activated kinase Pak2 in response to a large variety of stresses (Ling *et al.*, 2005). Pak2 phosphorylates several initiation factors including eIF4G, eIF3, and eIF4B, as well as Mnk1 (Orton *et al.*, 2004; Tuazon *et al.*, 1989). Phosphorylation of Mnk1 by Pak2 does not alter phosphorylation of eIF4E but decreases the phosphorylation of eIF4G as a consequence of the reduced binding of Mnk1 to eIF4G (Orton *et al.*, 2004).

During certain viral infections, eIF4G can be cleaved into two fragments. The leader proteinase of foot-and-mouth disease virus (FMDV) can cleave two isoforms of eIF4G (Kirchweger *et al.*, 1994; Lamphear *et al.*, 1993). eIF4G cleavage impairs the cap-dependent translation of host proteins whereas IRES-dependent viral translation is unaffected (Borman *et al.*, 1997).

1.2.1.4 Phosphorylation of eEF2

eEF2 is an elongation factor catalyzing translocation. Phosphorylation at its N-terminal GTP-binding domain inhibits its binding to the ribosome, thereby inhibiting elongation (Proud, 2000). eEF2 kinase is a Ca^{2+}/calmodulin-dependent protein kinase containing multiple sites which can be phosphorylated (Ryazanov *et al.*, 1988). The phosphorylation of eEF2 kinase inhibits its activity, which results in the dephosphorylation of eEF2 and promotes elongation. The activity of eEF2 kinase is controlled through the insulin and rapamycin sensitive mTOR pathway (Wang *et al.*, 2001), the stress-activated protein kinase SAPK4 pathway (Knebel *et al.*, 2001), as well as others.

1.2.2 mRNA-Specific Translational Control

Individual mRNAs can contain multiple structural elements that may be involved in regulating its translation (Figure 1.1) (Mignone *et al.*, 2002). These elements include the 5′-cap structure, stable secondary structures in the 5′- and 3′-UTRs, short upstream

open reading frames (uORFs), internal ribosome entry sites (IRES), protein or microRNA (miRNA) binding sites, and the poly(A) tail. mRNA-specific regulation provides quick and accurate control over certain groups of genes related to specific biological functions without affecting the overall protein synthesis in cells.

1.2.2.1 5'-untranslated region (5'-UTR)

The majority of eukaryotic translation is cap-dependent (Kozak, 1989b; Kozak, 1999). The 5′-cap structure not only plays an important role in translation initiation, but also is critical for maintaining mRNA stability (Wilusz *et al.*, 2001). Decapped mRNA is quickly degraded by 5′ – 3′ exonucleases (Beelman *et al.*, 1996; LaGrandeur and Parker, 1998). eIF4E binding to the cap structure can protect it from the attack of decapping enzymes (Vilela *et al.*, 2000).

The length of the 5′-UTR typically varies from 100 – 200 nucleotides (Mignone *et al.*, 2002; Pesole *et al.*, 2001). Initiation becomes inefficient at AUG codons fewer than 20 nucleotides or more than 200 nucleotides away from the 5′-terminus. Extensive secondary structure in the 5′-UTR also decreases the initiation efficiency, especially when it is close to the 5′-terminus.

The 5′-terminal oligopyrimidine (TOP) tract is a special element occurring in the 5′-UTR of all mammalian mRNAs that encode ribosomal proteins and other proteins involved in translation (Meyuhas and Hornstein, 2000). TOP consists of a short stretch of 5 – 14 pyrimidines. The translation of TOP mRNA is growth-dependent and is controlled through the mTOR-S6 kinases (S6Ks) pathway (Anthony *et al.*, 2001a; Anthony *et al.*, 2001b; Fumagalli and Thomas, 2000). Translational control of TOP mRNAs enables cells to rapidly express the proteins required for biosynthesis of the translational machinery in response to amino acid starvation or other stress conditions (Meyuhas and Hornstein, 2000). However, recent studies suggest that the mTOR-mediated phosphorylation of ribosomal protein S6 is essential for regulating the size of at least some cell types, but is dispensable for translational control of TOP mRNAs (Ruvinsky *et al.*, 2005; Ruvinsky and Meyuhas, 2006).

As discussed in section *1.1.1.5*, internal initiation provides an alternative pathway to initiate translation. First discovered in picornaviral mRNA (Jang *et al.*, 1988;

Pelletier and Sonenberg, 1988), the internal ribosome entry site (IRES) is a long (400 – 500 nucleotides) and highly structured element located in the 5′-UTR (Le and Maizel, 1998). The 43S PIC binds directly to the initiation codon with or without the assistance of initiation factors (Hellen and Sarnow, 2001; Pestova *et al.*, 1996; Pestova *et al.*, 1998b; Vagner *et al.*, 2001). In cricket paralysis virus (CrPV), initiation factors and initiator tRNA are not even required for internal initiation (Pestova and Hellen, 2003; Sasaki and Nakashima, 1999; Sasaki and Nakashima, 2000; Wilson *et al.*, 2000). IRES elements are not necessarily in the 5′-UTR. There is evidence that HIV-2 genomic RNA can initiate translation in the absence of a 5′-UTR by using IRES downstream of the AUG codon (Herbreteau *et al.*, 2005).

1.2.2.2 3′-Untranslated Region (3′-UTR)

Translational control mediated by the 3′-UTR generally requires *trans*-acting factors and is coupled to polyadenylation and stabilization of mRNA. A cytoplasmic polyadenylation element (CPE) located in the 3′-UTR of *c-mos* mRNA is responsible for the repression of translation in immature *Xenopus* oocytes and for activating the translation upon maturation (Mendez and Richter, 2001; Wickens *et al.*, 2000). A protein called Maskin is brought to the CPE by CPE-binding protein (CPEB). Maskin contains a binding domain for eIF4E and competes with eIF4G for eIF4E binding. The formation of the CPE·CPEB·Maskin·eIF4E ternary complex inhibits the formation of the eIF4F complex and represses translation (de Moor and Richter, 1999; Stebbins-Boaz *et al.*, 1999). Upon maturation, the phosphorylation of CPEB triggers the elongation of the poly(A) tail and the binding of Pab1p, which leads to activation of translation.

The male specific lethal 2 (MSL-2) protein is required in *Drosophila melanogaster* to compensate the two-fold difference in gene dosage of the X chromosome between males and females. *msl-2* is transcribed in both male and female *Drosophila*. In females, the expression of MSL-2 protein is repressed by a female specific RNA-binding protein sexual lethal (SXL). SXL binds to both the 5′- and 3′-UTR of *msl-2* mRNA. Co-repressor(s) are recruited to the 3′-UTR (Grskovic *et al.*, 2003). One of them has been identified as the *Drosophila* homolog of “upstream of N-ras” (dUNR) (Duncan *et al.*, 2006). Interestingly, the inhibition of *msl-2* mRNA

expression is mediated by a combination of distinct effects of SXL at different binding sites. SXL binding to the 3′-UTR interferes with the recruitment of 43S PIC, while its binding to the 5′-UTR provides a scanning block to prevent the translation machinery from accessing the AUG codon, thus enhancing the inhibition (Beckmann *et al.*, 2005).

1.2.2.3 microRNA

MicroRNAs (miRNAs) are a new class of endogenously encoded small (~22 nt) noncoding RNAs. The first identified miRNAs are *lin-4* (Lee *et al.*, 1993; Wightman *et al.*, 1993) and *let-7* (Reinhart *et al.*, 2000; Slack *et al.*, 2000) in *Caenorhabditis elegans*. They are critical for the regulation of developmental timing (Carrington and Ambros, 2003). miRNAs are able to regulate translation and stability of target mRNAs based on imperfect base-pairing. The complementary sites for the known miRNAs reside in the 3′-UTRs of target mRNAs (Lee *et al.*, 1993; Wightman *et al.*, 1993). There is no evidence that miRNAs can directly up-regulate gene expression. An RNA-induced silencing complex (RISC) is directed by miRNA to target mRNAs to down-regulate gene expression by either cleaving mRNAs or repressing translation (Hammond *et al.*, 2001; Orban and Izaurralde, 2005; Rana, 2007; Soifer *et al.*, 2007). So far, hundreds of miRNAs have been discovered in plants, invertebrates, and vertebrates; their functions include control of cell proliferation, cell death, and fat metabolism in flies (Brennecke *et al.*, 2003; Xu *et al.*, 2003), modulation of hematopoietic lineage differentiation in mammals (Chen *et al.*, 2004), and control of leaf and flower development in plants (Aukerman and Sakai, 2003; Chen, 2004; Emery *et al.*, 2003; Palatnik *et al.*, 2003).

1.3 Translational Control by Upstream Open Reading Frames

The initiation step of eukaryotic translation involves the cap-dependent ribosome scanning to locate the appropriate codon, typically an AUG codon as translation start site (Kozak, 1999; Kozak, 2002). That means AUG codons residing in the 5′-UTR can also serve as translation initiation sites. Upstream AUG (uAUG) codons and associated upstream open reading frames (uORFs) are found in a variety of eukaryotic mRNAs.

Earlier surveys suggest that 11 – 42% of vertebrate mRNAs and 20 – 48% of human mRNAs contain at least one uAUG codon (Davuluri *et al.*, 2000; Kozak, 1987; Pesole *et al.*, 1997; Pesole *et al.*, 2000; Suzuki *et al.*, 2000). Current studies of mammalian cDNAs show that uORFs are more common than expected; 30% of human genes contain one or more uORFs (Churbanov *et al.*, 2005; Iacono *et al.*, 2005). The number could be underestimated due to the incomplete mRNA 5′-end mapping data. The percentage of mRNAs containing uAUGs is even higher in certain classes of genes, such as those encoding oncogenes, transcription factors, growth factors, and other proteins involved in cell growth and differentiation (Kozak, 1987; Kozak, 1991; Morris, 1995).

The mechanism for recognition and initiation of the uAUG codon is essentially the same as that for the main AUG codon. The efficiency of 43S PIC recognizing an AUG codon is affected by parameters such as the distance between the AUG codon and the mRNA 5′-end, secondary structures in the 5′-UTR, and the nucleotide sequence flanking the AUG codon (Cao and Geballe, 1995; Kozak, 1999; Ruan *et al.*, 1994; Wang and Sachs, 1997b; Werner *et al.*, 1987). The translation of uORF(s) generally decreases the efficiency of translation initiation at the AUG codon preceding the main ORF (Morris and Geballe, 2000). Two modifications of the ribosome scanning model have been made to explain the translation of the main ORF downstream of the uORF(s). The leaky scanning model suggests that the scanning 43S PIC do not recognize the uAUG codon effectively. Some PICs will pass the uAUG codon and initiate translation at the downstream AUG codon (Kozak, 1978). The reinitiation model suggests that the small ribosomal subunit remains associated with the mRNA after termination at the uORF and continues to scan (Kozak, 1999). Generally, reinitiation is very inefficient and is only possible after translation of a short uORF. The sequence flanking the termination codon can stimulate reinitiation in some cases (Miller and Hinnebusch, 1989). Increasing the distance between ORFs can also enhance the efficiency of reinitiation (Kozak, 1987).

The translation of uORFs can also mediate mRNA stability. The uORF termination codon mimics a premature stop codon in the mRNA which will trigger the nonsense-mediated mRNA decay (NMD) pathway to rapidly degrade the mRNA (Amrani *et al.*, 2006).

1.3.1 Genes Regulated by uORFs

A number of uORFs identified and known to be involved in translational control in a variety of organisms are listed in Table 1.1. Some well characterized examples will be discussed followed by a detailed discussion of the best-understood examples of regulatory uORFs in budding yeast and filamentous fungi.

1.3.1.1 Mammalian AdoMetDC

Mammalian *S*-adenosylmethionine decarboxylase (AdoMetDC) plays a key role in the biosynthesis of polyamines. The mRNA contains a uORF that encodes a hexapeptide with the sequence MAGDIS, which is essential for its regulatory function (Mize *et al.*, 1998). This uORF negatively regulates the translation of *AdoMetDC* upon increase in the cellular levels of polyamines (Ruan *et al.*, 1996) and the regulation is also cell-specific (Hill and Morris, 1992). In resting normal T cells and in T cell lines with normal cellular levels of polyamines, the translation of *AdoMetDC* is repressed (Hill and Morris, 1992), whereas in non-lymphoid cells, the translation of *AdoMetDC* is uninhibited because of the inefficient initiation at the uAUG codon (Ruan *et al.*, 1994). Studies show that the translation of the uORF causes ribosome stalling at its termination codon with high concentrations of spermidine and the half-life of the ribosome stall correlates with the concentration of polyamines (Law *et al.*, 2001). A high level of polyamine actually impairs the hydrolysis of the peptidyl-tRNA upon translation termination, thus slowing down the release of the ribosome from the mRNA (Raney *et al.*, 2002).

1.3.1.2 Mammalian mdm2

MDM2 plays an important role in regulating the activity of the tumor suppressor p53 (Freedman, 1999; Juven-Gershon and Oren, 1999). Overexpression of MDM2 protein will lead to oncogenesis. The expression of *mdm2* is controlled at both transcriptional and translational levels. The transcription of *mdm2* is regulated by p53 through a binding site within the first intron (Barak *et al.*, 1994; Wu *et al.*, 1993). The *mdm2* transcript has two forms that differ in their 5′-UTRs (Landers *et al.*, 1997). The

long form of *mdm2* mRNA in human and mouse contains two uORFs that act synergistically to inhibit translational initiation (Brown *et al.*, 1999). The contribution of each individual uORF is different. uORF2 has weaker influence on downstream translation than uORF1. The amino acid sequence of uORF1 is important to its inhibitory activity, especially the conserved Gly9 residue. Missense alterations within the coding region reduce the inhibitory activity (Jin *et al.*, 2003). The short form of the mRNA lacks these two uORFs and is efficient in translation (Brown *et al.*, 1999).

1.3.1.3 Human Cytomegalovirus UL4

Human cytomegalovirus *UL4* encodes a structural virion glycoprotein. Three uORFs are found in the 5′-UTR of the most abundant transcript and repress the translation of *UL4* (Chang *et al.*, 1989). The amino acid sequence of uORF2 is essential for this negative regulation. Evidence shows that upon termination, the hydrolysis of peptidyl-tRNA is impaired so the nascent peptide is not released from the ribosome. Instead they form a ribosome-peptidyl-tRNA complex and block the access of upstream scanning 43S PIC to the downstream AUG codon, which in turn represses the translation of *UL4* (Cao and Geballe, 1996). The AUG codon of uORF2 is poorly recognized by scanning ribosomes. Most ribosomes will pass it and initiate translation at a downstream AUG codon (Cao and Geballe, 1994; Cao and Geballe, 1995). But once the uORF is translated, the ribosomes will be stalled at the termination codon and the translation of *UL4* is strongly inhibited (Cao and Geballe, 1995).

1.3.1.4 Vertebrate C/EBPs

C/EBPs (CCATT/enhancer binding proteins) are a family of transcription factors important to regulating tissue-specific genes during differentiation and cell proliferation. Different isoforms of C/EBPα and C/EBPβ are generated by using different potential initiation codons in the same ORF (Descombes and Schibler, 1991; Lin *et al.*, 1993; Ossipow *et al.*, 1993). These isoforms have an identical C-terminal DNA binding domain but different N-termini, the short (truncated) isoforms do not contain the transcription activation domain and seem to act as antagonists of the full-length

transcription factors. Both C/EBPα and C/EBPβ mRNAs contain an out-of-frame uORF immediately upstream of the major translation initiation site B1 (Calkhoven *et al.*, 1994). This uORF is crucial for the regulation of the isoform ratio and consequently for the transcriptional regulation of C/EBP target genes. The regulatory mechanism involves a combination of leaky scanning past the uORF AUG codon and reinitiation after uORF translation. The translation of the uORF inhibits the initiation at site B1 while enhancing initiation at a further downstream site C, generating the truncated form of the transcription factor (Calkhoven *et al.*, 1994; Calkhoven *et al.*, 2000). Two initiation factors eIF2 and eIF4E are also capable of regulating the expression of various C/EBP isoforms. Both high eIF2α activity and eIF4E overexpression result in an increase in synthesis of the truncated isoforms and a decrease of the full-length protein (Calkhoven *et al.*, 2000). eIF4E stimulates initiation at the uORF, reducing initiation at site B1 and eIF2α increases reinitiation efficiency at site C.

1.3.1.5 Mammalian CD36

The mammalian protein CD36 is a macrophage scavenger receptor that mediates the uptake of oxidized low-density lipoprotein (OxLDL) (Febbraio *et al.*, 2000; Li *et al.*, 2000; Nagy *et al.*, 1998). The expression of CD36 is induced by glucose in a dose-dependent manner. The control of CD36 translation is mediated by three uORFs in the mRNA 5′-UTR. Studies suggest that translational regulation mediated by the CD36 uORFs involves a reinitiation mechanism that is modulated by glucose (Griffin *et al.*, 2001). Mutation of all three uORFs significantly increases CD36 basal expression and the expression becomes glucose independent. The first uORF is essential for glucose sensing. Mutations that eliminate the first AUG result in a loss of glucose responsiveness. It is possible that the reinitiation efficiency is regulated by eIF2α phosphorylation *via* a mechanism similar to C/EBPs.

1.3.1.6 Cauliflower Mosaic Virus (CaMV) 35S ORF VII

The long 5′-UTR of CaMV 35S mRNA contains nine uORFs that play an essential role in the translation of ORF VII (Pooggin *et al.*, 1998; Pooggin *et al.*, 2000).

Translation of ORF VII is mediated by a non-linear ribosomal scanning mechanism known as shunting (Futterer *et al.*, 1990; Futterer *et al.*, 1993). This is a cap-dependent form of translation initiation in which the 40S ribosomal subunit skips part of the 35S 5′-UTR. Shunting enables efficient initiation even when there is a hairpin structure that blocks normal scanning. Two *cis*-acting elements in 35S mRNA 5′-UTR are crucial for translation: the first uORF (uORF-A) and a stable hairpin that is located immediately downstream. After translation of uORF-A, the 40S ribosomal subunit, independent of the activities of eIFs 4A, 4B, and/or 1 and 1A required for scanning, is shunted to a site downstream of the hairpin and resumes scanning (Ryabova *et al.*, 2000). Either uORF-A alone or the hairpin alone inhibits the translation of ORF VII (Pooggin *et al.*, 2000). Shunting on 35S RNA is a special form of reinitiation (Hohn *et al.*, 2001; Ryabova and Hohn, 2000; Ryabova *et al.*, 2002).

1.3.1.7 Mammalian her-2

Oncogene *her-2* (*neu*, *erbB2*) encodes a transmembrane receptor tyrosine kinase responsible for transduction of mitogenic signals from a variety of growth factor receptors (Olayioye *et al.*, 2000; Yarden and Sliwkowski, 2001). Overexpression of Her-2 protein contributes to approximately 30% of primary human cancers. A uORF is present in *her-2* mRNA and represses the expression of Her-2 protein (Child *et al.*, 1999a). This uORF also affects the selection of the downstream initiation site by interfering with the interaction between the ribosome and the primary *her-2* AUG codon (Spevak *et al.*, 2006). The control is cell-specific. In certain types of cancer cells, the translation efficiency is increased. Recent evidence shows that a U-rich translational derepression element (TDE) identified in the 3′-UTR also functions in regulating the expression of *her-2*. In the absence of other elements and factors, the expression of *her-2* is inhibited by the uORF. Factors bound to *her-2* 3′-UTR can somehow alter uORF function, allowing efficient translation reinitiation to occur at the downstream ORF (Mehta *et al.*, 2006).

1.3.2 *S. cerevisiae GCN4* and *Mammalian ATF4*

Yeast *GCN4* mRNA is the best understood example of translational control *via* reinitiation (Hinnebusch, 1996; Hinnebusch, 1997; Hinnebusch, 2005). GCN4 is a transcription factor that activates the expression of a large group of genes involved in amino acid biosynthesis. Expression of GCN4 is repressed under normal cell conditions. During amino acid starvation, the translation of *GCN4* mRNA is enhanced.

Four small uORFs, containing 2 or 3 codons each, reside in the 590-nt 5′-UTR of *GCN4* mRNA. The translational control is generated through the combined activities of the different uORFs. Mutations eliminating all 4 uORFs cause the translation of GCN4 to be significantly increased and to no longer respond to the availability of amino acids. Not every uORF in the 5′-UTR is required for *GCN4* regulation. Transcripts containing only uORF1 and uORF4 in the 5′-UTR are regulated similarly to the wild-type (Hinnebusch, 1996; Hinnebusch, 1997). uORF1 regulates downstream initiation *via* reinitiation, and by itself reduces downstream translation efficiency moderately, since it promotes reinitiaiton. In contrast, the downstream uORF4 by itself inhibits the reinitiation at the *GCN4* AUG codon by approximately 99% (Hinnebusch, 1996). The coding sequences of the uORFs are not important for the regulation. The nucleotide sequences surrounding the termination codons of these two uORFs determine the efficiencies of reinitiation. Replacement of AU-rich sequence surrounding the uORF1 termination codon with GC-rich sequence in the same region of uORF4 generates strong inhibitory activity (Miller and Hinnebusch, 1989).

As discussed in section *1.2.1.1*, the initiation efficiency is affected by the activity of initiation factor eIF2 in the ternary complex and the distance between ORFs. Under normal cell conditions, the high level of eIF2 assures the active ternary complex is available soon enough for initiation at downstream uORFs after ribosomes have translated uORF1. Therefore, the translation of inhibitory uORFs is favored. Upon amino acid starvation, the high level of uncharged tRNAs activates the GCN2 kinase to phosphorylate eIF2α. The phosphorylation of eIF2α lowers the availability of active ternary complex. Ribosomes are not able to reacquire initiation capability soon enough to translate the downstream uORFs (Hinnebusch, 1996). Ribosomes pass the inhibitory

uORFs and continue to scan the mRNA. This gains time for ribosomes to reacquire the active ternary complex to become capable of reinitiating translation at the *GCN4* start codon. Consistent with this, extending the distance between uORF1 and uORF4 reduces the expression of *GCN4* under starvation conditions (Abastado *et al.*, 1991).

The mRNA encoding mammalian activating transcription factor 4 (*ATF4*) contains two evolutionarily conserved uORFs in its 5′-UTR (Harding *et al.*, 2000). The second uORF overlaps with the *ATF4* coding region. Translation of *ATF4* is activated upon eIF2α phosphorylation by PERK in response to unfolded proteins in the ER. Under normal cell conditions, the translation of *ATF4* is significantly repressed. The stress condition increases the *ATF4* mRNA translation by 3-fold (Harding *et al.*, 2000). The *ATF4* uORFs are presumed to function analogously to the *S. cerevisiae GCN4* uORFs. uORF1 controls downstream translation *via* reinitiation and facilitates the translation of *ATF4* in response to stress-induced eIF2α phosphorylation, whereas uORF2 inhibits *ATF4* expression in non-stressed cells (Vattem and Wek, 2004).

1.3.3 *N. crassa arg-2* and *S. cerevisiae CPA1*

Neurospora arg-2 gene encodes the small subunit of arginine-specific carbamoyl phosphate synthetase (CPS-A) located in mitochondria (Davis, 1986). This small subunit functions as a glutamine amidotransferase, transferring the glutamine amide nitrogen to the large subunit, where carbamoyl phosphate is synthesized. The 5′-UTR of *arg-2* mRNA contains an evolutionarily conserved uORF encoding a 24-amino-acid peptide, known as the arginine attenuator peptide (AAP) (Orbach *et al.*, 1990; Orbach and Sachs, 1991; Wang *et al.*, 1998). This uORF negatively controls the expression of *arg-2* in response to the level of Arg in the culture medium (Luo *et al.*, 1995). Other AAP-encoding uORFs are found in the *arg-2* 5′-UTRs of *Magnaporthe grisea* (Shen and Ebbole, 1996), *Trichoderma virens* (Baek and Kenerley, 1998) and *Aspergillus nidulans* (GenBank AJ223085), as well as in the *CPA1* 5′-UTR of *S. cerevisiae* (Nyunoya and Lusty, 1984).

Repression of *Neurospora arg-2* by Arg occurs at both the transcriptional and translational level. Short exposure of wild-type *Neurospora* to high levels of Arg decreased the synthesis of ARG-2 to ~50%, while the level of *arg-2* transcript was

unaffected. The average size of polysomes associated with *arg-2* transcripts was reduced (Luo *et al.*, 1995). Long exposure of *Neurospora* to high levels of Arg will eventually reduce the level of *arg-2* transcripts. The regulation at the translational level provides a quick response to Arg. Expression of *arg-2-lacZ* reporter genes with the wild-type *arg-2* uORF or a mutated uORF, in which the initiation codon was eliminated, showed that removing the initiation codon of *arg-2* uORF abolishes this Arg-specific repression *in vivo* (Luo *et al.*, 1995). A single nucleotide missense mutation from Asp to Asn at codon 12 (D12N) in the *arg-2* uORF eliminates Arg-specific regulation (Freitag *et al.*, 1996).

The Arg-specific translational control by *arg-2* uORF has been reconstituted in a homologous *Neurospora* cell-free *in vitro* translation system. Synthetic mRNA containing wild-type *arg-2* 5′-UTR upstream of the firefly luciferase (*LUC*) showed negative regulation of luciferase activity in response to Arg, whereas constructs containing ΔAUG or D12N uORF did not (Wang and Sachs, 1997b). Analyses with a primer extension inhibition assay ("toeprint assay") in the *Neurospora* cell-free system demonstrated that in the presence of high Arg, ribosomes stalled at the wild-type uORF termination codon but not at the corresponding region of ΔAUG or D12N uORF (Wang and Sachs, 1997b). At the same time, a reduction in ribosomes associated with the downstream *LUC* initiation codon was detected in the transcript with the wild-type uORF, indicating decreased translation initiation at the *LUC* start codon in response to Arg (Wang and Sachs, 1997a). These results are highly consistent with that from *in vivo* studies and suggest that the translation of *arg-2* uORF is required for Arg-induced ribosome stalling, which in turn reduces the access of ribosomes to downstream initiation codons. Based on its *cis*-acting ability to negatively regulate translation, the *arg-2* uORF-encoded peptide was designated the arginine attenuator peptide (AAP) (Wang *et al.*, 1998). Further analyses revealed that the *arg-2* intercistronic sequence and uORF termination codon are dispensable for the Arg-specific regulation. Direct in-frame fusion of the uORF coding sequence to the downstream *LUC* confers arginine-specific translational repression (Wang *et al.*, 1998). In this case, the ribosome stalls after the uORF coding region (Wang *et al.*, 1999). That the translation of the *arg-2* uORF stalls both terminating and elongating ribosome shows that the nascent peptide encoded by the

arg-2 uORF regulates the movement of translating ribosomes through a novel mechanism. The *N. crassa* and *S. cerevisiae* AAPs were shown to function in *N. crassa*, *S. cerevisiae* and wheat germ extracts (Wang *et al.*, 1999). Their function does not rely on the charging status of arginyl-tRNAs (Wang *et al.*, 1999).

The uORF in yeast *CPA1* mRNA was identified two decades ago. Its presence is necessary for the Arg-specific regulation of *CPA1* expression both *in vivo* and in an isolated *in vitro* system. The elimination of the uORF initiation codon by mutating AUG to UUG abolishes Arg-specific repression, as does a single nucleotide missense mutation from Asp to Asn at codon 13 (D13N), like its corresponding D12N mutation in the *arg-2* uORF. Other constitutive mutations causing loss of responsiveness to Arg include the introduction of premature stop codons either at Glu8 or Trp20, and the C11T missense mutation. Furthermore, when the uORF contains silent mutations that alter the mRNA sequence but leave unchanged the uORF encoded amino acid sequence, the Arg-induced regulatory activity is not affected (Delbecq *et al.*, 1994). These studies suggest that the Arg-specific regulation is dependent on the peptide sequence but not the nucleotide sequence. Further studies also revealed that the uORF could act in *cis* to repress other genes in high levels of Arg, either present as a uORF in front of *GCN4* or in-frame fused to *lacZ* (Delbecq *et al.*, 1994) and the active domain of the peptide spans from codon 6 to codon 23 (Delbecq *et al.*, 2000).

Destabilization of *CPA1* mRNA observed in wild-type yeast culture grown in Arg-supplemented medium indicated that the nonsense-mediated mRNA decay (NMD) pathway might be involved in the *CPA1* regulatory system (Crabeel *et al.*, 1990). Direct evidence that the *CPA1* transcript is a substrate for NMD was provided by transcription inhibition experiments that demonstrated the *CPA1* mRNA half-life to be 3 min in a wild-type strain and approximately 18 min in an isogenic *upf1Δ* (up-frameshfit suppressor) strain (Ruiz-Echevarria and Peltz, 2000). Recent studies show that the translation of uORF-encoded peptide is required for reducing *CPA1* expression by affecting both translation efficiency and mRNA stability (Gaba *et al.*, 2005).

1.4 Ribosome Movement Controlled by Nascent Peptides

It has been established that the arginine-specific translational regulation of *arg-2* and *CPA1* mediated by AAP is dependent on the peptide sequence of AAP but not its mRNA sequence (Fang *et al.*, 2000). Translation of the AAP is required since eliminating its initiation codon abolishes the regulatory ability. Extension of the AAP at its C-terminus or direct in-frame fusion of its coding sequence to downstream reporter genes did not affect the regulatory function of AAP. In both cases, AAP causes ribosomes to stall after its coding region (Fang *et al.*, 2000; Wang *et al.*, 1999). This means that the termination codon of the uORF is not necessary for AAP's regulatory function. The N-terminus of AAP can also be extended and its regulatory ability is not affected. This makes it easy to track the nascent AAP peptide on an SDS-PAGE gel by adding eight additional methionines at its N-terminus to highlight the signal without compromising its function (Fang *et al.*, 2002). Further analysis shows that AAP exerts its arginine-specific regulatory ability as an internal domain within a large polypeptide and stalls ribosomes during elongation (details in CHAPTER 2). All of these data indicate that the nascent AAP can control the movement of translating ribosomes through a novel mechanism that resembles other instances of nascent peptide regulated control in eukaryotic and prokaryotic genes (Fang *et al.*, 2004), as described below.

1.4.1 *Arabidopsis thaliana CGS1*

Arabidopsis CGS1 encodes a cystathionine γ-synthase which catalyzes the first committed step of methionine biosynthesis (Kim and Leustek, 1996; Mathews, 1999). Expression of the *CGS1* gene is negatively regulated in response to the cellular methionine level. A high level of methionine decreases the amount of full-length *CGS1* mRNA and increases the amount of 5′-truncated mRNA as a result of mRNA degradation from the 5′-end. A region designated MTO1, which encodes a short peptide stretch within *CGS1* exon 1, is responsible for this regulation at a post-transcriptional level. Mutations altering the MTO1 amino acid sequence or chemicals inhibiting translation abolish the regulation (Chiba *et al.*, 1999; Chiba *et al.*, 2003; Lambein *et al.*, 2003). *In vitro* studies in wheat germ extracts have shown that *S*-adenosyl-L-methionine (AdoMet)

could induce a ribosome elongation-arrest immediately downstream of the *MTO1*-coding region with Trp93 and Ser94 at the P and A sites respectively. The arrested nascent peptide is in its peptidyl-tRNA form attached to $tRNA^{Ser}$. The ribosome stalls at the step of translocation (Onouchi *et al.*, 2005). The position of ribosome stalling suggests that the nascent MTO1 peptide resides in the ribosome exit tunnel when elongation is arrested. This elongation-arrest event occurs prior to the formation of 5′-truncated mRNA, suggesting the ribosome stalling triggers the mRNA degradation. When the ribosome is stalled during elongation, the 5′-end of *CGS1* mRNA protrudes from the stalled ribosome and is accessible for the RNA decay machinery. Thus the nascent peptide-mediated elongation arrest is coupled with mRNA degradation (Onouchi *et al.*, 2005).

1.4.2 *Escherichia coli secM*

The *secM* (secretion monitor) gene of *E. coli* encodes a unique secretory protein that monitors the cellular protein export status. As the first open reading frame of the *secA* operon, translation of *secM* is responsible for the translational regulation of the downstream *secA* gene, which encodes a translocation ATPase driving the movement of preproteins and internal membrane complexes (Oliver *et al.*, 1998). The nascent SecM peptide causes the ribosome to stall at a site close to its C-terminus during elongation (Nakatogawa and Ito, 2001). The amino acid sequence $FXXXXXWIXXXXGIRAGP^{166}$ at the C-terminal region of SecM is important and the Pro166 residue is essential for the ribosome stalling. Further studies show that the C-terminal region of SecM acts as an independent elongation-arrest element. It can be fused in-frame into two unrelated peptide sequences and still confer its ribosome-stalling function. Mutations in ribosomal RNA (23S rRNA) and protein (L22) have been identified that can suppress the elongation-arrest function of SecM. These mutations are located close to the narrowest constriction of the ribosome exit tunnel. The ribosome stalling mediated by nascent SecM peptide may be a result of specific interaction of its residues with ribosomal components (Nakatogawa and Ito, 2002). The stalled ribosome is predicted to disrupt the secondary structure of the *secM-secA* messenger RNA, leading to the exposure of the *secA* initiation sequence for the entry of new ribosomes that translate *secA* (McNicholas *et al.*, 1997; Nakatogawa and Ito, 2001; Oliver *et al.*, 1998). Recent evidence reveals

that when ribosome is arrested during elongation, the Pro166 codon is positioned at the A site instead of at the P site, and the last amino acid of the arrested peptidyl-tRNA is Gly165, which is inefficiently transferred to Pro166. The arrested peptidyl-tRNA is completely resistant to puromycin when the A site is occupied by prolyl-tRNA^{Pro} (Muto *et al.*, 2006), indicating the proline within the PTC is able to increase the stability of the peptidyl-tRNA complex. Further study suggests the C-terminus of the nascent SecM peptide adopts a compact conformation induced by the ribosome when synthesis of the arrest motif is complete. This conformation change is required but not sufficient for the arrest. Under such a condition, key residues within the arrest motif interact with ribosomal components to induce a further change in the ribosome-nascent chain complex that in turn arrests elongation (Woolhead *et al.*, 2006).

1.4.3 *E. coli tnaC*

The tryptophanase (*tna*) operon of *E. coli* consists of a leader regulatory region followed by two coding regions for tryptophanase (*tnaA*) and tryptophan-specific permease (*tnaB*), respectively (Gong and Yanofsky, 2002). The 319-nt leader region contains a 24-amino-acid peptide coding region *tnaC*. In the spacer between the *tnaC* stop codon and the downstream *tnaA* start codon there are multiple potential Rho-dependent transcription termination sites (Gong and Yanofsky, 2002). Free tryptophan functions as inducer of transcription of *tnaA* and *tnaB*. Tryptophan binds to the ribosome translating *tnaC* and inhibits termination. The nascent TnaC-tRNA remains uncleaved within the ribosome and the stalled ribosome prevents Rho factor from accessing its binding sites immediately adjacent to the *tnaC* stop codon, thus inhibiting Rho-induced transcription termination. The downstream genes in the operon, *tnaA* and *tnaB*, will then be transcribed and expressed (Gong and Yanofsky, 2001). Trp12 of TnaC and the spacing between Trp12 and the last amino acid Pro24 are peculiarly crucial for the tryptophan induced ribosome stalling (Gollnick and Yanofsky, 1990; Kamath and Yanofsky, 1992; Stewart and Yanofsky, 1986). Trp12 of arrested TnaC-tRNA is located near the narrow section of the ribosome exit tunnel. Mutations in ribosomal 23S RNA and L3 and L22 proteins near the position occupied by Trp12 virtually eliminate tryptophan induction (Cruz-Vera *et al.*, 2005). The latest studies

suggest that Trp12 would introduce a specific change in the ribosome PTC that creates a free tryptophan-binding site, where bound tryptophan inhibits peptidyl transferase activity (Cruz-Vera *et al.*, 2006; Cruz-Vera *et al.*, 2007).

1.4.4 Aphthovirus FMDV 2A

Positive-strand RNA viruses usually encode some or all of their proteins in the form of polyproteins that would be cleaved by viral or host proteinases to produce mature proteins through co-translational or post-translational proteolytic processing (Ryan *et al.*, 1998; Seipelt *et al.*, 1999). Studies on the processing of the aphthovirus FMDV polyprotein suggest an 18-amino-acid-long 2A region mediates cleavage at its own C-terminus (Ryan *et al.*, 1991). The uncleaved polyproteins (~10% of total products), when synthesized in cell-free translation systems, do not convert to cleaved proteins posttranslationally, suggesting that the cleavage is co-translational (Ryan and Drew, 1994). 2A and 2A-like sequences identified in other viral species share the conserved DXEXNPG⇓P motif (⇓ indicates the cleavage site) (de Felipe *et al.*, 2006; Donnelly *et al.*, 2001a). It has been noticed that there is an imbalance in the accumulation of products in *in vitro* translation systems. Proteins encoded upstream of 2A accumulate to higher levels than those encoded downstream of 2A. Puromycin preferentially incorporated into the N-terminal product (Donnelly *et al.*, 2001b). These results suggest that 2A-mediated cleavage is not a proteolytic process. It has been suggested that the nascent 2A peptide modifies the activity of the elongating ribosome, promoting the hydrolysis of the ester bond between peptide and $tRNA^{Gly}$, thereby releasing the N-terminal peptide. Thus the ribosome skips a codon to the next without forming a peptide bond (de Felipe *et al.*, 2006; Donnelly *et al.*, 2001b; Ryan *et al.*, 1999). This model is supported by the discovery that the cleavage is intraribosomal. During co-translational translocation, the released C-terminal protein of 2A-containing polyprotein remains in the cytosol unless it contains its own signal sequence, indicating the cleavage occurs before the nascent C-terminal protein could reach the translocon and therefore within the ribosome (de Felipe *et al.*, 2003). The latest data show that the nascent 2A protein causes a temporary ribosome pause at its C-terminus with the last amino acid residue of the N-terminal product in the P site, and promotes the hydrolysis of

the peptidyl-tRNA bond. Details will be discussed in CHAPTER 3.

1.5 Thesis Research

The application of the *Neurospora* cell-free *in vitro* translation system and primer extension inhibition assay has provided key insights into the *arg-2* AAP-mediated translational control. Similar systems and assays have been used in the *in vitro* studies of a variety of uORF- or nascent peptide-mediated translational regulation.

In my thesis, the *N. crassa* cell-free translation system was used to study the ribosome stalling mediated by the nascent *arg-2* AAP as both N-terminal and internal domains within a long polypeptide, revealing that the nascent AAP is peptidyl-tRNA associated with the stalling ribosome (CHAPTER 2). The same *in vitro* translation system was also used to study the mechanism underlying the co-translational cleavage mediated by aphthovirus FMDV 2A peptide (CHAPTER 3). Primer extension inhibition assay was modified to map the binding site of translational repressor FljA on *fliC* mRNA 5′-UTR in *Salmonella*, demonstrating that it overlaps the ribosome binding site (CHAPTER 4).

Table 1.1 Genes with uORFs known to be involved in translational control.

Organism	Gene	References
Viruses	baculovirus gp64	(Chang and Blissard, 1997)
	cauliflower mosaic virus 35S RNA	(Futterer and Hohn, 1992; Pooggin *et al.*, 1998)
	cytomegalovirus UL4*	(Alderete *et al.*, 1999; Degnin *et al.*, 1993; Schleiss *et al.*, 1991)
	influenza NB/NA	(Williams and Lamb, 1989)
	reovirus S1	(Belli and Samuel, 1993; Fajardo and Shatkin, 1990)
	Rous sarcoma virus	(Donzé and Spahr, 1992; Moustakas *et al.*, 1993)
	SV40 16S and 19S RNAs	(Grass and Manley, 1987; Perez *et al.*, 1987; Sedman and Mertz, 1988; Sedman *et al.*, 1989)
Fungi	arg-2*	(Freitag *et al.*, 1996; Luo and Sachs, 1996)
	brlA	(Han *et al.*, 1993)
	CLN3	(Polymenis and Schmidt, 1997)
	CPA1*	(Werner *et al.*, 1987)
	cpcA	(Hoffmann *et al.*, 2001)
	cyc1-362	(Pinto *et al.*, 1992)
	GCN4	(Hinnebusch, 1997)
	HOL1	(Wright *et al.*, 1996)
	INO2	(Eiznhamer *et al.*, 2001)
	stuA	(Wu and Miller, 1997)
	YAP2	(Vilela *et al.*, 1998; Vilela *et al.*, 1999)
Plants	S-adenosylmethionine decarboxylase*	(Franceschetti *et al.*, 2001; Hanfrey *et al.*, 2002)
	arginine decarboxylase	(Chang *et al.*, 2000)
	AtbZIP11/ATB2	(Wiese *et al.*, 2005)
	Lc	(Wang and Wessler, 1998)
	Opaque-2	(Lohmer *et al.*, 1993)
	ornithine decarboxylase	(Kwak and Lee, 2001)
	pma1	(Michelet *et al.*, 1994)
	pma3	(Lukaszewicz *et al.*, 1998)
Worm	gna-2	(Lee and Schedl, 2004)
Amphibia	Cx41	(Meijer *et al.*, 2000)
Mammals	S-adenosylmethionine decarboxylase*	(Hill and Morris, 1992; Hill and Morris, 1993; Mize *et al.*, 1998)
	ADH5	(Kwon *et al.*, 2001)
	β2 adrenergic receptor*	(Parola and Kobilka, 1994)
	ATF4	(Harding *et al.*, 2000)
	bcl-2	(Harigai *et al.*, 1996)

Organism	Gene	References
Mammals	BTEB	(Imataka *et al.*, 1994)
	cat-1	(Yaman *et al.*, 2003)
	CD36	(Griffin *et al.*, 2001)
	C/EBPα	(Calkhoven *et al.*, 2000; Lincoln *et al.*, 1998)
	C/EBPβ	(Calkhoven *et al.*, 2000; Lincoln *et al.*, 1998)
	CHOP	(Jousse *et al.*, 2001)
	erythrocyte carbonic anhydrase inhibitor	(Bergenhem *et al.*, 1992)
	estrogen receptor alpha uORF	(Kos *et al.*, 2002)
	fibroblast growth factor 5	(Bates *et al.*, 1991)
	fli-1	(Sarrazin *et al.*, 2000)
	glucocorticoid receptor 1a	(Diba *et al.*, 2001)
	her-2/neu	(Child *et al.*, 1999a; Child *et al.*, 1999b)
	lck	(Marth *et al.*, 1988)
	major vault protein	(Holzmann *et al.*, 2001)
	mdm-2	(Brown *et al.*, 1999)
	c-mos	(Steel *et al.*, 1996)
	muscle acylphosphatase	(Fiaschi *et al.*, 1997)
	ornithine decarboxylase	(Manzella and Blackshear, 1990; Shantz and Pegg, 1999)
	placental growth factor	(Maglione *et al.*, 1993)
	PR65 (PP2A regulatory subunit)	(Wera *et al.*, 1995)
	retinoic acid receptor β2*	(Reynolds *et al.*, 1996; Zimmer *et al.*, 1994)
	p27 (kip1)	(Gopfert *et al.*, 2003)
	SCL	(Calkhoven *et al.*, 2003)
	serine hydroxymethyltransferase	(Byrne *et al.*, 1995)
	suppressor of cytokine signaling 1 (socs-1)	(Schluter *et al.*, 2000)
	transforming growth factor β3	(Arrick *et al.*, 1991)
	UCP2	(Hurtaud *et al.*, 2006)
	V(1b) vasopressin receptor	(Nomura *et al.*, 2001)

Deletion of uORF initiator codons alters gene expression; updated and expanded from (Geballe and Sachs, 2000).

