SPECIFIC REQUIREMENTS FOR TRANSLATIONAL REGULATION BY A NASCENT PEPTIDE THAT STALLS RIBOSOMES IN RESPONSE TO ARGININE

Christina C. Spevak
B.A., University of San Diego, 1998
M.S., Oregon Graduate Institute of Science and Technology, 2000

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

September 2006

The thesis “Specific Requirements for Translational Regulation by a Nascent Peptide that Stalls Ribosomes in Response to Arginine” by Christina C. Spevak has been examined and approved by the following Examination Committee:

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

Holly M. Simon, Ph.D.
Assistant Professor

James W. Whittaker, Ph.D.
Associate Professor

Adam P. Geballe, M.D.
Member, Fred Hutchinson Cancer Research Center
Professor, University of Washington

ACKNOWLEDGMENTS

I would like to express my appreciation and gratitude towards my thesis advisor, Dr. Matthew S. Sachs, for his inspiration, unconditional motivation, and his insights and suggestions, which made my Ph.D studies fulfilling and rewarding. A special thank you to my dissertation committee members: Drs. Holly M. Simon and James W. Whittaker. I am truly indepted to Dr. Adam P. Geballe for being part of my thesis committee and for his guidance and critical discussions during the Northwest Translational Control Group meetings. I would also like to thank Dr. Gebre Woldegeorgis and Dr. Keith Garlid for their insights and guidance. Thanks to Dr. Peter Zuber and Michiko Nakano for their support and encouragement throughout my Ph.D studies, and to their lab members for their support and technical assistance.

My sincere thanks to Dr. Peng Fang for his guidance with my research project and for his unconditional support and friendship. Thanks to members of the Sachs lab who have kindly assisted me with experiments: Cheng Wu, Kanan Lathia, Richard Feinberg, Paul ReFalo, Megan Kennedy, Ana Luiza Ramos-Crawford, Rex Flowers, and the current postdocs Drs. Meray Baştürkmen, Heather Hood and Yanzhen Cui. A special thank you to Dr. Anthony Gaba for being my closest labmate, friend and basketball buddy. Thanks to Nancy Christie and Terrie Hadfield for all their assistance and kindness throughout my graduate studies. I appreciate all the moral support from all my friends especially Seth Ettwein for his unconditional friendship and encouragement and Montira Leelakriangsak for her delicious meals, assistance both professionally and personally, and for her friendship that made graduate school a great experience.

A special thank you goes to my uncle Christopher for his moral support and involvement throughout my studies. I am forever indebted to my family especially my mother and father and my grandparents for their continuous love and support, which without, this thesis would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Specific Requirements for Translational Regulation by a Nascent Peptide that Stalls Ribosomes in Response to Arginine

Christina C. Spevak

Ph.D., OGI School of Science & Engineering
Oregon Health & Science University
September 2006

Thesis Advisor: Dr. Matthew S. Sachs

Neurospora crassa arg-2 gene encodes the small subunit of arginine specific carbamoyl phosphate synthethase, the first enzyme in fungal arginine (Arg) biosynthesis. The *arg-2* mRNA contains an upstream open reading frame (uORF) specifying an evolutionarily conserved 24-residue peptide called the arginine attenuator peptide (AAP). Synthesis of the AAP causes ribosomes to stall on the mRNA in the presence of high concentrations of Arg. The amino acid sequence of the AAP, and not the sequence of its coding region, is responsible for this regulation. Scanning mutagenesis within the evolutionarily conserved region revealed that some residues are more important than others for the AAP to function. While most known nascent peptides that regulate translation are found encoded as uORFs or as N-terminal leader peptides, the AAP can exert regulatory function whether placed near the N-terminus or internally within a large polypeptide. The AAP's peptide sensing features are conserved due to regulated stalling of fungal, plant, and animal ribosomes in response to Arg. That the AAP functions as an internal domain to regulate elongation in response to Arg establishes that such domains

can provide a means of controlling translational elongation. The minimal sequences required for AAP to function as an internal domain was revealed by systematic deletion of its natural N- and C-terminal regions. Comparative analysis of the AAP with other fungi showed that the evolutionarily conserved region of the peptide is required for regulation. This minimal domain functions when placed as a uORF as seen by toeprint assay. Analyses of Arg analogs provided key insights in the structural requirements for Arg's role in regulation. These results taken together provide a detailed picture of the requirements for Arg-specific regulation mediated by the AAP.

CHAPTER 1
INTRODUCTION

The genetic code is a guideline for conversion of DNA sequences to proteins in cells. The universality of the genetic code includes animals, plants, fungi, archaea, bacteria, and viruses. There exist small variations in the code in mitochondria and certain microbes. However, these variations only represent a fraction of cases and the code applies to the majority of known nuclear genes.

The conversion of messenger RNA (mRNA) to protein, also known as translation, is an essential process for proper development and maintenance of all organisms. Translational controls regulate a variety of biological processes. The mechanism of translational regulation modulates gene expression in processes involved in early embryonic development, cell differentiation and metabolism. Recently, there has been a link between the regulation of protein synthesis and disruption of cell behavior that may result in diseases such as cancer.

This chapter summarizes the processes underlying translation and control mechanisms in eukaryotes. It also explains the role of upstream open reading frames (uORFs) in gene expression. The study of uORFs is significant since these are seen in ~10% of eukaryotic transcripts, products involved in growth control and development, and are common in certain classes of genes, including two-thirds of cancer-causing genes (proto-oncogenes). The two examples of genes that contain these uORFs will be discussed further in the following chapters. These are the *Neurospora crassa arg-2* gene and the human *her-2* gene. The *arg-2* uORF encodes an evolutionarily conserved peptide called the arginine attenuator peptide (AAP), involved in arginine-specific regulation. The AAP is shown to maintain regulatory function in fungal, plant, and animal cell-free translation systems when placed as a domain near the N-terminus or internally within a large polypeptide. Our detailed studies on *arg-2* are documented in Chapters 2-4. The

human *her-2* gene, involved in ~30% of breast cancers, contains a uORF that inhibits downstream translation of mRNA in reticulocyte lysates, wheat germ, and *Neurospora* cell-free systems as well as after RNA transfection into HeLa cells as indicated by luciferase reporter assays. Most importantly, the *her-2* uORF has an effect on the selection of downstream initiation codons which is discussed in detail in Chapter 5.

1.1 THE MOLECULAR MECHANICS OF EUKARYOTIC TRANSLATION

Protein synthesis is fundamental for gene expression. There are multiple elements that influence translation of mRNA (Figure 1.1) (Gebauer and Hentze, 2004). Most eukaryotic mRNAs have a distinctive and essential modification called the m^7G cap structure at the 5' end that is critical for efficient translation (Hershey and Merrick, 2000; Kozak, 1999). This cap structure is involved in various cellular processes such as pre-mRNA splicing, nucleocytoplasmic RNA export and localization and translation initiation (Cougot et al., 2004). Poly (A) tails are found at the 3' ends of almost all eukaryotic mRNAs. The discovery and characterization of the poly (A)-binding protein (PAB) supported the idea that the poly(A) tail was involved in translation (Adam et al., 1986; Sachs and Kornberg, 1985). PAB contains four RNA recognition motifs (RRMs) and has a minimal RNA site size of ~10 nucleotides (Deo et al., 1999). The average length of poly(A) tails in yeast is ~70 nucleotides and is several fold higher in mammals (~200-250 nucleotides) (Jacobson, 1996) Between two and ten PAB molecules will generally be bound to the end of an mRNA (Kapp and Lorsch, 2004b).

The cap and poly (A) tail on eukaryotic mRNA serve as translational enhancers for the cap-dependent protein synthesis. PAB interacts with eIF4G, a translation initiation factor, that interacts with eIF4E (the mRNA 5' cap-binding protein) and acts as an anchor for further translation initiation factors. This results in the circularization of eukaryotic mRNAs due to the cap and poly(A) tail interaction resulting in the stimulation of translation (Gallie, 1991; Munroe and Jacobson, 1990; Preiss and Hentze, 1998; Tarun and Sachs, 1995; Tarun and Sachs, 1996).

Although an internal ribosome entry site (IRES) can also serve as a specific point of translation initiation (Prats and Prats, 2002), the majority of eukaryotic translation occurs in a "cap-dependent" manner and this mode of regulation will be the focus in this thesis.

1.1.1 Cap-dependent initiation

In eukaryotes, initiation of protein synthesis occurs in the cytoplasm. The fundamental objective of translation initiation is to assemble the initiator tRNA and the mRNA onto the small ribosomal subunits to form the initiation complex (Figure 1.2). The 43S complex, which is composed of eIF2·GTP·Met-tRNA and the 40S subunit, is loaded onto the 5' end cap structure and locates the initiation codon via the scanning mechanism (Kozak, 2002). Recognition of the first AUG triggers GTP hydrolysis which later facilitates the joining of the large (60S) ribosomal subunit to the 40S·Met·tRNAi·mRNA complex which is the final step of translation initiation. The machinery required for initiation includes at least 12 initiation factors which are composed of more than 23 different polypeptides (Kapp and Lorsch, 2004b). The molecular mechanics of the key steps in translation initiation in eukaryotes are discussed below.