*The uORF peptide sequence is established to be important for controlling gene

expression.

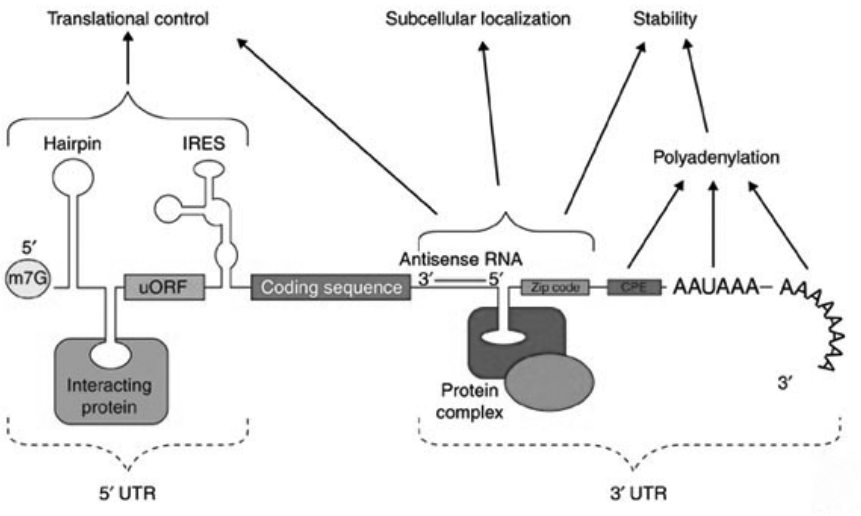

Figure 1.1 *cis*-acting elements on mRNA that affect gene expression.

See the text for further details. Adapted by the permission from BioMed Central Ltd (Mignone *et al.*, 2002).

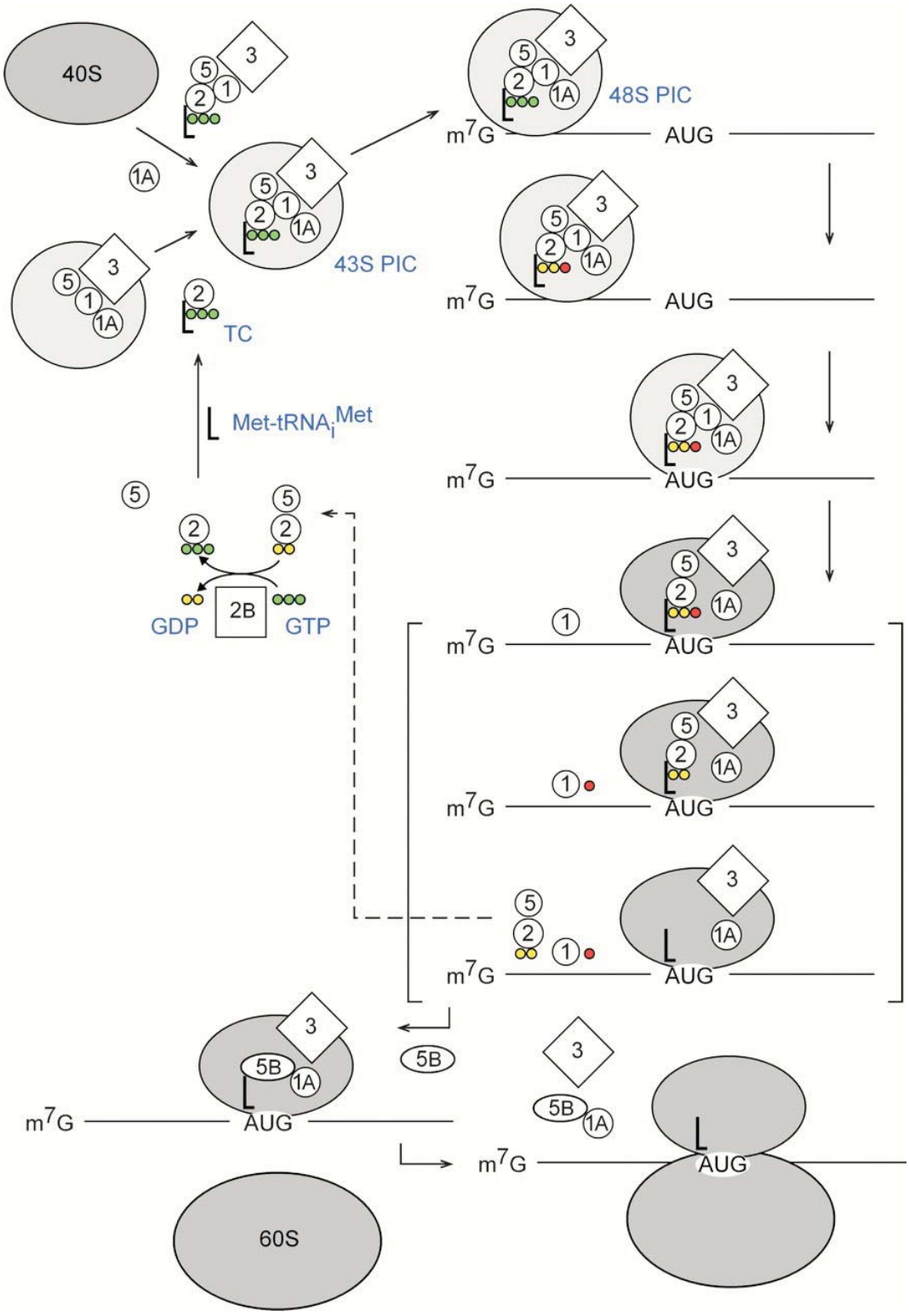

Figure 1.2 Simplified model for initiation of protein synthesis in eukaryotes. eIF2, GTP, and Met-tRNA$_i^{Met}$ is assembled into a ternary complex (TC). The 40S

ribosomal subunit and the TC are brought together by multiple initiation factors to form a 43S pre-initiation complex (PIC). The 43S PIC is loaded at the mRNA 5′-end with the assistance of eIF4F (not shown) to create the 48S PIC. During the 43S PIC scanning the mRNA to locate AUG codon in the appropriate context, the GTP (three green circles) bound to eIF2 is hydrolyzed to GDP-P_i (two yellow circles and one red). The recognition of AUG codon triggers the release of P_i (red circle) and other initiation factors. Factors, GDP (two yellow circles), and P_i that leave the ribosome are indicated to the *left* of the ribosome. The bracketed intermediates represent the changes that occur when the PIC converts from an open to a closed conformation; the conformational change is indicated by alterations to the 40S subunit's shape and shading. The large (60S) ribosomal subunit joins the 40S subunit to form the 80S elongating ribosome. The exchange of GDP bound to eIF2 for GTP, with release of eIF5, by eIF2B is also indicated, as is the reformation of TC by binding of eIF2·GTP to Met-$tRNA_i^{Met}$. Adapted by the permission from Cold Spring Harbor Laboratory Press (Asano and Sachs, 2007).

CHAPTER 2 THE AAP REGULATES TRANSLATION ELONGATION IN RESPONSE TO ARGININE AS A NASCENT POLYPEPTIDE DOMAIN*

2.1 Introduction

Nascent polypeptides can control translation. A variety of peptides specified by uORFs in eukaryotic and prokaryotic mRNAs can regulate the movement of ribosomes translating them (Gaba *et al.*, 2001; Geballe and Sachs, 2000; Gong and Yanofsky, 2002; Lovett and Rogers, 1996; Morris and Geballe, 2000; Raney *et al.*, 2002; Tenson and Ehrenberg, 2002). The evolutionarily conserved arginine attenuator peptide (AAP), encoded by a uORF, is able to negatively regulate the expression of the small subunit of fungal carbamoyl phosphate synthetase in response to the cellular arginine level (Luo *et al.*, 1995). The amino acid sequence of the AAP but not the sequence of mRNA is crucial for this regulation (Fang *et al.*, 2000). The regulation only responds to the level of Arg but not the level of aminoacylated arginyl-tRNA (Wang *et al.*, 1999). The nascent AAP causes ribosomes to stall at the uORF termination codon when a high level of Arg is present, thereby blocking the translating ribosomes from accessing the downstream initiation codon used for synthesis of enzyme (Gaba *et al.*, 2001). Genetic selection for mutations deficient in Arg-specific regulation in *Neurospora crassa* and *Saccharomyces cerevisiae* independently identified an Asp residue of the AAP, which is conserved in other AAP-encoding uORFs identified across many fungi (Hood *et al.*, 2007), as critical for this negative *cis*-acting regulation (Freitag *et al.*, 1996; Werner *et al.*, 1987). Translational control by Arg is eliminated when there is a substitution of this Asp residue with Asn (Freitag *et al.*, 1996; Gaba *et al.*, 2005; Wang *et al.*, 1999; Werner

*Part of this material has been published in similar form in the Proceedings of the National Academy of Sciences and is used here with permission of The National Academy of Sciences.

Fang, P., Spevak, C.C., Wu, C., and Sachs, M.S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A*. **101** (12): 4059-4064.

et al., 1987). These mutations also eliminate the AAP's capacity to stall ribosomes in cell-free translation systems (Wang and Sachs, 1997a; Wang *et al.*, 1999). Unlike other uORF-encoded peptides that only cause stalling during termination, the AAP also causes Arg-regulated ribosome stalling during elongation immediately downstream of the AAP coding region when fused in-frame to the N-terminus of a reporter gene (Wang *et al.*, 1998; Wang *et al.*, 1999).

The evidence accumulated thus far suggests that ribosome stalling requires the translation of the AAP domain and that the nascent AAP interacts with the ribosome to cause elongation-arrest in the presence of high Arg. How this is accomplished is still poorly understood. An internal domain in the prokaryotic regulatory protein SecM can stall ribosome during elongation. Studies indicate that the nascent peptide is not released from the ribosome after being synthesized. Instead it is arrested in the ribosome exit tunnel in peptidyl-tRNA form. The specific interaction between the nascent peptide and ribosomal component causes ribosome to stall (Nakatogawa and Ito, 2001; Nakatogawa and Ito, 2002). The characteristics of the AAP suggest it may function in the same way. To test this possibility, the regulatory function of the AAP has been examined in a *N. crassa* cell-free translation system first as a domain near the N-terminus and then internally within a longer polypeptide. To facilitate protein detection, extra Met residues were placed at the extreme N-terminus and all internally located Met residues were removed. Analyses of the radiolabeled translation products indicate that wild-type AAP functions at either location to stall polypeptide synthesis in response to arginine (Fang *et al.*, 2004). Pulse-chase analyses also reveal that ribosomes stalled during elongation resume translation instead of dissociating from the mRNA (Fang *et al.*, 2004). The appearance of arrested peptidyl-tRNA intermediates was directly detected using the NuPAGE® gel system from Invitrogen. These findings demonstrate that an internal polypeptide domain within a nascent chain can regulate eukaryotic translation by functioning as a *cis*-acting regulator of polypeptide elongation modulating ribosome movement in response to changes in the concentration of a small molecule.

2.2 Materials and Methods

2.2.1 DNA Templates and mRNA Synthesis

The plasmids designed to produce capped and polyadenylated synthetic mRNAs are listed in Table 2.1. They were derived from previous constructs by using described procedures (Fang *et al.*, 2002). These synthetic mRNAs encode polypeptides containing firefly LUC with *arg-2* AAP sequences fused in-frame to the N-termini (Figure 2.1), or polypeptides containing a domain from rabbit α-globin and a domain from LUC with AAP domains placed at both N-termini and between the globin and LUC domains (Figure 2.2). A string of eight additional methionine codons was added to the N-termini. Site-specific mutagenesis was used to remove every ATG codon (except for the nine at the N-termini) in the forward open reading frames in constructs containing double AAP domains (Figure 2.2). Plasmid DNA templates were purified by equilibrium centrifugation (Wang and Sachs, 1997a) or by Promega Wizard® midi-prep.

Capped and polyadenylated mRNA was synthesized with T7 RNA polymerase from plasmid templates linearized with *Eco*RI and the yield of mRNA was quantified as described (Fang *et al.*, 2000; Wang and Sachs, 1997a).

2.2.2 Cell-free Translation of Synthetic mRNA

The preparation of *N. crassa* cell-free extract and the reaction conditions for *in vitro* translation using *N. crassa* extracts were as described (Wu *et al.*, 2007). Translation reaction mixtures were programmed with synthetic mRNAs at a final concentration of 6 ng/μL; [^{35}S]Met was used at a final concentration of 0.5 μCi/μL. Radiolabeled translation products were examined by using 16% tricine SDS-PAGE gels as described (Schägger and von Jagow, 1987), or by using Invitrogen 12% NuPAGE® Bis-Tris gels in 1× MES SDS running buffer with 2.5 mM of DTT. Gels were dried and exposed to screens of a GE Healthcare Typhoon Trio^{+} PhosphorImager for ~24 h. All image data shown are representatives of multiple experiments.

2.3 Results

2.3.1 AAP-mediated Ribosome Stalling Is Associated with the Accumulation of Nascent AAP and the Delayed Synthesis of Full-length Polypeptide

In several cases it has been reported impossible to directly detect the [^{35}S]Met-labeled peptides encoded by short uORFs without immunoprecipitation or further purification (Cao and Geballe, 1996; Raney *et al.*, 2000). We were unable to detect the AAP with a single Met residue by [^{35}S]Met labeling *in vitro*. Since it has been established that the N-terminus of the AAP can be extended without affecting function, we constructed AAP coding regions with a string of eight additional methionine codons at the N-termini of both wild-type and nonfunctional D12N *N. crassa* AAPs (Fang *et al.*, 2002). The translation products of these uORFs are readily detectable on a tricine SDS-PAGE gel without immunoprecipitation (Fang *et al.*, 2002).

Previous toeprint data have shown that the AAP, when fused to the N-terminus of a reporter gene, can stall the ribosome at a location immediately following its C-terminus without a termination codon present (Wang *et al.*, 1998). To understand the relationship between ribosome stalling and the synthesis of nascent AAP, translation reactions were performed in cell-free systems with mRNAs specifying in-frame fusion of wild-type (AAP_w) or D12N AAP (AAP_m, in which the critical residue corresponding to *arg-2* AAP Asp12 was substituted by Asn), with extra Met residues at their N-termini, and firefly LUC coding regions (Figure 2.1). Equal amounts of capped and polyadenylated mRNAs were translated in *N. crassa* cell-free extracts containing different concentrations of arginine for 30 minutes and the [^{35}S]Met-labeled products were analyzed by using a 16% tricine SDS-PAGE gel (Figure 2.3). The major translation products had a migration consistent with the 585-residue full-length Met_9AAP-LUC polypeptide (Figure 2.3, open arrow). With a high amount of arginine present, mRNA encoding the Met_9AAP_w-LUC fusion also produced a short intermediate translation product whose migration was consistent with the 33-residue Met_9AAP peptide (Figure 2.3, closed arrow). The Met_9AAP_w peptide accumulation increased as the concentration of arginine was increased. As the amount of Met_9AAP_w peptide increased, the amount of full-length

Met_9AAP-LUC polypeptide decreased. There was no small Met_9AAP_m peptide detectable at any concentration of arginine in translation reactions with mRNA encoding Met_9AAP_m-LUC fusion, and the accumulation of full-length Met_9AAP_m-LUC polypeptide was not affected by the amount of arginine (Figure 2.3). These results are consistent with previous toeprint assay data indicating that increases in the Arg concentration in translation reactions caused more ribosome stalling when the Met_1AAP_w-LUC construct was translated (Wang *et al.*, 1999).

The relationship between the synthesis of Met_9AAP products and full-length Met_9AAP-LUC polypeptides was examined through time-course studies. Equal amounts of synthetic mRNAs were used to program *N. crassa* cell-free extracts containing Arg at low (10 μM) or high (2000 μM) concentrations. Aliquots were removed from translation at intervals and the [^{35}S]Met-labeled products were analyzed by using 16% tricine SDS-PAGE gels (Figure 2.4). In the translation reaction with mRNA encoding Met_9AAP_m-LUC fusion there was no small Met_9AAP_m product detectable at any time point. Full-length polypeptide synthesis was not affected by the amount of arginine and was essentially complete by 8 min (Figure 2.4B). In contrast, with mRNA encoding Met_9AAP_w-LUC fusion, the small Met_9AAP_w product was detected in the presence of high Arg. It was first observed after 2 min, reached its highest level at 6 – 8 min, and then decreased. Synthesis of full-length Met_9AAP_w-LUC product was delayed until 10 min in high Arg, suggesting ribosome stalling (Figure 2.4A).

These data suggest that the AAP-mediated ribosome stalling in response to high Arg is associated with the accumulation of Met_9AAP_w product and the delayed synthesis of full-length polypeptide.

2.3.2 Polypeptide Synthesis Appears to Resume after Stalling

To obtain further insight into the mechanism of nascent AAP-mediated ribosome movement, pulse-chase experiments in cell-free translation system were performed with a series of synthetic mRNAs designed to assess whether the AAP would function as an internal domain and to determine whether ribosomes resumed protein synthesis after stalling. These synthetic mRNAs encode fused polypeptides containing a domain from rabbit α-globin and a domain from firefly LUC. AAP sequences were placed (i) near

the N-termini following the eight additional Met residues and (ii) internally between the globin and LUC domains. All internally positioned Met residues were removed so that radiolabeled translation products would contain isotope only at their N-termini (Figure 2.2).

Equal amounts of synthetic mRNAs were used to program *N. crassa* cell-free extracts containing Arg at high (2000 μM) concentrations. Edeine (NIH/NSC #153112-K/L) was added to translation reactions at a final concentration of 1 mM after 2 min of incubation to block subsequent rounds of initiation. Aliquots were removed at intervals and analyzed with 16% tricine SDS-PAGE gels to monitor the polypeptide synthesis (Figure 2.5A and B). In the translation with mRNA containing two AAP_m coding regions there was no major intermediate product detected, consistent with the D12N AAP's inability to stall ribosomes. Full-length polypeptide synthesis was complete by 4 min and the product stopped accumulating after 5 min, indicating all translation initiated before the addition of edeine was completed (Figure 2.5B). In contrast, with mRNA containing two AAP_w coding regions, two intermediate peptides were observed with high Arg present, migrating with sizes consistent with products formed by stalled translation after synthesis of the N-terminal and internal AAP_w domains respectively (Figure 2.5A, closed and open triangles). In high Arg, synthesis of full-length polypeptide was delayed and the product continued accumulating after 12 min (Figure 2.5A, asterisk). Each wild-type AAP independently elicited stalling as determined by pulse-chase analyses with constructs containing a single AAP_w and a single AAP_m (data not shown). The toeprint assay used to directly map the positions of stalled ribosomes detected Arg-dependent stalling immediately downstream of AAP coding regions in translation reactions with mRNA containing AAP_w domains but not AAP_m domains (Fang *et al.*, 2004). This result verifies the intermediate products observed by [^{35}S]Met labeling from ribosome stalling as a consequence of synthesizing each AAP_w domain.

Since all nascent peptides only have their N-termini labeled with [^{35}S]Met during incubation before edeine addition, the radiolabel incorporated into the short N-terminal intermediate products should be quantitatively recovered in full-length polypeptides at later time points if translation resumed after ribosome stalling. This is supported by

quantitative analysis of radiolabel in major intermediate translation products between 2- and 12-min incubation, showing radiolabel is conserved during conversion of small N-terminal product to full-length polypeptide with a transient accumulation of intermediate products (Figure 2.5C).

2.3.3 The Arrested Nascent AAP Appears to Be in Peptidyl-tRNA Form

Since the short products are intermediates resulting from stalled translation and ribosomes resume elongation after stalling, these short products may be present as peptidyl-tRNAs. The observation that these products but not the full-length polypeptides were associated with the stalled ribosomes (data not shown) also suggests these are peptidyl-tRNAs trapped in the exit tunnel of the stalled ribosomes. The traditional alkaline gel system such as tricine SDS-PAGE could destroy the peptidyl-tRNA ester bonds, thus making it impossible to directly detect whether a nascent translation product is in peptidyl-tRNA form or is a released peptide. Therefore, we switched to the neutral pH NuPAGE® gel system which should preserve the ester bonds (Muto *et al.*, 2006). Pulse-chase experiments were performed as described above with Arg at low (10 μM) and high (2000 μM) concentrations. Radiolabeled products from different time points were examined by using 12% NuPAGE® Bis-Tris gels (Figure 2.6A).

In translation reactions programmed with mRNA containing AAP_w domains (Figure 2.6A, WT-WT), the delayed appearance of full-length products was observed as usual in the presence of high Arg. Products corresponding to intermediates formed by stalled translation after synthesis of the N-terminal and internal AAP_w domains were not detected in either low or high Arg. Instead, products indicated by arrows were detected. Several observations provide evidence that these products represent the elongation-arrested peptidyl-tRNA forms of intermediates resulting from translation stalling after synthesis of N-terminal and internal AAP_w domains. First, upon RNaseA treatment, these products, but not the full-length polypeptides, disappeared with concomitant appearance of products whose migrations were consistent with the predicted mass of the intermediate products observed in Figure 2.5A (Figure 2.6B, WT-WT), suggesting the upshift in migration is due to their attachment to the ~20 kDa tRNA.

Second, these products were downshifted to the same relative positions observed in 16% tricine SDS-PAGE gels when aliquots of the same samples were analyzed with the traditional alkaline gel system (Figure 2.6C). Third, the products indicated by the upper arrow accumulate faster in low Arg than in high Arg, and the amount of these products decrease as the full-length polypeptides start to accumulate. This is consistent with the finding that the gain of full-length products is inversely correlated to the loss of small products, with transient accumulation of intermediate products by stalled translation following the internal AAP_w domain. Finally, other data show that these products but not the full-length polypeptides could be precipitated with cetyltrimethylammonium bromide (CTABr), a reagent that precipitates nucleic acids (Fang *et al.*, 2004).

In translation reactions programmed with mRNA containing AAP_m domains, similar products were also detected with 12% NuPAGE® Bis-Tris gels (Figure 2.6A, D12N-D12N). But nothing similar to the intermediate products observed in the case of AAP_w was produced by RNaseA treatment (Figure 2.6B, D12N-D12N) or by analyzing translation products with 16% tricine SDS-PAGE gels (Figure 2.6C). Based on the fact that these products appear to be a variety of species on the gel, they might be peptidyl-tRNA forms of non-specific intermediate products caught during elongation when incubation was stopped.

All these results suggest that (i) peptide synthesis resumes after ribosome stalling at a wild-type AAP; (ii) the products associated with ribosome stalling are arrested peptidyl-tRNA intermediates in ribosomes instead of free peptides released from ribosome after translation is complete.

2.4 Discussion

In previous studies, it was demonstrated that the AAP negatively regulates the expression of downstream genes by causing ribosome stalling at its termination codon when it is encoded by a uORF, or at the C-terminal region immediately following the AAP coding region when it is encoded by an N-terminal coding motif in the mRNA (Wang *et al.*, 1998). The amino acid sequence of the AAP, not the nucleotide sequence of its coding region, was responsible for regulation (Fang *et al.*, 2000). Here we provide

evidence supporting the conclusion that the nascent AAP is an independent elongation-arrest element that causes regulated stalling of eukaryotic ribosomes in response to Arg when placed either upstream or downstream of two different domains. Pulse-chase analysis of the radiolabel products obtained during cell-free translation directly indicates that the ribosome stalling responsible for the delayed synthesis of the full-length polypeptide is associated with the accumulation of the nascent AAP. That the radiolabel contained in these short peptides as a consequence of AAP-mediated stalling is quantitatively recovered in the full-length polypeptides suggests these are intermediates in polypeptide synthesis and translation is resumed after the ribosome stalling. These intermediates are confirmed to be in their peptidyl-tRNA form transiently associated with the ribosome.

The nascent AAP-tRNA trapped in the ribosome exit tunnel may interact with ribosomal components to constrain the nascent peptide movement, thereby pausing the elongation, or to directly affect the PTC activity by changing ribosome conformation. The toeprint data indicate that the stalling occurs when the C-terminal residue of the AAP is at or near the ribosome P site. This would place the critical Asp12 residue (Asp13 in *S. cerevisiae*) at a location ~12-residue from the P site, which happens to be near the narrowest section of the exit tunnel (Ban *et al.*, 2000; Berisio *et al.*, 2003; Cruz-Vera *et al.*, 2005; Nakatogawa and Ito, 2002). Amino acid residues in the polypeptide that are identified to be crucial for stalling prokaryotic ribosomes during the synthesis of SecM and TnaC are also located ~12-residue from the arrested ribosome's P site. Ribosomal L22 protein in this narrowest constriction may act as a discrimination gate which interacts with the nascent peptide to pause ribosomes in elongation (Cruz-Vera *et al.*, 2005; Nakatogawa and Ito, 2002). Studies on the AAP suggest that the eukaryotic ribosome might have a similar constriction gate in its exit tunnel.

The elongation-arrest caused by the nascent peptide as an internal domain within a polypeptide in concert with a small molecule can contribute to the control of gene expression in many different ways. The regulated instability of the *A. thaliana CGS1* transcript in response to the availability of methionine or *S*-adenosyl-L-methionine is coupled with the synthesis of a specific internal MTO1 domain. Translational stalling as a consequence of synthesis of this domain is responsible for this regulatory effect (Chiba

et al., 2003; Lambein *et al.*, 2003; Onouchi *et al.*, 2005). A translational pause that occurs when a ribosome encounters a rare codon during the translation of *c-myc* mRNA facilitates binding of a factor that regulates mRNA stability (Lemm and Ross, 2002). In *E. coli*, ribosome-arrest at the C-terminal of nascent TnaC mediated by free tryptophan inhibits transcription termination so downstream genes will be expressed to hydrolyze tryptophan (Gong and Yanofsky, 2001). RNA structures such as pseudoknots can contribute to ribosome pausing during elongation and thus contribute to programmed frame-shifting (Kontos *et al.*, 2001; Lopinski *et al.*, 2000; Somogyi *et al.*, 1993). Thus regulated stalling mediated by internal polypeptide domains not only regulate gene expression at the translational level but also could potentially influence mRNA transcription, mRNA stability, or frame-shifting.

How arginine functions in the ribosome stalling event is still unclear. Studies on *E. coli* TnaC, which can stall ribosomes in response to tryptophan, have revealed that free tryptophan functions by occupying the A site (Gong and Yanofsky, 2002). Substituting the termination codon with a Trp codon causes constitutive stalling. The aminoacylated tRNA appears to place tryptophan in the proper spot to exert its effect. If this is the same mechanism underlying the arginine and AAP mediation of ribosome stalling, replacement of the stop codon of AAP with an Arg codon would have the same effect observed in TnaC. Data from our laboratory indicate that this substitution is unable to cause ribosome stalling in the *N. crassa* system unless high Arg is present. Arginine may directly interact with the nascent AAP and/or the translational machinery to cause conformational changes that control polypeptide synthesis. Alternatively, it may interact with the ribosome or associated translation factors to render the ribosome sensitive to the nascent peptide sequence. Other small molecules have been shown to be capable of acting in concert with nascent leader peptides and/or directly interacting with mRNA to control translational events (Geballe and Sachs, 2000; Morris and Geballe, 2000; Raney *et al.*, 2002; Sachs and Geballe, 2002). Study on the control of fungal translation by the AAP in response to arginine provides an example of how a nascent polypeptide and an amino acid can affect eukaryotic gene expression.

Table 2.1 Constructs used in this research.

Construct	Structure
pKL401	Met_9AAP_w-LUC fusion (WT)
pKLS401	Met_9AAP_m-LUC fusion (D12N)
pGL201	Met_9AAP_w-globin-AAP_w-LUC fusion (WT-WT)
pGL202	Met_9AAP_w-globin-AAP_m-LUC fusion (WT-D12N)
pGL203	Met_9AAP_m-globin-AAP_w-LUC fusion (D12N-WT)
pGL204	Met_9AAP_m-globin-AAP_m-LUC fusion (D12N-D12N)

Plasmids are derived from previous constructs by using described procedures (Fang *et al.*, 2002). Site-specific mutagenesis was used to remove every ATG codon (except for the nine at the N-termini) in the forward open reading frames.

The rabbit α-globin domain used was obtained by PCR from plasmid pSPα (Jobling and Gehrke, 1987) (from Dr. Uttam L. RajBhandary at MIT).

pKL401 is a wild-type Met_9AAP-LUC fusion construct, originally named pKL205 (Fang *et al.*, 2002).

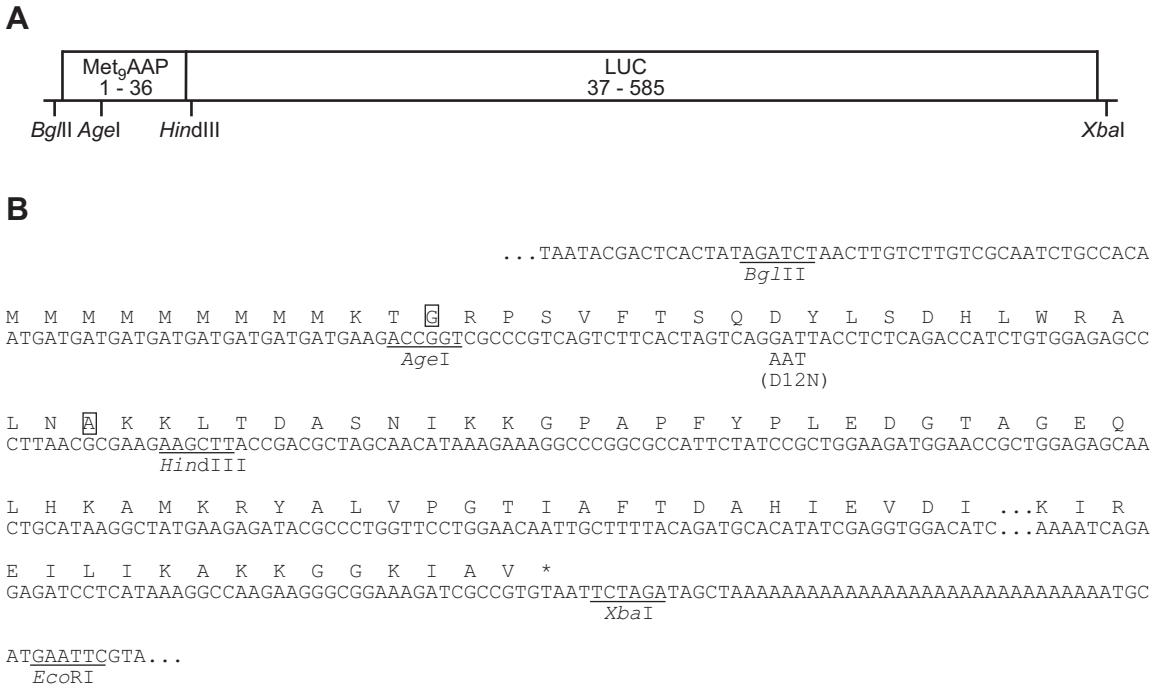

Figure 2.1 Schematic of the Met$_9$AAP-LUC transcripts.

(A) The coding region of the Met$_9$AAP-LUC fusion polypeptide. The *N. crassa arg-2* AAP coding region was fused in-frame to the N-terminus of firefly LUC (from its third orginal codon) with three bridge codons (codons 34 – 36 in fusion polypeptide) to separate these two domains. Nine Met codons were present at the extreme N-terminus. (B) Sequences of the DNA templates for making Met$_9$AAP-LUC mRNA. The sequence shown begins with the T7 RNA polymerase binding site and ends with the *Eco*RI site that follows the poly(A) tail, which is used for linearizing the DNA template. The amino acid sequence of the Met$_9$AAP-LUC fusion is indicated. The N- and C-terminal amino acid residues corresponding to the AAP domain are boxed. The D12N mutation, which eliminates regulation, is shown below the wild-type sequence. The restriction sites in this region are underlined; names of the corresponding restriction enzymes are indicated below the underlined sequences.

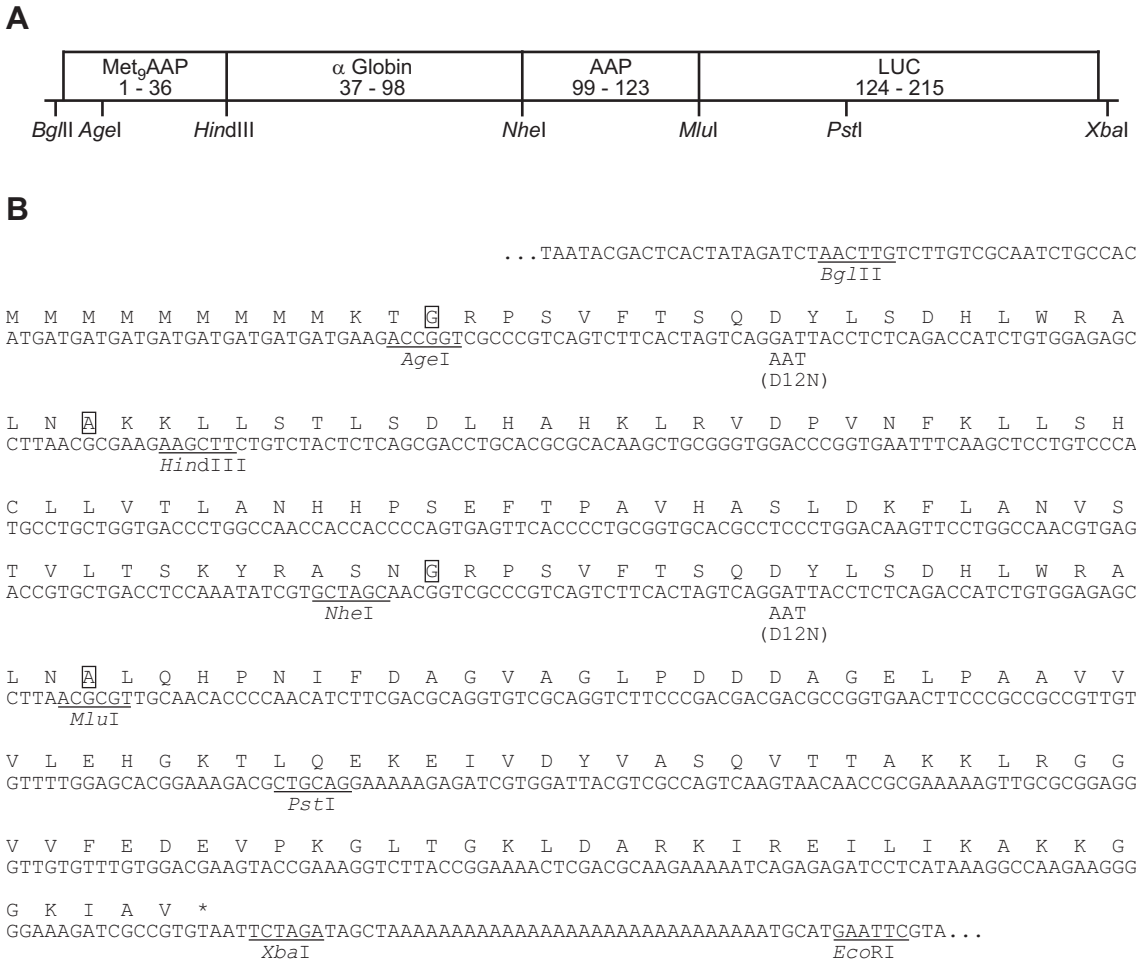

Figure 2.2 Schematic of the Met_9AAP-globin-AAP-LUC transcripts.

(A) The coding region of the Met_9AAP-globin-AAP-LUC fusion polypeptide. The polypeptide contains a domain from rabbit α-globin (codons 80 – 142 in the original context; codons 37 – 98 in the fusion polypeptide) and a domain from firefly LUC (codons 460 – 551 in the original context; codons 124 – 215 in the fusion polypeptide). *Neurospora crassa arg-2* AAP coding regions were placed at two different positions: one at the N-terminus of α-globin, and the other internally between the globin and LUC domains. Nine Met codons were placed at the extreme N-terminus. All Met codons in the forward open reading frame were removed, and so were out-of-frame ones. We constructed all possible combinations containing wild-type AAP (AAP_w) and nonfunctional D12N AAP (AAP_m) at either or both positions. (B) Sequences of the DNA templates for synthesizing Met_9AAP-globin-AAP-LUC transcripts. The sequence

shown begins with the T7 RNA polymerase binding site and ends with the *Eco*RI site that follows the poly(A) tail, which is used for linearizing the DNA template. The amino acid sequence of the Met_9AAP-globin-AAP-LUC fusion is indicated. The N- and C-terminal amino acid residues corresponding to each AAP domain are boxed. The D12N mutation, which eliminates regulation, is shown below the wild-type sequence. The restriction sites in this region are underlined; names of the corresponding restriction enzymes are indicated below the underlined sequences.

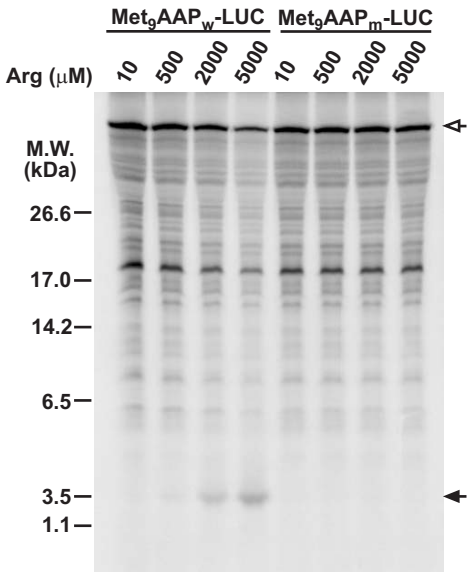

Figure 2.3 Effect of increasing Arg concentration on translation of Met_9AAP-LUC transcripts in *N. crassa* cell-free extracts.

Indicated synthetic mRNAs were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, 10 μM of all 19 amino acids, and Arg at 10 μM, 500 μM, 2000 μM or 5000 μM, and were incubated for 30 min at 25°C. Reactions were stopped by adding an equal volume of 2× SDS loading buffer and examined by using a 16% tricine SDS-PAGE gel. Positions of protein products corresponding to the full-length Met_9AAP-LUC and the short Met_9AAP are indicated by open and closed arrows respectively.

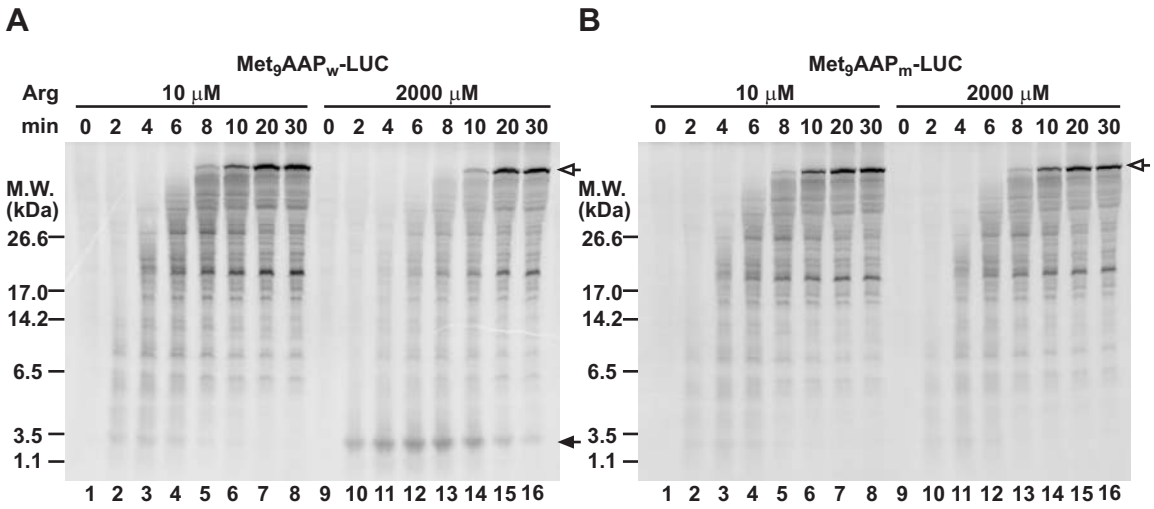

Figure 2.4 Time-course translation of Met_9AAP-LUC transcripts in *N. crassa* cell-free extracts.

Synthetic mRNAs specifying fused polypeptides containing (A) the wild-type AAP or (B) the nonfunctional D12N AAP were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL in the presence of either 10 μM or 2000 μM of Arg as indicated and 10 μM of the other 19 amino acids. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, and were incubated at 25°C. 10-μL aliquots were removed into an equal volume of 2× SDS loading buffer at the indicated time points, and were subsequently examined bt using 16% tricine SDS-PAGE gels. Positions of translation products corresponding to the full-length Met_9AAP-LUC and the short Met_9AAP are indicated by open and closed arrows respectively.

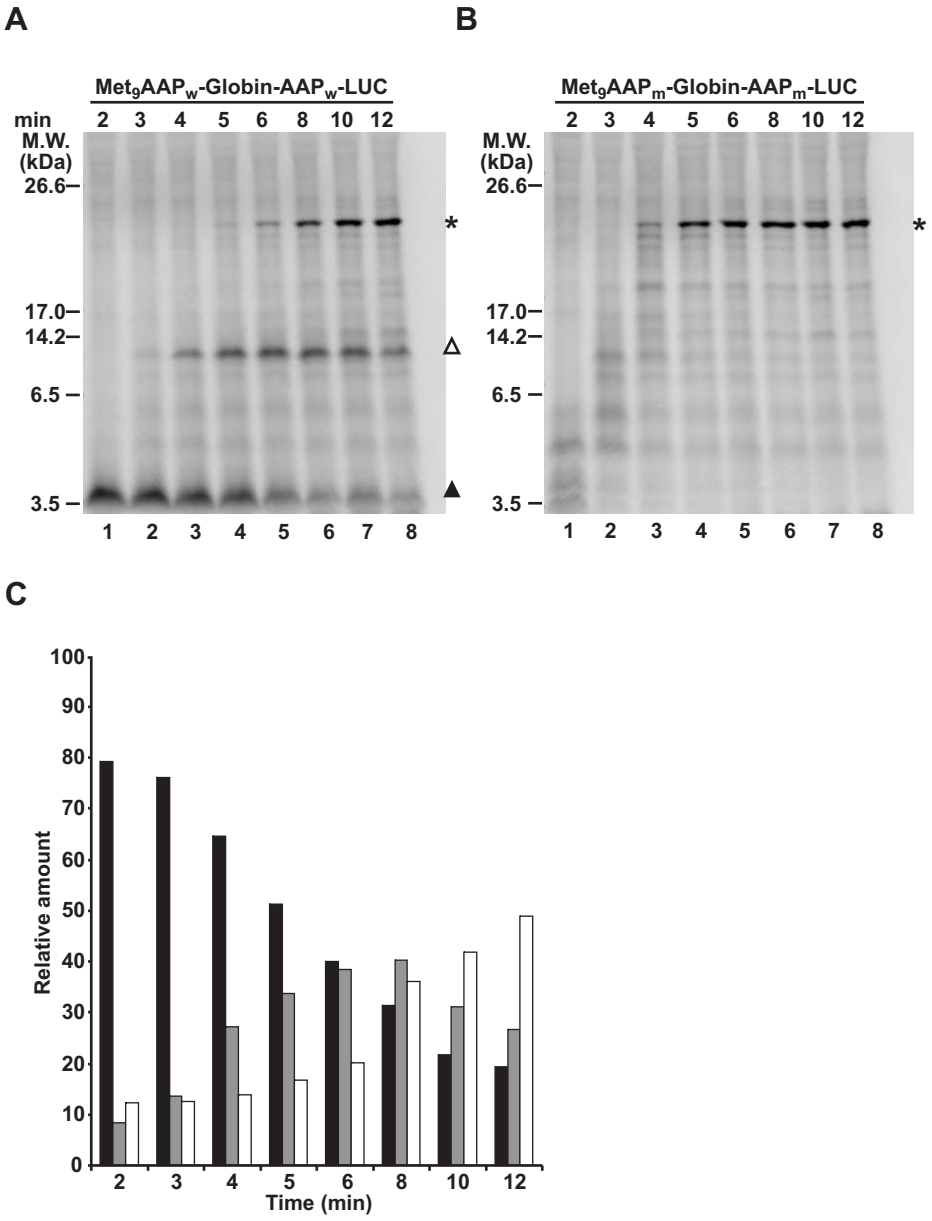

Figure 2.5 Pulse-chase analyses of polypeptide synthesis in *N. crassa* cell-free extracts. Synthetic mRNAs specifying Met_9AAP-globin-AAP-LUC fusions containing (A) the wild-type AAP or (B) the nonfunctional D12N AAP at both N-terminal and internal locations were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, 2000 μM of Arg, 10 μM of the other 19 amino acids, and were incubated at 25°C. Edeine was added at a final concentration of 1 mM at 2 min to block subsequent rounds of initiation. 10-μL aliquots were removed into 2 μL of 5× SDS loading buffer

at the indicated time points, and were subsequently examined by using 16% tricine SDS-PAGE gels. Positions of major translation products corresponding to the full-length Met_9AAP-globin-AAP-LUC, the intermediate Met_9AAP-globin-AAP and the short Met_9AAP are indicated by asterisk, open and closed triangles, respectively. (C) Quantitative analysis of translation products obtained from data in A. The radiolabel in each product was determined by ImageQuant 5.2 (Molecular Dynamics). The total amount of radiolabel in three major products at 2 min was normalized to 100%, and the radiolabel in each band at each time point was calculated as a fraction of this value. Black, radiolabel in N-terminal products corresponding to Met_9AAP; gray, radiolabel in intermediate products corresponding to Met_9AAP-globin-AAP; white, radiolabel in full-length products.

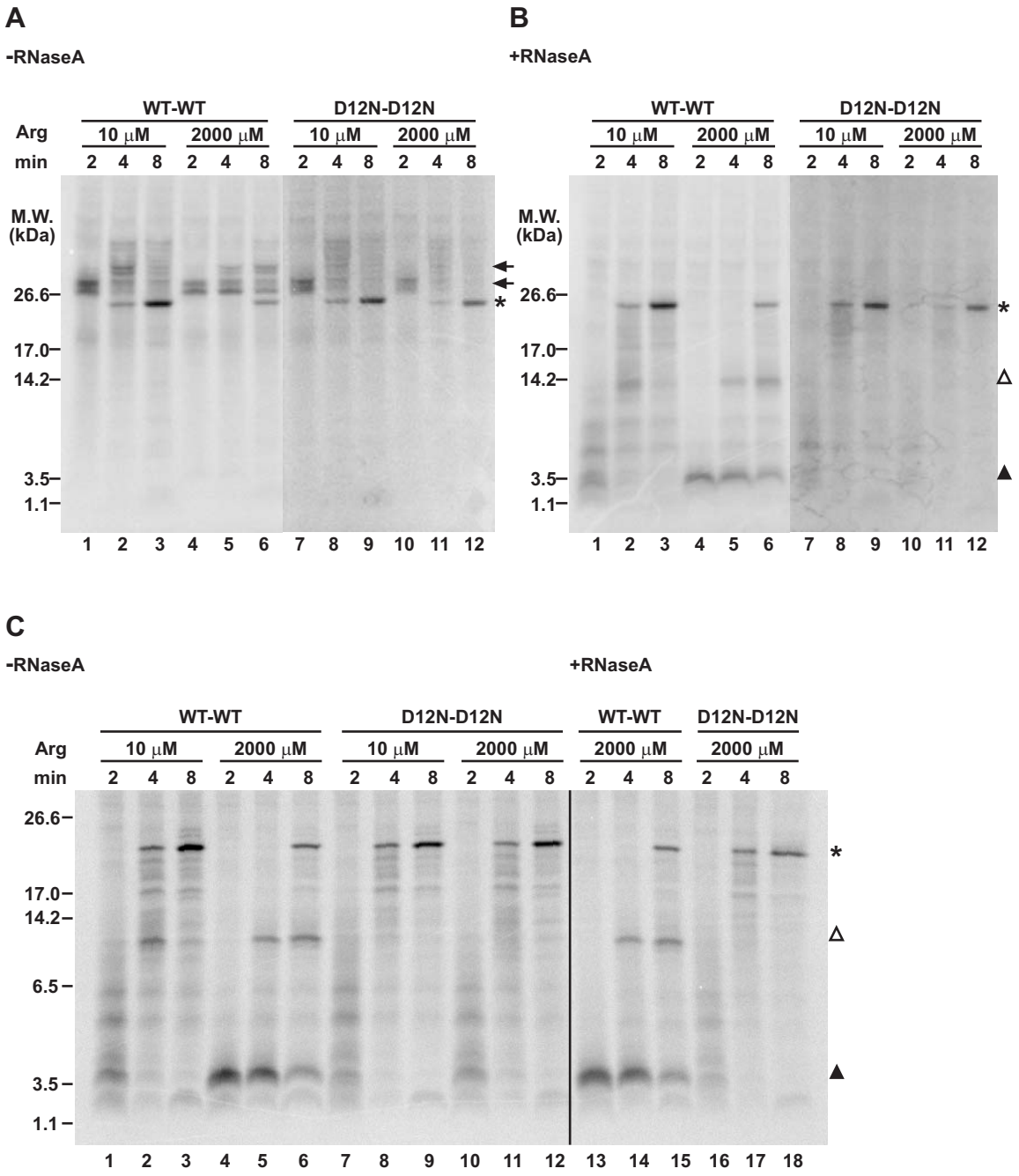

Figure 2.6 *In vitro* reproduction of AAP elongation arrest.

(A) Synthetic mRNAs specifying Met_9AAP-globin-AAP-LUC fusions were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL in the presence of low (10 μM) or high (2000 μM) Arg. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, 10 μM of the other 19 amino acids, and were incubated at 25°C. Edeine was added at a final concentration of 1 mM at 2 min to block

subsequent rounds of initiation. 10-μL aliquots were removed into 2 μL of 5× SDS loading buffer at the indicated time points and half of each sample was examined by using 12% NuPAGE® Bis-Tris gels in 1× MES SDS running buffer containing 2.5 mM of DTT. Positions of products appear to be arrested peptidyl-tRNA intermediates are indicated by arrows. (B) The other half of each aliquot was treated with 10 units of RNaseA for 2 min at room temperature and then examined by using 12% NuPAGE® Bis-Tris gels. (C) Replicate experiment was examined by using traditional 16% tricine SDS-PAGE gels. Positions of major translation products corresponding to the full-length Met_9AAP-globin-AAP-LUC, the intermediate Met_9AAP-globin-AAP and the short Met_9AAP are indicated by asterisk, open and closed triangles, respectively.

CHAPTER 3 THE FMDV 2A-MEDIATED CLEAVAGE IS A TRANSLATIONAL RECODING EVENT SEPARATING THE NASCENT POLYPEPTIDE CHAIN*

3.1 Introduction

Positive-strand RNA viruses typically encode multiple proteins in a single long ORF. Polyproteins translated from this single ORF will be processed by proteinases into distinct mature proteins by viral or host-encoded proteinases. The processing may be co-translational (in *cis*) or post-translational (in *trans*) (Ryan *et al.*, 1998; Seipelt *et al.*, 1999). In picornaviruses, the polyprotein undergoes a co-translational cleavage mediated by the 2A region within the polyprotein itself. The 2A region in entero- and rhinoviruses has been characterized as a proteinase that acts in *cis* to cleave the nascent polyprotein at its own N-terminus (Sommergruber *et al.*, 1989). The 2A region in aphtho- and cardioviruses mediates 2A/2B cleavage at its C-terminus, and the amino acid sequence of this 2A region shows no similarity to proteinase 2A in entero- and rhinoviruses or any other characterized proteinases (Donnelly *et al.*, 1997).

In aphthovirus FMDV, the 2A region is only 18-residues long (Ryan *et al.*, 1991), and is functional when placed between various reporter proteins, suggesting no viral sequences outside this region are required for the cleavage (Ryan and Drew, 1994). The amino acid sequence of 2A (together with the N-terminal proline residue of 2B), but not the nucleotide sequence encoding it is essential for the cleavage (Donnelly *et al.*, 2001b). 2A and 2A-like sequences identified in other viral species are all active and share the conserved DXEXNPG⇓P motif (⇓ indicates the cleavage site) (de Felipe *et al.*, 2006; Donnelly *et al.*, 2001a). The FMDV 2A region is active in a wide range of eukaryotic

*Collaboration with the laboratories of Dr. Jeremy D. Brown at Newcastle University and Dr. Martin D. Ryan at the University of St. Andrews. Part of this material in this or similar form is contained in the manuscript entitled *A translational recoding event that generates separated proteins from one open reading frame* by Doronina, V. A., Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D., and Brown, J. D..

systems (Ryan and Drew, 1994), making it a powerful tool for coordinated protein expression (de Felipe *et al.*, 2006). Synthesis of polyproteins containing the 2A region in eukaryotic cell-free translation systems showed that the uncleaved products (~10% of total products) cannot be converted into cleaved products posttranslationally, indicating that the cleavage mediated by 2A occurs only co-translationally (Ryan and Drew, 1994).

An imbalance in the accumulation of products has been noticed in cell-free translation systems. Proteins encoded upstream of 2A accumulate to higher levels than those encoded downstream of 2A (Donnelly *et al.*, 2001b). Puromycin preferentially incorporates into the N-terminal product ending with the 2A region, indicating a significant population of ribosomes is paused at the C-terminus of 2A (Donnelly *et al.*, 2001b). These results suggest that 2A-mediated cleavage is not a proteolytic process. It has been suggested that the nascent 2A peptide interacts with the ribosome exit tunnel to modify the activity of the elongating ribosome. This change inhibits the formation of the next peptide bond and promotes the hydrolysis of the ester bond between peptide and $tRNA^{Gly}$, thereby releasing the N-terminal peptide (Donnelly *et al.*, 2001a; Donnelly *et al.*, 2001b; Ryan *et al.*, 1999). Thus the FMDV 2A-mediated cleavage is termination and reinitiation occurring at a normal sense codon. The ribosome skips a codon to the next without forming a peptide bond (de Felipe *et al.*, 2006; Donnelly *et al.*, 2001b; Ryan *et al.*, 1999). The imbalance in the accumulation of N-terminal and C-terminal products implies that some ribosomes dissociate from the mRNA after the peptidyl-tRNA hydrolysis while others continue translating the downstream protein. Consistent with this model, the cleavage has been proved to be intraribosomal. During co-translational translocation, the released C-terminal protein of 2A-containing polypeptide remains in the cytosol unless it contains its own signal sequence, indicating the cleavage occurs before the nascent C-terminal protein triggers translocation and therefore within the ribosome (de Felipe *et al.*, 2003).

Here we provide more evidence supporting the proposed model. Consistent with the puromycin incorporation data, a ribosome pause is detected at the C-terminus of 2A when polypeptide containing the wild-type 2A region is translated in *N. crassa* cell-free extracts. A pulse-chase experiment indicates that the majority of the N-terminal products are free peptides rather than in peptidyl-tRNA form when the ribosome is

paused, indicating the occurrence of peptidyl-tRNA hydrolysis. The simultaneous appearance of the peptidyl-tRNA and the released peptide suggests that the hydrolysis of the ester bond is very rapid, if the hydrolysis and ribosome pausing do not occur simultaneously.

3.2 Materials and Methods

3.2.1 DNA Templates and mRNA Synthesis

The plasmids designed to produce capped and polyadenylated synthetic mRNAs (from Dr. Jeremy D. Brown at Newcastle University) are listed in Table 3.1. These synthetic mRNAs encode polypeptides containing a domain from yeast prepro-α-factor without signal sequence and a domain from bovine preprolactin. Either a functional FMDV 2A domain (together with the 2B proline residue, designated Pro19) or an inactive 2A* domain (in which Pro17 was substituted by Ala) was placed between these two domains. To provide a positive control, the *S. cerevisiae CPA1* AAP domain (as an *Xba*I-*Apa*I fragment) was placed in the same context as 2A. Plasmid DNA templates were purified by equilibrium centrifugation (Wang and Sachs, 1997a) or by Promega Wizard® midi-prep.

Capped polyadenylated RNA was synthesized with SP6 RNA polymerase from plasmid templates linearized with *Sal*I and the yield of RNA was quantified as described (Fang *et al.*, 2000; Wang and Sachs, 1997a).

3.2.2 Cell-free Translation of Synthetic mRNA

The preparation of *N. crassa* and *S. cerevisiae* cell-free extract and the reaction conditions for *in vitro* translation using *N. crassa* and *S. cerevisiae* extract were as described (Wu *et al.*, 2007). Translation reaction mixtures were programmed with synthetic mRNAs at a final concentration of 6 ng/μL; [^{35}S]Met was used at a final concentration of 0.5 μCi/μL for pulse-chase analyses. Radiolabeled translation products were examined by using 16% tricine SDS-PAGE gels as described (Schägger and von

Jagow, 1987), or by using 12% NuPAGE® Bis-Tris gels in 1× MES SDS running buffer containing 2.5 mM of DTT. The gels were dried and exposed to screens of a GE Healthcare Typhoon Trio+ PhosphorImager for ~24 h. All image data shown are representatives of multiple experiments.

3.2.3 Primer Extension Inhibition (Toeprint) Assay

Toeprint assays were accomplished as described (Wang and Sachs, 1997b) using a ^{32}P-labeled primer oCW100 (5′-GGACTTCATGGTGGGTCTG-3′). For each sample, 8 μL of the aqueous phase was extracted with phenol/chloroform, mixed with 6 μL of loading buffer, and half of the mixture was analyzed with denaturing 6% polyacrylamide gels. Gels were dried and exposed to screens of a GE Healthcare Typhoon Trio+ PhosphorImager for ~24 h. All image data shown are representatives of multiple experiments.

3.3 Results

3.3.1 Recapitulation of 2A-mediated Cleavage in *N. crassa* Cell-free Extract

Synthetic mRNAs were prepared that specified polypeptides containing domains from yeast alpha factor and bovine prolactin. Functional FMDV 2A or the inactive 2A* coding region was fused in-frame between these two domains. A positive control for ribosome pausing was provided by placing a sequence encoding *S. cerevisiae CPA1* AAP in the same context as 2A (Table 3.1 and Figure 3.1). When implanted into a longer polypeptide, the AAP directs a strong pause in translation in response to a high concentration of arginine (Fang *et al.*, 2004).

Equal amounts of synthetic mRNAs were used to program *N. crassa* cell-free extracts. Edeine was added at a final concentration of 1 mM to translation reactions after 2 min of incubation to block subsequent rounds of initiation. Aliquots were removed at intervals and analyzed with the tricine SDS-PAGE gel system to monitor the polypeptide synthesis. When the AAP-containing mRNA was used to program the *N.*

crassa cell-free extracts, the arginine-induced pause was revealed by the appearance of a translation intermediate migrating at a position consistent with the predicted mass of synthesis stalled after the AAP and a delay in the appearance of full-length product (Figure 3.2A). Analysis of translation products generated from the mRNA containing 2A revealed that the 2A-mediated cleavage was recapitulated in the *N. crassa* cell-free extract, producing separated N- and C-terminal fragments of the predicted sizes (Figure 3.2B). A small proportion of full-length product was also observed, which is consistent with data from previous experiments (Ryan and Drew, 1994). Time-course analyses revealed that the N-terminal product accumulated first, prior to the accumulation of either the C-terminal or full-length products. This result is consistent with the requirement for synthesis of the N-terminal product for the C-terminal product to be synthesized (Donnelly *et al.*, 2001b). As expected, translation of the synthetic mRNA containing inactive 2A* mutant gave rise only to the full-length product.

3.3.2 A Pause in Translation at the C-terminus of 2A when Glycyl-tRNA^{Gly} in the Ribosome P Site

A single major species corresponding to the N-terminal product ending with 2A was detected in the previously published puromycin incorporation studies, indicating that ribosomes paused at the end of 2A (Donnelly *et al.*, 2001b). To explore this further, primer extension inhibition (toeprint) analysis (Sachs *et al.*, 2002) was used to map the position of ribosomes translating mRNAs encoding 2A- or 2A*-containing polypeptide. Due to the size of the ribosome, toeprint signals corresponding to ribosomes accumulating at certain positions on the mRNA are ~16-nucleotides downstream distal to the first base of the codon in the ribosome P site.

To provide a positive control for ribosome pausing, the *CPA1* AAP-containing mRNA was used to program *N. crassa* cell-free extracts in the presence of low (10 μM) or high (2000 μM) Arg. Primer extension on this mRNA produced arginine-inducible toeprint signals corresponding to ribosomes stalled at positions immediately downstream of the AAP coding region (Figure 3.3; lanes 1 and 4). These toeprint signals were not significantly strengthened by addition of cycloheximide after 10 min of translation (Figure 3.3; lanes 4 and 6), but were eliminated when cycloheximide was added at time 0

to block translation elongation events subsequent to initiation (Figure 3.3, lane 5). These data are consistent with previous analyses (Fang *et al.*, 2004; Wang *et al.*, 1999) and the pulse-chase experiment (Figure 3.2A).

Translation of the 2A-containing mRNA yielded a toeprint signal in the absence of cycloheximide (Figure 3.3, lane 9). This signal was ~16-nucleotides downstream of the first base of the Gly18 codon, ribosomes translating 2A-containing mRNA thus pause at the C-terminus of 2A with the Gly18 codon in the P site and the Pro19 codon in the A site. The signals were translation-specific, as they were not seen when the reaction was initiated in the presence of cycloheximide or in the absence of extract (Figure 3.3; lanes 10 and 12). No signals were detected at these positions with the inactive 2A*-containing mRNA (Figure 3.3; lanes 13 – 15) and these therefore appeared specific to the synthesis of functional 2A peptide. The 2A-specific signals were not significantly enhanced when cycloheximide was added to the reactions (Figure 3.3, lane 11), suggesting that ribosomes paused at 2A are in the post-translocation state. The observation that ribosome pause at the FMDV 2A with its Gly18 codon in the P site is consistent with the 2A-mediated cleavage generating an N-terminal product ending with the glycine residue (Donnelly *et al.*, 2001b; Ryan *et al.*, 1991). Mass spectrometry analyses of peptides derived from the N-terminal products of the 2A-mediated cleavage only detected a peptide ending at Gly18 (data not shown). All these data confirm the 2A peptide sequence and place the termination position of 2A at Gly18.

3.3.3 The Majority of the N-terminal Products Were Free Peptides Resulting from Termination

As described in section 2.3.3, when the ribosome stalls after synthesizing the wild-type AAP region, the nascent AAP is in its peptidyl-tRNA form arrested in the exit tunnel. When translation resumes the AAP will be incorporated in the full-length product. In the case of 2A-mediated cleavage, a similar ribosome pause signal at the cleavage site was detected by toeprint assay (Figure 3.3). But, unlike the case for the AAP, hydrolysis of peptidyl-tRNA is hypothesized to happen. Would the N-terminal product corresponding to the pause be detected as a released peptide resulting from termination, or as a translation intermediate in its peptidyl-tRNA form waiting for the

recruitment of release factors? To answer this question, pulse-chase experiments were performed as described above. Radiolabeled products from different time points were examined by using 12% NuPAGE® Bis-Tris gels (Figure 3.4), in which peptidyl-tRNA species are stable.

In translation reactions programmed with mRNA containing the *CPA1* AAP domain, the delayed appearance of full-length products was observed as usually in the presence of high Arg. The intermediate product resulting from ribosome stalling after synthesizing AAP was detected in its peptidyl-tRNA form, as an upshift in migration was detected due to its attachment to the tRNA. Upon RNaseA treatment, the peptide ending with the AAP domain was released from tRNA. Thus, the migration of this product was downshifted to the position consistent with its predicted mass (Figure 3.4A and B; lanes 10 – 12).

In translation reactions programmed with mRNA containing a functional 2A domain, a signal potentially corresponding to peptidyl(2A)-tRNA was also detected. However, the majority of the N-terminal product is released peptide whose migration was not affected by RNaseA treatment (Figure 3.4A and B; lanes 1 – 3). This result suggests that the hydrolysis of peptidyl-tRNA and the release of the N-terminal product are happening when ribosome is paused at the cleavage site.

To obtain a better understanding of the relationship between ribosome pause and hydrolysis, pulse-chase experiments in a short time course were performed with 2A- and 2A*-containing mRNAs. Edeine was added at a final concentration of 1 mM to translation reactions after 30 sec of incubation to block subsequent rounds of initiation. Aliquots were removed at intervals and analyzed by using 12% NuPAGE® Bis-Tris gels (Figure 3.4C and D). The accumulation of the peptidyl(2A)-tRNA was not observed earlier than the accumulation of released N-terminal product, suggesting that either the hydrolysis is very rapid or that hydrolysis and ribosome pausing occur simultaneously. In both long and short pulse-chase experiments, only the single product corresponding to the full-length polypeptide was observed when mRNA containing inactive 2A* mutant was used to program *N. crassa* cell-free extracts, consistent with the sequence requirement for producing two peptides from one coding region.

3.4 Discussion

A model of the 2A-mediated cleavage consistent with the data from this and previous studies is presented in Figure 3.5. In pathway (A) the nascent 2A peptide interacts with the ribosome exit tunnel and PTC causing a pause in translation with the final Gly18 codon of 2A in the P site. This interaction may also modify the PTC activity, leading to the recruitment of release factors to hydrolyze the peptidyl-tRNA, thereby releasing the N-terminal peptide ending with 2A. In this case, the Pro19 codon in the A site is recoded as a termination codon. Following the dissociation of release factors from the ribosome, translation of the downstream product is reinitiated by the entry of prolyl-tRNA^{Pro} into the A site, and a translocation reaction that moves it to the P site. In pathway (B), the nascent peptide does not engage with the ribosome and translation proceeds as normal with incorporation of Pro19 into the growing nascent chain.

After the hydrolysis is complete, the ribosome may continue translation from the Pro19 codon or dissociate from the mRNA, depending on whether the ribosome possesses all the factors and energy that are required for the ingress of prolyl-tRNA^{Pro} and translocation. A large excess of products upstream of 2A observed using *in vitro* translation systems (Donnelly *et al.*, 2001b) may be due to the limited availability of translation elongation factors such as eEF2, which provides the driving force for translocation by hydrolyzing the GTP bound to it (Rodnina *et al.*, 1997; Taylor *et al.*, 2007).

The toeprint signal corresponding to ribosome pause at 2A cleavage site was not significantly affected by the presence of cycloheximide (Figure 3.3), consistent with the idea that ribosomes that generated them are in the post-translocation state, available to interact with release factors, prolyl-tRNA^{Pro}, or puromycin as observed previously (Donnelly *et al.*, 2001b). The pause is consistent with determination of the C-terminal amino acid in the N-terminal product of 2A cleavage as Gly18 using mass spectrometry. All codons of 2A are decoded and the corresponding amino acids are present in the final translation products. Therefore, the 2A-mediated cleavage is not due to special ribosomal movement on the mRNA as in frame-shifting and shunting events.

Pulse-chase of translation products revealed that the majority of N-terminal

product was released peptide instead of intermediate peptidyl(2A)-tRNA. The accumulation of N-terminal product started no later than the accumulation of peptidyl(2A)-tRNA (Figure 3.4). These data suggest that the hydrolysis of peptidyl(2A)-tRNA is quite fast, or that the ribosome pause and hydrolysis occur at the same time. The ribosome pause may be prolonged by the interaction of the nascent 2A peptide with the ribosome exit tunnel and PTC to promote restart of translation at Pro19 codon.

Interestingly, 2A is not the only element that inhibits the formation of glycine-proline peptide bond. The arrest peptide of *E. coli* SecM (Nakatogawa and Ito, 2002) also stalls ribosomes with SecM-tRNA^{Gly} in the P site and a proline codon in the A site (Muto *et al.*, 2006). However, in this case the prolyl-tRNA^{Pro} in the A site stabilizes the SecM-tRNA^{Gly}-ribosome complex, making it resistant to puromycin, even though it does not form a peptide bond to the SecM-tRNA^{Gly}, and RF is excluded from the complex (Muto *et al.*, 2006). 2A does not function in *E. coli* (Donnelly *et al.*, 1997). An explanation may be that similar exclusion of RF at 2A when it is translated by the *E. coli* ribosome may preclude the termination event necessary for the reaction.

How could 2A promote termination without stop codon recognition? Ribosomes undergo conformational changes as they cycle through each round of peptide bond synthesis. A number of antibiotics, relatively small molecules, and the nascent peptide chain, can individually or coordinately alter the progress or fidelity of translation and in some cases achieve their effects by driving conformational changes within the ribosome PTC (Berisio *et al.*, 2003; Ogle *et al.*, 2002). To obviate the necessity for stop codon decoding, the 2A peptide may direct the ribosome into a conformation similar to that which it takes once release factors have bound productively to the A site. Candidate movements might be at the PTC towards the conformation in which the hydrolytic termination reaction takes place and in the decoding region to mimic productive proofreading of incoming RF. 2A represents the founding member of a new class of recoding element that dictates a "stop – carry on" form of translation. Recoding of a sense codon to nonsense has not been previously documented. Furthermore, 2A is the first peptide recoding element to be characterized.

Table 3.1 Constructs used in *in vitro* studies on the FMDV 2A protein.

Construct	Structure
pJN177	ssΔαF-2A-ppLactin fusion
pJN178	ssΔαF-2A*-ppLactin fusion
pJN179	ssΔαF-AAP-ppLactin fusion

Plasmids listed in this table were obtained from Dr. Jeremy D. Brown at Newcastle University.

Sequence encoding amino acids 66 – 101 of prepro-α-factor (ssΔαF) was PCR amplified from pDJ100 (Hansen *et al.*, 1986) as a *Pst*I-*Xba*I fragment. The 19-amino acid 2A or 2A* fragments were excised from pMR90 (Ryan and Drew, 1994) as an *Xba*I-*Apa*I fragments. The ppLactin fragment encoding amino acids 51 – 128 of bovine preprolactin appended by SHRGTGYERSPDAIMSM was amplified from pSPBP4 (Siegel and Walter, 1988) as an *Apa*I-*Sal*I fragment. These fragments were cloned into pDJ100 between the *Pst*I site covering residues 8 – 10 of the α-factor and the *Sal*I site. Once the final construct was made, a PCR product was generated from it using the SP6 primer and a primer complementary to the 3′-end of the fragment. After the stop codon, an *Xho*I site and then a poly(A) tail and a *Sal*I site were incorporated. The PCR product was then cloned back into pSP65.

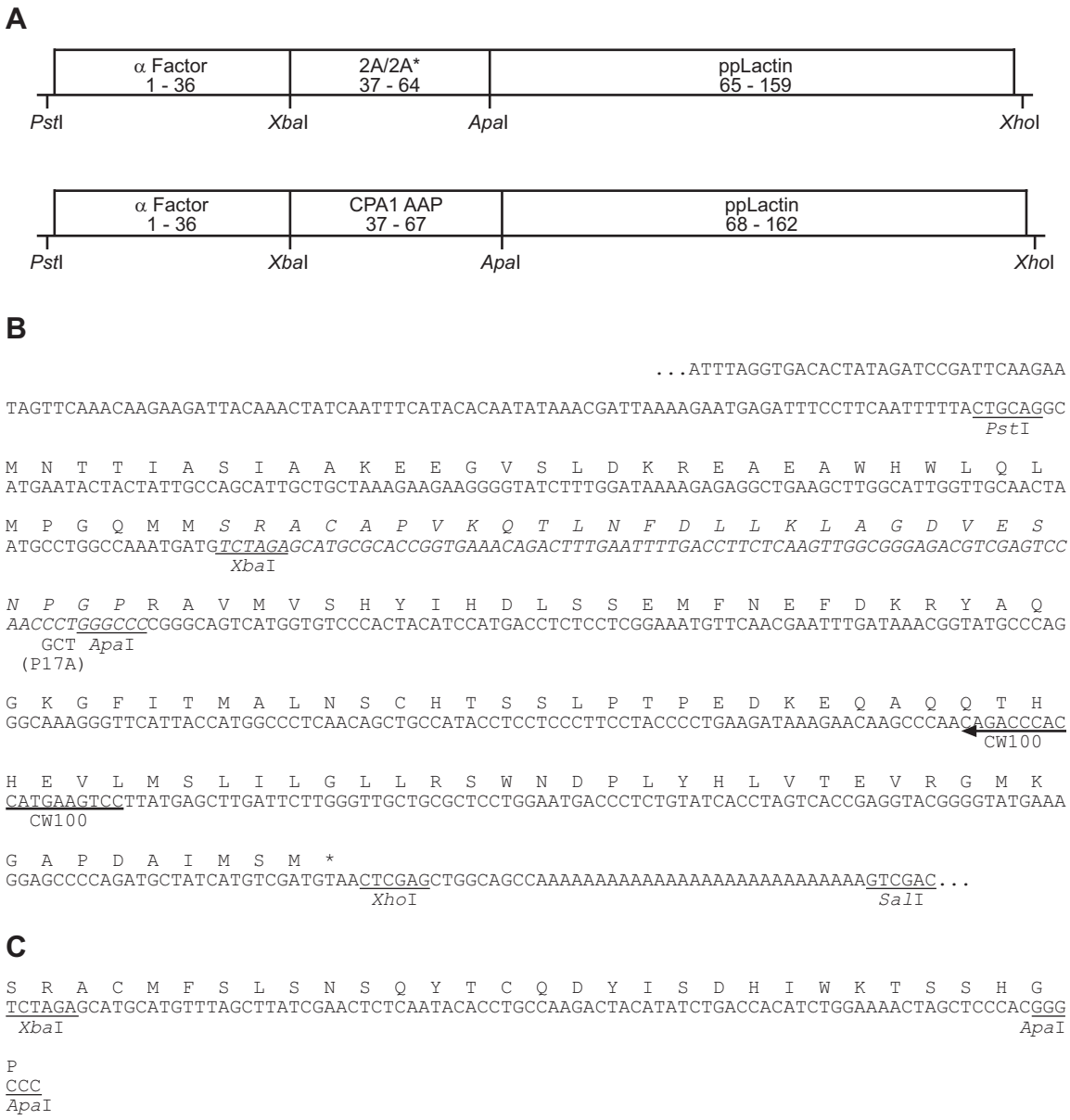

Figure 3.1 Schematic of 2A- or *CPA1* AAP-containing transcripts.

(A) The coding region of the ssΔαF-2A-ppLactin and ssΔαF-AAP-ppLactin fusion polypeptides. The polypeptides contain a domain from yeast prepro-α-factor (lacking its signal sequence, codons 1 – 36 in the fusion polypeptide) and a domain from bovine preprolactin (codons 65 – 159 in the fusion polypeptide). FMDV 2A coding region (including the proline codon of 2B) was placed between these two domains (codons 37 – 64 in the fusion polypeptide). Constructs incorporating 2A* or *CPA1* AAP were generated by exchange of an *Xba*I-*Apa*I fragment. (B) Sequences of the DNA templates

for making ssΔαF-2A-ppLactin mRNA. The sequence shown begins with the SP6 RNA polymerase binding site and ends with the *Sal*I site that follows the poly(A) tail, which is used for linearizing the DNA template. The amino acid sequences of ppαF-2A-ppLactin fusion are indicated. The 2A sequence and the downstream proline codon are italicized. The P17A mutation, which generates an inactive 2A*, is shown below the wild-type sequence. The restriction sites in this region are underlined; names of the corresponding restriction enzymes are indicated below the underlined sequences. The sequence for which the reverse complement was synthesized and used as primer oCW100 for toeprint assay is indicated by a horizontal arrow below the sequence. (C) Sequence of the *CPA1* AAP region as an *Xba*I-*Apa*I fragment to replace the 2A region within the ssΔαF-2A-ppLactin fusion. The sequence indicated replaces the italicized sequence in (B).

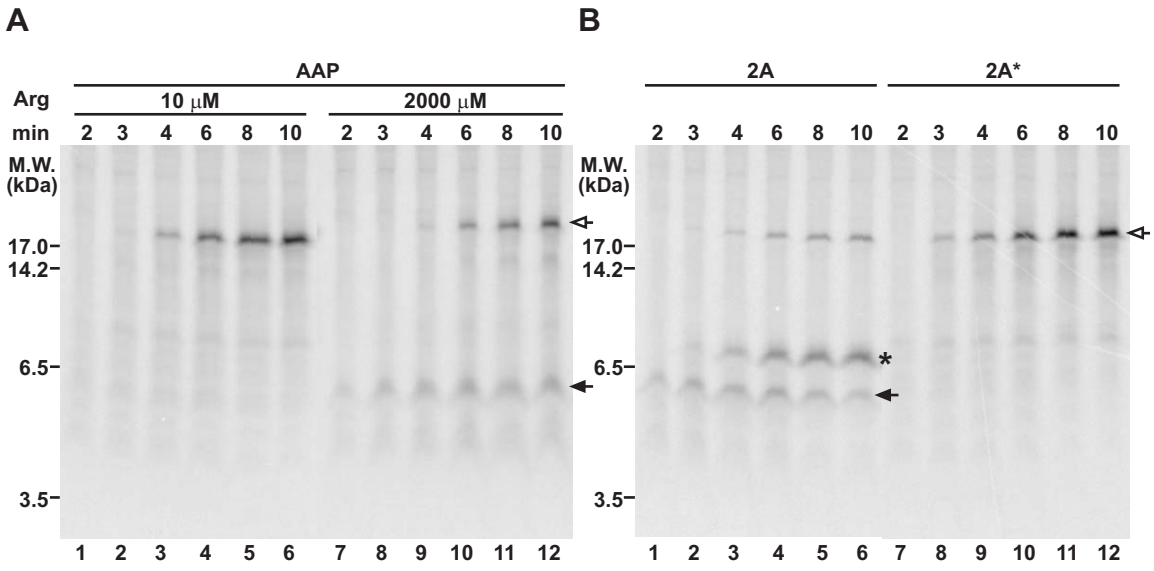

Figure 3.2 Pulse-chase analyses of polypeptide synthesis in *N. crassa* cell-free extracts. Synthetic mRNAs specifying fused polypeptides containing (A) *CPA1* AAP or (B) 2A or 2A* were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL in the presence of indicated amount of Arg and 10 μM of the other 19 amino acids. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, and were incubated at 25°C. Edeine was added at a final concentration of 1 mM at 2 min to block subsequent rounds of initiation. 10-μL aliquots were removed into an equal volume of 2× SDS loading buffer at the indicated time points, and were subsequently examined by using 16% tricine SDS-PAGE gels. Positions of translation products corresponding to the full-length polypeptide and the separated N- and C-terminal products are indicated by open and closed arrows, and asterisk, respectively.

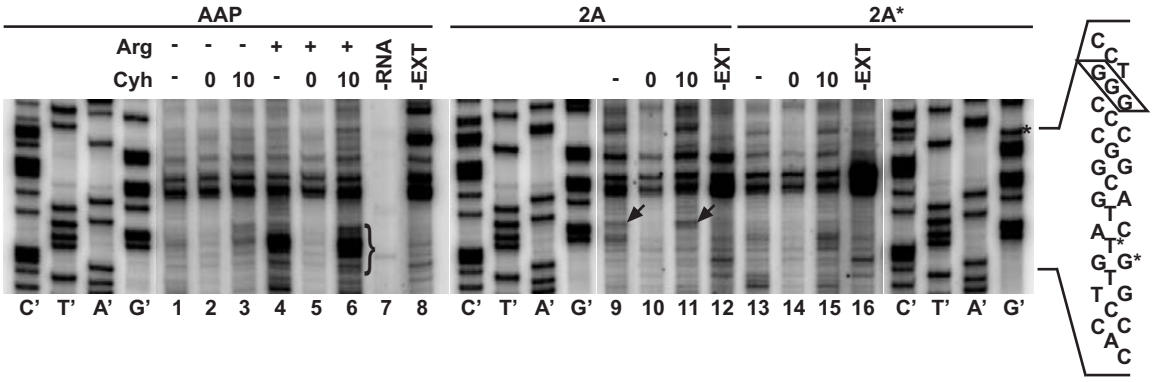

Figure 3.3 Detection of ribosome stalling at 2A cleavage site.

Primer extension inhibition analyses were performed on translation reactions initiated on the indicated synthetic mRNAs using ^{32}P-labeled primer oCW100 (Figure 3.1). Where indicated, 2000 μM of arginine (Arg) and/or 0.5 mg/mL of cycloheximide (Cyh) were added to the reactions. Cycloheximide was added either at the beginning (no translation control) or after 10 minutes of translation. Controls without template (-RNA) or extract (-EXT) are also shown. Positions of specific pauses on mRNA are marked: AAP (bracket) and 2A (arrows). Dideoxynucleotide sequencing reactions using ^{32}P-labeled primer oCW100 were run alongside the toeprint reactions. The mutation that yields the P17A change in 2A* is marked with asterisk. The nucleotide complementary to the dideoxynucleotide added to each sequencing reaction is indicated below the corresponding lane so that the sequence of the template can be directly deduced; the 5′ – 3′ sequence reads from top to bottom. The relevant portion of the 2A sequence is to the right of the figure as a series of triplets corresponding to codons in the mRNA. The Gly18 codon is boxed, and the positions of the main 2A-specific toeprints are asterisked.

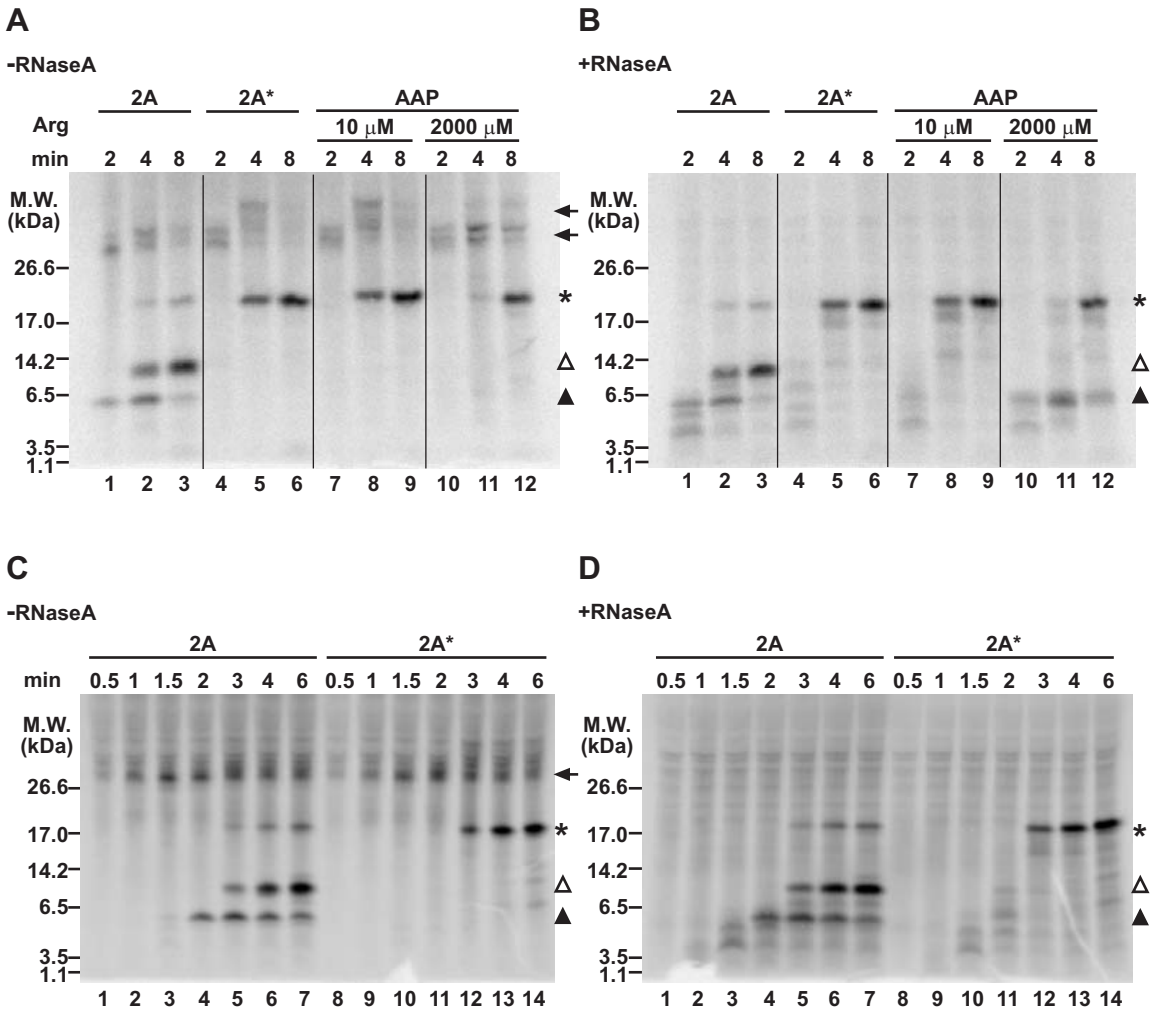

Figure 3.4 Pulse-chase analyses of polypeptide synthesis in *N. crassa* cell-free extracts. (A) Synthetic mRNAs containing 2A, 2A*, or *CPA1* APP were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL in the presence of indicated amount of Arg. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, 10 μM of the other 19 amino acids, and were incubated at 25°C. Edeine was added at a final concentration of 1 mM at 2 min to block subsequent rounds of initiation. 10-μL aliquots were removed into 2 μL of 5× SDS loading buffer at the indicated time points and half of each sample was examined by using 12% NuPAGE® Bis-Tris gels in 1× MES SDS running buffer containing 2.5 mM of DTT. Positions of products that appear to be arrested peptidyl-tRNA intermediates are indicated by arrows. Positions of major translation products corresponding to the full-length fusion

polypeptide and the separated N- and C-terminal products are indicated by asterisks, closed and open triangles, respectively. (B) The other half of each aliquot was treated with 10 units of RNaseA for 2 min at room temperature and then examined by using 12% NuPAGE® Bis-Tris gels. (C) Synthetic mRNA transcripts containing 2A or 2A* were used to program micrococcal nuclease-treated *N. crassa* cell-free extracts at a final concentration of 6 ng/μL in the presence of 10 μM of all 20 amino acids. Reaction mixtures also contained 0.5 μCi/μL of [^{35}S]Met, and were incubated at 25°C. Edeine was added at 0.5 min to block subsequent rounds of initiation. 10-μL aliquots were removed into 2 μL of 5× SDS loading buffer at the indicated time points and half of each sample was examined by using 12% NuPAGE® Bis-Tris gels. Positions of products that appear to be arrested peptidyl-tRNA intermediates are indicated by arrows. Positions of major translation products corresponding to the full-length fusion polypeptide and the separated N- and C-terminal products are indicated by asterisks, closed and open triangles, respectively. (D) The other half of each aliquot was treated with 10 units of RNaseA for 2 min at room temperature and then examined by using 12% NuPAGE® Bis-Tris gels.

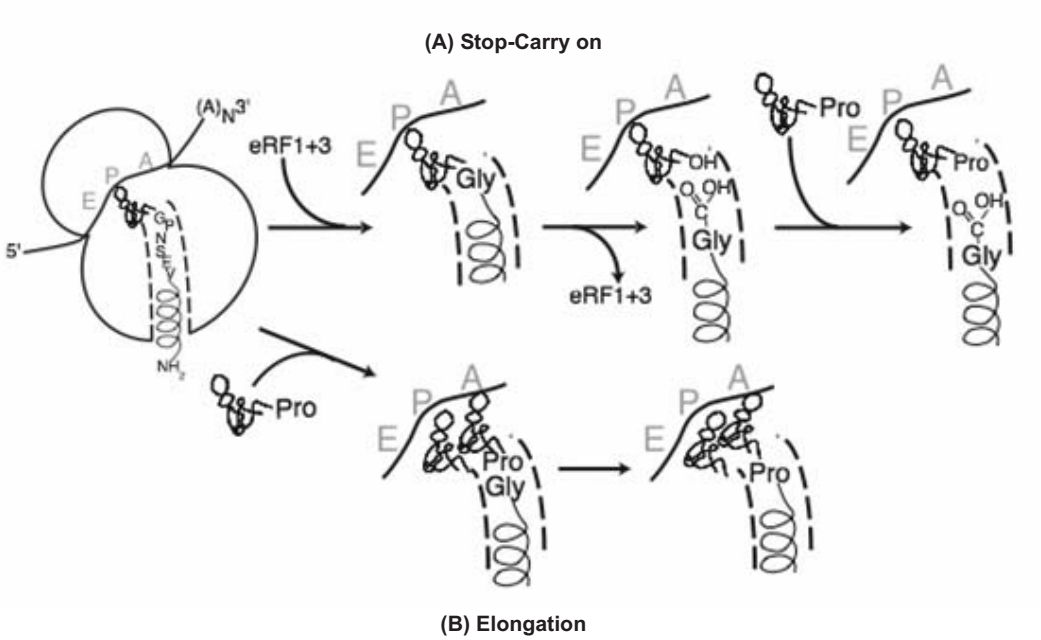

Figure 3.5 Model for recoding at 2A.[†]

See the text for further details.

[†]This figure is adapted from the manuscript entitled *A translational recoding event that generates separated proteins from one open reading frame* by Doronina, V. A., Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D., and Brown, J. D..

CHAPTER 4 FLJA INTERACTS WITH RNA STRUCTURES WITHIN THE 5′-UTR OF THE *SALMONELLA FLIC* GENE*

4.1 Introduction

Bacterial flagella are used for the motility of the organism (Blair, 1995). Each individual flagellum is composed of three substructures: (i) a basal body as a transmembrane motor; (ii) a hook connecting the filament to the basal body; and (iii) the filament as a propeller (Aizawa, 1996). The filament is composed of approximately 20,000 subunits of flagellin protein. Flagellin protein itself acts as a potent antigen that stimulates immune responses in many plants and animals (Hayashi *et al.*, 2001). The pathogen *Salmonella enterica* alternately expresses two different flagellin proteins, FljB and FliC (Stocker, 1949). This switching, known as flagellar phase variation (Andrewes, 1922), changes the structural features of the cell surface which leads to antigenic variation, therefore providing a mechanism for the bacteria to temporarily avoid cellular immunity (Fierer and Guiney, 2001; Ikeda *et al.*, 2001). Most *Salmonella*-specific $CD4^+$ T cells generated in response to *Salmonella* infection are directed at flagellin epitopes (Cookson and Bevan, 1997).

The alternative expression of the two flagellin proteins in *S. enterica* is achieved by a site-specific inversion of a chromosome region (Haykinson *et al.*, 1996). The *Hin* recombinase and the Fis protein (factor for inversion stimulation) mediate a reversible recombination reaction that leads to the inversion of a 996-bp DNA segment containing the *fljBA* promoter. In one orientation, the *fljBA* operon is transcribed to produce one type of flagellin FljB. A gene downstream of *fljB* encoding FljA, which is a repressor of

*Collaboration with the laboratory of Dr. Kelly T. Hughes at the University of Utah. This material has been published in similar form in the Proceedings of the National Academy of Sciences and is used here with permission of The National Academy of Sciences.

Aldridge, P. D., Wu C., Gnerer, J., Karlinsey, J. E., Hughes, K. T., and Sachs, M. S. (2006). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proc Natl Acad Sci U S A*. **103** (30): 11340-11345.

the unlinked *fliC* gene, is coexpressed along with FljB (Fujita *et al.*, 1973; Suzuki and Iino, 1973). Therefore the expression of the other type of flagellin, FliC, is inhibited (Kutsukake and Iino, 1980). In the opposite orientation, the transcription of the *fljBA* operon is turned off. FliC is then expressed, since FljA is not.

Previous studies indicated that FljA is a transcriptional repressor of the *fliC* gene (Simon and Silverman, 1983). However, a recent study indicates FljA regulates the production of FliC protein at the posttranscriptional level (Bonifield and Hughes, 2003). A greater reduction in protein levels than in mRNA levels was observed in cells locked in the $fljBA^{ON}$ orientation when compared to $fljBA^{OFF}$ cells. Depletion of FljA in $fljBA^{ON}$ cells causes a greater increase in protein levels than in mRNA levels. All these data suggest that the FljA protein is a translational regulator of *fliC* expression in addition to inhibiting *fliC* transcription. The translational control provides a quick response to inhibit FliC expression from existing transcripts when *fljBA* operon is turned on, thereby minimizing the production of flagella with mixed flagellin types.

The potential stem-loop region in the 5′-UTR of the *fliC* transcript contains sequences that are crucial for the efficient translation and assembly of FliC into the growing flagella structure (Aldridge *et al.*, 2006a). However, in a separate study, the stem-loop structure called SL2 in the *fliC* 5′-UTR was demonstrated to function to inhibit *fliC* mRNA translation. Deleting SL2 causes an increase in *fliC* mRNA translation. Strengthening of the interaction of SL2 with the *fliC* coding region enhances the inhibitory effect (Rosu *et al.*, 2006). *Cis*-acting mutations that were identified because they bypass FljA regulation are also located in the 5′-UTR of the *fliC* transcript, in a region immediately adjacent to and overlapping the Shine-Dalgarno sequence. This suggests that FljA regulates FliC synthesis by interacting with the mRNA to inhibit ribosome binding to the translational initiation site. This could also lead to the decreased stability of the *fliC* transcript to further reduce translation. In this research, we demonstrated that FljA binds to the 5′-UTR of the *fliC* mRNA but not the DNA template that specifies it. The FljA-binding site was determined, using a toeprint assay, to be within the ribosome binding site. However, characterization of mutations that circumvent the inhibition by FljA indicated that the FljA regulation is through a more complicated mechanism than simply occluding ribosomes from binding the RNA.

4.2 Materials and Methods

4.2.1 Strains

Bacterial strains (Table 4.1) used in this research were obtained from Dr. Kelly T. Hughes at University of Utah.

4.2.2 DNA Templates and mRNA Synthesis

Primer sets T7UTRFLIC1 (5′-TAATACGACTCACTATAGGGACGGTGAGAAA CCGTGGG-3′) and FliC+289R (5′-CCGCCAGTTCACGCAC-3′) were used to PCR-amplify the 5′-UTR region and part of the coding region (nt -62 – +289) of *S. typhimurium fliC*. The forward primer appends the T7 promoter 5′ to the 5′-UTR. The T7-tagged UTR-*fliC* PCR products were subsequently used as a template for a T7 RNA polymerase-dependent *in vitro* transcription reaction using [α-^{32}P]UTP to generate ^{32}P-labeled synthetic mRNAs.

4.2.3 Toeprint Assay

General procedures for primer extension inhibition are as described (Wang and Sachs, 1997b). Purified His-tagged FljA (70 μg) (from Dr. Kelly T. Hughes at University of Utah) was incubated with 20 μL of Ni-NTA·agarose (50% slurry; Qiagen) at 4°C for 60 min in 200 μL of 50 mM Tris·HCl, pH 7.5/150 mM NaCl/1% Triton X-100. The resin was washed once with 200 μL of wash buffer (50 mM sodium phosphate, pH 7.8/300 mM NaCl) and resuspended in 20 μL of binding buffer (10 mM Tris·HCl, pH 7.2/1 mM $MgCl_2$/2.5 mM DTT/0.1 U/μL RNasin/1.6 μg/μL yeast tRNA). For controls, FljA was omitted from the procedure. Synthetic mRNA (300 fmol) was preannealed for 5 min to ^{32}P-labeled oligonucleotide primer FliC+69R (5′-GGACTGGATTTGTTC AGG-3′) (9×10^6 cpm) in 5 μL of RT buffer (50 mM Tris·HCl, pH 8.3/75 mM KCl/10 mM $MgCl_2$/10 mM DTT/0.25 mM dNTP/1 U/μL RNasin) and added to Ni-NTA·agarose containing or lacking bound FljA. After 20-min incubation at room temperature, unbound material was removed, and the Ni-NTA·agarose was washed once with 200 μL of wash buffer. As controls to evaluate primer extension in the presence of

Ni-NTA·agarose, RNA preannealed with primer was added directly to washed Ni-NTA·agarose at this point. Toeprint reactions were performed directly on the Ni-NTA·agarose samples by the addition of 10 μL of RT buffer containing 100 units of SuperScript II reverse transcriptase from Invitrogen and incubated for 30 min at 37°C. For each sample, 8 μL of aqueous phase was extracted with phenol/chloroform, mixed with 6 μL of loading buffer, and half of the mixture was analyzed with denaturing 6% polyacrylamide gels. Gels were dried and exposed to screens of a Molecular Dynamics PhosphorImager for ~24 h. All image data shown are representatives of multiple experiments.

4.3 Results

4.3.1 FljA Binds to *fliC* mRNA but not the DNA with the Same Sequences

So far, all bypass mutations that have been identified that allow synthesis of FliC flagellin protein in the presence of repressor FljA are located within the 5′-UTR of the *fliC* transcript, adjacent to and overlapping the ribosome binding site. It is proposed that FljA binds to the 5′-UTR to inhibit ribosome binding and translation. To test this hypothesis, purified His_6-FljA was incubated with a synthetic radiolabeled mRNA fragment containing the 5′-UTR of *fliC* and with a DNA fragment containing the *fliC* promoter, 5′-UTR, and N-terminal coding region. Samples were analyzed using 8% polyacrylamide gels. The presence of FljA did not cause any shift to the migration of the DNA fragment, indicating there was no interaction between FljA and *fliC* DNA (Figure 4.2A). In contrast, a mobility shift was observed when FljA was incubated with *fliC* mRNA containing 5′-UTR (Figure 4.2B). This result confirms that FljA only binds to the *fliC* mRNA but not to the DNA for the corresponding region.

The binding of FljA to the *fliC* mRNA is sequence-specific. The addition of an excess of unlabeled RNA of the same sequence competes for the FljA binding, thereby decreasing the amount of radiolabeled RNA bound to FljA, as evidenced by decrease in the signal corresponding to the FljA-bound mRNA and an increase in the signal corresponding to the free mRNA. In contrast, an excess of tRNA has no effect on the

mobility shift caused by the interaction between FljA and *fliC* mRNA (Figure 4.2C).

4.3.2 Binding of FljA to the 5′-UTR of Bypass Mutants

Based on the model that FljA binds to the *fliC* 5′-UTR to inhibit translation, it was thought that bypass mutants are defective in binding FljA so that they allow FliC expression whether or not FljA is present. However, this did not appear to be the case. One remarkable mutation, a large duplication including both the 5′-UTR and the first 15 bases of *fliC* coding sequence, DUP(-13 – +15), allows FliC expression only when FljA is present. This suggests that FljA binds to this mutated 5′-UTR but acts as an activator rather than a repressor. Five potential stem-loop structures are predicted to be present in the 5′-UTR (Figure 4.1B). The SL2 structure has been demonstrated to have inhibitory effect on *fliC* translation (Rosu *et al.*, 2006). These data suggest that the mutation may change the RNA structure, thereby changing the role of bound FljA. If this is true, the bound FljA may also have to interact with the secondary structures within the wild-type 5′-UTR to inhibit translation.

Filter-binding assays were performed to further check the capability of FljA to bind to different bypass mutants. Synthetic radiolabeled mRNA fragments containing wild-type or mutant *fliC* 5′-UTR were incubated with purified His_6-FljA and then filtered onto a nitrocellulose filter. Radioactivity that did not pass through the filter was measured to determine the amount of mRNA bound to FljA. All bypass mutants derived from single-base substitution or deletion are defective in binding FljA relative to the wild-type *fliC* 5′-UTR (Figure 4.3B). As expected, the mRNA fragment containing the large duplication DUP(-13 – +15) still binds to FljA (Figure 4.3A). Because DUP(-13 – +15) is a longer fragment extended into coding region, a second wild-type mRNA fragment with similar length was used as control. This longer wild-type fragment binds to FljA with a higher affinity than the short one (Figure 4.3A). This might be explained by the binding of the N-terminal coding sequences to FljA or by enhanced FljA binding to the 5′-UTR based on the secondary structure conformation within that region. A second mutant derived from a short duplication, DUP(-19 – -15), also shows the ability to bind FljA (Figure 4.3A). However, this mutant allows FliC expression no matter whether FljA is present or not, suggesting the binding of FljA can

neither inhibit nor activate the expression.

4.3.3 Binding of FljA to the Ribosome Binding Site

Toeprint assays were performed to determine the FljA binding site in the *fliC* 5′-UTR. Reverse transcriptase is used to map the transcription start sites by primer extension that stops at the 5′-end of the mRNA. Proteins bound at one or more discrete positions on the mRNA will inhibit primer extension, creating toeprint products that are smaller than the full-length product. When accompanied by a parallel sequence ladder for the mRNA, the sizes of primer extension products can be determined, enabling mapping of the boundaries of proteins bound to the mRNA that block primer extension. These boundaries yield "toeprints".

The first attempt to toeprint the FljA binding site by using solution-hybridization failed. There appeared not to be enough template which had FljA bound to it to produce a detectable signal. We modified the procedure by first concentrating the His_6-FljA protein using Ni-NTA·agarose. The resin with bound His_6-FljA was then incubated with synthetic *fliC* mRNA fragments pre-annealed to radiolabeled primer. After washing off unbound RNA, primer extension was directly performed on the Ni-NTA·agarose-associated material. Thus the toeprint signal relative to signal from unbound RNA should be significantly increased. Two control reactions in which either mRNA or FljA were absent were also performed. A third control reaction was performed on the *fliC* mRNA added to Ni-NTA·agarose in the absence of FljA after the final washing step.

When FljA was absent in the reaction, toeprint signals were weak (Figure 4.4; lanes 2, 5, 8, and 11). This establishes that *fliC* mRNA does not interact with Ni-NTA·agarose, and that the primer extension products in reactions in which FljA is present are from the interaction of mRNA and FljA, and not from Ni-NTA·agarose. In reactions in which FljA was present, a toeprint ~4-bases upstream of the AUG initiation codon was detected in all transcripts except that containing the Δ-13U mutation, suggesting (i) the two DUP mutants bind FljA as efficiently as wild-type, (ii) the binding site overlaps the Shine-Dalgarno site, and (iii) Δ-13U is defective in binding FljA (Figure 4.4; lanes 1, 4, 7, and 10; closed arrows). These results are consistent with filter-binding

assays. These toeprint signals are FljA-specific because no corresponding signals were detected in reactions containing only Ni-NTA·agarose and *fliC* mRNA (Figure 4.4; lanes 3, 6, 9, and 12). Duplicated toeprint signals were detected in mutant DUP(-13 – +15) since it has duplicated binding sites (Figure 4.4, closed arrows). These data support the model that the FljA binds to the translation initiation region of the *fliC* 5′-UTR to regulate translation.

Toeprints corresponding to the 3′ boundaries of two stem-loops, SL1 and SL2 (Figure 4.1B), were observed as well (Figure 4.4; lanes 1, 4, 7, and 10; open arrows). However the signal of SL2 in Δ-13U mutant is weaker when compared to the others, indicating that the SL2 structure is unstable. This result implies that SL2 may have a role in FljA binding to the mRNA, and the unstable SL2 may be the reason why mutation Δ-13U is defective in binding FljA. Alternatively, the inability to bind FljA may cause an unstable SL2, thereby removing its inhibitory effect on translation (Rosu *et al.*, 2006).

The similar strength of detectable signals for all four mRNAs indicates a similar amount of Δ-13U *fliC* mRNA was bound to FljA as with the other mRNAs (Figure 4.4; lanes 1, 4, 7, and 10). This is an unexpected result since this mutant is relatively defective in binding FljA. Possibly, FljA has nonspecific as well as specific RNA-binding properties.

4.4 Discussion

The biogenesis of the bacterial flagellum is a highly ordered process. Expression of numerous genes is tightly controlled to ensure the assembly and function of these structures (Kalir *et al.*, 2001). *Salmonella enterica* undergoes a process known as phase variation. Two antigenically distinct flagellin proteins, FljB and FliC, are alternatively expressed. Coexpressed with FljB, the FljA protein has been historically characterized as a transcriptional repressor of the unlinked *fliC* gene (Simon and Silverman, 1983). Thus only one type of flagellin is expressed at a time. Recent evidence shows that FljA-dependent inhibition of FliC expression is at both transcriptional and translational levels (Bonifield and Hughes, 2003). The decrease in the FliC protein levels is greater than that in the *fliC* transcript levels when FljA is present.

Furthermore, it seems to block the assembly of FliC protein as well since FliC is still detectable in cells locked in the *fljBA*^{ON} orientation. Bypass mutations allowing synthesis of FliC flagellin in the presence of inhibitory FljA have been isolated. These mutations are all mapped within the 5′-UTR of the *fliC* mRNA, clustering to a 15-bp sequence close to and overlapping the Shine-Dalgarno sequence, which is a site for ribosome binding to initiate translation instead of a site used for transcriptional regulation. A model is proposed that FljA regulates the expression of *fliC* by binding to the 5′-UTR of the mRNA, competing with the ribosome for the Shine-Dalgarno sequence, thereby inhibiting ribosome binding and thus translation (Bonifield and Hughes, 2003). Bypass mutants may be defective in binding FljA, allowing ribosomes to come in and initiate translation.

The ability of FljA to bind to the 5′-UTR of the wild-type *fliC* mRNA was demonstrated for mRNA but not for the DNA with corresponding sequences. The binding is specific to the mRNA sequence because the binding of FljA to radiolabeled mRNA was eliminated only by cognate competing unlabeled mRNA (Figure 4.2). Further study shows that bypass mutants having single-base deletion or substitution mutations lose their ability to bind FljA, consistent with the proposed model (Figure 4.3B). However, in two bypass mutants having duplications of 5 and 27 bases, respectively, within the potential FljA-binding area, FljA was still able to bind to the mRNA (Figure 4.3A). In these cases, binding does not inhibit the synthesis of FliC protein. More surprisingly, the mutant containing a large duplication, DUP(-13 – +15), is dependent upon FljA for FliC synthesis. In this case, FljA binding to the mRNA activates the translation rather than inhibiting it.

Toeprint assays were performed to map the positions of bound FljA on both wild-type and mutant *fliC* mRNA 5′-UTRs. FljA-specific toeprints ~4-bases upstream of the AUG initiation codon were detected in wild-type and both Dup mutants but not in the Δ-13U mutant. The DUP(-13 – +15) mutant has a duplicated toeprint due to it having duplicated potential binding site (Figure 4.4). The results of toeprint assays (i) confirm the FljA binding site overlaps the Shine-Dalgarno sequence and (ii) support that bound FljA inhibits ribosome binding to inhibit translation.

But is it possible that the binding of FljA inhibits *fliC* translation in wild-type yet

is required for FliC synthesis in DUP(-13 – +15)? Moreover, in the case of DUP(-19 – -15), why does FljA not affect FliC translation at all? Studies on the 5′-UTR of *fliC* mRNA revealed that this region contains sequences that are crucial for the efficient translation and assembly of FliC into the growing flagella structure (Aldridge *et al.*, 2006a; Rosu *et al.*, 2006). Five potential stem-loop structures were predicted to be present in this region (Figure 4.1B). SL2 has been proved to be important for the regulation of FliC expression. SL2 itself and its interaction with the downstream coding region have negative effects on *fliC* translation (Rosu *et al.*, 2006). Two of the predicted stem-loops, SL3 and SL4, overlap each other and possess equal predicted energies, would compete in their formation. Isolated bypass mutations are all located either in regions that would affect the SL3 and SL4 or between these two SLs and the downstream *fliC* AUG codon. In the case of Δ-13U, toeprint assays suggest a weaker SL2 structure is present. The weaker SL2 may cause deficiency in binding FljA to bypass its negative effect. Alternatively, defects in FljA-binding may cause an unstable SL2, thereby removing its inhibition. Other mutants may also change the RNA secondary structures and these changes may change the role of bound FljA. In the case of DUP(-13 – +15), the large duplicated regions may interact with each other and block the entry of ribosome. The binding of FljA to either binding location may interrupt this interaction between the duplicated regions and allow the ribosome access to the translation initiation site. In the case of DUP(-19 – -15), the change in the mRNA secondary structure may activate *fliC* translation constitutively so the synthesis of FliC is always active no matter whether FljA binds to the mRNA or not. Thus FljA and the RNA structures within the 5′-UTR of *fliC* could work individually and/or coordinately to regulate the translation of FliC. These results suggests signals intrinsic to the *fliC* mRNA that have both positive and negative effects on *fliC* translation involving both RNA structure and interacting proteins.

Although all the bypass mutations restore the ability to synthesize FliC in the presence of FljA, two of these were found incapable to provide motility (Aldridge *et al.*, 2006b). This observation suggests that FljA not only inhibits the expression of FliC but also inhibits the assembly of FliC into the flagellum. At the same time, some bypass mutations were found to impair the expression and assembly of FljB. All these data indicate that *fljB* and *fliC* compete for the expression and assembly into the flagellum.

The presence of FljA ensures only one type of flagellin is expressed at a time so that the chance of producing flagella with a mixture of flagellin types is minimized, even when mRNAs encoding both proteins are present especially during the phase variation.

Table 4.1 Bacterial strains used in this research.

Strain	Genotype
TH437	*S. enterica* serovar Typhimurium LT2 (wild-type) (Bunny *et al.*, 2002)
TH5802	Δ*hin-132*::FRT *fliC5469*::Mu*d*K *fliC5751* [FljA-BP Δ(-13U) from AUG]
TH5806	Δ*hin-132*::FRT *fliC5469*::Mu*d*K *fliC5751* [FljA-BP DUP(-15 – -19) from AUG]
TH8980	*fliC6323* [DUP(-13 – +15) relative to AUG]

Salmonella strains were obtained from Dr. Kelly T. Hughes at the University of Utah. Mutations listed in this table were selected based on their ability to express a *fliC-lac* fusion gene as a criterion for expressing FliC protein in the presence of inhibitory FljA.

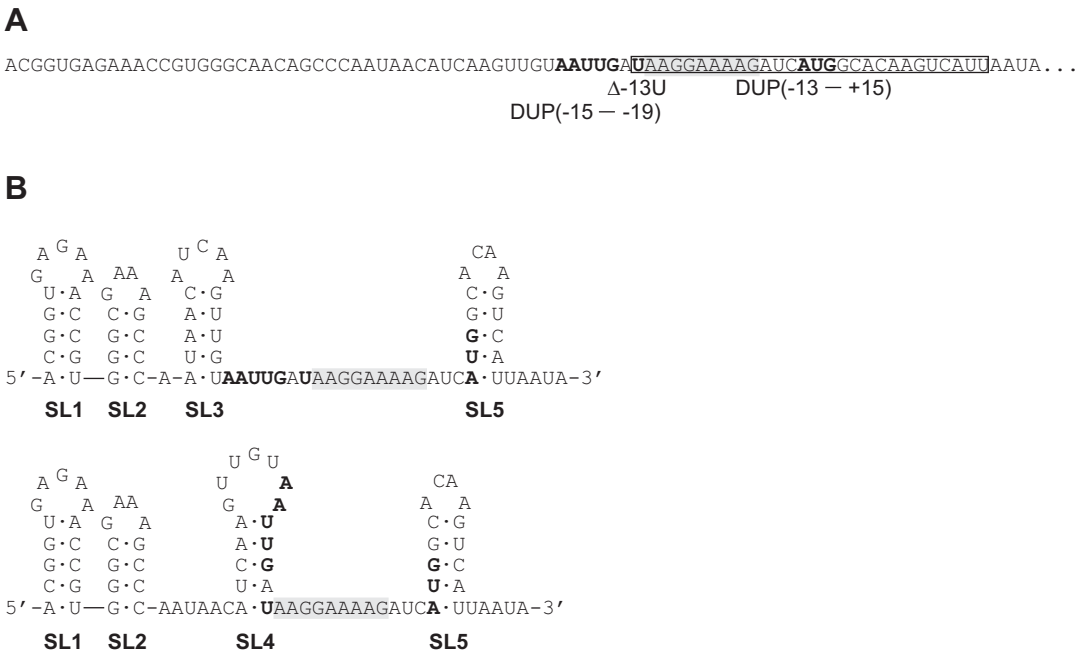

Figure 4.1 Sequence of the 5′-UTR of *S. enterica* serovar Typhimurium *fliC* transcript. (A) Sequence of the *fliC* mRNA 5′-UTR. The sequence shown here starts with the transcription initiation site and ends with 19 bases into the *fliC* coding region. (B) The potential secondary structures located within the *fliC* 5′-UTR predicted by mfold RNA-folding program of Zucker (http://www.bioinfo.rpi.edu/applications/mfold). Five stem-loop structures were identified. SL3 and SL4 are two competing stem-loop structures that are illustrated in the upper and lower diagrams, respectively. The Shine-Dalgarno sequence is shadowed. *fliC* AUG codon is in bold font. The bases that are affected by mutations Δ-13U and DUP(-15 – -19) are also in bold font. The bases affected by mutation DUP(-13 – +15) are boxed. Adapted from (Aldridge *et al.*, 2006b).

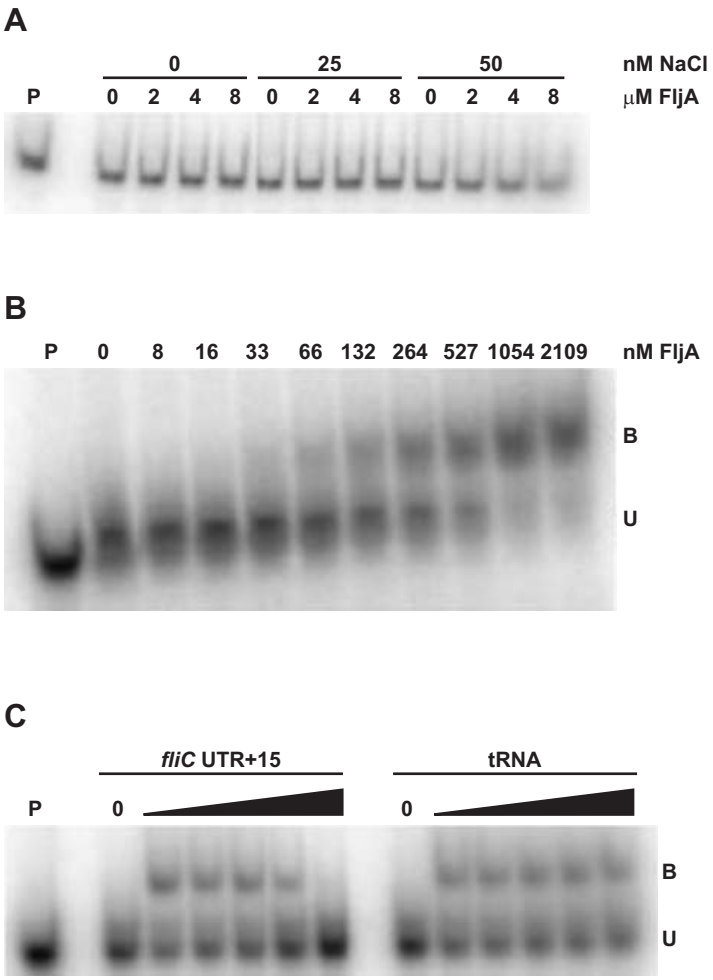

Figure 4.2 Gel-shift assays of purified FljA binding to DNA and RNA.

Indicated amounts of purified His_6-FljA were used to incubate with (A) a radiolabeled DNA fragment containing the *fliC* promoter, 5′-UTR, and N-terminal coding sequences from -160 to +100 relative to the ATG codon; and (B) a synthetic radiolabeled mRNA fragment including the 5′-UTR and the first 15 bases of the coding sequences of *fliC*. (C) 264 nM of purified His_6-FljA were used to incubate with the same mRNA fragment in B in the presence of either unlabeled mRNA of the same sequences (left) or tRNA (right) as the chase. Excess unlabeled mRNA and tRNA were added at 20-, 200-, 2,000-, and 20,000-fold molar excess with respect to the concentration of radiolabeled *fliC* mRNA. Reactions of 10 μL were incubated for 15 min at room temperature and then analyzed by using 8% polyacrylamide gels. Adapted from (Aldridge *et al.*, 2006b).

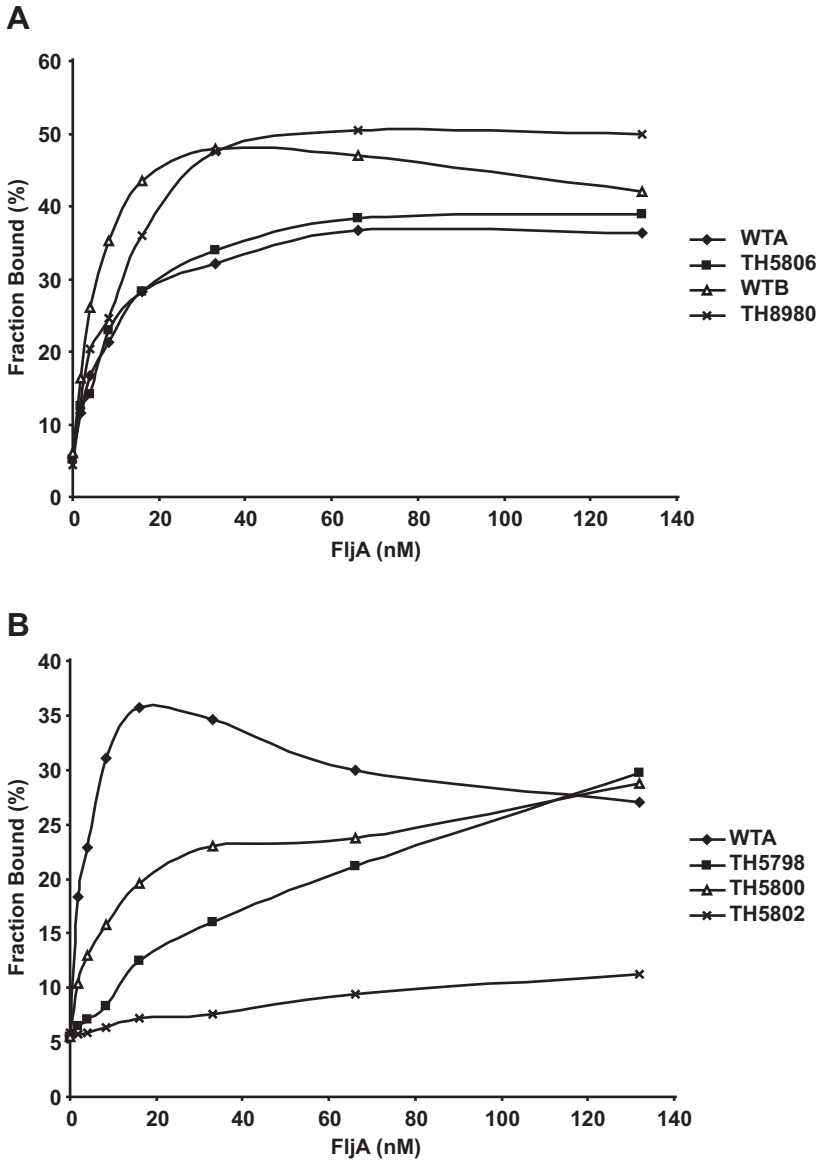

Figure 4.3 Filter-binding assays of FljA binding to the *fliC* 5′-UTR.

Assays of purified His_6-FljA binding to synthetic radiolabeled *fliC* 5′-UTR from (A) wild-type (-62 – +15 bases relative to AUG, WTA), DUP(-15 – -19) (TH5806), DUP(-13 – +15) (TH8980); and from (B) WTA, -24G:C (TH5798), -4G:U (TH5800), and Δ-13U (TH5802). Because TH8980 is a longer fragment, a second wild-type transcript (-62 – +26, WTB) was used as control. Binding reactions were performed in triplicate containing equal amount of ^{32}P-labeled *fliC* mRNA and His_6-FljA, and 200-fold molar excess of *E. coli* tRNA. Reactions were incubated for 15 min at room temperature, and then filtered onto prewetted 0.2-μm Protran BA83 nitrocellulose filters from Schleicher

& Schull. The filters were washed and dried. Radioactivitiy was measured by using a liquid-scintillation counter. Adapted from (Aldridge *et al.*, 2006b).

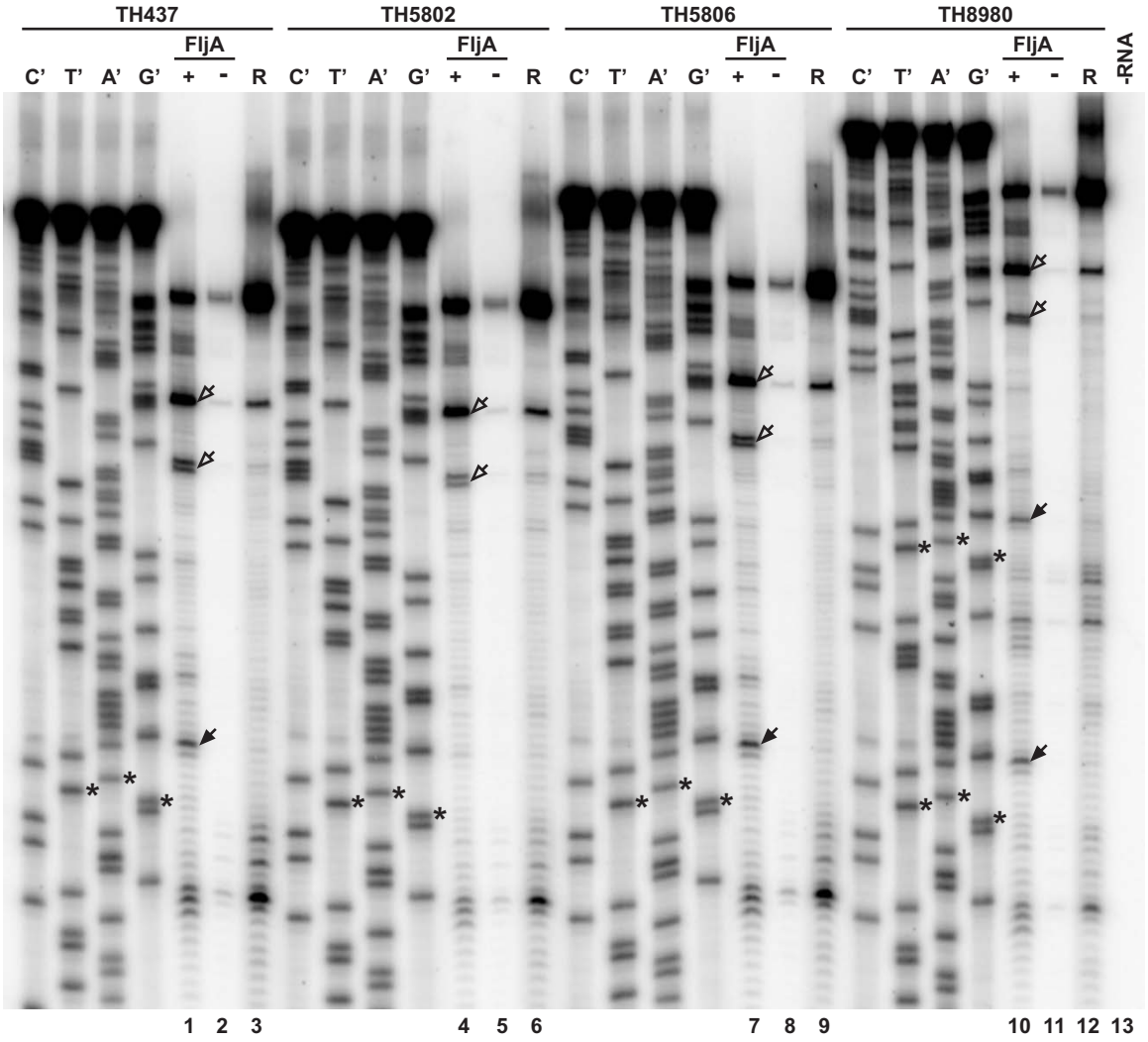

Figure 4.4 Toeprint analyses of FljA binding to the *fliC* 5′-UTR.

Equal amounts of synthetic mRNA and ^{32}P-labeled oligonucleotide primer were incubated with FljA bound to Ni-NTA·agarose (+FljA; lanes 1, 4, 7, and 10) or Ni-NTA·agarose alone (-FljA; lanes 2, 5, 8, and 11) and the associated RNA was toeprinted. The RNAs analyzed contained the wild-type *fliC* 5′-UTR (TH437, lanes 1 – 3), and mutant 5′-UTRs from Δ-13U (TH5802, lanes 4 – 6), DUP(-15 – -19) (TH5806, lanes 7 – 9) and DUP(-13 – +15) (TH8980, lanes 10 – 12). A reaction lacking RNA but containing radiolabeled oligonucleotide and FljA bound to Ni-NTA·agarose (-RNA, lane 13) showed all signals in lanes 1 – 9 were RNA-specific. Primer extension analyses of RNA in the presence of Ni-NTA·agarose (R; lanes 3, 6, 9, and12) was used as a control to identify protein-specific signals. The DNA templates used to synthesize mRNAs were

sequenced with the same oligonucleotide primer used for primer extension so that toeprint sites could be precisely mapped. The nucleotide complementary to the dideoxynucleotide added to each sequencing reaction is indicated above the corresponding lane so that the sequence of the template can be directly deduced; the 5′ – 3′ sequence reads from top to bottom. The positions of identically placed FljA-dependent signals observed with TH437, 5806, and 8980 are indicated with closed arrows. FliC start codon positions were indicated with asterisks. The duplicated toeprint and start codon in the duplicated region of TH8980 are indicated in the same way. Open arrows indicate the toeprints corresponding to the 3′ boundaries at SL1 and SL2, respectively

CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Summary of Research

5.1.1 The Nascent *N. crassa arg-2* AAP Transiently Stalls the Translating Ribosome in Response to Arg

Translation of the Met_9AAP-LUC fusion gene in *N. crassa* cell-free extracts demonstrated that Arg-specific ribosome stalling is associated with the synthesis of AAP and a delayed synthesis of full-length polypeptide. Increasing the amount of Arg increased the accumulation of AAP and the time delay in synthesizing the full-length polypeptide. Pulse-chase analyses indicated that AAP, as an internal domain within a polypeptide, also caused the ribosome to stall at its C-terminus in the presence of high Arg. Quantitative analysis of the radiolabel in the major translation products demonstrated that Arg-stalled intermediate products will end up in the full-length polypeptides. These intermediate products were proved to be in their peptidyl-tRNA form by analyzing with NuPAGE® gels. An up-shift in migration was observed in Arg-stalled intermediate products due to being attached to the tRNA with a mass of ~20 kDa, but not in the full-length polypeptides. This up-shift was eliminated upon RNaseA treatment. All of these data strongly suggest that the ribosome stalling at the C-terminus of AAP is transient and the ribosome resumes translation after released from stalling, and all these are mediated by the nascent AAP in response to Arg.

5.1.2 The FMDV 2A Peptide Mediates A Translational Recoding Event at Its C-terminus

The FMDV 2A-mediated cleavage was recapitulated in the *N. crassa* cell-free translation system. Primer extension inhibition (toeprint) assays demonstrated that the ribosome pauses at the C-terminus of 2A with Gly18 codon in the P site. This result is

consistent with results from the previous puromycin incorporation assay and from mass spectrometry that Gly18 is the last amino acid residue in the N-terminal product. It also demonstrated that the pausing ribosome was in the post-translocation state, available to interact with release factors, prolyl-tRNA^{Pro}, or puromycin. Further analyses by using [^{35}S]Met pulse-chase experiments revealed the majority of N-terminal product at the moment the ribosome paused was released peptide instead of intermediate peptidyl-tRNA as observed in AAP. The observation of the simultaneous accumulation of peptidyl(2A)-tRNA and N-terminal product suggests ribosome pause and hydrolysis occur at the same time. All these data support the idea that the nascent 2A peptide may interact with the ribosomal components to cause a ribosome pause after synthesizing the 2A domain, and that this interaction may also change the ribosome PTC activity to induce the hydrolysis.

5.1.3 *Salmonella* FljA Binds to the Ribosome Binding Site in *fliC* mRNA 5′-UTR to Inhibit Translation

Salmonella FljA is coexpressed with FljB flagellin to prevent the expression of FliC flagellin, transcriptionally and translationally. FljA-bypass mutations identified are all located close to and overlapping the ribosome binding site, suggesting FljA binds to the same region of *fliC* mRNA to inhibit translation. The ability of FljA to bind the critical regions of the *fliC* mRNA was demonstrated by gel-shift assays for RNA but not for the DNA with corresponding sequences. The binding specificity was demonstrated by that the binding of FljA to radiolabeled mRNA was eliminated only by unlabeled cognate competing mRNA. Single-base mutations in *fliC* 5′-UTR are defective in binding FljA, while mutants that have duplications can still bind FljA. These results were further confirmed by toeprint assays. Toeprint assays also demonstrated the FljA binding site to be within the Shine-Dalgarno sequence, thereby supporting the hypothesis that bound FljA would inhibit ribosome binding to block translation. RNA structures in the *fliC* 5′-UTR are also important in regulating translation, especially the stem-loop SL2. Toeprint assay with Δ-13U bypass mutant, which is deficient in binding FljA, indicated an unstable SL2 is present in its 5′-UTR. These results suggest the formation of RNA structures and the binding of FljA may have effects on each other, and they may work

individually and coordinately to regulate *fliC* translation.

5.2 Future Directions

5.2.1 To Determine How the Nascent AAP Interacts with Ribosomal Components to Mediate Stalling

There are many examples of ribosome stalling caused by the nascent peptide. Studies on *E. coli* TnaC and SecM demonstrated that both of them can stall ribosome at the C-terminus. In both cases, the amino acid residues critical for stalling ribosome are at a location ~12-residues from the P site of the arrested ribosome. That location happens to be near the narrowest section of the exit tunnel (Ban *et al.*, 2000; Berisio *et al.*, 2003; Cruz-Vera *et al.*, 2005; Nakatogawa and Ito, 2002). It has been established that ribosomal components in this narrowest constriction acts as a discrimination gate which interacts with the nascent peptide to pause ribosomes in elongation (Cruz-Vera *et al.*, 2005; Nakatogawa and Ito, 2002). The critical Asp12 residue in AAP (Asp13 in *S. cerevisiae*) is also located ~12-residues form the arrested ribosome's P site. The eukaryotic ribosome might have a similar constriction gate in its exit tunnel. Thus the nascent AAP would interact with ribosomal components to constrain the nascent peptide movement, thereby pausing the elongation. One approach is to directly detect the components inside the ribosome interacting with the nascent AAP. Photocrosslinking is a good choice because it has been proved successful in detecting components in translational machinery with which nascent peptides interact. This technique will be adapted to investigate what components the nascent AAP interacts with inside the ribosome in the presence of high Arg. Synthetic mRNAs with amber codon mutations introduced into different codons in the coding region would be used to perform translation. Photoaffinity crosslinkers can be incorporated into the nascent chain by adding amber-suppressor tRNA charged with appropriate amino acid analogs. After translation in the dark, the reaction mixture will be irradiated with UV light to obtain crosslinked products, which will be used to determine what components are present. Due to the AAP's functional similarity to TnaC, another approach would be using the

bacterial translation system to study AAP. Shine-Dalgarno sequence could be put at 5′ proximal to the Met_9AAP-LUC-AAP-LUC coding region, which will be used for *in vitro* translation in *E. coli* S30 extracts.

5.2.2 To Determine How Arginine Functions in Ribosome Stalling

How arginine functions in the AAP-mediated ribosome stalling event is poorly understood. It was first thought to function as tryptophan does in the case of *E. coli* TnaC, which occupies the ribosome A site and alter the activity of PTC (Gong and Yanofsky, 2002). Substituting the termination codon with a Trp codon can cause constitutive stalling because the aminoacylated tRNA places tryptophan in the proper spot to exert its effect. But our data show that substituting the termination codon with an Arg codon in the AAP did not cause ribosome stalling in the *N. crassa* system unless high Arg is present. Arginine may directly interact with the nascent AAP and/or the translational machinery to cause conformational changes that control polypeptide synthesis. We attempted the detection of the direct binding of arginine to the synthetic AAP *in vitro*, but this was not successful. This suggests that if there was interaction between Arg and the AAP, the presence of the other components and/or the conformation of these components and AAP itself are important. One approach to detect this binding will be using radiolabeled L-arginine analogs covalently attached to a crosslinker to obtain crosslinked products.

5.2.3 To Determine How 2A Mediates Ribosome Pause and Hydrolysis

The present data suggest 2A-mediated cleavage at its C-terminus is actually a translational recoding event including both termination and initiation at Pro19 codon. A ribosome pause was also detected during this event. It has been thought that the nascent 2A peptide forms a helix with a tight turn at the C-terminus (Donnelly *et al.*, 2001b; Ryan *et al.*, 1999). In this conformation, movement of the nascent peptide may be restricted. The restricted conformation of 2A within the PTC may block the attack of the peptidyl(2A)-$tRNA^{Gly}$ ester bond by prolyl-$tRNA^{Pro}$ (Donnelly *et al.*, 2001a; Ryan *et al.*, 1999). Unpublished data from Dr. Jeremy D. Brown's lab implies that release factors

are involved in the 2A-mediated cleavage. Overexpression of 2A-containing polypeptides in yeast strains with limited release factor activity inhibits the cell growth. Altering the activity of release factors affects the outcome of the 2A reaction both *in vivo* and *in vitro*, indicating the Pro19 codon in the A site when the ribosome is paused is decoded as a termination codon. *E. coli* SecM is another example of inhibiting the formation of glycine-proline peptide bond. However in this case, prolyl-tRNA is in the A site and excludes the attack of release factors. This may explain why 2A does not function in prokaryotic system. It was thought that the C-terminus of nascent SecM peptide adopts a compact conformation that would facilitate key residues within the arrest motif to interact with ribosomal components to induce a further change in the ribosome-nascent chain complex that in turn arrests elongation (Woolhead *et al.*, 2006). 2A and 2A-like peptides share a conserved DXEXNPG$^{\Downarrow}$P motif. Mutations changing any of these residues would cause 2A to lose its ability to cleave, and maybe the ability to pause ribosome as well, as what has been observed in 2A*. But if we were able to identify some mutations either in 2A or in the ribosome that would cause ribosome to stop but not able to promote the hydrolysis, we might be able to use cryo-EM to study these mutations to learn how 2A mediates the ribosome pausing. At the same time, if we were able to use anti-RF antibodies to pull down 2A-containing message, this would further confirm the termination event at proline codon.

5.2.4 To Uncover Features Involved in FljA-dependent Regulation of FliC Translation

The present data suggest that FljA inhibits the translation of the *fliC* gene by binding to the ribosome binding site upstream of the *fliC* AUG codon, thereby blocking the ribosome access to the translation initiation site. The *fliC* mutants defective in binding FljA allow the expression of FliC in the presence of FljA. However, *fliC* mutant requiring FljA to translate FliC suggests in this case the bound FljA acts as an activator rather than as an inhibitor. Another *fliC* mutant can also bind FljA but the bound FljA has no effect on the translation of FliC. These data suggest something else is involved in regulating the translation of FliC. The stem-loop structures within the 5′-UTR of *fliC* mRNA have been proved to have effects on translation (Aldridge *et al.*, 2006a; Rosu *et*

al., 2006). These features may interact with the bound FljA to work coordinately to affect translation. The *fliC* mutations may change the secondary structures, which in turn may change the role of the bound FljA in regulating translation. The effects of FljA bypass mutations on the RNA structures should be studied. Mutations in FljA should also be identified that could restore its inhibitory effect in the FljA-dependent bypass mutant. Thus we might learn how FljA and RNA structures affect each other and how this effect has an impact on their functions to regulate FliC translation.

LITERATURE CITED

Abastado, J. P., Miller, P. F., Jackson, B. M. and Hinnebusch, A. G. (1991). Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. *Mol Cell Biol 11*, 486-496.

Aizawa, S.-I. (1996). Flagellar assembly in *Salmonella typhimurium*. *Mol Microbiol 20*, 1-4.

Alderete, J. P., Jarrahian, S. and Geballe, A. P. (1999). Translational effects of mutations and polymorphisms in a repressive upstream open reading frame of the human cytomegalovirus UL4 gene. *J Virol 73*, 8330-8337.

Aldridge, P., Gnerer, J., Karlinsey, J. E. and Hughes, K. T. (2006a). Transcriptional and translational control of the *Salmonella fliC* gene. *J Bacteriol 188*, 4487-4496.

Aldridge, P. D., Wu, C., Gnerer, J., Karlinsey, J. E., Hughes, K. T. and Sachs, M. S. (2006b). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proc Natl Acad Sci U S A 103*, 11340-11345.

Algire, M. A., Maag, D., Savio, P., Acker, M. G., Tarun, S. Z., Jr., Sachs, A. B., Asano, K., Nielsen, K. H., Olsen, D. S., Phan, L., Hinnebusch, A. G. and Lorsch, J. R. (2002). Development and characterization of a reconstituted yeast translation initiation system. *RNA 8*, 382-397.

Algire, M. A., Maag, D. and Lorsch, J. R. (2005). P_i release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol Cell 20*, 251-262.

Alkalaeva, E. Z., Pisarev, A. V., Frolova, L. Y., Kisselev, L. L. and Pestova, T. V. (2006). *In vitro* reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell 125*, 1125-1136.

Amrani, N., Sachs, M. S. and Jacobson, A. (2006). Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol 7*, 415-425.

Anand, M., Chakraburtty, K., Marton, M. J., Hinnebusch, A. G. and Kinzy, T. G. (2003). Functional interactions between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. *J Biol Chem 278*, 6985-6991.

Andersen, C. B. F., Becker, T., Blau, M., Anand, M., Halic, M., Balar, B., Mielke, T., Boesen, T., Pedersen, J. S., Spahn, C. M. T., Kinzy, T. G., Andersen, G. R. and Beckmann, R. (2006). Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature 443*, 663-668.

Andrewes, F. W. (1922). Studies on group-agglutination. I. The *Salmonella* group and its antigenic structure. *J Pathol Bacteriol 25*, 1509-1514.

Anthony, T. G., Anthony, J. C., Yoshizawa, F., Kimball, S. R. and Jefferson, L. S. (2001a). Oral administration of leucine stimulates ribosomal protein mRNA translation but not global rates of protein synthesis in the liver of rats. *J Nutr 131*, 1171-1176.

Anthony, T. G., Reiter, A. K., Anthony, J. C., Kimball, S. R. and Jefferson, L. S. (2001b). Deficiency of dietary EAA preferentially inhibits mRNA translation of ribosomal proteins in liver of meal-fed rats. *Am J Physiol Endocrinol Metab 281*, E430-439.

Arkov, A. L., Freistroffer, D. V., Ehrenberg, M. and Murgola, E. J. (1998). Mutations in RNAs of both ribosomal subunits cause defects in translation termination. *EMBO J 17*, 1507-1514.

Arkov, A. L., Hedenstierna, K. O. F. and Murgola, E. J. (2002). Mutational evidence for a functional connection between two domains of 23S rRNA in translation termination. *J Bacteriol 184*, 5052-5057.

Arrick, B. A., Lee, A. L., Grendell, R. L. and Derynck, R. (1991). Inhibition of translation of transforming growth factor-β3 mRNA by its 5′ untranslated region. *Mol Cell Biol 11*, 4306-4313.

Asano, K., Krishnamoorthy, T., Phan, L., Pavitt, G. D. and Hinnebusch, A. G. (1999). Conserved bipartite motifs in yeast eIF5 and eIF2Bε, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J 18*, 1673-1688.

Asano, K., Clayton, J., Shalev, A. and Hinnebusch, A. G. (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.

Asano, K. and Sachs, M. S. (2007). Translation factor control of ribosome conformation during start codon selection. *Genes Dev 21*, 1280-1287.

Aukerman, M. J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell 15*, 2730-2741.

Avdulov, S., Li, S., Van, M., Burrichter, D., Peterson, M., Perlman, D. M., Manivel, J. C., Sonenberg, N., Yee, D., Bitterman, P. B. and Polunovsky, V. A. (2004). Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant

phenotype in human mammary epithelial cells. *Cancer Cell 5*, 553-563.

Baek, J.-M. and Kenerley, C. M. (1998). The *arg2* gene of *Trichoderma virens*: cloning and development of a homologous transformation system. *Fungal Genet Biol 23*, 34-44.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science 289*, 905-920.

Barak, Y., Gottlieb, E., Juven-Gershon, T. and Oren, M. (1994). Regulation of *mdm2* expression by p53: alternative promoters produce transcripts with nonidentical translation potential. *Genes Dev 8*, 1739-1749.

Basu, U., Si, K., Warner, J. R. and Maitra, U. (2001). The *Saccharomyces cerevisiae TIF6* gene encoding translation initiation factor 6 is required for 60S ribosomal subunit biogenesis. *Mol Cell Biol 21*, 1453-1462.

Bates, B., Hardin, J., Zhan, X., Drickamer, K. and Goldfarb, M. (1991). Biosynthesis of human fibroblast growth factor-5. *Mol Cell Biol 11*, 1840-1845.

Beckmann, K., Grskovic, M., Gebauer, F. and Hentze, M. W. (2005). A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. *Cell 122*, 529-540.

Beelman, C. A., Stevens, A., Caponigro, G., LaGrandeur, T. E., Hatfield, L., Fortner, D. M. and Parker, R. (1996). An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature 382*, 642-646.

Belli, B. A. and Samuel, C. E. (1993). Biosynthesis of reovirus-specified polypeptides: identification of regions of the bicistronic reovirus S1 mRNA that affect the efficiency of translation in animal cells. *Virology 193*, 16-27.

Benne, R., Brown-Luedi, M. L. and Hershey, J. W. (1978). Purification and characterization of protein synthesis initiation factors eIF-1, eIF-4C, eIF-4D, and eIF-5 from rabbit reticulocytes. *J Biol Chem 253*, 3070-3077.

Bergenhem, N. C. H., Venta, P. J., Hopkins, P. J., Kim, H. J. and Tashian, R. E. (1992). Mutation creates an open reading frame within the 5′ untranslated region of macaque erythrocyte carbonic anhydrase (CA) I mRNA that suppresses CA I expression and supports the scanning model for translation. *Proc Natl Acad Sci U S A 89*, 8798-8802.

Berisio, R., Schluenzen, F., Harms, J., Bashan, A., Auerbach, T., Baram, D. and Yonath, A. (2003). Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat Struct Biol 10*, 366-370.

Blair, D. F. (1995). How bacteria sense and swim. *Annu Rev Microbiol 49*, 489-520.

Bonetti, B., Fu, L., Moon, J. and Bedwell, D. M. (1995). The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol 251*, 334-345.

Bonifield, H. R. and Hughes, K. T. (2003). Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *J Bacteriol 185*, 3567-3574.

Borman, A. M., Kirchweger, R., Ziegler, E., Rhoads, R. E., Skern, T. and Kean, K. M. (1997). elF4G and its proteolytic cleavage products: effect on initiation of protein synthesis from capped, uncapped, and IRES-containing mRNAs. *RNA 3*, 186-196.

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. and Cohen, S. M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell 113*, 25-36.

Brown, C. Y., Mize, G. J., Pineda, M., George, D. L. and Morris, D. R. (1999). Role of two upstream open reading frames in the translational control of oncogene *mdm2*. *Oncogene 18*, 5631-5637.

Bunny, K., Liu, J. and Roth, J. (2002). Phenotypes of *lexA* mutations in *Salmonella enterica*: evidence for a lethal *lexA* null phenotype due to the Fels-2 prophage. *J Bacteriol 184*, 6235-6249.

Byrne, P. C., Sanders, P. G. and Snell, K. (1995). Translational control of mammalian serine hydroxymethyltransferase expression. *Biochem Biophys Res Commun 214*, 496-502.

Calkhoven, C. F., Bouwman, P. R. J., Snippe, L. and Ab, G. (1994). Translation start site multiplicity of the CCAAT/enhancer binding protein α mRNA is dictated by a small 5′ open reading frame. *Nucl Acids Res 22*, 5540-5547.

Calkhoven, C. F., Muller, C. and Leutz, A. (2000). Translational control of C/EBPα and C/EBPβ isoform expression. *Genes Dev 14*, 1920-1932.

Calkhoven, C. F., Muller, C., Martin, R., Krosl, G., Hoang, T. and Leutz, A. (2003). Translational control of SCL-isoform expression in hematopoietic lineage choice. *Genes Dev 17*, 959-964.

Cao, J. and Geballe, A. P. (1994). Mutational analysis of the translational signal in the human cytomegalovirus gpUL4 (gp48) transcript leader by retroviral infection. *Virology 205*, 151-160.

Cao, J. and Geballe, A. P. (1995). Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *J Virol 69*, 1030-1036.

Cao, J. and Geballe, A. P. (1996). Inhibition of nascent-peptide release at translation termination. *Mol Cell Biol 16*, 7109-7114.

Carrington, J. C. and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science 301*, 336-338.

Caskey, C. T., Beaudet, A. L., Scolnick, E. M. and Rosman, M. (1971). Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc Natl Acad Sci U S A 68*, 3163-3167.

Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L. A., Offenhauser, N., Marchisio, P. C. and Biffo, S. (2003). Release of eIF6 ($p27^{BBP}$) from the 60S subunit allows 80S ribosome assembly. *Nature 426*, 579-584.

Chakravarti, D. and Maitra, U. (1993). Eukaryotic translation initiation factor 5 from *Saccharomyces cerevisiae*. Cloning, characterization, and expression of the gene encoding the 45,346-Da protein. *J Biol Chem 268*, 10524-10533.

Chang, C. P., Vesole, D. H., Nelson, J., Oldstone, M. B. and Stinski, M. F. (1989). Identification and expression of a human cytomegalovirus early glycoprotein. *J Virol 63*, 3330-3337.

Chang, K. S., Lee, S. H., Hwang, S. B. and Park, K. Y. (2000). Characterization and translational regulation of the arginine decarboxylase gene in carnation (*Dianthus caryophyllus* L.). *Plant J 24*, 45-56.

Chang, M. J. and Blissard, G. W. (1997). Baculovirus gp64 gene expression: negative regulation by a minicistron. *J Virol 71*, 7448-7460.

Chaudhuri, J., Chowdhury, D. and Maitra, U. (1999). Distinct functions of eukaryotic translation initiation factors eIF1A and eIF3 in the formation of the 40 S ribosomal preinitiation complex. *J Biol Chem 274*, 17975-17980.

Chen, C.-Z., Li, L., Lodish, H. F. and Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science 303*, 83-86.

Chen, J.-J. (2000). Heme-regulated eIF2α kinase. In *Translational Control of Gene Expression*, pp. 529-546. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Chen, X. (2004). A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science 303*, 2022-2025.

Cheung, Y.-N., Maag, D., Mitchell, S. F., Fekete, C. A., Algire, M. A., Takacs, J. E., Shirokikh, N., Pestova, T., Lorsch, J. R. and Hinnebusch, A. G. (2007). Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection *in vivo*. *Genes*

Dev 21, 1217-1230.

Chiba, Y., Ishikawa, M., Kijima, F., Tyson, R. H., Kim, J., Yamamoto, A., Nambara, E., Leustek, T., Wallsgrove, R. M. and Naito, S. (1999). Evidence for autoregulation of cystathionine γ-synthase mRNA stability in *Arabidopsis*. *Science 286*, 1371-1374.

Chiba, Y., Sakurai, R., Yoshino, M., Ominato, K., Ishikawa, M., Onouchi, H. and Naito, S. (2003). *S*-adenosyl-L-methionine is an effector in the posttranscriptional autoregulation of the cystathionine γ-synthase gene in *Arabidopsis*. *Proc Natl Acad Sci U S A 100*, 10225-10230.

Child, S. J., Miller, M. K. and Geballe, A. P. (1999a). Translational control by an upstream open reading frame in the HER-2/*neu* transcript. *J Biol Chem 274*, 24335-24341.

Child, S. J., Miller, M. K. and Geballe, A. P. (1999b). Cell type-dependent and -independent control of HER-2/*neu* translation. *Int J Biochem Cell Biol 31*, 201-213.

Churbanov, A., Rogozin, I. B., Babenko, V. N., Ali, H. and Koonin, E. V. (2005). Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucl Acids Res 33*, 5512-5520.

Cookson, B. T. and Bevan, M. J. (1997). Identification of a natural T cell epitope presented by *Salmonella*-infected macrophages and recognized by T cells from orally immunized mice. *J Immunol 158*, 4310-4319.

Cougot, N., van Dijk, E., Babajko, S. and Seraphin, B. (2004). 'Cap-tabolism'. *Trends Biochem Sci 29*, 436-444.

Crabeel, M., Lavalle, R. and Glansdorff, N. (1990). Arginine-specific repression in *Saccharomyces cerevisiae*: kinetic data on *ARG1* and *ARG3* mRNA transcription and stability support a transcriptional control mechanism. *Mol Cell Biol 10*, 1226-1233.

Cruz-Vera, L. R., Rajagopal, S., Squires, C. and Yanofsky, C. (2005). Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of *tna* operon expression. *Mol Cell 19*, 333-343.

Cruz-Vera, L. R., Gong, M. and Yanofsky, C. (2006). Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. *Proc Natl Acad Sci U S A 103*, 3598-3603.

Cruz-Vera, L. R., New, A., Squires, C. and Yanofsky, C. (2007). Ribosomal features essential for *tna* operon induction: tryptophan binding at the peptidyl transferase center. *J Bacteriol 189*, 3140-3146.

Das, S. and Maitra, U. (2000). Mutational analysis of mammalian translation initiation

factor 5 (eIF5): role of interaction between the β subunit of eIF2 and eIF5 in eIF5 function *in vitro* and *in vivo*. *Mol Cell Biol 20*, 3942-3950.

Dasmahapatra, B. and Chakraburtty, K. (1981). Protein synthesis in yeast. I. Purification and properties of elongation factor 3 from *Saccharomyces cerevisiae*. *J Biol Chem 256*, 9999-10004.

Davis, R. H. (1986). Compartmental and regulatory mechanisms in the arginine pathways of *Neurospora crassa* and *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev 50*, 280-313.

Davuluri, R. V., Suzuki, Y., Sugano, S. and Zhang, M. Q. (2000). CART classification of human 5′ UTR sequences. *Genome Res 10*, 1807-1816.

de Felipe, P., Hughes, L. E., Ryan, M. D. and Brown, J. D. (2003). Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. *J Biol Chem 278*, 11441-11448.

de Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C. and Ryan, M. D. (2006). *E unum pluribus*: multiple proteins from a self-processing polyprotein. *Trends Biotech 24*, 68-75.

de Moor, C. H. and Richter, J. D. (1999). Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. *EMBO J 18*, 2294-2303.

Decker, C. J. and Parker, R. (1994). Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem Sci 19*, 336-340.

Degnin, C. R., Schleiss, M. R., Cao, J. and Geballe, A. P. (1993). Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol 67*, 5514-5521.

Delbecq, P., Werner, M., Feller, A., Filipkowski, R. K., Messenguy, F. and Pierard, A. (1994). A segment of mRNA encoding the leader peptide of the *CPA1* gene confers repression by arginine on a heterologous yeast gene transcript. *Mol Cell Biol 14*, 2378-2390.

Delbecq, P., Calvo, O., Filipkowski, R. K., Piérard, A. and Messenguy, F. (2000). Functional analysis of the leader peptide of the yeast gene *CPA1* and heterologous regulation by other fungal peptides. *Curr Genet 38*, 105-112.

Descombes, P. and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell 67*, 569-579.

Dever, T. E., Feng, L., Wek, R. C., Cigan, A. M., Donahue, T. F. and Hinnebusch, A. G.

(1992). Phosphorylation of initiation factor 2α by protein kinase GCN2 mediates gene-specific translational control of *GCN4* in yeast. *Cell 68*, 585-596.

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

Dever, T. E., Dar, A. C. and Sicheri, F. (2007). The eIF2α kinases. In *Translational Control in Biology and Medicine*, pp. 319-344. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Diba, F., Cheryl, S. and Gametchu, W. B. (2001). 5′UTR sequences of the glucocorticoid receptor 1A transcript encode a peptide associated with translational regulation of the glucocorticoid receptor. *J Cell Biochem 81*, 149-161.

Donahue, T. (2000). Genetic approaches to translation initiation in *Saccharomyces cerevisiae*. In *Translational Control of Gene Expression*, pp. 467-502. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Donnelly, M. L., Gani, D., Flint, M., Monaghan, S. and Ryan, M. D. (1997). The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol 78*, 13-21.

Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M. D. (2001a). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol 82*, 1027-1041.

Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D. and Ryan, M. D. (2001b). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J Gen Virol 82*, 1013-1025.

Dontsova, M., Frolova, L., Vassilieva, J., Piendl, W., Kisselev, L. and Garber, M. (2000). Translation termination factor aRF1 from the archaeon *Methanococcus jannaschii* is active with eukaryotic ribosomes. *FEBS Lett 472*, 213-216.

Donzé, O. and Spahr, P. F. (1992). Role of the open reading frames of Rous sarcoma virus leader RNA in translation and genome packaging. *EMBO J 11*, 3747-3757.

Dorris, D. R., Erickson, F. L. and Hannig, E. M. (1995). Mutations in *GCD11*, the structural gene for eIF-2γ in yeast, alter translational regulation of *GCN4* and the selection of the start site for protein synthesis. *EMBO J 14*, 2239-2249.

Doudna, J. and Rath, V. (2002). Structure and function of the eukaryotic ribosome: the next frontier. *Cell 109*, 153-156.

Doudna, J. A. and Sarnow, P. (2007). Translation initiation by viral internal ribosome entry sites. In *Translational Control in Biology and Medicine*, pp. 129-154. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Duncan, K., Grskovic, M., Strein, C., Beckmann, K., Niggeweg, R., Abaza, I., Gebauer, F., Wilm, M. and Hentze, M. W. (2006). Sex-lethal imparts a sex-specific function to UNR by recruiting it to the *msl-2* mRNA 3′ UTR: translational repression for dosage compensation. *Genes Dev 20*, 368-379.

Eiznhamer, D. A., Ashburner, B. P., Jackson, J. C., Gardenour, K. R. and Lopes, J. M. (2001). Expression of the *INO2* regulatory gene of *Saccharomyces cerevisiae* is controlled by positive and negative promoter elements and an upstream open reading frame. *Mol Microbiol 39*, 1395-1405.

Elroy-Stein, O. and Merrick, W. C. (2007). Translation initiation *via* cellular internal ribosome entry sites. In *Translational Control in Biology and Medicine*, pp. 155-172. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F. and Bowman, J. L. (2003). Radial Patterning of *Arabidopsis* Shoots by Class III HD-ZIP and KANADI Genes. *Curr Biol 13*, 1768-1774.

Erickson, F. L. and Hannig, E. M. (1996). Ligand interactions with eukaryotic translation initiation factor 2: role of the γ-subunit. *EMBO J 15*, 6311-6320.

Fajardo, J. E. and Shatkin, A. J. (1990). Translation of bicistronic viral mRNA in transfected cells: regulation at the level of elongation. *Proc Natl Acad Sci U S A 87*, 328-332.

Fang, P., Wang, Z. and Sachs, M. S. (2000). Evolutionarily conserved features of the arginine attenuator peptide provide the necessary requirements for its function in translational regulation. *J Biol Chem 275*, 26710-26719.

Fang, P., Wu, C. and Sachs, M. S. (2002). *Neurospora crassa* supersuppressor mutants are amber codon-specific. *Fungal Genet Biol 36*, 167-175.

Fang, P., Spevak, C. C., Wu, C. and Sachs, M. S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A 101*, 4059-4064.

Febbraio, M., Podrez, E. A., Smith, J. D., Hajjar, D. P., Hazen, S. L., Hoff, H. F., Sharma, K. and Silverstein, R. L. (2000). Targeted disruption of the class B scavenger receptor CD36 protects against atherosclerotic lesion development in mice. *J Clin Invest 105*, 1049-1056.

Fekete, C. A., Applefield, D. J., Blakely, S. A., Shirokikh, N., Pestova, T., Lorsch, J. R. and Hinnebusch, A. G. (2005). The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection *in vivo*. *EMBO J 24*, 3588-3601.

Fiaschi, T., Marzocchini, R., Raugei, G., Veggi, D., Chiarugi, P. and Ramponi, G. (1997). The 5′-untranslated region of the human muscle acylphosphatase mRNA has an inhibitory effect on protein expression. *FEBS Letters 417*, 130-134.

Fierer, J. and Guiney, D. G. (2001). Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. *J Clin Invest 107*, 775-780.

Fingar, D. C., Salama, S., Tsou, C., Harlow, E. and Blenis, J. (2002). Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev 16*, 1472-1487.

Fingar, D. C., Richardson, C. J., Tee, A. R., Cheatham, L., Tsou, C. and Blenis, J. (2004). mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E. *Mol Cell Biol 24*, 200-216.

Franceschetti, M., Hanfrey, C., Scaramagli, S., Torrigiani, P., Bagni, N., Burtin, D. and Michael, A. J. (2001). Characterization of monocot and dicot plant *S*-adenosyl-L-methionine decarboxylase gene families including identification in the mRNA of a highly conserved pair of upstream overlapping open reading frames. *Biochem J 353*, 403-409.

Freedman, D. A. (1999). Regulation of the p53 protein by the MDM2 oncoprotein – Thirty-eighth G. H. A. Clowes Memorial Award lecture. *Cancer Res 59*, 1-7.

Freitag, M., Dighde, N. and Sachs, M. S. (1996). A UV-induced mutation in *Neurospora* that affects translational regulation in response to arginine. *Genetics 142*, 117-127.

Frolova, L., Le Goff, X., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A. L., Celis, J. E., Philippe, M. and et al. (1994). A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature 372*, 701-703.

Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M. and Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA 2*, 334-341.

Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V. M., Tatkov, S. I. and Kisselev, L. L. (1999). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA 5*, 1014-1020.

Fujita, H., Yamaguchi, S. and Iino, T. (1973). Studies on H-O variants in *Salmonella* in

relation to phase variation. *J Gen Microbiol 76*, 127-134.

Fumagalli, S. and Thomas, G. (2000). S6 phosphorylation and signal transducion. In *Translational Control of Gene Expression*, pp. 695-718. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Futterer, J., Gordon, K., Sanfaçon, H., Bonneville, J. M. and Hohn, T. (1990). Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. *EMBO J 9*, 1697-1707.

Futterer, J. and Hohn, T. (1992). Role of an upstream open reading frame in the translation of polycistronic mRNAs in plant cells. *Nucl Acids Res 20*, 3851-3857.

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

Gaba, A., Wang, Z., Krishnamoorthy, T., Hinnebusch, A. G. and Sachs, M. S. (2001). Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *EMBO J 20*, 6453-6463.

Gaba, A., Jacobson, A. and Sachs, M. S. (2005). Ribosome occupancy of the yeast *CPA1* upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell 20*, 449-460.

Gallie, D. R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev 5*, 2108-2116.

Gaspar, N. J., Kinzy, T. G., Scherer, B. J., Humbelin, M., Hershey, J. W. and Merrick, W. C. (1994). Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the γ-subunit. *J Biol Chem 269*, 3415-3422.

Geballe, A. P. and Sachs, M. S. (2000). Translational control by upsteam open reading frames. In *Translational Control of Gene Expression*, pp. 595-614. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gingras, A.-C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R. and Sonenberg, N. (1999). Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev 13*, 1422-1437.

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

Gollnick, P. and Yanofsky, C. (1990). $tRNA^{Trp}$ translation of leader peptide codon 12 and other factors that regulate expression of the tryptophanase operon. *J Bacteriol 172*, 3100-3107.

Gong, F. and Yanofsky, C. (2001). Reproducing *tna* operon regulation *in vitro* in an S-30 System. Tryptophan induction inhibits cleavage of TnaC peptidyl-tRNA. *J Biol Chem 276*, 1974-1983.

Gong, F. and Yanofsky, C. (2002). Instruction of translating ribosome by nascent peptide. *Science 297*, 1864-1867.

Gopfert, U., Kullmann, M. and Hengst, L. (2003). Cell cycle-dependent translation of p27 involves a responsive element in its 5′-UTR that overlaps with a uORF. *Hum Mol Genet 12*, 1767-1779.

Grass, D. S. and Manley, J. L. (1987). Selective translation initiation on bicistronic simian virus 40 late mRNA. *J Virol 61*, 2331-2335.

Green, R. and Noller, H. F. (1997). Ribosomes and translation. *Annu Rev Biochem 66*, 679-716.

Griffin, E., Re, A., Hamel, N., Fu, C., Bush, H., McCaffrey, T. and Asch, A. S. (2001). A link between diabetes and atherosclerosis: glucose regulates expression of CD36 at the level of translation. *Nat Med 7*, 840-846.

Grifo, J. A., Tahara, S. M., Morgan, M. A., Shatkin, A. J. and Merrick, W. C. (1983). New initiation factor activity required for globin mRNA translation. *J Biol Chem 258*, 5804-5810.

Grskovic, M., Hentze, M. W. and Gebauer, F. (2003). A co-repressor assembly nucleated by Sex-lethal in the 3′UTR mediates translational control of *Drosophila msl-2* mRNA. *EMBO J 22*, 5571-5581.

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

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science 293*, 1146-1150.

Han, S., Navarro, J., Greve, R. A. and Adams, T. H. (1993). Translational repression of *brlA* expression prevents premature development in *Aspergillus*. *EMBO J 12*, 2449-2457.

Hanfrey, C., Franceschetti, M., Mayer, M. J., Illingworth, C. and Michael, A. J. (2002). Abrogation of upstream open reading frame-mediated translational control of a plant

S-adenosylmethionine decarboxylase results in polyamine disruption and growth perturbations. *J Biol Chem 277*, 44131-44139.

Hansen, W., Garcia, P. D. and Walter, P. (1986). *In vitro* protein translocation across the yeast endoplasmic reticulum: ATP-dependent post-translational translocation of the prepro-α-factor. *Cell 45*, 397-406.

Harashima, S. and Hinnebusch, A. G. (1986). Multiple *GCD* genes required for repression of *GCN4*, a transcriptional activator of amino acid biosynthetic genes in *Saccharomyces cerevisiae*. *Mol Cell Biol 6*, 3990-3998.

Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. and Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell 6*, 1099-1108.

Harigai, M., Miyashita, T., Hanada, M. and Reed, J. C. (1996). A cis-acting element in the *BCL-2* gene controls expression through translational mechanisms. *Oncogene 12*, 1369-1374.

Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M. and Aderem, A. (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature 410*, 1099-1103.

Haykinson, M. J., Johnson, L. M., Soong, J. and Johnson, R. C. (1996). The Hin dimer interface is critical for Fis-mediated activation of the catalytic steps of site-specific DNA inversion. *Curr Biol 6*, 163-177.

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

Hentze, M. W., Gebauer, F. and Preiss, T. (2007). *Cis*-regulatory sequences and *trans*-acting factors in translational control. In *Translational Control in Biology and Medicine*, pp. 269-296. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Herbreteau, C. H., Weill, L., Décimo, D., D., P., Darlix, J. L., Sargueil, B. and Ohlmann, T. (2005). HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat Struct Mol Biol 12*, 1001-1007.

Hershey, J. W. (1989). Protein phosphorylation controls translation rates. *J Biol Chem 264*, 20823-20826.

Hershey, J. W. B. (1991). Translational control in mammalian cells. *Annu Rev Biochem 60*, 717-755.

Hershey, J. W. B. and Merrick, W. C. (2000). Pathway and mechanism of initiation of

protein synthesis. In *Translational Control of Gene Expression*, pp. 33-88. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hill, J. R. and Morris, D. R. (1992). Cell-specific translation of *S*-adenosylmethionine decarboxylase mRNA. Regulation by the 5′ transcript leader. *J Biol Chem 267*, 21886-21893.

Hill, J. R. and Morris, D. R. (1993). Cell-specific translational regulation of *S*-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the *cis*-acting upstream open reading frame. *J Biol Chem 268*, 726-731.

Hinnebusch, A. G. (1996). Translational control of *GCN4*: gene-specific regulation by phosphorylation of eIF2. In *Translational Control*, pp. 199-244. Edited by J. W. B. Hershey, M. B. Mathews and N. Sonenberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

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

Hinnebusch, A. G. (2000). Mechanism and regulation of initiator methionyl-tRNA binding to ribosomes. In *Translational Control of Gene Expression*, pp. 185-244. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hinnebusch, A. G. (2005). Translational regulation of *GCN4* and the general amino acid control of yeast. *Annu Rev Microbiol 59*, 407-450.

Hinnebusch, A. G. (2006). eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci 31*, 553-562.

Hinton, T. M., Coldwell, M. J., Carpenter, G. A., Morley, S. J. and Pain, V. M. (2007). Functional analysis of individual binding activities of the scaffold protein eIF4G. *J Biol Chem 282*, 1695-1708.

Hoffmann, B., Valerius, O., Andermann, M. and Braus, G. H. (2001). Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Mol Biol Cell 12*, 2846-2857.

Hohn, T., Corsten, S., Dominguez, D., Futterer, J., Kirk, D., Hemmings-Mieszczak, M., Pooggin, M., Scharer-Hernandez, N. and Ryabova, L. (2001). Shunting is a translation strategy used by plant pararetroviruses (*Caulimoviridae*). *Micron 32*, 51-57.

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

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

Holzmann, K., Ambrosch, I., Elbling, L., Micksche, M. and Berger, W. (2001). A small upstream open reading frame causes inhibition of human major vault protein expression from a ubiquitous mRNA splice variant. *FEBS Letters 494*, 99-104.

Hood, H. M., Spevak, C. C. and Sachs, M. S. (2007). Evolutionary changes in the fungal carbamoyl-phosphate synthetase small subunit gene and its associated upstream open reading frame. *Fungal Genet Biol 44*, 93-104.

Hoshino, S., Imai, M., Kobayashi, T., Uchida, N. and Katada, T. (1999). The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3′-poly(A) tail of mRNA. Direct association of eRF3/GSPT with polyadenylate-binding protein. *J Biol Chem 274*, 16677-16680.

Huang, H. K., Yoon, H., Hannig, E. M. and Donahue, T. F. (1997). GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev 11*, 2396-2413.

Hurtaud, C., Gelly, C., Bouillaud, F. and Lévi-Meyrueis, C. (2006). Translation control of UCP2 synthesis by the upstream open reading frame. *Cell Mol Life Sci 63*, 1780-1789.

Iacono, M., Mignone, F. and Pesole, G. (2005). uAUG and uORFs in human and rodent 5′untranslated mRNAs. *Gene 349*, 97-105.

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

Ikeda, J. S., Schmitt, C. K., Darnell, S. C., Watson, P. R., Bispham, J., Wallis, T. S., Weinstein, D. L., Metcalf, E. S., Adams, P., O'Connor, C. D. and O'Brien, A. D. (2001). Flagellar phase variation of *Salmonella enterica* serovar Typhimurium contributes to virulence in the murine typhoid infection model but does not influence *Salmonella*-induced enteropathogenesis. *Infect Immun 69*, 3021-3030.

Imataka, H., Nakayama, K., Yasumoto, K., Mizuno, A., Fujii-Kuriyama, Y. and Hayami, M. (1994). Cell-specific translational control of transcription factor BTEB expression. The role of an upstream AUG in the 5′-untranslated region. *J Biol Chem 269*, 20668-20673.

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

Jacobson, A. (1996). Poly(A) metabolism and translation: the closed-loop model. In

Translational Control, pp. 451-480. Edited by J. W. B. Hershey, M. B. Mathews and N. Sonenberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Jacobson, A. and Peltz, S. W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem 65*, 693-739.

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

Jin, X., Turcott, E., Englehardt, S., Mize, G. J. and Morris, D. R. (2003). The two upstream open reading frames of oncogene *mdm2* have different translational regulatory properties. *J Biol Chem 278*, 25716-25721.

Jivotovskaya, A. V., Valasek, L., Hinnebusch, A. G. and Nielsen, K. H. (2006). Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. *Mol Cell Biol 26*, 1355-1372.

Jobling, S. A. and Gehrke, L. (1987). Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. *Nature 325*, 622-625.

Jousse, C., Bruhat, A., Carraro, V., Urano, F., Ferrara, M., Ron, D. and Fafournoux, P. (2001). Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the *chop* 5′UTR. *Nucl Acids Res 29*, 4341-4351.

Juven-Gershon, T. and Oren, M. (1999). *Mdm2*: the ups and downs. *Mol Med 5*, 71-83.

Kahvejian, A., Roy, G. and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol 66*, 293-300.

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M. G. and Alon, U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science 292*, 2080-2083.

Kamath, A. V. and Yanofsky, C. (1992). Characterization of the tryptophanase operon of Proteus vulgaris. Cloning, nucleotide sequence, amino acid homology, and *in vitro* synthesis of the leader peptide and regulatory analysis. *J Biol Chem 267*, 19978-19985.

Kapp, L. D. and Lorsch, J. R. (2004). The molecular mechanics of eukaryotic translation. *Annu Rev Biochem 73*, 657-704.

Kaufman, R. J. (2000). Double-stranded RNA-activated protein kinase PKR. In *Translational Control of Gene Expression*, pp. 503-528. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Press.

Kaziro, Y. (1978). The role of guanosine 5′-triphosphate in polypeptide chain elongation. *Biochim Biophys Acta 505*, 95-127.

Kim, J. and Leustek, T. (1996). Cloning and analysis of the gene for cystathionine γ-synthase from *Arabidopsis thaliana*. *Plant Mol Biol 32*, 1117-1124.

Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. *Int J Biochem Cell Biol 31*, 25-29.

Kimball, S. R. and Jefferson, L. S. (2000). Regulation of translation initiation in mammalian cells by amino acids. In *Translational Control of Gene Expression*, pp. 561-580. Edited by N. Sonenberg, J. W. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H. D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D. and Rhoads, R. E. (1994). Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4γ. *J Virol 68*, 5677-5684.

Knebel, A., Morrice, N. and Cohen, P. (2001). A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38δ. *EMBO J 20*, 4360-4369.

Kolupaeva, V. G., Unbehaun, A., Lomakin, I. B., Hellen, C. U. T. and Pestova, T. V. (2005). Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA 11*, 470-486.

Konecki, D. S., Aune, K. C., Tate, W. and Caskey, C. T. (1977). Characterization of reticulocyte release factor. *J Biol Chem 252*, 4514-4520.

Kontos, H., Napthine, S. and Brierley, I. (2001). Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol 21*, 8657-8670.

Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C., Merrick, W. C. and Rhoads, R. E. (2001). Characterization of the two eIF4A-binding sites on human eIF4G-1. *J Biol Chem 276*, 2872-2879.

Korneeva, N. L., First, E. A., Benoit, C. A. and Rhoads, R. E. (2005). Interaction between the NH2-terminal domain of eIF4A and the central domain of eIF4G modulates RNA-stimulated ATPase activity. *J Biol Chem 280*, 1872-1881.

Kos, M., Denger, S., Reid, G. and Gannon, F. (2002). Upstream open reading frames regulate the translation of the multiple mRNA variants of the estrogen receptor$_{\alpha}$. *J Biol*

Chem 277, 37131-37138.

Kovalchuke, O., Kambampati, R., Pladies, E. and Chakraburtty, K. (1998). Competition and cooperation amongst yeast elongation factors. *Eur J Biochem 258*, 986-993.

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

Kozak, M. (1980). Role of ATP in binding and migration of 40S ribosomal subunits. *Cell 22*, 459-467.

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

Kozak, M. (1989a). Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol 9*, 5073-5080.

Kozak, M. (1989b). The scanning model for translation: an update. *J Cell Biol 108*, 229-241.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol 115*, 887-903.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene 234*, 187-208.

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

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

Kurland, C. G. (1992). Translational accuracy and the fitness of bacteria. *Annu Rev Genet 26*, 29-50.

Kutsukake, K. and Iino, T. (1980). Inversions of specific DNA segments in flagellar phase variation of *Salmonella* and inversion systems of bacteriophages P1 and Mu. *Proc Natl Acad Sci U S A 77*, 7338-7341.

Kwak, S.-H. and Lee, S. H. (2001). The regulation of ornithine decarboxylase gene expression by sucrose and small upstream open reading frame in tomato (*Lycopersicon esculentum* Mill). *Plant Cell Physiol 42*, 314-323.

Kwon, H.-S., Lee, D.-K., Lee, J.-J., Edenberg, H. J., Ahn, Y.-h. and Hur, M.-W. (2001). Posttranscriptional regulation of human *ADH5*/*FDH* and *Myf6* gene expression by upstream AUG codons. *Arch Biochem Biophys 386*, 163-171.

LaGrandeur, T. E. and Parker, R. (1998). Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme. *EMBO J 17*, 1487-1496.

Lambein, I., Chiba, Y., Onouchi, H. and Naito, S. (2003). Decay kinetics of autogenously regulated *CGS1* mRNA that codes for cystathionine γ-synthase in *Arabidopsis thaliana*. *Plant Cell Physiol 44*, 893-900.

Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H. D., Klump, H., Kuechler, E., Skern, T. and Rhoads, R. E. (1993). Mapping the cleavage site in protein synthesis initiation factor eIF-4γ of the 2A proteases from human coxsackievirus and rhinovirus. *J Biol Chem 268*, 19200-19203.

Lamphear, B. J., Kirchweger, R., Skern, T. and Rhoads, R. E. (1995). Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. *J Biol Chem 270*, 21975-21983.

Landers, J. E., Cassel, S. L. and George, D. L. (1997). Translational enhancement of *mdm2* oncogene expression in human tumor cells containing a stabilized wild-type p53 protein. *Cancer Res 57*, 3562-3568.

Laurino, J. P., Thompson, G. M., Pacheco, E. and Castilho, B. A. (1999). The β subunit of eukaryotic translation initiation factor 2 binds mRNA through the lysine repeats and a region comprising the C_2-C_2 motif. *Mol Cell Biol 19*, 173-181.

Law, G. L., Raney, A., Heusner, C. and Morris, D. R. (2001). Polyamine regulation of ribosome pausing at the upstream open reading frame of *S*-adenosylmethionine decarboxylase. *J Biol Chem 276*, 38036-38043.

Lazaris-Karatzas, A. and Sonenberg, N. (1992). The mRNA 5′ cap-binding protein, eIF-4E, cooperates with v-*myc* or E1A in the transformation of primary rodent fibroblasts. *Mol Cell Biol 12*, 1234-1238.

Le, S. Y. and Maizel, J. V. J. (1998). Evolution of a common structural core in the internal ribosome entry sites of picornavirus. *Virus Genes 16*, 25-38.

Lee, J. H., Pestova, T. V., Shin, B. S., Cao, C., Choi, S. K. and Dever, T. E. (2002). Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proc Natl Acad Sci U S A 99*, 16689-16694.

Lee, M.-H. and Schedl, T. (2004). Translation repression by GLD-1 protects its mRNA targets from nonsense-mediated mRNA decay in *C. elegans*. *Genes Dev 18*, 1047-1059.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell 75*, 843-854.

Leeds, P., Wood, J. M., Lee, B. S. and Culbertson, M. R. (1992). Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol Cell Biol 12*, 2165-2177.

Lemm, I. and Ross, J. (2002). Regulation of c-*myc* mRNA decay by translational pausing in a coding region instability determinant. *Mol Cell Biol 22*, 3959-3969.

Lewis, J. D. and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem 247*, 461-469.

Lewis, J. D. and Tollervey, D. (2000). Like attracts like: getting RNA processing together in the nucleus. *Science 288*, 1385-1389.

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

Li, A. C., Brown, K. K., Silvestre, M. J., Willson, T. M., Palinski, W. and Glass, C. K. (2000). Peroxisome proliferator-activated receptor γ ligands inhibit development of atherosclerosis in LDL receptor-deficient mice. *J Clin Invest 106*, 523-531.

Lin, F., MacDougald, O. A., Diehl, A. M. and Lane, M. D. (1993). A 30-kDa alternative translation product of the CCAAT/enhancer binding protein α message: transcriptional activator lacking antimitotic activity. *Proc Natl Acad Sci U S A 90*, 9606-9610.

Lincoln, A. J., Monczak, Y., Williams, S. C. and Johnson, P. F. (1998). Inhibition of CCAAT/enhancer-binding protein α and β translation by upstream open reading frames. *J Biol Chem 273*, 9552-9560.

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

Lohmer, S., Maddaloni, M., Motto, M., Salamini, F. and Thompson, R. D. (1993). Translation of the mRNA of the maize transcriptional activator *Opaque-2* is inhibited by upstream open reading frames present in the leader sequence. *Plant Cell 5*, 65-73.

Lopinski, J. D., Dinman, J. D. and Bruenn, J. A. (2000). Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol 20*, 1095-1103.

Lovett, P. S. and Rogers, E. J. (1996). Ribosome regulation by the nascent peptide. *Microbiol Rev 60*, 366-385.

Lukaszewicz, M., Jerouville, B. and Boutry, M. (1998). Signs of translational regulation within the transcript leader of a plant plasma membrane H^+-ATPase gene. *Plant J 14*, 413-423.

Luo, Z., Freitag, M. and Sachs, M. S. (1995). Translational regulation in response to

changes in amino acid availability in *Neurospora crassa*. *Mol Cell Biol 15*, 5235-5245.

Luo, Z. and Sachs, M. S. (1996). Role of an upstream open reading frame in mediating arginine-specific translational control in *Neurospora crassa*. *J Bacteriol 178*, 2172-2177.

Lynch, M., Fitzgerald, C., Johnston, K. A., Wang, S. and Schmidt, E. V. (2004). Activated eIF4E-binding protein slows G1 progression and blocks transformation by c-*myc* without inhibiting cell growth. *J Biol Chem 279*, 3327-3339.

Maag, D., Fekete, C. A., Gryczynski, Z. and Lorsch, J. R. (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.

Mader, S., Lee, H., Pause, A. and Sonenberg, N. (1995). The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4γ and the translational repressors 4E-binding proteins. *Mol Cell Biol 15*, 4990-4997.

Maglione, D., Guerriero, V., Rambaldi, M., Russo, G. and Persico, M. G. (1993). Translation of the placenta growth factor mRNA is severely affected by a small open reading frame localized in the 5′ untranslated region. *Growth Factors 8*, 141-152.

Major, L. L., Poole, E. S., Dalphin, M. E., Mannering, S. A. and Tate, W. P. (1996). Is the in-frame termination signal of the *Escherichia coli* release factor-2 frameshift site weakened by a particularly poor context? *Nucl Acids Res 24*, 2673-2678.

Majumdar, R., Bandyopadhyay, A. and Maitra, U. (2003). Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J Biol Chem 278*, 6580-6587.

Mamane, Y., Petroulakis, E., Rong, L., Yoshida, K., Ler, L. W. and Sonenberg, N. (2004). eIF4E – from translation to transformation. *Oncogene 23*, 3172-3179.

Mamane, Y., Petroulakis, E., LeBacquer, O. and Sonenberg, N. (2006). mTOR, translation initiation and cancer. *Oncogene 25*, 6416-6422.

Manzella, J. M. and Blackshear, P. J. (1990). Regulation of rat ornithine decarboxylase mRNA translation by its 5′-untranslated region. *J Biol Chem 265*, 11817-11822.

Marcotrigiano, J., Gingras, A., Sonenberg, N. and Burley, S. (1999). Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol Cell 3*, 707-716.

Marcotrigiano, J., Lomakin, I., Sonenberg, N., Pestova, T., Hellen, C. and Burley, S. (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell 7*, 193-203.

Marth, J. D., Overell, R. W., Meier, K. E., Krebs, E. G. and Perlmutter, R. M. (1988). Translational activation of the *lck* proto-oncogene. *Nature 332*, 171-173.

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

Mathews, B. F. (1999). Lysine, thereonine, and methionine biosynthesis. In *Plant amino acids: Biochemistry and biotechnology*, pp. 205-226. Edited by B. K. Singh. New York, NY: Marcel Dekker Inc.

Mathews, M. B., Sonenberg, N. and Hershey, J. W. B. (2007). Origins and principles of translational control. In *Translational Control in Biology and Medicine*, pp. 1-40. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J. and Tate, W. P. (1995). Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc Natl Acad Sci U S A 92*, 5431-5435.

McKendrick, L., Morley, S. J., Pain, V. M., Jagus, R. and Joshi, B. (2001). Phosphorylation of eukaryotic initiation factor 4E (eIF4E) at Ser209 is not required for protein synthesis *in vitro* and *in vivo*. *Eur J Biochem 268*, 5375-5385.

McNicholas, P., Salavati, R. and Oliver, D. (1997). Dual regulation of *Escherichia coli secA* translation by distinct upstream elements. *J Mol Biol 265*, 128-141.

Mehta, A., Trotta, C. R. and Peltz, S. W. (2006). Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes Dev 20*, 939-953.

Meijer, H. A., Dictus, W. J., Keuning, E. D. and Thomas, A. A. (2000). Translational control of the *Xenopus laevis* connexin-41 5′-untranslated region by three upstream open reading frames. *J Biol Chem 275*, 30787-30793.

Mendez, R. and Richter, J. D. (2001). Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol 2*, 521-529.

Merrick, W. C. and Hershey, J. W. B. (1996). The pathway and mechanism of eukaryotic protein synthesis. In *Translational Control*, pp. 31-69. Edited by J. W. B. Hershey, M. B. Mathews and N. Sonenberg. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Meyuhas, O. and Hornstein, E. (2000). Translational control of TOP mRNAs. In *Translational Control of Gene Expression*, pp. 671-694. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Press.

Michelet, B., Lukaszewicz, M., Dupriez, V. and Boutry, M. (1994). A plant plasma membrane proton-ATPase gene is regulated by development and environment and shows signs of a translational regulation. *Plant Cell 6*, 1375-1389.

Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol 3*, reviews0004.0001 - reviews0004.0010.

Mikuni, O., Ito, K., Moffat, J., Matsumura, K., McCaughan, K., Nobukuni, T., Tate, W. and Nakamura, Y. (1994). Identification of the *prfC* gene, which encodes peptide-chain-release factor 3 of *Escherichia coli*. *Proc Natl Acad Sci U S A 91*, 5798-5802.

Miller, P. F. and Hinnebusch, A. G. (1989). Sequences that surround the stop codons of upstream open reading frames in *GCN4* mRNA determine their distinct functions in translational control. *Genes Dev 3*, 1217-1225.

Minich, W. B., Balasta, M. L., Goss, D. J. and Rhoads, R. E. (1994). Chromatographic resolution of *in vivo* phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF-4E: increased cap affinity of the phosphorylated form. *Proc Natl Acad Sci U S A 91*, 7668-7672.

Miron, M., Lasko, P. and Sonenberg, N. (2003). Signaling from Akt to FRAP/TOR targets both 4E-BP and S6K in *Drosophila melanogaster*. *Mol Cell Biol 23*, 9117-9126.

Mize, G. J., Ruan, H., Low, J. J. and Morris, D. R. (1998). The inhibitory upstream open reading frame from mammalian *S*-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J Biol Chem 273*, 32500-32505.

Moazed, D. and Noller, H. F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature 342*, 142-148.

Moldave, K. (1985). Eukaryotic protein synthesis. *Annu Rev Biochem 54*, 1109-1149.

Moore, P. B. and Steitz, T. A. (2003). After the ribosome structures: how does peptidyl transferase work? *RNA 9*, 155-159.

Morley, S. J. and Naegele, S. (2002). Phosphorylation of eukaryotic initiation factor (eIF) 4E is not required for *de novo* protein synthesis following recovery from hypertonic stress in human kidney cells. *J Biol Chem 277*, 32855-32859.

Morris, D. R. (1995). Growth control of translation in mammalian cells. *Prog Nucleic Acid Res Mol Biol 51*, 339-363.

Morris, D. R. and Geballe, A. P. (2000). Upstream open reading frames as regulators of

mRNA translation. *Mol Cell Biol 20*, 8635-8642.

Moustakas, A., Sonstegard, T. S. and Hackett, P. B. (1993). Alterations of the three short open reading frames in the Rous sarcoma virus leader RNA modulate viral replication and gene expression. *J Virol 67*, 4337-4349.

Munroe, D. and Jacobson, A. (1990). mRNA poly(A) tail, a 3′ enhancer of translational initiation. *Mol Cell Biol 10*, 3441-3455.

Muto, H., Nakatogawa, H. and Ito, K. (2006). Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell 22*, 545-552.

Nagy, L., Tontonoz, P., Alvarez, J. G. A., Chen, H. and Evans, R. M. (1998). Oxidized LDL regulates macrophage gene expression through ligand activation of PPARγ. *Cell 93*, 229-240.

Nakatogawa, H. and Ito, K. (2001). Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol Cell 7*, 185-192.

Nakatogawa, H. and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell 108*, 629-636.

Naranda, T., Sirangelo, I., Fabbri, B. J. and Hershey, J. W. (1995). Mutations in the NKXD consensus element indicate that GTP binds to the γ-subunit of translation initiation factor eIF2. *FEBS Lett 372*, 249-252.

Nierhaus, K. H. (1990). The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry 29*, 4997-5008.

Nomura, A., Iwasaki, Y., Saito, M., Aoki, Y., Yamamori, E., Ozaki, N., Tachikawa, K., Mutsuga, N., Morishita, M., Yoshida, M., Asai, M., Oiso, Y. and Saito, H. (2001). Involvement of upstream open reading frames in regulation of rat V1b vasopressin receptor expression. *Am J Physiol Endocrinol Metab 280*, E780-787.

Nyunoya, H. and Lusty, C. J. (1984). Sequence of the small subunit of yeast carbamyl phosphate synthetase and identification of its catalytic domain. *J Biol Chem 259*, 9790-9798.

Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P. and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science 292*, 897-902.

Ogle, J. M., Murphy, F. V., Tarry, M. J. and Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell 111*, 721-732.

Olayioye, M. A., Neve, R. M., Lane, H. A. and Hynes, N. E. (2000). The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J 19*, 3159-3167.

Oliver, D., Norman, J. and Sarker, S. (1998). Regulation of *Escherichia coli secA* by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. *J Bacteriol 180*, 5240-5242.

Onouchi, H., Nagami, Y., Haraguchi, Y., Nakamoto, M., Nishimura, Y., Sakurai, R., Nagao, N., Kawasaki, D., Kadokura, Y. and Naito, S. (2005). Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the *CGS1* gene of *Arabidopsis*. *Genes Dev 19*, 1799-1810.

Orbach, M. J., Sachs, M. S. and Yanofsky, C. (1990). The *Neurospora crassa arg-2* locus. Structure and expression of the gene encoding the small subunit of arginine-specific carbamoyl phosphate synthetase. *J Biol Chem 265*, 10981-10987.

Orbach, M. J. and Sachs, M. S. (1991). The Orbach/Sachs cosmid library of *N. crassa* DNA sequences (pMOcosX). *Fungal Genet Newsl 38*, 97.

Orban, T. I. and Izaurralde, E. (2005). Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA 11*, 459-469.

Oridate, N., Kim, H.-J., Xu, X. and Lotan, R. (2005). Growth inhibition of head and neck squamous carcinoma cells by small interfering RNAs targeting eIF4E or cyclin D1 alone or combined with cisplatin. *Cancer Biol Ther 4*, 318-323.

Orton, K. C., Ling, J., Waskiewicz, A. J., Cooper, J. A., Merrick, W. C., Korneeva, N. L., Rhoads, R. E., Sonenberg, N. and Traugh, J. A. (2004). Phosphorylation of Mnk1 by caspase-activated Pak2/γ-PAK inhibits phosphorylation and interaction of eIF4G with Mnk. *J Biol Chem 279*, 38649-38657.

Ossipow, V., Descombes, P. and Schibler, U. (1993). CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials. *Proc Natl Acad Sci U S A 90*, 8219-8223.

Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C. and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature 425*, 257-263.

Pape, T., Wintermeyer, W. and Rodnina, M. V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J 17*, 7490-7497.

Pape, T., Wintermeyer, W. and Rodnina, M. (1999). Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J 18*, 3800-3807.

Parola, A. L. and Kobilka, B. K. (1994). The peptide product of a 5′ leader cistron in the β_2 adrenergic receptor mRNA inhibits receptor synthesis. *J Biol Chem 269*, 4497-4505.

Parra-Palau, J.-L., Scheper, G. C., Wilson, M. L. and Proud, C. G. (2003). Features in the N and C termini of the MAPK-interacting kinase Mnk1 mediate its nucleocytoplasmic shuttling. *J Biol Chem 278*, 44197-44204.

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

Perez, L., Wills, J. W. and Hunter, E. (1987). Expression of the Rous sarcoma virus *env* gene from a simian virus 40 late-region replacement vector: effects of upstream initiation codons. *J Virol 61*, 1276-1281.

Pesole, G., Liuni, S., Grillo, G. and Saccone, C. (1997). Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene 205*, 95-102.

Pesole, G., Gissi, C., Grillo, G., Licciulli, F., Liuni, S. and Saccone, C. (2000). Analysis of oligonucleotide AUG start codon context in eukariotic mRNAs. *Gene 261*, 85-91.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F. and Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. *Gene 276*, 73-81.

Pestova, T. V., Hellen, C. U. and Shatsky, I. N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol 16*, 6859-6869.

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

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

Pestova, T. V. and Hellen, C. U. (2000). The structure and function of initiation factors in eukaryotic protein synthesis. *Cell Mol Life Sci 57*, 651-674.

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

Pestova, T. V. and Kolupaeva, V. G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev 16*, 2906-2922.

Pestova, T. V. and Hellen, C. U. T. (2003). Translation elongation after assembly of ribosomes on the cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev 17*, 181-186.

Pestova, T. V., Lorsch, J. R. and Hellen, C. U. T. (2007). The mechanism of translation initiation in eukaryotes. In *Translational Control in Biology and Medicine*, pp. 87-128. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Pinto, I., Na, J. G., Sherman, F. and Hampsey, M. (1992). *Cis*- and *trans*-acting suppressors of a translation initiation defect at the *cyc1* locus of *Saccharomyces cerevisiae*. *Genetics 132*, 97-112.

Polymenis, M. and Schmidt, E. V. (1997). Coupling of cell division to cell growth by translational control of the G_1 cyclin *CLN3* in yeast. *Genes Dev 11*, 2522-2531.

Pooggin, M. M., Hohn, T. and Futterer, J. (1998). Forced evolution reveals the importance of short open reading frame A and secondary structure in the cauliflower mosaic virus 35S RNA leader. *J Virol 72*, 4157-4169.

Pooggin, M. M., Hohn, T. and Futterer, J. (2000). Role of a short open reading frame in ribosome shunt on the cauliflower mosaic virus RNA leader. *J Biol Chem 275*, 17288-17296.

Prats, A. C. and Prats, H. (2002). Translational control of gene expression: role of IRESs and consequences for cell transformation and angiogenesis. *Prog Nucleic Acid Res Mol Biol 72*, 367-413.

Preiss, T. and Hentze, M. W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. *Nature 392*, 516-520.

Preiss, T. and Hentze, M. W. (2003). Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays 25*, 1201-1211.

Prevot, D., Darlix, J.-L. and Ohlmann, T. (2003). Conducting the initiation of protein synthesis: the role of eIF4G. *Biol Cell 95*, 141-156.

Proud, C. (2000). Control of the elongation phase of protein synthesis. In *Translational Control of Gene Expression*, pp. 719-740. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Proud, C. G. (2005). The eukaryotic initiation factor 4E-binding proteins and apoptosis. *Cell Death Differ 12*, 541-546.

Qin, S. L., Xie, A. G., Bonato, M. C. and McLaughlin, C. S. (1990). Sequence analysis of the translational elongation factor 3 from *Saccharomyces cerevisiae*. *J Biol Chem 265*,

1903-1912.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell 108*, 557-572.

Rana, T. M. (2007). Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol 8*, 23-36.

Raney, A., Baron, A. C., Mize, G. J., Law, G. L. and Morris, D. R. (2000). *In vitro* translation of the upstream open reading frame in the mammalian mRNA encoding *S*-adenosylmethionine decarboxylase. *J Biol Chem 275*, 24444-24450.

Raney, A., Law, G. L., Mize, G. J. and Morris, D. R. (2002). Regulated translation termination at the upstream open reading frame in *S*-adenosylmethionine decarboxylase mRNA. *J Biol Chem 277*, 5988-5994.

Raychaudhuri, P., Stringer, E. A., Valenzuela, D. M. and Maitra, U. (1984). Ribosomal subunit antiassociation activity in rabbit reticulocyte lysates. Evidence for a low molecular weight ribosomal subunit antiassociation protein factor (Mr = 25,000). *J Biol Chem 259*, 11930-11935.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R. and Ruvkun, G. (2000). The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature 403*, 901-906.

Reynolds, K., Zimmer, A. M. and Zimmer, A. (1996). Regulation of RAR β_2 mRNA expression: evidence for an inhibitory peptide encoded in the 5′-untranslated region. *J Cell Biol 134*, 827-835.

Rodnina, M. V., Savelsbergh, A., Katunin, V. I. and Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature 385*, 37-41.

Rodnina, M. V. and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu Rev Biochem 70*, 415-435.

Rogers, G. W., Jr., Richter, N. J., Lima, W. F. and Merrick, W. C. (2001). Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J Biol Chem 276*, 30914-30922.

Ron, D. and Harding, H. P. (2000). PERK and translational control by stress in endoplasmic reticulum. In *Translational Control of Gene Expression*, pp. 547-560. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Ron, D. and Harding, H. P. (2007). eIF2α phosphorylation in cellular stress responses and

diseases. In *Translational Control in Biology and Medicine*, pp. 345-368. Edited by M. B. Mathews, N. Sonenberg and J. W. B. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Rosu, V., Chevance, F. F. V., Karlinsey, J. E., Hirano, T. and Hughes, K. T. (2006). Translation inhibition of the *Salmonella fliC* gene by the *fliC* 5′ untranslated region, *fliC* coding sequences, and FlgM. *J Bacteriol 188*, 4497-4507.

Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. and Sonenberg, N. (1990). Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol Cell Biol 10*, 1134-1144.

Ruan, H., Hill, J. R., Fatemie-Nainie, S. and Morris, D. R. (1994). Cell-specific translational regulation of *S*-adenosylmethionine decarboxylase mRNA. Influence of the structure of the 5′ transcript leader on regulation by the upstream open reading frame. *J Biol Chem 269*, 17905-17910.

Ruan, H., Shantz, L. M., Pegg, A. E. and Morris, D. R. (1996). The upstream open reading frame of the mRNA encoding *S*-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. *J Biol Chem 271*, 29576-29582.

Ruiz-Echevarria, M. J. and Peltz, S. W. (2000). The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell 101*, 741-751.

Russell, D. W. and Spremulli, L. L. (1979). Purification and characterization of a ribosome dissociation factor (eukaryotic initiation factor 6) from wheat germ. *J Biol Chem 254*, 8796-8800.

Ruvinsky, I., Sharon, N., Lerer, T., Cohen, H., Stolovich-Rain, M., Nir, T., Dor, Y., Zisman, P. and Meyuhas, O. (2005). Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. *Genes Dev 19*, 2199-2211.

Ruvinsky, I. and Meyuhas, O. (2006). Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. *Trends Biochem Sci 31*, 342-348.

Ryabova, L. A. and Hohn, T. (2000). Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. *Genes Dev 14*, 817-829.

Ryabova, L. A., Pooggin, M. M., Dominguez, D. I. and Hohn, T. (2000). Continuous and discontinuous ribosome scanning on the cauliflower mosaic virus 35S RNA leader is controlled by short open reading frames. *J Biol Chem 275*, 37278-37284.

Ryabova, L. A., Pooggin, M. M. and Hohn, T. (2002). Viral strategies of translation initiation: ribosomal shunt and reinitiation. *Prog Nucleic Acid Res Mol Biol 72*, 1-39.

Ryan, M. D., King, A. M. Q. and Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol 72*, 2727-2732.

Ryan, M. D. and Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J 13*, 928-933.

Ryan, M. D., Monaghan, S. and Flint, M. (1998). Virus-encoded proteinases of the *Flaviviridae*. *J Gen Virol 79*, 947-959.

Ryan, M. D., Donnelly, M. L. L., Lewis, A., Mehrotra, A., Wilkie, J. and Gani, D. (1999). A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorg Chem 27*, 55-79.

Ryazanov, A. G., Natapov, P. G., Shestakova, E. A., Severin, F. F. and Spirin, A. S. (1988). Phosphorylation of the elongation factor 2: The fifth Ca^{2+}/calmodulin-dependent system of protein phosphorylation. *Biochimie 70*, 619-626.

Sachs, A. (2000). Physical and functional interactions between the mRNA cap structure and the poly(A) tail. In *Translational Control of Gene Expression*, pp. 447-466. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sachs, M. S. and Geballe, A. P. (2002). Biochemistry. Sense and sensitivity – controlling the ribosome. *Science 297*, 1820-1821.

Sachs, M. S., Wang, Z., Gaba, A., Fang, P., Belk, J., Ganesan, R., Amrani, N. and Jacobson, A. (2002). Toeprint analysis of the positioning of translation apparatus components at initiation and termination codons of fungal mRNAs. *Methods 26*, 105-114.

Salas-Marco, J. and Bedwell, D. M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol Cell Biol 24*, 7769-7778.

Sarrazin, S., Starck, J., Gonnet, C., Doubeikovski, A., Melet, F. and Morle, F. (2000). Negative and translation termination-dependent positive control of FLI-1 protein synthesis by conserved overlapping 5′ upstream open reading frames in *Fli-1* mRNA. *Mol Cell Biol 20*, 2959-2969.

Sasaki, J. and Nakashima, N. (1999). Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus *in vitro*. *J Virol 73*, 1219-1226.

Sasaki, J. and Nakashima, N. (2000). From the cover: methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci U S A 97*, 1512-1515.

Schägger, H. and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem 166*, 368-379.

Schleiss, M. R., Degnin, C. R. and Geballe, A. P. (1991). Translational control of human cytomegalovirus gp48 expression. *J Virol 65*, 6782-6789.

Schluter, G., Boinska, D. and Nieman-Seyde, S.-C. (2000). Evidence for translational repression of the SOCS-1 major open reading frame by an upstream open reading frame. *Biochem Biophys Res Commun 268*, 255-261.

Schmidt, E. V. (2004). The role of c-*myc* in regulation of translation initiation. *Oncogene 23*, 3217-3221.

Scolnick, E., Tompkins, R., Caskey, T. and Nirenberg, M. (1968). Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A 61*, 768-774.

Sedman, S. A. and Mertz, J. E. (1988). Mechanisms of synthesis of virion proteins from the functionally bigenic late mRNAs of simian virus 40. *J Virol 62*, 954-961.

Sedman, S. A., Good, P. J. and Mertz, J. E. (1989). Leader-encoded open reading frames modulate both the absolute and relative rates of synthesis of the virion proteins of simian virus 40. *J Virol 63*, 3884-3893.

Seipelt, J., Guarne, A., Bergmann, E., James, M., Sommergruber, W., Fita, I. and Skern, T. (1999). The structures of picornaviral proteinases. *Virus Res 62*, 159-168.

Seit-Nebi, A., Frolova, L., Justesen, J. and Kisselev, L. (2001). Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucl Acids Res 29*, 3982-3987.

Shantz, L. M. and Pegg, A. E. (1999). Translational regulation of ornithine decarboxylase and other enzymes of the polyamine pathway. *Int J Biochem Cell Biol 31*, 107-122.

Shen, W. C. and Ebbole, D. J. (1996). Cross-pathway and pathway-specific control of amino acid biosynthesis in *Magnaporthe grisea*. *Fungal Genet Biol 21*, 40-49.

Shin, B. S., Maag, D., Roll-Mecak, A., Arefin, M. S., Burley, S. K., Lorsch, J. R. and Dever, T. E. (2002). Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. *Cell 111*, 1015-1025.

Si, K. and Maitra, U. (1999). The *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 6 does not function as a translation initiation factor. *Mol Cell Biol 19*, 1416-1426.

Siegel, V. and Walter, P. (1988). Each of the activities of signal recognition particle (SRP)

is contained within a distinct domain: analysis of biochemical mutants of SRP. *Cell 52*, 39-49.

Simon, M. and Silverman, M. (1983). Recombinational regulation of gene expression in bacteria. In *Gene Function in Prokaryotes*, pp. 211-227. Edited by J. Beckwith, J. Davies and J. A. Gallant. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Singh, C. R., Lee, B., Udagawa, T., Mohammad-Qureshi, S. S., Yamamoto, Y., Pavitt, G. D. and Asano, K. (2006). An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. *EMBO J 25*, 4537-4546.

Skogerson, L. and Wakatama, E. (1976). A ribosome-dependent GTPase from yeast distinct from elongation factor 2. *Proc Natl Acad Sci U S A 73*, 73-76.

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R. and Ruvkun, G. (2000). The *lin-41* RBCC Gene Acts in the *C. elegans* Heterochronic Pathway between the *let-7* Regulatory RNA and the LIN-29 Transcription Factor. *Mol Cell 5*, 659-669.

Soifer, H. S., Rossi, J. J. and Sætrom, P. (2007). MicroRNAs in disease and potential therapeutic applications. *Mol Ther 15*, 2070-2079.

Sommergruber, W., Zorn, M., Blaas, D., Fessl, F., Volkmann, P., Maurer-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M., Skern, T. and Keuchler, E. (1989). Polypeptide 2A of human rhinovirus type 2: identification as a protease and characterization by mutational analysis. *Virology 169*, 68-77.

Somogyi, P., Jenner, A. J., Brierley, I. and Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol 13*, 6931-6940.

Sonenberg, N., Morgan, M. A., Merrick, W. C. and Shatkin, A. J. (1978). A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5′-terminal cap in mRNA. *Proc Natl Acad Sci U S A 75*, 4843-4847.

Song, H., Mugnier, P., Das, A., Webb, H., Evans, D., Tuite, M., Hemmings, B. and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1 – mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell 100*, 311-321.

Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G. and Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae* – tRNA-ribosome and subunit-subunit interactions. *Cell 107*, 373-386.

Spevak, C. C., Park, E.-H., Geballe, A. P., Pelletier, J. and Sachs, M. S. (2006). *her-2* upstream open reading frame effects on the use of downstream initiation codons. *Biochem Biophys Res Commun 350*, 834-841.

Srivastava, S., Verschoor, A. and Frank, J. (1992). Eukaryotic initiation factor 3 does not

prevent association through physical blockage of the ribosomal subunit-subunit interface. *J Mol Biol 226*, 301-304.

Stansfield, I., Jones, K. M., Kushnirov, V. V., Dagkesamanskaya, A. R., Poznyakovski, A. I., Paushkin, S. V., Nierras, C. R., Cox, B. S., Ter-Avanesyan, M. D. and Tuite, M. F. (1995). The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in *Saccharomyces cerevisiae*. *EMBO J 14*, 4365-4373.

Stebbins-Boaz, B., Cao, Q., de Moor, C. H., Mendez, R. and Richter, J. D. (1999). Maskin is a CPEB-associated factor that transiently interacts with elF-4E. *Mol Cell 4*, 1017-1027.

Steel, L. F., Telly, D. L., Leonard, J., Rice, B. A., Monks, B. and Sawicki, J. A. (1996). Elements in the murine c-*mos* messenger RNA 5′-untranslated region repress translation of downstream coding sequences. *Cell Growth Differ 7*, 1415-1424.

Stewart, V. and Yanofsky, C. (1986). Role of leader peptide synthesis in tryptophanase operon expression in *Escherichia coli* K-12. *J Bacteriol 167*, 383-386.

Stocker, B. A. D. (1949). Measurement of the rate of mutation of flagellar antigenic phase in *Salmonella typhimurium*. *J Hyg 47*, 398-413.

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

Stortchevoi, A. A. (2006). Misacylation of tRNA in prokaryotes: a re-evaluation. *Cell Mol Life Sci 63*, 820-831.

Suzuki, H. and Iino, T. (1973). *In vitro* synthesis of phase-specific flagellin of *Salmonella*. *J Mol Biol 81*, 57-70.

Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., Suyama, A. and Sugano, S. (2000). Statistical analysis of the 5′ untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. *Genomics 64*, 286-297.

Tarun, S. Z., Jr. and Sachs, A. B. (1995). A common function for mRNA 5′ and 3′ ends in translation initiation in yeast. *Genes Dev 9*, 2997-3007.

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

Taylor, D. J., Nilsson, J., Merrill, A. R., Andersen, G. R., Nissen, P. and Frank, J. (2007). Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J 26*, 2421-2431.

Tee, A. R. and Proud, C. G. (2002). Caspase cleavage of initiation factor 4E-binding protein 1 yields a dominant inhibitor of cap-dependent translation and reveals a novel regulatory motif. *Mol Cell Biol 22*, 1674-1683.

Tenson, T. and Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. *Cell 108*, 591-594.

Thompson, R. C., Dix, D. B. and Karim, A. M. (1986). The reaction of ribosomes with elongation factor Tu•GTP complexes. Aminoacyl-tRNA-independent reactions in the elongation cycle determine the accuracy of protein synthesis. *J Biol Chem 261*, 4868-4874.

Tuazon, P. T., Merrick, W. C. and Traugh, J. A. (1989). Comparative analysis of phosphorylation of translational initiation and elongation factors by seven protein kinases. *J Biol Chem 264*, 2773-2777.

Uchida, N., Hoshino, S.-i., Imataka, H., Sonenberg, N. and Katada, T. (2002). A novel role of the mammalian GSPT/eRF3 associating with poly(A)-binding protein in cap/poly(A)-dependent translation. *J Biol Chem 277*, 50286-50292.

Vagner, S., Galy, B. and Pyronnet, S. (2001). Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep 2*, 893-898.

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

Vestergaard, B., Van, L. B., Andersen, G. R., Nyborg, J., Buckingham, R. H. and Kjeldgaard, M. (2001). Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol Cell 8*, 1375-1382.

Vilela, C., Linz, B., Rodrigues-Pousada, C. and McCarthy, J. E. (1998). The yeast transcription factor genes *YAP1* and *YAP2* are subject to differential control at the levels of both translation and mRNA stability. *Nucl Acids Res 26*, 1150-1159.

Vilela, C., Ramirez, C. V., Linz, B., Rodrigues-Pousada, C. and McCarthy, J. E. (1999). Post-termination ribosome interactions with the 5′UTR modulate yeast mRNA stability. *EMBO J 18*, 3139-3152.

Vilela, C., Velasco, C., Ptushkina, M. and McCarthy, J. E. (2000). The eukaryotic mRNA decapping protein Dcp1 interacts physically and functionally with the eIF4F translation initiation complex. *EMBO J 19*, 4372-4382.

Wang, L. and Wessler, S. R. (1998). Inefficient reinitiation is responsible for upstream open reading fame-mediated translational repression of the maize *R* gene. *Plant Cell 10*,

1733-1746.

Wang, X., Li, W., Williams, M., Terada, N., Alessi, D. R. and Proud, C. G. (2001). Regulation of elongation factor 2 kinase by p90 (RSK1) and p70 S6 kinase. *EMBO J 20*, 4370-4379.

Wang, X., Li, W., Parra, J.-L., Beugnet, A. and Proud, C. G. (2003). The C terminus of initiation factor 4E-binding protein 1 contains multiple regulatory features that influence its function and phosphorylation. *Mol Cell Biol 23*, 1546-1557.

Wang, Z. and Sachs, M. S. (1997a). Arginine-specific regulation mediated by the *Neurospora crassa arg-2* upstream open reading frame in a homologous, cell-free *in vitro* translation system. *J Biol Chem 272*, 255-261.

Wang, Z. and Sachs, M. S. (1997b). Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. *Mol Cell Biol 17*, 4904-4913.

Wang, Z., Fang, P. and Sachs, M. S. (1998). The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol Cell Biol 18*, 7528-7536.

Wang, Z., Gaba, A. and Sachs, M. S. (1999). A highly conserved mechanism of regulated ribosome stalling mediated by fungal arginine attenuator peptides that appears independent of the charging status of arginyl-tRNAs. *J Biol Chem 274*, 37565-37574.

Weng, Y., Czaplinski, K. and Peltz, S. W. (1996a). Identification and characterization of mutations in the *UPF1* gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Mol Cell Biol 16*, 5491-5506.

Weng, Y., Czaplinski, K. and Peltz, S. W. (1996b). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol Cell Biol 16*, 5477-5490.

Wera, S., Fernandez, A., Lamb, N. J. C., Turowski, P., Hemmings-Mieszczak, M., Mayer-Jaekel, R. E. and Hemmings, B. A. (1995). Deregulation of translational control of the 65-kDa regulatory subunit (PR65α) of protein phosphatase 2A leads to multinucleated cells. *J Biol Chem 270*, 21374-21381.

Werner, M., Feller, A., Messenguy, F. and Pierard, A. (1987). The leader peptide of yeast gene *CPA1* is essential for the translational repression of its expression. *Cell 49*, 805-813.

Wickens, M., Goodwin, E. B., Kimble, J., Strickland, S. and Hentze, M. W. (2000). Translational control of developmental decisions. In *Translational Control of Gene Expression*, pp. 295-370. Edited by N. Sonenberg, J. W. B. Hershey and M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Wiese, A., Elzinga, N., Wobbes, B. and Smeekens, S. (2005). Sucrose-induced translational repression of plant bZIP-type transcription factors. *Biochem Soc Trans 33*, 272-275.

Wightman, B., Ha, I. and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell 75*, 855-862.

Williams, M. A. and Lamb, R. A. (1989). Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins. *J Virol 63*, 28-35.

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

Wilusz, C. J., Wormington, M. and Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol 2*, 237-246.

Wintermeyer, W., Savelsbergh, A., Semenkov, Y. P., Katunin, V. I. and Rodnina, M. V. (2001). Mechanism of elongation factor G function in tRNA translocation on the ribosome. *Cold Spring Harb Symp Quant Biol 66*, 449-458.

Woolhead, C. A., Johnson, A. E. and Bernstein, H. D. (2006). Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell 22*, 587-598.

Wright, M. B., Howell, E. A. and Gaber, R. F. (1996). Amino acid substitutions in membrane-spanning domains of Hol1, a member of the major facilitator superfamily of transporters, confer nonselective cation uptake in Saccharomyces cerevisiae. *J Bacteriol 178*, 7197-7205.

Wu, C., Amrani, N., Jacobson, A., Sachs, M. S. and Jon, L. (2007). The use of fungal *in vitro* systems for studying translational regulation. In *Methods Enzymol*, pp. 203-225: Academic Press.

Wu, J. and Miller, B. L. (1997). *Aspergillus* asexual reproduction and sexual reproduction are differentially affected by transcriptional and translational mechanisms regulating stunted gene expression. *Mol Cell Biol 17*, 6191-6201.

Wu, X., Bayle, J. H., Olson, D. and Levine, A. J. (1993). The p53-mdm-2 autoregulatory feedback loop. *Genes Dev 7*, 1126-1132.

Xu, P., Vernooy, S. Y., Guo, M. and Hay, B. A. (2003). The *Drosophila* microRNA *Mir-14* suppresses cell death and is required for normal fat metabolism. *Curr Biol 13*, 790-795.

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

Yarden, Y. and Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol 2*, 127-137.

Zavialov, A. V., Buckingham, R. H. and Ehrenberg, M. (2001). A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell 107*, 115-124.

Zavialov, A. V., Mora, L., Buckingham, R. H. and Ehrenberg, M. (2002). Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol Cell 10*, 789-798.

Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L. and Philippe, M. (1995). Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J 14*, 4065-4072.

Zimmer, A., Zimmer, A. M. and Reynolds, K. (1994). Tissue specific expression of the retinoic acid receptor-β2: regulation by short open reading frames in the 5′-noncoding region. *J Cell Biol 127*, 1111-1119.

APPENDIX

CHARACTERIZATION OF CHROMOSOME ENDS IN THE FILAMENTOUS FUNGUS *NEUROSPORA CRASSA*[*]

Abstract

Telomeres and subtelomere regions have vital roles in cellular homeostasis and can facilitate niche adaptation. However, information on telomere/subtelomere structure is still limited to a small number of organisms. The availability of genome sequence data promised to rectify this situation but, unfortunately, telomeres tend to be poorly represented in genome sequence databases. Prior to this project, the *Neurospora crassa* genome assembly contained only seven of the fourteen telomeres. The missing telomeres were first identified through bioinformatic mining of raw sequence data, and were then assigned chromosomal locations based on paired-end read information, or by RFLP mapping. One of the *Neurospora* telomeres is attached to the ribosomal repeat array, which appears to be prone to terminal truncation. The remaining chromosome ends have atypical structures because they lack distinct subtelomere domains and are devoid of telomere-associated tandem repeats. Many of the chromosome ends terminate in highly AT-rich sequences that appear to be the products of repeat-induced point mutation yet, surprisingly, most are not repeated sequences. Several chromosome termini in the Oak Ridge strain were compared to their homologous counterpart in Mauriceville. This revealed that, in nearly all cases, the sequences immediately adjacent to the telomeres are genome-specific.

[*]Collaboration with the laboratory of Dr. Eric U. Selker at the University of Oregon, and the laboratories of Dr. Chuck Staben and Dr. Mark L. Farman at the University of Kentucky. Material in this or similar form is contained in the manuscript entitled *Characterization of chromosome ends in the filamentous fungus Neurospora crassa* by Wu, C., Kim, Y.-S., Smith, K., Li, W., Hood, H. M., Staben, C., Selker, E. U., Sachs, M. S. and Farman, M. L..

Introduction

Eukaryotic chromosomes are linear molecules that terminate in specialized sequences known as telomeres. Telomeres are added on to the 3′-end of the chromosome to prevent loss of DNA from the lagging strand during replication (Henderson *et al.*, 1987; Nakamura *et al.*, 1997; Zakian, 1989). In most eukaryotes, telomeres consist of tandem arrays of simple sequence repeats (Blackburn, 1990; Zakian, 1989). However, notable exceptions are *Drosophila* and some other Dipterans, which instead possess tandem arrays of retrotransposons at their chromosome ends (Shpiz and Kalmykova, 2007). Telomeres made up of simple sequence repeats vary in sequence among organisms, although the strand that reads 5′ to 3′ toward the chromosome ends tends to be G-rich. For example, the *Saccharomyces cerevisiae* chromosome ends contain a TG-rich repeat sequence (Walmsley *et al.*, 1984), plants typically have $(TTTAGGG)_n$ (Richards and Ausubel, 1988), while the telomeres of humans (Moyzis *et al.*, 1988), other mammals (Meyne *et al.*, 1989), and filamentous fungi (Schechtman, 1987) are made up of the repeating unit, TTAGGG.

The 3′-strand extends as an overhang and is a capable of base-pairing with itself using non-Watson-Crick interactions (Henderson *et al.*, 1987). These fold-backs make telomeres refractory to cloning unless the 3′-tails are removed by enzymatic treatment. Consequently, for most organisms, there is very little information on the organization of chromosome ends. Nevertheless, the characterization of terminal chromosome regions in a few model organisms has revealed striking similarities in organization. In most cases, the sequences that reside adjacent to the telomere repeats tend to be duplicated at other chromosome ends. As such, these sequences define a specific subtelomere region. Two domains are often discernable within the subtelomere. Immediately adjacent to the telomere repeats lie the distal subtelomere domains, which are usually dispersed among several chromosome ends and contain various types of tandem repeat motifs. Next, one finds the proximal domains. These tend to be duplicated at fewer ends but they often contain clusters of related genes (Flint *et al.*, 1997; Nickles and McEachern, 2004; Pryde *et al.*, 1997).

Intriguingly, in microbial eukaryotes, subtelomeric gene clusters often have roles

in niche adaptation. For example, the *S. cerevisiae* subtelomeres contain families of genes involved in sugar utilization, and the types of genes that are amplified in these regions depend on the niche from which a given strain is isolated (Denayrolles *et al.*, 1997; Ness and Aigle, 1995). Other terminally amplified sequences in *S. cerevisiae* include the *FLO* and *PAU* genes, which are involved in flocculation and anaerobic growth, respectively (Pryde and Louis, 1997). These too are traits that are likely to be adaptive. In microbial pathogens of humans, such as the protists, *Plasmodium falciparum* (malaria) and *Trypanosoma brucei* (sleeping sickness), and the fungus *Pneumocystis carinii*, the subtelomeres contain families of variant genes coding for surface proteins (Barry *et al.*, 2003). These organisms use various mechanisms to switch expression among different gene copies — a strategy that allows them to evade the immune system (Barry *et al.*, 2003; Donelson *et al.*, 1998; Hernandez-Rivas *et al.*, 1996; Wada and Nakamura, 1996). Interestingly, humans and chickens have large families of olfactory genes, many of which are encoded by subtelomeric gene clusters (Riethman *et al.*, 2004; Trask *et al.*, 1998).

From this small number of examples, it appears that the terminal chromosome regions are frequently co-opted for gene amplification and diversification, and that some microbes have taken this a step further by developing active mechanisms for switching expression among the subtelomeric genes. Nevertheless, characterization of chromosome ends in additional eukaryotic microbes is needed to determine if this situation is exceptional, or the norm. It should be noted that telomeres have been cloned from many different microbes, including filamentous fungi such as *N. crassa* (Schechtman, 1987; Schechtman, 1990), *Aspergillus nidulans* (Bhattacharyya and Blackburn, 1997; Connelly and Arst, 1991), *A. oryzae* (Kusumoto *et al.*, 2003), *Cladosporium fulvum* (Coleman *et al.*, 1993), and *Fusarium oxysporum* (Powell and Kistler, 1990). However, in most studies, the number of chromosome ends that were characterized was very small, as was the amount of sequence information obtained for each chromosome end. Therefore, these studies were insufficient to provide evidence for subtelomeric gene amplifications.

N. crassa is a model organism for genetic, biochemical and cell biology studies, and was the first filamentous fungus to have a publicly-available genome sequence (Galagan *et al.*, 2003). Prior to the genome sequencing project, two *N. crassa* telomeres

had been cloned (Schechtman, 1987; Schechtman, 1990). These consisted of tandem arrays of the motif, TTAGGG. One telomere mapped at the right-hand end of chromosome *V* (*TEL_VR)* and a putative transposable element was identified immediately adjacent to the telomere repeats (Schechtman, 1990). This element, known as Pogo, exists in at least nine copies in the Oak Ridge strain, however only the one copy is linked to a telomere (Schechtman, 1987). A second telomere, whose identity was not known at the time (TC8), was also characterized. Apart from the telomere repeats, this end exhibited no similarity to the *TEL_VR* sequence (Schechtman, 1990). Use of a $(TTAGGG)_4$ oligomer probe to perform a Southern hybridization analysis of DNA from the multicent-2 and Mauriceville strains revealed restriction fragment length polymorphisms at all 14 telomeres.

The *N. crassa* genome contains very little repetitive DNA due to the repeat induced point (RIP) mutation process, which acts during the sexual cycle and cumulatively obliterates duplicated sequences (Cambareri *et al.*, 1991; Galagan and Selker, 2004). Therefore, it seemed unlikely that *N. crassa* would possess subtelomeric gene duplications. However, there was also the possibility that subtelomeres might persist in RIP'ed forms, as has been observed in the related fungus, *M. oryzae* (Ikeda *et al.*, 2002; Thon *et al.*, 2004). The genome sequence of *N. crassa* promised to provide valuable new insights into the organization and gene content of chromosome ends. BLAST searches of the genome assembly revealed only two telomeres out of the expected 14. Therefore, the telomere-mining program TERMINUS was used to identify new telomeres among the raw, unassembled sequence reads, and to link these new ends to the genome sequence (Li *et al.*, 2005). Cosmid clones containing the newly-identified telomeres were then sequenced to close gaps. Through these efforts, we were able to identify genome sequence contigs that are close to telomeres, which in turn provided insight into the organization and gene content of these important chromosome regions. Finally, to determine the molecular basis for the high levels of polymorphism at *N. crassa* chromosome ends, we cloned and characterized several telomeres from the Mauriceville strain, and compared them to their homologous counterparts in Oak Ridge.

Materials and Methods

Fungal Strains

N. crassa wild-type strains used were Oak Ridge 74-OR23-IVA (FGSC 2489), obtained from Dr. D. Perkins, Stanford University, and Mauriceville 1c-A (FGSC 2225), obtained from the Fungal Genetics Stock Center.

Preparation of ^{32}P-Labeled Telomere Probes

Concatemers of the telomeric repeat sequence $(TTAGGG)_n$ were obtained by using the primers TTAGGG and CCCTAA (Table 1) to perform PCR in the absence of template. Reactions were set up as follows: 1× buffer (Takara); 12.5 pmol of TTAGGG primer; 12.5 pmol of CCCTAA primer; 400 μM of dNTPs; 1.25 U of LA Taq polymerase (Takara). The amplification conditions were: 94°C, 5 min; followed by 35 cycles of 94°C, 30 s; 55°C, 30 s and 72°C, 1 min. Finally, a 5 min extension was performed at 72°C. PCR products were resolved by electrophoresis in a 1% agarose gel and the DNA products that were 1.5 to 2 kb in size were excised and purified using a gel extraction kit from Qiagen. The purified concatemers were then used as template in a modified primer-directed DNA synthesis reaction (Feinberg and Vogelstein, 1983) to generate ^{32}P-labeled telomere probe using an equivalent concentration of oligonucleotide $(TAACCC)_3$ instead of random hexanucleotides.

Construction of Cosmid Libraries for Telomere Enrichment

A minimal cosmid vector was constructed by digesting 100 ng of pMLF2 (An *et al.*, 1998) with *Eco*RI and *Xba*I, filling in with Klenow polymerase, followed by ligation in a large volume (100 μL) to promote recircularization. The resulting deletion derivative was then digested with *Bam*HI and ligated to an adaptor created by annealing together the oligonucleotide primers, pMLF/*Eco*RV_1 and pMLF/*Eco*RV_2 (Table 1). The ligation mix was then transformed into *E. coli* DH5 competent cells. The resulting vector is 5.5 kb in size and was named pMLF4.

The pMLF4 vector was linearized with *Eco*RV, and then treated with shrimp

alkaline phosphatase following the manufacturer's protocol (Promega). Prior to cloning, the genomic DNA was end repaired with the End-It kit from Epicentre. Polished chromosomal DNA (500 ng) and linearized vector (100 ng) were ligated with T4 DNA ligase for overnight at 16°C. Half of the ligation mixture (5 µL) was then packaged *in vitro* with lambda phage packaging extract from Epicentre, following the manufacturer's instructions. Aliquots of the packaging extract then were transfected into *E. coli* XL-10 cells, using the protocol provided in the packaging kit. Cells were plated on LB agar plus 100 µg/mL of ampicillin at a density of 400 to 500 transformants per plate. Recombinant colonies were replica-plated onto sterile Whatman 541 filter paper discs and the agar plates were kept at 25°C overnight to allow the colonies to re-grow. They were then stored at 4°C until positive clones were identified. Filter discs with colony replica facing up were placed on 0.5 M NaOH-soaked filter paper for 5 min, and then transferred onto 1 M Tris·HCl, pH 7.5-soaked filter paper for 5 min. Discs were then washed with 2× SSC followed by 95% ethanol for 5 min, respectively. After discs were completely dry, they were incubated at 65°C for 1 h in 5× SSC/0.5% SDS. On-filter screening was then performed by hybridizing to ^{32}P-labeled telomere probe as described (Farman and Leong, 1995). Filter discs were dried and exposed to Kodak X-Omat films for ~24 h to locate the positions of positive colonies. Positive colonies were re streaked on fresh LB amp plates and re-probed to obtain pure clones, which were then cultured in 100 mL of LB plus 100 µg/mL of ampicillin. Cosmid DNA was extracted using Wizard midi-prep kit.

Creation of Telomeric Contigs and Paired Reads

Sequencing primers MLF4-T3 and MLF4-T7 (Table 1) were used to sequence cosmid clones from the opposite ends of the inserts to create paired reads. The telomeric sequences and the paired reads were then assembled into telomeric contigs (TelContigs) and subtelomeric contigs, respectively, creating a set of telomeric scaffolds. Each scaffold represented a separate chromosome end. The scaffolds were then linked to the genome assembly using BLASTn searches. Primer walking was performed to close the gaps between the genome contig sequence and the telomere ends.

Sequencing and Assembly of Telomeric Cosmids

Cosmid DNA was prepared by alkaline lysis and sheared to ~2 kb fragments using a Hydroshear machine fitted with the standard sized shearing assembly (Genemachines). The fragments were end repaired using the End-It kit, ligated to the pHCamp vector (Lucigen Technologies), and electroporated into the *E. coli* EPI300 electrocompetent cells. Recombinant DNAs were prepared by alkaline lysis and sequenced with Big Dye V3 chemistry (Applied Biosystems), using the SL1 and SR2 primers provided with the vector. Sequences were assembled using the Phred/Phrap software packages (Ewing and Green, 1998; Ewing *et al.*, 1998) and manually checked using Consed (Gordon *et al.*, 1998) (http://www.phrap.org/phredphrapconsed.html). Small gaps (≤ 5 kb) between the telomeres and the genome assembly were filled by primer walking, using cosmids as templates.

Southern Hybridization Analysis

Genomic DNA isolated from *N. crassa* and cosmid DNA samples were digested with *Eco*RV/*Not*I or *Hin*dIII/*Not*I and fragments were separated by electrophoresis in a 1% agarose gel. After electrophoresis, DNA samples in the gel were first depurinated for 10 min in 200 mL of 0.25 M HCl and then denatured for 1 h in 200 mL of 0.5 M NaOH/1.5 M NaCl. The gel was finally neutralized for 1 h in 200 mL of 0.5 M Tris·HCl, pH 7.5/1.5 M NaCl. DNA samples were then transferred onto Zeta-Probe membranes from Bio-Rad by capillary transfers, and covalently linked using a UV crosslinker. Hybridization to ^{32}P-labeled telomere probe was performed as described (Farman and Leong, 1995). Membranes were dried and exposed to screens of a Molecular Dynamics PhosphorImager for ~24 h.

Generation of Probes for Telomere Adjacent Sequences

Telomere adjacent sequences were PCR amplified from Oak Ridge genomic DNA, using the primers listed in Table 1. The amplification reaction contained 80 ng of genomic DNA, 12.5 pmol of each primer, 1× buffer (Takara), 400 μM of dNTPs and 1.25

U of LA Taq polymerase (Takara). Cycling conditions were as follows: 94°C for 1 min, followed by 35 cycles of 94°C, 30 s; 55°C, 30 s; 72°C, 1 min. The final extension was at 72°C for 5 min. Probes were purified by agarose gel electrophoresis and QiaQuick extraction. All probes were sequenced to check their veracity.

RFLP Mapping

Mapping was performed using a set of 38 segregants from ordered asci derived from a cross between the Oak Ridge laboratory strain and the Mauriceville field strain. This set was established for RFLP mapping (Metzenberg and Grotelueschen, 1995; Metzenberg *et al.*, 1984) and is distributed by the Fungal Genetics Stock Center. Genomic DNA was isolated from 7-day old cultures (Freitag *et al.*, 2002). 500 ng of DNA from each progeny were digested overnight with 0.5 units of restriction endonuclease, followed by electrophoresis through 0.8% agarose gels. DNA samples were transferred to nylon membranes (Pall Corp) and probed with radioactive probes prepared by oligolabeling (Feinberg and Vogelstein, 1983). The hybridization and washing conditions were as described previously (Kouzminova and Selker, 2001).

Sliding Window Analysis of GC Content and RIP Indices

To determine how the nucleotide compositions of the terminal sequences change as a function of the position relative to the telomere, the values of interest were calculated in a "window" of 200 nucleotides. The window was slid in 20-bp increments (in a centromere to telomere direction) and values were recalculated for each position. This was reiterated until the right hand edge of the window met the telomeric end of the sequence. GC content was measured as the percent of G or C nucleotides in each window. RIP index I was calculated as ApT/TpA, and RIP index II was (CpA+TpG)/(ApC+GpT).

Terminal Gene Analysis

We identified genes that were contained within 30 kb of DNA sequence proximal to each telomere by inspecting gene predictions from the *Neurospora crassa* genome

database (Assembly 7, version 3) at the Broad Institute (http://www.broad.mit.edu/annotation/genome/neurospora). Most of the predicted genes were categorized as either predicted or hypothetical and therefore no known function was associated with them. To gain insight into the potential role these genes may play in *N. crassa* biology, we BLASTed the predicted protein sequences against the SwissProt database at NCBI using the default parameters and simultaneously queried the conserved domain database (http://www.ncbi.nlm.nih.gov/BLAST). We also BLASTed the Gene Ontology (GO) database (GOst; http://www.godatabase.org/cgi-bin/gost/gost.cgi). This database contains genes that have been annotated with GO terms by curators from several model organism databases (e.g., *Saccharomyces* Genome Database, Wormbase, and Flybase). We used the most stringent threshold available (0.001) and then manually inspected the results for significance using the same thresholds of the SwissProt BLAST. GO terms that had "inferred from structural similarity" (ISS) as their evidence code were excluded from consideration unless there was additional experimental data to support the assignment.

Results

N. crassa Telomeres

N. crassa has seven chromosomes (Perkins *et al.*, 2001). However, Southern hybridization analysis of genomic DNA from Oak Ridge 74-OR23-IVA (OR) digested with various restriction enzymes produced at least 16 visible fragments (Figure 5). The majority of hybridizing fragments were of a similar intensity, except for two which were noticeably fainter. This suggested that the Oak Ridge culture used in this project contained some nuclei with variant telomeres.

Screening of the Version 3 genome assembly revealed only four telomere sequences. Release of the version 6 and 7 assemblies allowed the identification of five additional telomeres, although one telomere that was present in the version 3 assembly was omitted from the latter sequences. Therefore, analysis of the genome sequence identified a total of nine telomeres, of which one, by virtue of its omission from later

assemblies, was questionable.

To determine if additional telomeres might have been captured in the genome sequencing project but had escaped assembly, we used TERMINUS to identify and assemble telomere containing sequences identified among the raw sequence reads. This produced 14 contigs (TelContigs) that started with the telomere sequence $(CCCTAA)_n$. TelContigs TC6 to TC14 were very robust as they were derived from large numbers of independent sequences, while TelContigs TC1 to TC5 were less certain because they were represented by far fewer reads (Table 2). As shown in Table 2, the reads making up TCs 1, 6, 8, 11, 12, 13 and 14 had already been incorporated into the *Neurospora* genome sequence. Using mate pair sequence information, TERMINUS allowed additional linkages to be established for TCs 5, 7, 9 and 10. Thus, analysis of the genome sequence data alone provided reliable telomere-to-genome linkage information for 11 of the 14 TelContigs.

TC2 and TC3 did exhibit a single match to the genome assembly but the respective alignments exhibited numerous G to A and C to T transition mutations, characteristic of RIP mutations (Cambareri *et al.*, 1989). Therefore, it is clear that there were not true links to the genome sequence. BLAST searches using the SubTelContigs supported the linkage between TC4 and genomic contig 7.141. However, the alignments obtained for the SubTelContigs corresponding to TC2 and TC3 all exhibited numerous RIP-like transitions, indicating that the genome assembly does not contain the specific copies of these subtelomeric sequences.

Identification of Plasmid and Cosmid Clones Containing *Neurospora* Chromosome Ends

Although TERMINUS identified seven telomeres that were not present in the genome assembly and established physical linkages for four of them, there still remained a number of sequence gaps between these telomeres and the genome sequence. Therefore, in order to close these gaps, and to try and link up TC2 and TC3, we used a ^{32}P-labeled $(TTAGGG)_n$ probe to identify telomere containing clones in end enriched plasmid libraries, as well a blunt ended cosmid library of Oak Ridge 74-OR23-IVA genomic DNA. Telomere containing plasmid clones were subjected to end sequencing

and the resulting reads were assembled.

This resulted in a total of seventeen different sequences, of which nine precisely matched TelContigs that had been identified by TERMINUS. Two of the sequences matched TC14 but the starting positions of the telomere repeats were different, suggesting that these clones contained truncated version of *TEL_VL* (Figure 3). This left six sequences that lacked matches to the TERMINUS TelContigs. Similarly, BLAST searches of the whole genome failed to detect corresponding sequences and, therefore, we were unable to ascertain the origins of these telomeres.

Cosmid clones were grouped based on their *Not*I+*Eco*RV restriction patterns, and end sequences were determined for one clone representative of each group. This revealed that nine different telomeres had been captured in cosmid clones (Table 2). Interestingly, the telomeres that were captured in cosmid clones were precisely the same as the ones that were present in the version 7 genome assembly. Consequently, the plasmid and cosmid clones failed to provide any additional linkage information.

Nevertheless, the targeted cloning of chromosome ends in plasmids and cosmids providing independent confirmation of veracity of thirteen of the fourteen telomeres identified by TERMINUS. TC3 was the only TelContig for which we were unable to provide independent support of its validity. However, TC3 is the most likely candidate for the elusive 14th telomere because it is derived from five independent sequence reads. By contrast the six “cryptic” telomeres identified among the plasmid clones are all based on singleton sequences.

Despite the fact that the plasmid/cosmid clones had failed to identify new telomere-genome links, four gaps between TelContigs and the assembly had been captured in cosmid clones. Therefore, we used primer walking on cosmid templates to complete the sequences of four chromosome ends, resulting in the closure of gaps ranging in size from 1.2 kb to 4.7 kb (Table 2).

RFLP Mapping of Telomere Sequences to Linkage Groups

To confirm the linkage assignments suggested by TERMINUS, and to determine the chromosomal locations of the unlinked TelContigs/genome assembly contigs, we used telomere adjacent sequences as probes to identify restriction fragment length

polymorphisms (RFLPs), and then monitored the segregation of those RFLPs among the *N. crassa* mapping population. The telomere-adjacent probe from TC3 hybridized to a single restriction fragment in each of the six restriction digests of Oak Ridge DNA but did not to hybridize to Mauriceville DNA. This polymorphism segregated 22 present to 16 absent and exhibited complete linkage to markers, nuo21.3c, DA122, and the previously mapped *TEL_VIR* (Figure 2).

The TC4 probe also hybridized exclusively to the Oak Ridge DNA and identified a single restriction fragment in each of the *Dra*I, *Eco*RV and *Hin*dIII digests. The marker exhibited a segregation ratio of 20 present to 18 absent and was tightly linked to marker NP4A9, which maps in the region that would correspond to *TEL_VIIR* (Figure 2).

The probe from the telomere adjacent region of TC2 exhibited weak hybridization to multiple loci in the genomes of both the Oak Ridge and Mauriceville strains, resulting in a background smear in each lane. However, three digests (*Bam*HI, *Eco*RV and *Hin*dIII) yielded distinct bands in Oak Ridge DNA and these were missing in Mauriceville. Segregation analysis of the strongly hybridizing fragment using progeny DNA samples digested with *Eco*RV revealed that 29 progeny possessed the fragment, while nine lacked it. This approximates to a 3:1 ratio, which suggested that the hybridization signal actually comes from two unlinked loci. Therefore, we analyzed TC2 segregation using a *Bam*HI digest which produced two hybridization signals. One of the *Bam*HI fragments exhibited perfect cosegregation with markers on the right arm of chromosome I, indicating that TC2 corresponds to *TEL_IR* (Figure 2).

Nucleotide Composition at the *N. crassa* Chromosome Termini

Visual inspection of the telomere sequences revealed a superabundance of A and T nucleotides in the regions adjacent to the TTAGGG repeats. A sliding window analysis of GC content revealed that eight of the nine chromosome ends analyzed had a region with >75% AT within the terminal 1 kb (Figure 4). In most cases, the AT-rich region was restricted to the chromosome tip, as there was only one example of an AT-rich sequence in the subterminal regions (*TEL_IIL*; Figure 4). As a further test of the significance of the telomeric AT-richness, we analyzed the GC content for groups of nine 20-kb sequences sampled randomly from the *N. crassa* genome. Based on 1000

samplings, the average number of 20-kb sequences that contained a region with a GC content lower than 25% was 1.1 per group of nine sequences. Furthermore, none of the 1000 groups that were analyzed contained more than five sequences with 25% or lower GC. Therefore, we can conclude that AT-rich sequences are significantly over represented in the telomere adjacent regions ($P >> 0.001$).

In *N. crassa*, the presence of AT-rich sequences is often indicative of the action of repeat induced point mutation, a process which causes G to A and C to T transition mutations in repeated sequences. To determine if the AT-richness at the chromosome termini is likely to have been caused by RIP, we used a sliding window procedure to calculate the RIP indices [I: TpA/ApT and II: (CpA+TpG)/(ApC+GpT)] across the terminal regions. Regions with a RIP index I > 1, and a RIP index II < 1, are likely to have been mutated by RIP (Margolin *et al.*, 1998). As shown in Figure 4, the AT-rich sequences near to the telomeres all exhibited a RIP index I that was greater than 1 and a RIP index II less than 1. There were also numerous internal sequences that had high RIP I and low RIP II indices. However, nearly all of the internal RIP positive regions spanned far fewer than 400 nucleotides. The only exception was the previously mentioned AT-rich sequence that lies centromere proximal to *TEL_IIL*.

Repeated Sequences in the Terminal Regions

To determine if the terminal AT-richness and RIP-type mutations are due to the presence of repeated sequences, we identified within the terminal regions all sequences (> 100 bp) with two or more copies in the *Neurospora* genome assembly. Twenty different repeats were identified, and their copy numbers ranged from two to 498. Only two showed similarity to transposable elements, one being the Pogo element previously identified at *TEL_VR* (Schechtman, 1987), and the other bore similarity to retrotransposons from *Cryptococcus* and *Magnaporthe* (Farman *et al.*, 1996; Goodwin and Poulter, 2001). The positions of these repeats are shown as gray boxes in Figure 4. Surprisingly, there was a poor correspondence between the locations of AT-rich/RIP-positive DNA and the positions of repeats. In most cases, the AT-rich/RIP DNA was either not associated with a repeat at all, or it extended well beyond the repeat's boundary. Two good examples of this are near *TEL_IIL*, and *TEL_IIR*, where there are

large tracts of AT-rich/RIP-positive sequences that are single copy (Figure 4). Interestingly, the internal AT-rich/RIP-positive sequence at *TEL_IIL* did coincide perfectly with repeated sequences (Figure 4).

Terminally Located Genes

We identified predicted genes 30 kb upstream of the telomeres using the *Neurospora* genome database. These genes were evaluated for potential biological processes using BlastP queries of the SwissProt database along with a simultaneous conserved domains search and by using GOst, a web-based program that blasts proteins against a database containing genes previously annotated using Gene Ontology.

Fungal subtelomeric regions commonly harbor genes related to secondary metabolism or detoxification and *N. crassa* does not appear to be an exception. Our analyses identified a cluster of genes in contig 7.79 (TC8) that may be involved in secondary metabolism. The strongest evidence comes from the conserved domains identified. Two genes (NCU10285.3 and NCU09636.3) contain conserved domains for cytochrome P450, which is involved in secondary metabolite biosynthesis. Furthermore, other domains that are associated with secondary metabolism were identified, including FAD/FMN-contain dehydrogenesase (NCU09635.3), O-methytransferase (NCU09637.3), polyketide synthase (NCU09638.3), and fungal tricholthecene efflux pump (NCU9640.3). Results from BLAST queries show that NCU09638.3 is most closely related to the *Aspergillus terreus* lovastatin nonaketide synthase. According to Gene Ontology results obtained from GOst, the lovastatin nonaketide synthase gene from *A. terreus* has been experimentally shown to have polyketide synthase activity. NCU09640.3 is related to a putative HC-toxin efflux carrier (TOXA) from *Cochliobolus carbonum*. GOst data suggests that TOXA from *C. carbonum* has toxin transport activity. Taken together, conserved domains, BLAST and GOst results indicate that this gene cluster is potentially involved in secondary metabolism.

Molecular Basis for Strain-to-strain Variation in Telomere Structure

Neurospora telomeres are highly polymorphic. For example, the Oak Ridge

(OR) and Mauriceville (MV) strains have completely different telomeric restriction fragment profiles (Figure 5), and this allowed RFLP mapping of twelve of the fourteen telomeres using a single restriction enzyme (Schechtman, 1989). By comparison, internal loci tend to exhibit much less polymorphism, so that multiple restriction enzymes are often necessary to identify RFLPs. To determine the basis for this extreme telomeric variation, we cloned telomeres from the MV strain, so that we could compare the sequences of homologous chromosome ends. Saturation screening of a cosmid library of MV DNA using a telomere probe resulted in the recovery of 115 clones. Restriction analysis with *Eco*RV+*Not*I allowed these clones to be partitioned into 10 distinct groups. Eight of the groups were very robust, comprising multiple clones with almost identical restriction profiles. There were also two singleton clones whose restriction profiles were either unique, or exhibited partial matches with other cosmids, indicative of the presence of chimeric inserts. Representative clones from seven of the robust groups were subjected to end sequencing.

As expected, all of the cosmid clones contained TTAGGG repeats at one end of the insert. In order to match each of the MV chromosome ends to its homologous counterpart in OR, the end sequences were used to search the V.7 genome assembly using BLAST. The repeat filtering option was selected, so that matches involving only telomere repeats were not reported. Cosmid MV752 lacked matches at its telomeric end, while four (MV750, 756, MV789 and MV791) exhibited matches but the alignments started near to the ends of the sequence read, suggesting that the sequences immediately adjacent to the telomere repeat are not present in the OR genome. The sixth MV clone, MV762, had a repetitive element adjacent to the telomere repeat; however, its centromere proximal end matched sequences near *TEL_IIIR*. The sequences derived from the subtelomeric ends of the inserts all exhibited unique matches to the genome assembly. For four of the cosmids, the subtelomeric end matched the same contig as the telomeric end. The exceptions were MV762, whose subtelomeric end was homologous to contig 7.17 (a known telomeric contig); and MV750, whose subtelomeric and telomeric ends matched 7.78 and 7.251 respectively. This arrangement is consistent with the results from TERMINUS (Li *et al.*, 2005), which also placed 7.251 distal to 7.78 in Oak Ridge. Thus, for each of the cloned MV telomeres, we were successful in identifying the

homologous end in OR.

Alignments of homologous chromosome ends are shown in Figure 6. The figure shows that, for all homologous pairs, the sequences found adjacent to the telomere in OR are replaced by different sequences in MV. For example, an ~8.9 kb region next to *TEL_VIL* is replaced by an ~700 bp sequence in the MV homologue (Figure 6, top alignment). Although the lengths of the missing sequences and their replacements vary among the homologous telomeres, BLAST searches indicated that all but one of the telomere adjacent sequences in MV are not present in the OR genome. The only exception was MV762 which contained the high copy repeat. Interestingly, this repeat was not present at the homologous OR telomere (data not shown).

As was the case in OR, four of the seven chromosome ends characterized from MV transitioned into highly AT-rich DNA in the telomere adjacent regions (Figure 7). This led us to question whether the sequence divergence might be the result of recurrent RIP cycles having acted on a once shared sequence. This possibility was addressed by using Clustalw to search for weak sequence similarities beyond the presumed divergence points. This revealed only one case — involving *TEL_VIL* — where the homologues could be aligned beyond the initial divergence point detected by BLAST. Interestingly, the new alignments did reveal a predominance of the transition mutations that are characteristic of RIP, as well as two large insertion/deletions (indels) (shown in Figure 6). Nevertheless, despite the discovery of extended homology and evidence for RIP, there were still ~280 bp of sequence adjacent to *MV_TEL_VIL* that were not present in the OR homologue (Figure 7). For the remaining pairs of homologous chromosome ends, we were unable to detect additional stretches of sequence similarity in the telomere adjacent regions, indicating that the telomere adjacent sequences in MV are completely unrelated to their counterparts in OR.

Calculation of RIP indices for the MV chromosome ends revealed a similar pattern to what was observed in OR. Specifically, the telomere adjacent sequences showed hallmarks of having undergone RIP, but in most cases, the sequences involved were not repetitive (Figure 7). We also examined where the RIP positive sequences started relative to the points of homologue divergence. Interestingly, these positions tended to be quite close to one another (< 1 kb), although there was rarely a perfect

correspondence. For example, in *OR_TEL_IIL*, *OR_TEL_VIL*, *MV_TEL_IIL*, *MV_TEL_VIL* and *MV_TEL_VIIL*, the RIP positive regions start before the points of sequence divergence, and in *MV_TEL_IIR* it starts afterward (Figure 7).

Genome-specific Sequences in the Telomere-adjacent Regions

As noted above, analysis of the sequences that were immediately adjacent to the Mauriceville telomeres revealed no matches to the *Neurospora* genome assembly, which was derived from the Oak Ridge strain. Therefore, we asked whether the reciprocal situation is true. Probes were derived from sequences adjacent to each of the fourteen Oak Ridge telomeres and were used to probe Southern blots of both Mauriceville and Oak Ridge DNA. As summarized in Table 3, of the telomere-adjacent sequences from Oak Ridge failed to hybridize to Mauriceville DNA. Significantly, all of the exceptional cases involved probes that hybridized to more than one locus in Oak Ridge.

Discussion

Here, we present a comprehensive analysis of the organization of chromosome ends in the model filamentous fungus *N. crassa*. Through a combination of bioinformatic analysis, targeted telomere cloning and RFLP mapping, we determined the sequences of all fourteen of the *N. crassa* telomeres, and assigned them to their respective chromosome ends. For four telomeres, we were able to contribute to the genome finishing effort by closing gaps between telomeric contigs and the neighboring sequences in the genome assembly. Unfortunately, we were unable to close sequence gaps at five chromosome ends. In this regard, it is perhaps significant that the chromosome termini that were captured in our deep plasmid and cosmid libraries corresponded to ones that had already been cloned as part of the genome sequencing project. This strongly suggests that the remaining gaps consist of sequences that are uncloneable in plasmid, cosmid and fosmid vectors. Based on our characterization of the available terminal sequence data, it would appear that these regions could be highly AT-rich and, therefore, retractable to cloning in standard cloning vehicles.

Telomere Structure in *N. crassa*

Analysis of the chromosome termini for which complete sequence information is now available revealed that the organization of the *Neurospora* chromosome ends is quite different to those of other fungi. *S. cerevisiae* (Louis and Haber, 1990), *Ustilago maydis* (Guzman and Sanchez, 1994), *P. carinii* (Keely *et al.*, 2005), *Kluyveromyces lactis* (Fairhead and Dujon, 2006), *M. oryzae* (Farman and Kim, 2005; Gao *et al.*, 2002; Rehmeyer *et al.*, 2006), *A. nidulans* (Clutterbuck and Farman, 2007) *Nectria haematococca* (Farman, unpublished data) and *Cercospora zeae maydis* (Dunkle and Farman, unpublished data) all possess distinct subtelomere regions consisting of sequences that are duplicated at several chromosome ends. By contrast, none of the fully assembled *Neurospora* chromosome ends have any similarity to one another for at least 20 kb in from the telomere repeat. Nor were there any matches to the smaller, unassembled, telomere-linked contigs. Therefore, it appears that *N. crassa* lacks a distinct subtelomere domain — or at least one that is defined by specific sequences.

Consistent with the absence of subtelomeres, *N. crassa* lacks the telomere-linked helicase (*TLH*) genes that are present in the subtelomere regions of diverse fungi (Gao *et al.*, 2002; Inglis *et al.*, 2005; Mandell *et al.*, 2004; Mandell *et al.*, 2005; Sanchez-Alonso and Guzman, 1998). *TEL_VL* is attached to ribosomal RNA gene sequences. This particular telomere appears to be prone to terminal truncations, as a truncated version of this chromosome end was identified in the genome sequence data, and two additional truncations were identified in our plasmid libraries. Several instances of telomeric rDNA instability were also observed in *M. oryzae* (Rehmeyer, Starnes, Zhou and Farman, unpublished data). These findings suggest that some feature(s) of the rDNA array compromises the protective functions of the telomere.

In addition, to novel "rDNA" ends, we also identified three plasmid clones that contained telomere repeats linked to sequences with no matches to the genome. It is possible that these are also truncated telomeres, but they were resected far enough back that the remaining sequences correspond to regions that remain as gaps in the assembly.

AT-rich Sequences in the Telomere-adjacent Regions and Their Relationship to RIP

One of the most striking findings of this study is that the *N. crassa* telomere-adjacent regions are almost universally comprised of highly AT-rich DNA. The sliding window analyses indicate that the majority of sequence within 20 kb of the chromosome ends has a GC content of >50%, but this value drops significantly in the terminal 1 – 2 kb. The only exceptions to this rule are *OR_TEL_IIL*, *OR_TEL_IIR* and *OR_TEL_IIIL*, where the GC content drops well before this distance from the telomere. AT-richness is commonly seen in the telomere-adjacent regions of other fungi, including *K. lactis* (Nickels and McEachern, 2004), *P. carinii* (Keely *et al.*, 2005) and *M. oryzae* (Rehmeyer *et al.*, 2006). However, in these organisms, the AT-rich DNA is part of a distinct subtelomere sequence that is duplicated at multiple chromosome ends. By contrast, all of the AT-rich sequences that adjoin the *N. crassa* telomeres are quite distinct from one another and, therefore, appear to have arisen independently. One possibility is that the AT-richness has been selected because it is important for telomere function and/or maintenance.

Another possible reason for AT-rich sequences at the *N. crassa* chromosome termini is that these regions might be unusually prone to RIP. The RIP indices for these regions were certainly consistent with their having experienced the ravages of this process. However, there were very few repetitive sequences in these regions, which is surprising because there is generally a good correspondence between AT-rich/RIP-positive DNA and the presence of repeated sequences across the rest of the *Neurospora* genome (Galagan *et al.*, 2003). One possibility is that the *N. crassa* chromosome ends might have once had organizations resembling those of other organisms, with abundant repetitive elements and gene duplications, and that the repeats are no longer recognizable, either because copies have undergone unusually intensive RIP, or they have been deleted from the genome altogether. Alternatively, perhaps RIP operates on chromosome termini in a repeat-independent manner.

Novel Sequences at the Telomeres of a Second *N. crassa* Strain

Several years ago, it was documented that there are numerous telomeric RFLPs between the OR and MV strains (Schechtman, 1989). Here, we show that this polymorphism is the result of abrupt divergence in the terminal regions, with the homologous chromosomes having completely different sequences beyond these transition points. In four out of five such instances, the OR chromosome contains a repeated sequence at the point of divergence. The repeat at the divergence point in *TEL_IIL* is a retroelement and, therefore, the most economical explanation for the difference in terminal sequence composition is that a retrotransposon inserted into the OR homologue, and an ectopic recombination event at some time afterward caused the translocation of a new sequence into the telomere-adjacent position. The repeats at *OR_TEL_IIR* and *OR_TEL_VIIL* are also highly repetitive (44 and 19 copies, respectively) which suggests that they might too be transposons, and that a similar mechanism might have led to the divergence of these chromosome ends. By contrast, *TEL_IIIL* and *TEL_VIL* do not have repeats at the points where the sequences of the homologous chromosome ends diverge, although there is still a marked increase in AT-richness and evidence of RIP having mutated in the novel sequences beyond. It is not clear what is the basis for sequence divergence in these homologous chromosome pairs. However, as noted above, it is possible that these regions contain sequences that were once highly repeated but have experienced RIP to the extent that they are no longer recognizable as such.

Table 1 Oligonucleotide primers used in this study.

Oligonucleotide ID	Sequence (5′ to 3′)
pMLF/EcoRV_1	GATCCCGGGATATCT
pMLF/EcoRV_2	GATCAGATATCCCGG
pMLF4_T3	CGCAATTAACCCTCACTAAAGG
pMLF4_T7	ATAATACGACTCACTATAGG
TTAGGG	TTAGGGTTAGGGTTAGGG
CCCTAA	CCCTAACCCTAACCC
CWO45	GGAAACGTACTTAAAAGGG
CWO46	TAAACGGGAATGATATACTA
CWO49	AAGCCTTCTAGAAGTTAACC
CWO50	CTTCTTCATTTTGTAATTCTAC
CWO51	CCGGGTGGATCTTTATAC
CWO52	AGAACTTTAATAGTGCGTTT
CWO57	GGATTTCTTTTATAATAAGGC
CWO58	GTCTAAGCGGTTAAAGTTAG
CWO59	AATCCTCAATAATCGCCG
CWO60	CAGCCCCTTAAAGTTAGAGG
CWO61	AGAGATGCTTATACTTAGGG
CWO62	GGGGTCGGTAGTATAATTAA
CWO65	CGTACCCGTACCTAAACTAG
CWO66	AAGAGCGTAGCGTCCTAG
CWO67	AAGGTCCTCCTTTAATTCTC
CWO68	GGGGTAGGGACATAGAG
CWO69	TTCCTTCCTAATTACCTTCC
CWO70	AGGGAGGGGGTAATAAATAT
CWO75	CCCTAAAGTATAGGGTATTGCAGAG
CWO76	CAATACTGGCGAAGCTTGTG
CWO77	CTTGCCCTAGCTCTAGC
CWO78	CTCAGGGTTCTAGTAGTTAC

Oligonucleotide ID	Sequence (5′ to 3′)
CWO79	CTCTGGCTCTAGTTCTAGCTC
CWO80	GCTCTTTTTCCCTTTATAAGG
CWO83	CCTGCTCTGGATTCTTTTCT
CWO84	TTTTCCTTCTGCTTCCTCTC
CWO85	CTCGGATCAGGTAGGAATACCC
CWO86	GAGGGTATTACGCCTAGGGC

Table 2 Physical and genetic mapping of TelContigs.

TC	Linked to assembly contig[a]	TERMINUS linkage	Plasmid clone(s)[b]	Cosmid clone(s)[b]	RFLP mapping	Map position inferred	Gap size (kb)
1	7.247	7.247, 7.246	CWOR010	–	IVR		
2	–	–	CWOR012	–	IR		
3	–	–	YSK_A3	–	VIR		
4	7.141	7.141		–	VIIR		
5	–	7.74	JMOR004	OR708	IIIL		4.69
6	–	7.77	CWOR009	OR711	IIR		
7	–	7.17	YSK_B8	OR713	IIIR		1.74
8	7.79	7.79	YSK_A10	OR712		IVL	
9	–	7.81	OR078	OR709	IIL		1.19
10	–	7.37		OR703		VR	1.96
11	7.34	7.34	YSK_A9	OR706	VIL		
12	7.93	7.93	YSK_A2	OR726	IL		
13	7.251	7251	JMOR001	OR710	VIIL		
14	7.162	7.162	CWOR004	–	VL		

[a]Shows TelContigs (TCs) that were already incorporated into the genome assembly. A minus sign indicates that the TelContig (TC) was not assembled.

[b]Listed are the IDs of individual clones that represent each TC. A minus sign indicates that a corresponding clone was not identified.

Table 3 Cross hybridization between Oak Ridge telomere-adjacent sequences and Mauriceville DNA.

TelContig probe	Fragments in Oak Ridge	Fragments in Mauriceville	Allele missing from MV
TC1	1	0	yes
TC2	1 – 2[a]	0[a]	yes
TC3	1	0	yes
TC4	1	0	yes
TC5	1	0	yes
TC6	3 – 4	3[b]	yes
TC7	1	0	yes
TC8	2	0[a]	yes
TC9	>10[c]	>10[d]	yes
TC10	7	6[b]	yes
TC11	6+	5+	unclear
TC12	3+	1 – 2[b]	yes
TC13	6	6	unclear
TC14[e]	4	3	unclear

[a]A smear of background hybridization was observed.

[b]Hybridization in MV was to weakly-hybridizing secondary fragments.

[c]One fragment hybridized much more strongly than the others. This was probably the fragment from which the probe was derived.

[d]The most strongly hybridizing fragment was clearly not present in MV.

[e]rDNA telomere.

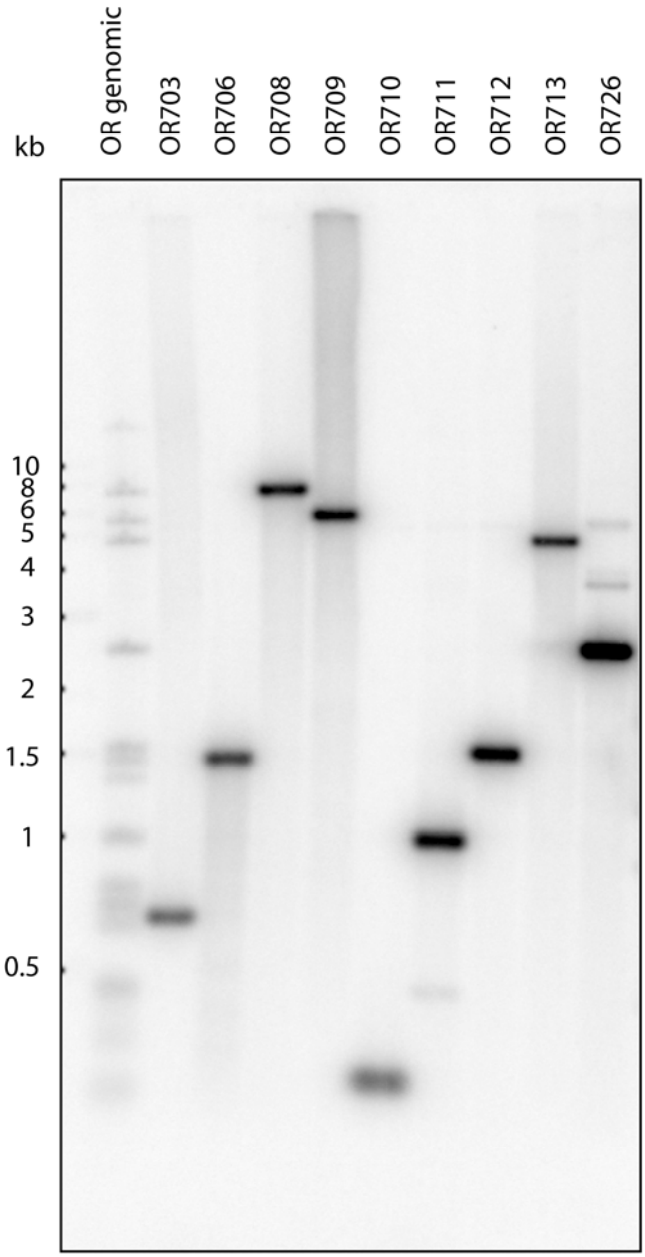

Figure 1 Southern hybridization analysis of representative telomeric cosmid clones. Genomic DNA and cosmid DNA samples were treated with *Not*I/*Hin*dIII in the reaction buffer supplied by the manufacturer. DNA fragments were separated by electrophoresis in a 1% agarose gel. A ^{32}P-labeled telomere probe was used to highlight restriction fragments containing telomere repeats. Size markers are indicated on the left-hand side.

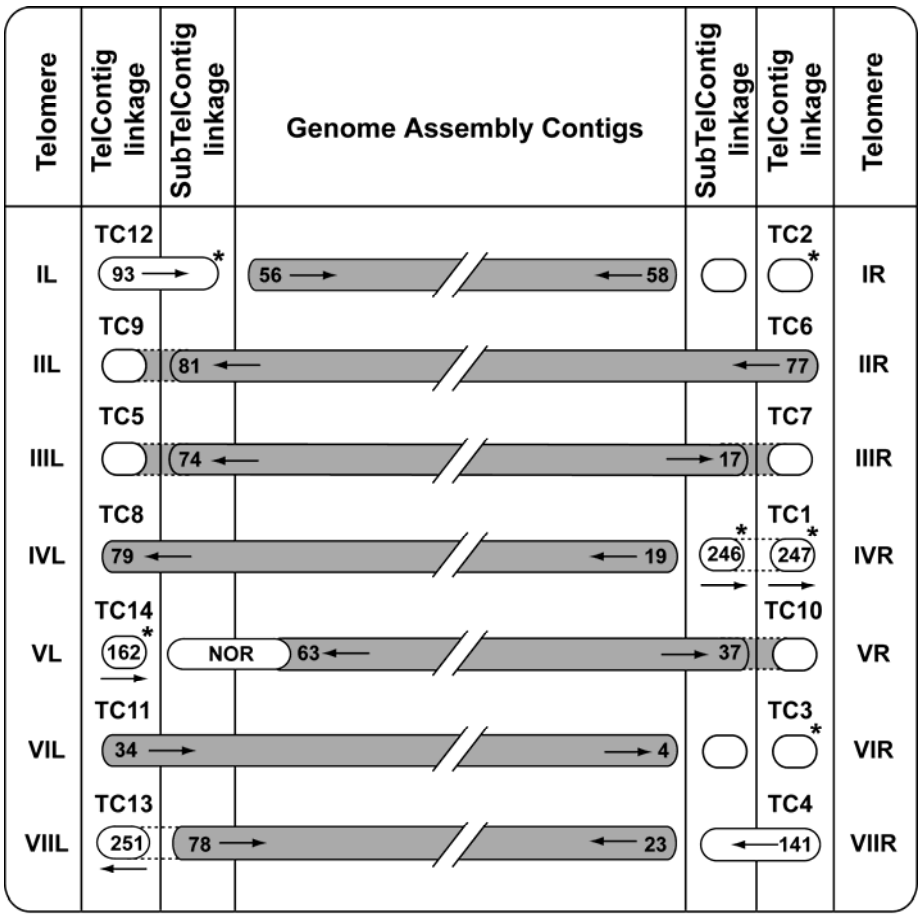

Figure 2 Status of the mapping of telomeres to the *N. crassa* genome assembly. Boxes that contain numbers represent genome assembly contigs and the arrows indicate sequence orientation. The shaded boxes show contigs whose chromosomal locations are known, while unshaded boxes represent contigs that had not been linked to the genetic or physical maps. Boxes without numbers depict TelContigs and SubTelContigs that lacked overlaps with the genome assembly. Genomic contigs that are shown extending into the TelContig and/or SubTelContig columns exhibited physical overlaps with the respective sequences. The nucleolus organizer region (NOR) spans several sequence contigs which are not listed due to space limitations. Physical linkages established by TERMINUS are represented by dotted lines, and shading between the lines indicates that the gap was closed by targeted sequencing. Contigs whose chromosomal locations were established or confirmed by RFLP mapping are highlighted with asterisks. The outer columns give the telomere identities.

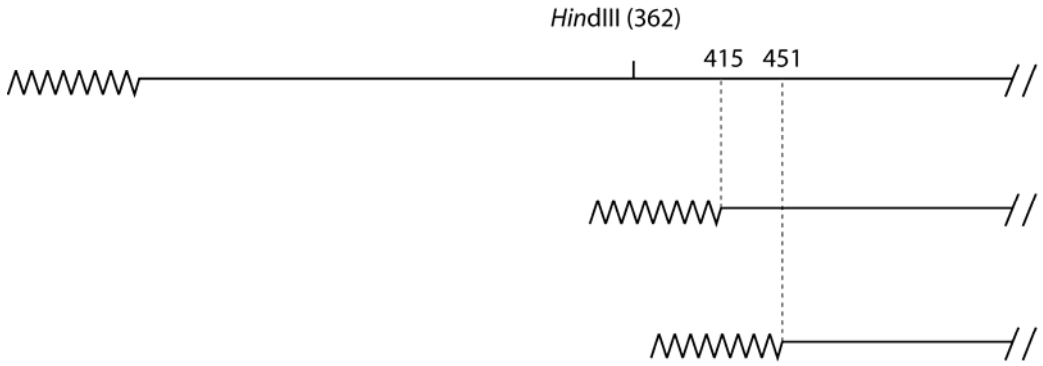

Figure 3 Truncations of *TEL_VL*.

The truncated versions of the rDNA telomere are shown aligned with the intact end. The zig zag line represents the telomere repeat. Dotted lines and the numbers in brackets show the locations of the newly-formed telomeres, relative to the original one.

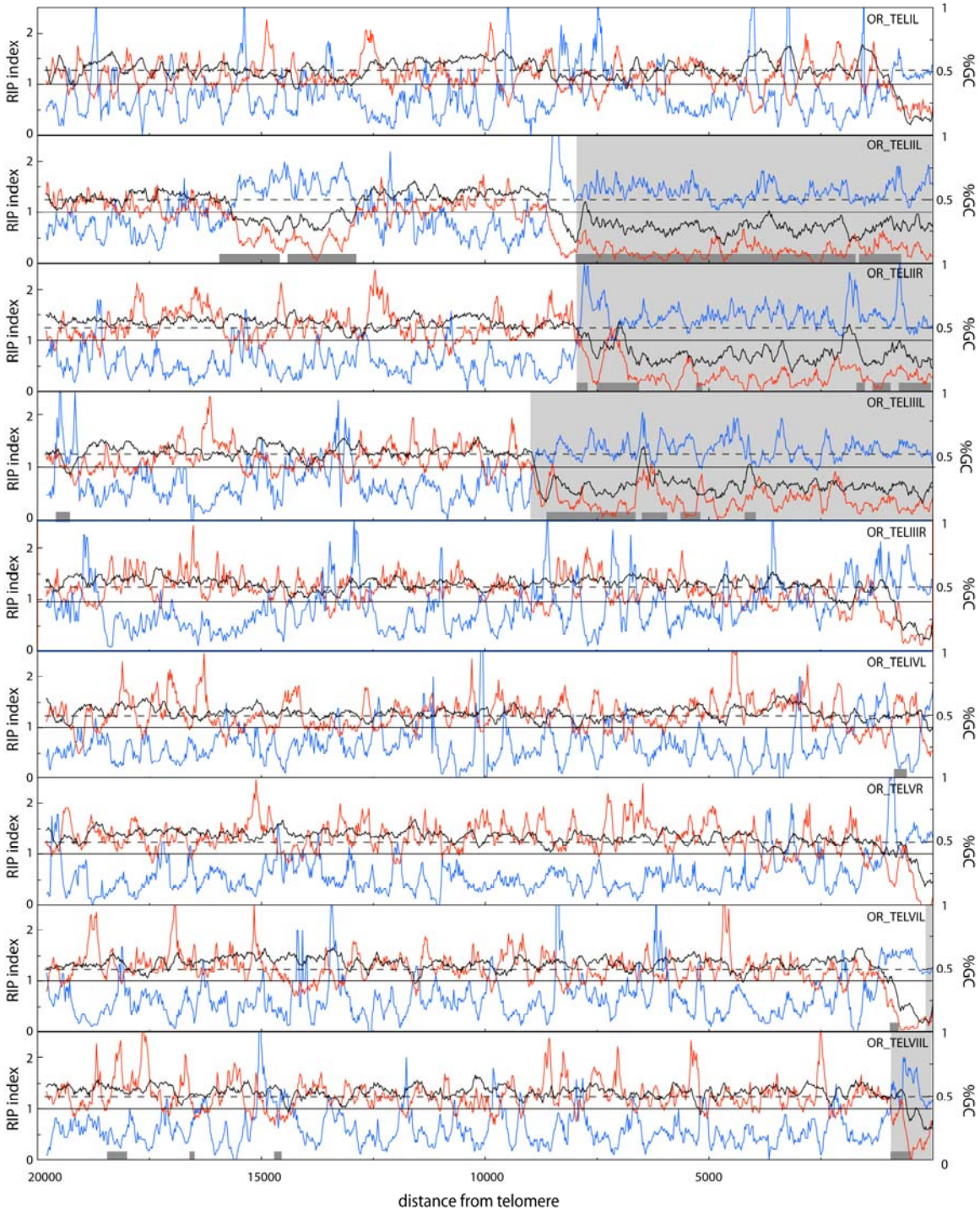

Figure 4 Sliding window analysis of nucleotide composition at the chromosome ends of the Oak Ridge strain.

GC content (%GC) and RIP indices I (TpA/ApT) and II [(CpA+TpG)/(ApC+GpT)] were calculated in a 200-bp window, which was slid in 20-bp increments across the terminal

20 kb of each chromosome end. The telomere repeats were excluded from the analysis. The black line represents %GC, RIP index I is plotted in blue and RIP index II in red. The x-axis shows the distance from the right-hand edge of the window to the start of the telomere array. The gray area highlights sequence regions that are not present at the homologous telomeres of the MV strain.

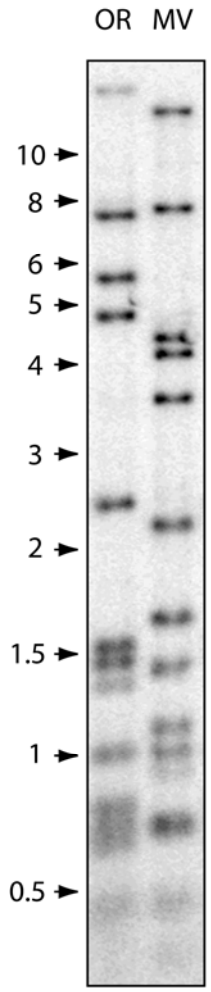

Figure 5 Telomeric RFLPs between the Oak Ridge and Mauriceville strains. Approximately 200 ng of DNA from each strain were digested with *Hin*dIII and *Not*I, electrophored through a 0.7% agarose and electroblotted to a nylon membrane. After alkaline fixation, the membrane was probed with a ^{32}P-labeled telomere probe for 20 h. The membrane was then washed to high stringency (0.1× SSC/0.1% SDS at 65°C) and then exposed to a phosphorimager screen. The resulting phosphorimage is shown. The positions of size markers (in kilobases) in the original gel are shown on the left-hand side.

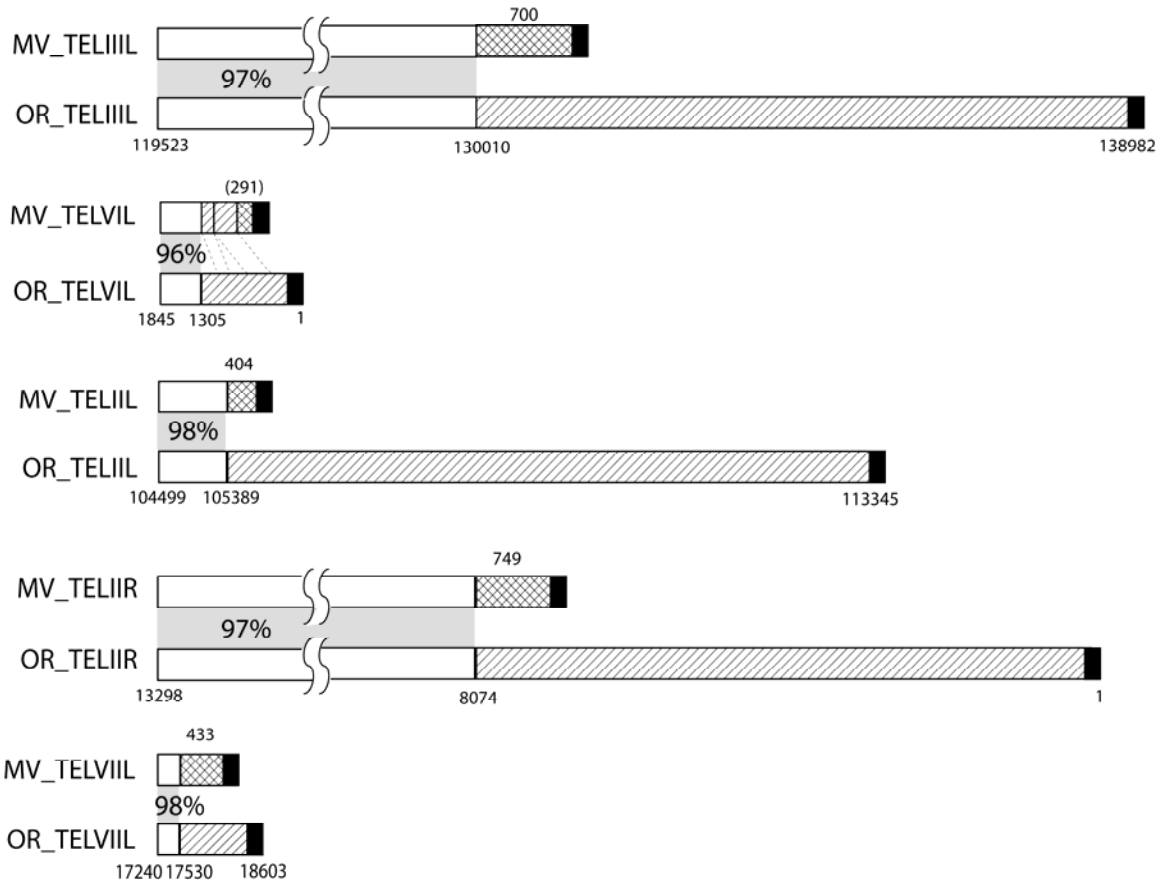

Figure 6 Alignment of Mauriceville chromosome ends with their homologous counterparts in Oak Ridge.

Telomere repeats $(TTAGGG)_n$ are shown as black rectangles. Rectangles with diagonal shading represent sequences unique to the OR telomeres, and those with crosshatching are unique to MV. The numbers above the MV telomeres correspond to the lengths of the novel sequences. The coordinates beneath the OR chromosome ends correspond to the positions within the respective sequence contigs. White, open-ended rectangles connected by gray shading represent regions of shared terminal sequence and the percent similarity is noted in the shaded areas. The amount of shared sequence that was compared can be calculated from the coordinates.

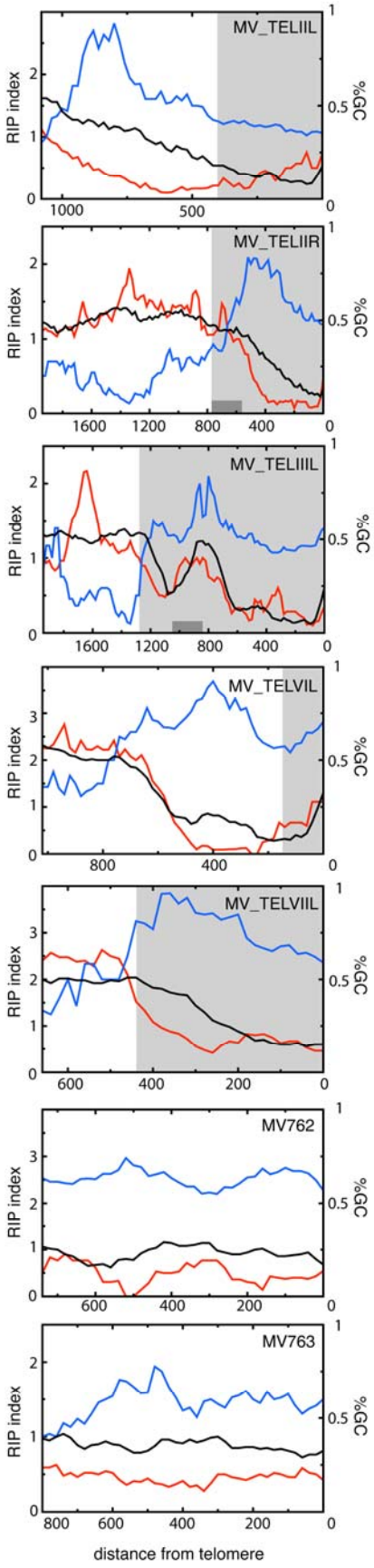
MV_TELIIL
MV_TELIIR
MV_TELIIIL
MV_TELVIL
MV_TELVIIL
MV762
MV763
RIP index
%GC
distance from telomere

Figure 7 Sliding window analysis of nucleotide composition at the chromosome termini of Mauriceville.

Values for %GC [(G+C)/(A+T+G+C)], RIP index I (TpA/ApT) and RIP index II [(CpA+TpG)/(ApC+GpT)] were calculated across a 200-bp window, which was slid in 20-bp increments across the terminal sequence (excluding the TTAGGG repeats). The x-axis shows the distance from the telomere to the right-hand edge of the window. The gray area highlights the region beyond the point of sequence divergence between the MV and OR chromosome termini. Black boxes show the positions of repeated sequences

Supplementary Table 1 List of probes used for RFLP mapping.

TelContig	probe	Polymorphism enzyme
TC2	PCR product (primers CWO49 & CWO50)	*Bam*HI
TC3	PCR product (primers CWO51 & CWO52)	*Eco*RV
TC4	PCR product (primers CWO57 & CWO58)	*Eco*RV
TC5	OR708, 819 bp *Dde*I fragment (341 – 1160)	*Rsa*I
TC9	OR709, 2.7 kb *Spe*I fragment (3050 – 5731)	*Alu*I
TC11	pJMOR005 insert	*Dde*I
TC12	pCWOR011 insert	*Dde*I
TC13	pJMOR001 insert	*Rsa*I

Supplementary Table 2 Segregation patterns of telomeric RFLPs.

TelContig	Segregation pattern[a]	Telomere
TC2	MOMOMMMOMOMOMMMOOMMMMOOOOOOOOMOMOMMOOO	IR
TC3	MOMOOOOOOMOOOOOOOMOOOOOOOMMOOOOOMOMOMM	VIR
TC4	MMOMMOOOMOMOOMM??OMOOMOMOMOOOMOMMOOMOM	VIIR
TC5	OMMOMMMMOMOMMMOMMMOOMOOMOMMMOMMOOOMOOM	IIIL
TC9	MOOMMOOMMOOMOMMMOMOMOMOMOMOOOMMOOMOOMO	IIL
TC11	MOOMMOOMOMOMOMMOOOMOOOOMOMOMOMMOMOMOOM	VIL
TC12	MOOMOMMOOMOMMOMOOMMOMOMOOMMMOMMOOMMOOO	IL
TC13	MOOMOMMOOMOMMOMOOMMOMOMOOMOMOMMOOMMOOM	VIIL

[a]"O" represents the band pattern of the Oak Ridge parent and "M" represents the band pattern of the Mauriceville parent. "?" indicates uncertainty (M vs O).

Supplementary Table 3 BLAST matches and Pfam domains for genes that occur within ~50 kb of *N. crassa* telomeres.

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
I L	NCU10129				
	NCU09901	putative arginine/serine rich protein	$3e^{-28}$	PF00076.13	RRM_1
	NCU09923	mRNA-nucleus export ATPase (Elf1), putative	0	PF00385.15	Chromo
				PF00005.18	ABC_tran
	NCU11134				
	NCU09904	glucan 3-β-glucosidase	$1e^{-168}$	PF00722.12	Glyco_hydro_16
	NCU09905				
	NCU09906				
	NCU09907	proteophosphoglycan ppg4	$4e^{-13}$		
	NCU09908	glycoside hydrolase, family	$6e^{-06}$		
	NCU09909	urea transporter (Dur3), putative	0	PF00474.8	SSF
	NCU09910	related to tol protein	$2e^{-95}$	PF06985.2	HET

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
II L	NCU09634	related to 2′, 3′-cyclic-nucleotide 3′-phosphodiesterase	$1e^{-100}$	PF00702.17	Hydrolase
	NCU09632	alkaline phosphatase, putative	$9e^{-49}$		
	NCU09631	alkaline phosphatase, putative	$1e^{-125}$	PF09423.1	PhoD
	NCU09630				
	NCU09629	Tol protein	0	PF06985.2	HET
	NCU09627				
	NCU09626				
	NCU09625	related to RNA binding protein	0		
	NCU09624				
	NCU09623				
	NCU09621	related to guanine deaminase	0	PF01979.11	Amidohydro_1
	NCU09620				
	NCU09619				
	NCU09617	related to homeotic protein CDP2	$2e^{-07}$		

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
II L	NCU09616			PF09637.1	Med18
	NCU09615	related to negative acting factor	0		
II R	NCU09495	Mcg1p	$1e^{-21}$		
	NCU09494				
	NCU09492	related to dock 180 protein	0	PF00018.19	SH3_1
	NCU09491	ferulic acid esterase, type B	$1e^{-170}$		
	NCU09490				
	NCU11088	6-O-methylguanine DNA methyltransferase	$4e^{-14}$	PF01035.11	DNA_binding_1
	NCU09489	phosphoglycerate mutase family protein	$3e^{-39}$	PF00300.13	PGAM
	NCU09488	related to SDA1 protein, required for normal orgs	0	PF08158.3	NUC130_3NT
				PF05285.3	SDA1
	NCU09487	MUC5A_HUMAN Mucin-5AC precursor (Mucin-5 subtype AC)	$3e^{-09}$		

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
II R	NCU09486	related to α-amylase	0	PF00128.15	α-amylase
	NCU09485	heat shock protein-like protein	0	PF00012.11	HSP70
	NCU09484	related to podosporapepsin papA (aspartyl protease)	0	PF00026.14	Asp
III L	NCU09526				
	NCU09525	secreted protein		PF09352.1	DUF1994
	NCU09524	flocculin	$1e^{-10}$		
	NCU09523	clock	$3e^{-29}$		
	NCU09522				
	NCU09521	ribosome biogenesis protein (Rrb1), putative	0	PF00400.23	WD40
	NCU09520	TPA_exp: putative two-component histidine kinases	0	PF02518.17	HATPase_c
				PF00512.16	HisKA
				PF01590.17	GAF
				PF00072.15	Response_reg

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
III L	NCU09519	aldehyde reductase, putative	$1e^{-100}$	PF00248.12	Aldo_ket_red
	NCU09518	glucooligosaccharide oxidase	$1e^{-101}$	PF01565.14	FAD_binding_4
				PF08031.3	BBE
	NCU09517	predicted: similar to putative phytoene synthase	$9e^{-41}$		
	NCU09516	RAD5_NEUCRsDNA repair protein rad-5s	0	PF00097.16	zf-C3HC4
				PF00176.14	SNF2_N
				PF00271.22	Helicase_C
				PF08797.2	HIRAN
	NCU09515				
	NCU09514	arrestin (or S-antigen)	$4e^{-69}$		
	NCU09513	GTP binding protein (GTPBP1), putative	0	PF03144.16	GTP_EFTU_D2
				PF03143.8	GTP_EFTU_D3
				PF00009.18	GTP_EFTU

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
III R	NCU11292				
	NCU11291			PF09351.1	DUF1993
	NCU08752	carboxylesterase family protein	$1e^{-124}$	PF00135.19	COesterase
	NCU08753				
	NCU08754				
	NCU08755	β-glucosidase		PF00933.12	Glyco_hydro_3
				PF01915.13	Glyco_hydro_3_C
	NCU08756				
	NCU08757	Zinc-containing alcohol dehydrogenase superfamily	$2e^{-06}$	PF08240.3	ADH_N
				PF00107.17	ADH_zinc_N
	NCU10427				
	NCU08758				
	NCU08759				
	NCU08760	endoglucanase, putative	$6e^{-60}$	PF00734.9	CBM_1

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
III R	NCU08761	vacuolar sorting receptor (Mrl1), putative	$1e^{-62}$	PF09451.1	ATG27
	NCU08762	C-14 sterol reductase (sterol C14-reductase)	0	PF01222.8	ERG4_ERG24
	NCU08763	LipA and NB-ARC domain protein	$7e^{-52}$		
	NCU08764				
	NCU08765				
	NCU10641				
	NCU08766	ATG17_MAGGR autophagy-related protein 17	$1e^{-120}$		
IV L	NCU10285	benzoate-monooxygenase cytochrome P450	$3e^{-29}$		
	NCU09635	FAD binding domain protein	$3e^{-72}$	PF01565.14	FAD_binding_4
				PF08031.3	BBE
	NCU09636	cytochrome P450	$1e^{-107}$	PF00067.13	p450
	NCU09637	O-methyltransferase, putative	$1e^{-88}$	PF00891.9	Methyltransf_2

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
IV L	NCU09638	polyketide synthase	0	PF08240.3	ADH_N
				PF00698.12	Acyl_transf_1
				PF02801.13	Ketoacyl-synt_C
				PF08242.3	Methyltransf_12
				PF00109.17	ketoacyl-synt
				PF08659.1	KR
	NCU09639			PF07247.3	AATase
	NCU09640	MFS toxin efflux pump (AflT), putative	$1e^{-156}$	PF07690.7	MFS_1
	NCU09641			PF04082.9	Fungal_trans
	NCU10572	short chain oxidoreductase (CsgA), putative	$2e^{-26}$	PF00106.16	adh_short
	NCU09642	sulfate transporter, putative	0	PF01740.12	STAS
				PF00916.11	Sulfate_transp
	NCU09643	phosphatidate cytidylyltransferase	$1e^{-164}$	PF01148.11	CTP_transf_1
	NCU09644	DNA repair protein Rad26, putative	$2e^{-33}$		

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
V R	NCU07159	subtilisin-like serine protease precursor	$1e^{-133}$	PF00082.13	Peptidase_S8
				PF05922.7	Subtilisin_N
	NCU07161				
	NCU07144	proteophosphoglycan	$3e^{-09}$		
	NCU07145				
	NCU07146	viral A-type inclusion protein, putative	$4e^{-44}$		
	NCU07147				
	NCU07148				
	NCU07149				
	NCU07150				
	NCU07151	ORF	$1e^{-29}$		
	NCU07152				
	NCU07153	glutamate carboxypeptidase	0	PF01546.19	Peptidase_M20
				PF07687.5	M20_dimer

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
VR	NCU07154	Yippee putative zinc-binding protein	$3e^{-32}$	PF03226.5	Yippee
	NCU07155	related to ornithine decarboxylase antizyme	$1e^{-32}$		
	NCU07156	imidazole glycerol phosphate synthase hisHF	0	PF00977.12	His_biosynth
				PF00117.19	GATase
	NCU07157				
	NCU07158	putative vacuolar targeting protein	$3e^{-45}$	PF05437	azlD
				PF00227	Proteasome
				PF08566	Pam17
VI L	NCU07143				
	NCU07142				
	NCU07141				
	NCU07140	MFS transporter, putative	$2e^{-29}$	PF07690.7	MFS_1
	NCU07139	fungal specific transcription factor	$8e^{-13}$	PF04082.9	Fungal_transcrip

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
VI L				PF00172.9	Zn_clus
	NCU11341	translation initiation factor eIF-2B	$1e^{-110}$	PF01008.8	IF-2B
				PF00293.19	NUDIX
	NCU07138				
	NCU11342	MFS hexose transporter, putative	$1e^{-179}$	PF00083.15	Sugar_tr
	NCU07136				
	NCU07135				
	NCU07134	putative xylanase	$3e^{-76}$	PF00722.12	Glyco_hydro_16
	NCU07133	metallo-β-lactamase superfamily protein	$4e^{-73}$	PF00753.18	Lactamase_B
	NCU07132	TAM domain methyltransferase, putative	$5e^{-24}$	PF08242.3	Methyltransf_12
	NCU07131	FF domain protein	$1e^{-113}$		
	NCU07130	putative xylanase	$7e^{-76}$	PF00331.11	Glyco_hydro_10
VI R	NCU04991	AF125094_ peptide transporter MTD1	$1e^{-141}$	PF03169.6	OPT

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
VI R	NCU04992				
	NCU04994	leucine aminopeptidase	$7e^{-58}$	PF04389.8	Peptidase_M28
	NCU10537				
	NCU04996	GNAT family acetyltransferase, putative	$2e^{-32}$		
	NCU04997	xylanase	$1e^{-143}$	PF00331.11	Glyco_hydro_10
				PF00734.9	CBM_1
	NCU04998	aspartic acid-rich protein aspolin2	$9e^{-46}$		
	NCU04999				
	NCU05000	L-aminoadipate-semialdehyde dehydrogenase	$1e^{-176}$	PF00501.19	AMP-binding
				PF07993.3	NAD_binding_4
	NCU05001	cytochrome P450-like protein	$1e^{-104}$	PF00067.13	p450
	NCU11221	CI-1	$1e^{-33}$		
	NCU10597	proteophosphoglycan ppg4	$2e^{-15}$		

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
VII L	NCU11407	UDP-glucose:sterol glycosyltransferase	0	PF03033.11	Glyco_transf_28
	NCU11408	RNA Polymerase II CTD phosphatase Fcp1	$1e^{-152}$	PF00533.17	BRCT
				PF03031.9	NIF
	NCU09299	NADH:ubiquinone oxidoreductase 14 kD subunit	$3e^{-66}$	PF06212.3	GRIM-19
	NCU11409	DNA helicase, putative	$1e^{-159}$	PF06733.6	DEAD_2
	NCU11410	cell cycle protein kinase, putative	$1e^{-137}$	PF00069.16	Pkinase
	NCU09297	condensin complex subunit	0	PF02985.13	HEAT
	NCU09296				
	NCU09295				
	NCU09294	conserved T+S rich domain protein	$5e^{-08}$		
	NCU09293				
	NCU09292	MFS transporter, putative	$8e^{-24}$		

Telomere[a]	Gene ID[b]	BLAST match[c]	e-value	Pfam ID	Pfam domain
VII L	NCU09291	FAD-linked sulfhydryl oxidase ERV2	$3e^{-48}$	PF04777.4	Evr1_Alr
	NCU09290	proteasome component PRE3 precursor	$1e^{-130}$	PF00227.17	Proteasome
	NCU09289				
	NCU09288	cyclin-dependent protein kinase complex	$3e^{-80}$	PF08613.2	Cyclin

[a]Gene content is provided for terminal regions whose sequence terminated in a telomere, and for regions that were shown by TERMINUS to be closely linked to a telomere.

[b]Genes are listed in order of their chromosomal positions (top = telomere-proximal; bottom = centromere proximal).

[c]The first meaningful[*] BLAST match is listed ([*]i.e. not predicted protein/hypothetical protein, etc.).

References

Barry, J. D., Ginger, M. L., Burton, P. and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int J Parasitol 33*, 29-45.

Bhattacharyya, A. and Blackburn, E. H. (1997). *Aspergillus nidulans* maintains short telomeres throughout development. *Nucleic Acids Res 25*, 1426-1431.

Blackburn, E. H. (1990). Telomeres: structure and synthesis. *J Biol Chem 265*, 5919-5921.

Blackburn, E. H. and Szostak, J. W. (1984). The molecular structure of centromeres and telomeres. *Annu Rev Biochem 53*, 163-194.

Cambareri, E. B., Jensen, B. C., Schabtach, E. and Selker, E. U. (1989). Repeat-induced G – C to A – T mutations in *Neurospora*. *Science 244*, 1571-1575.

Cambareri, E. B., Singer, M. J. and Selker, E. U. (1991). Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. *Genetics 127*, 699-710.

Coleman, M. J., McHale, M. T., Arnau, J., Watson, A. and Oliver, R. P. (1993). Cloning and characterisation of telomeric DNA from *Cladosporium fulvum*. *Gene 132*, 67-73.

Connelly, J. C. and Arst, H. N. Jr. (1991). Identification of a telomeric fragment from the right arm of chromosome III of *Aspergillus nidulans*. *FEMS Microbiol Lett 64*, 295-297.

Denayrolles, M., de Villechenon, E. P., Lonvaud-Funel, A. and Aigle, M. (1997). Incidence of *SUC-RTM* telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet 31*, 457-461.

Donelson, J. E., Hill, K. L. and El-Sayed, N. M. (1998). Multiple mechanisms of immune

evasion by African trypanosomes. *Mol Biochem Parasitol 91*, 51-66.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res 8*, 186-194.

Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res 8*, 175-185.

Fairhead, C. and Dujon, B. (2006). Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. *FEMS Yeast Res 6*, 428-441.

Farman, M. L. and Kim, Y.-S. (2005). Telomere hypervariability in *Magnaporthe oryzae Mol Plant Pathol 6*, 287-298.

Farman, M. L. and Leong, S. A. (1995). Genetic and physical mapping of telomeres in the rice blast fungus, *Magnaporthe grisea*. *Genetics 140*, 479-492.

Farman, M. L., Tosa, Y., Nitta, N. and Leong, S. A. (1996). MAGGY, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. *Mol Gen Genet 251*, 665-674.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem 132*, 6-13.

Flint, J., Bates, G. P., Clark, K., Dorman, A., Willingham, D., Roe, B. A., Micklem, G., Higgs, D. R. and Louis, E. J. (1997). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet 6*, 1305-1314.

Freitag, M., Williams, R. L., Kothe, G. O. and Selker, E. U. (2002). A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc Natl Acad Sci U S A 99*, 8802-8807.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., Fitzhugh, W., Ma, L.-J., Smirnov, S., Purcell, S. *et al.* (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature 422*, 859-868.

Galagan, J. E. and Selker, E. U. (2004). RIP: the evolutionary cost of genome defense. *Trends Genet 20*, 417-423.

Gao, W., Khang, C. H., Park, S. Y., Lee, Y.H. and Kang, S. (2002). Evolution and organization of a highly dynamic, subtelomeric helicase gene family in the rice blast fungus *Magnaporthe grisea*. *Genetics 162*, 103-112.

Goodwin, T. J. and Poulter, R. T. (2001). The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. *Yeast 18*, 865-880.

Gordon, D., Abajian, C. and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res 8*, 195-202.

Guzman, P. A. and Sanchez, J. G. (1994). Characterization of telomeric regions from *Ustilago maydis*. *Microbiol 140*, 551-557.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco Jr., I. and Blackburn, E. H. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell 51*, 899-912.

Hernandez-Rivas, R., Hinterberg, K. and Scherf, A. (1996). Compartmentalization of genes coding for immunodominant antigens to fragile chromosome ends leads to dispersed subtelomeric gene families and rapid gene evolution in *Plasmodium falciparum*. *Mol Biochem Parasitol 78*, 137-148.

Ikeda, K., Nakayashiki, H., Kataoka, T., Tamba, H., Hashimoto, Y., Tosa, Y. and Mayama, S. (2002). Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for

its sexual cycle in the natural field context. *Mol Microbiol 45*, 1355-1364.

Inglis, P. W., Rigden, D. J., Mello, L. V., Louis, E. J. and Valadares-Inglis, M. C. (2005) Monomorphic subtelomeric DNA in the filamentous fungus, *Metarhizium anisopliae*, contains a RecQ helicase-like gene. *Mol Genet Genom 274*, 79-90.

Keely, S. P., Renauld, H., Wakefield, A. E., Cushion, M. T., Smulian, A. G., Fosker, N., Fraser, A., Harris, D., Murphy, L., Price, C., *et al.* (2005). Gene arrays at *Pneumocystis carinii* telomeres. *Genetics 170*, 1589-1600.

Kouzminova, E. and Selker, E. U. (2001). *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J 20*, 4309-4323.

Kusumoto, K. I., Suzuki, S. and Kashiwagi, Y. (2003). Telomeric repeat sequence of *Aspergillus oryzae* consists of dodeca-nucleotides. *Appl Microbiol Biotechnol 61*, 247-251.

Li, W., Rehmeyer, C. J., Staben, C. and Farman, M. L. (2005). TERMINUS — Telomeric End-Read Mining IN Unassembled Sequences. *Bioinformatics 21*, 1695-1698.

Louis, E. J. and Haber, J. E. (1990). The subtelomeric Y′ repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. *Genetics 124*, 533-545.

Mandell, J. G., Goodrich, K. J., Bahler, J. and Cech, T. R. (2004). Expression of a RecQ helicase homolog affects progression through crisis in fission yeast lacking telomerase. *J Biol Chem 280*, 5249-5257.

Mandell, J. G., Bahler, J., Volpe, T. A., Martienssen, R. A. and Cech, T. R. (2005). Global expression changes resulting from loss of telomeric DNA in fission yeast. *Genome Biol 6*, R1.

Margolin, B. S., Garrett-Engele, P. W., Stevens, J. N., Fritz, D. Y., Garrett-Engele, C., Metzenberg, R. L. and Selker, E. U. (1998). A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics 149*, 1787-1797.

Metzenberg, R. L. and Grotelueschen, J. (1995). *Fungal Genet Newsl 42*, 82-90.

Metzenberg, R. L., Stevens, J. N., Selker, E. U. and Morzycka-Wroblewska, E. (1984). A method for finding genetic map position of cloned DNA fragments. *Neurospora Newsl 31*, 35-39.

Metzenberg, R. L., Stevens, J. N., Selker, E. U. and Morzycka-Wroblewska, E. (1985) Identification and chromosomal distribution of 5S rRNA genes in *Neurospora crassa*. *Proc Natl Acad Sci U S A 82*, 2067-2071.

Meyne, J., Ratliff, R. L. and Moyzis, R. K. (1989). Conservation of the hunman telomere sequence $(TTAGGG)_n$ among vertebrates. *Proc Natl Acad Sci U S A 86*, 7049-7053.

Moyais, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. and Wu, J.-R. (1988). A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc Natl Aced Sci U S A 85*, 6622-6626.

Nakamura, T. M., Morin, G. B., Chapman, K. B., Weirich, S. L., Andrews, W. H., Lingner, J., Harley, C. B. and Cech, T. R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. *Science 277*, 955-959.

Ness, F. and Aigle, M. (1995). *RTM1*: a member of a new family of telomeric repeated genes in yeast. *Genetics 140*, 945-956.

Nickles, K. and McEachern, M. J. (2004). Characterization of *Kluyveromyces lactis* subtelomeric sequences including a distal element with strong purine/pyrimidine strand bias. *Yeast 21*, 813-830.

Perkins, D. D., Radford, A. and Sachs, M. S. (2001). *The Neurospora Compendium: Chromosomal loci*. New York, NY: Academic Press.

Powell, W. A. and Kistler, H. C. (1990). *In vivo* rearrangement of foreign DNA by *Fusarium oxysporum* produces linear self-replicating plasmids. *J Bacteriol 172*, 3163-3171.

Pryde, F. E., Gorham, H. C. and Louis, E. J. (1997). Chromosome ends: all the same under their caps? *Curr Opin Genet Dev 7*, 822-828.

Pryde, F. E. and Louis, E. J. (1997). *Saccharomyces cerevisiae* telomeres. A review. *Biochemistry 62*, 1232-1241.

Rehmeyer, C. J., Li, W., Kusaba, M., Kin, Y. S., Brown, D., Staben, C., Dean, R. A. and Farman, M. L. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Res 34*, 4685-4701.

Richards, E. J. and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell 53*, 127-136.

Riethman, H., Ambrosini, A., Castaneda, C., Finklestein, J., Hu, X. L., Mudunuri, U., Paul, S. and Wei, J. (2004). Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res 14*, 18-28.

Sanchez-Alonso, P. and Guzman, P. A. (1998). Organization of chromosome ends in *Ustilago maydis*. RecQ-like helicase motifs at telomeric regions. *Genetics 148*, 1043-1054.

Schechtman, M. G. (1987). Isolation of telomere DNA from *Neurospora crassa*. *Mol Cell Biol 7*, 3168-3177.

Schechtman, M. G. (1989). Segregation patterns of *Neurospora* chromosome ends: mapping chromosome tips. *Fungal Genet Newl 36*, 71.

Schechtman, M. G. (1990). Characterization of telomere DNA from *Neurospora crassa*. *Gene 88*, 159-165.

Shampay, J., Szostak, J. W. and Blackburn, F. H. (1984). DNA sequences of telomeres maintained in yeast. *Nature 310*, 154-157.

Shpiz, S. G. and Kalmykova, A. I. (2007). Structure of telomeric chromatin in *Drosophila*. *Biochem 72*, 618-630.

Thon, M. R., Martin, S. L., Goff, S., Wing, R. A. and Dean, R. A. (2004). BAC end sequences and a physical map reveal transposable element content and clustering patterns in the genome of *Magnaporthe grisea*. *Fungal Genet Biol 41*, 657-666.

Trask, B. J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., *et al.* (1998). Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet 7*, 13-26.

Wada, M. and Nakamura, Y. (1996). Unique telomeric expression site of major-surface-glycoprotein genes of *Pneumocystis carinii*. *DNA Res 3*, 55-64.

Walmsley, R. W., Chan, C. S., Tye, B. K. and Petes, T. D. (1984). Unusual DNA sequences associated with the ends of yeast chromosome. *Nature 310*, 157-160.

Zakian, V. A. (1989). Structure and function of telomeres. *Annu Rev Genet 23*, 579-604.

BIOGRAPHICAL SKETCH

Cheng Wu was born in Wuxi, Jiangsu Province, People's Republic of China on March 17, 1977. In 1999, he received his B.S. in Genetics from Fudan University. He decided to pursue his Ph.D. degree in Biochemistry and Molecular Biology at the Oregon Health & Science University under the supervision of Dr. Matthew S. Sachs. Following the successful defense of his dissertation, Cheng will begin his postdoctoral research in the laboratory of Dr. Matthew S. Sachs at Texas A&M University.

Publications

Wu, C., Kim, Y.-S., Smith, K., Li, W., Hood, H. M., Staben, C., Selker, E. U., Sachs, M. S. and Farman, M. L.. Characterization of chromosome ends in the filamentous fungus *Neurospora crassa* (manuscript in preparation).

Doronina, V. A., **Wu, C.**, de Felipe, P., Sachs, M. S., Ryan, M. D. and Brown, J. D.. A translational recoding event that generates separated proteins from one open reading frame (manuscript in revision).

Wu, C., Amrani, N., Jacobson, A. and Sachs, M. S. (2007). The use of fungal *in vitro* systems for studying translational regulation. *Methods Enzymol* **429**: 203-225.

Aldridge, P. D., **Wu, C.**, Gnerer, J., Karlinsey, J. E., Hughes, K. T. and Sachs, M. S. (2006). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proc Natl Acad Sci U S A* **103** (30): 11340-11345.

Fang, P., Spevak, C. C., **Wu, C.** and Sachs, M. S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A* **101** (12): 4059-4064.

Fang, P., **Wu, C.** and Sachs, M. S. (2002). *Neurospora crassa* supersuppressor mutants are amber codon-specific. *Fungal Genet Biol* **36** (3): 167-175.