1.1.1.1 Ternary complex formation. Initiation factor, eIF2 (~125kDa), is a heteromultimer, composed of three subunits: α, β, and γ. Previous genetic and biochemical experiments indicated that the γ subunit is implicated in both GTP and Met-tRNAi binding (Erickson and Hannig, 1996; Kurzchalia et al., 1984; Naranda et al., 1995). The γ subunit is homologous in sequence to elongation factor eEF1A/EF-Tu (Kyrpides and Woese, 1998) and protein structure confirmed its similarity to EF-Tu/eEF2A (Schmitt et al., 2002) which is also a GTP-dependent carrier of aminoacylated tRNAs. eIF2 has more specificity than EF-Tu/eEF1A, however, since it binds only the initiator tRNA. The elements in the initiator tRNA that contribute to its specificity are the anticodon stem, which is important for ribosome binding along with conserved sequences around the T-loop, a phosphoribosyl modification at position 64 (Åström et al., 1993) and an A1:U72 base pair at the end of the acceptor stem which is the important for identification of the initiator tRNA (Farruggio et al., 1996; Kapp and Lorsch, 2004a; von

Pawel-Rammingen et al., 1992). Complex formation requires GTP and is inhibited by GDP. This GTP-dependent recognition of the methionine moiety may be involved in the prevention of unacylated $tRNA_i$ from entering the initiation pathway and may be important in the release of initiator tRNA from eIF2 after recognition of the initiation codon (Kapp and Lorsch, 2004b). eIF2, after one round of initiation, leaves the ribosome as a binary complex with GDP, its inactive form. For it to bind another Met- $tRNA_i$, it must be in its active form bound to GTP. GDP is replaced by GTP through the action of the exchange factor, eIF2B, which contains five non-identical subunits and is involved in complex formation with eIF2, GDP, and GTP (Sonenberg et al., 2000). The α subunit of eIF2 is subject to phosphorylation on residue Ser 51 under a variety of extracellular conditions, such as nutrient deprivation and viral infection. In the phosphorylated state, eIF2α subunit does not form the ternary complex (eIF2·GTP·Met-tRNA) necessary for translational initiation.

1.1.1.2 Preinitiation Complex (43S) Formation. The binding of Met tRNA to the 40S ribosomal subunit is a common step in translation of mRNAs. Regulation of this step controls global rates of protein synthesis. Numerous components of the initiation machinery are involved in the preinitiation (43S) complex formation. This complex can be isolated and detected via sucrose gradient centrifugation (Sonenberg et al., 2000). eIF1A and eIF3 enhance ternary complex binding to the small (40S) ribosomal subunit in a reconstituted mammalian initiation system (Benne et al., 1978a; Thomas et al., 1980). In the same system, eIF1, eIF1A, eIF3 are all required to stabilize binding of the ternary complex to the 40S subunit (Chaudhuri et al., 1999; Majumdar et al., 2003). However, in a reconstituted yeast system, only eIF1 and eIF1A are critical (Algire et al., 2002). Studies in vivo show that both eIF1A and eIF3 affect the binding of the ternary complex (Olsen et al., 2003; Valasek et al., 2002). In vitro translation experiments using yeast cell extracts indicate that eIF5 affects recruitment of ternary complex and/or mRNA binding to 40S subunit. Since eIF5 is a central component of the multifactor complex, its interaction with the ternary complex along with eIF1 and eIF3, is likely to stabilize the ternary complex binding to the 40S subunit via the chelate effect that allowed further complex binding (He et al., 2003).

1.1.1.3 Preinitiation complex association with the mRNA. The recognition of the 5' cap structure by eIF4E is the first step in the association of the 40S subunit with the mRNA. The significance of the interaction of eIF4E with cap has been shown in vivo and in vitro (Iizuka et al., 1994; Michel et al., 2000; Preiss and Hentze, 1998; Tarun and Sachs, 1995). eIF4E, is part of a complex called eIF4F which is comprised of eIF4G, and eIF4A. The cap-eIF4E interaction is important since eIF4E is a key factor in selecting mRNAs for translation.

The eIF4E can associate with the 5' cap when there is a stable secondary structure present, but a functional interaction requires association of subunits of eIF4F with the 5' end of the mRNA (Muckenthaler et al., 1998). Additionally, the eIF4F-cap complex has a role in the recruitment of eIF4A to the 5' end of the mRNA which leads to the unwinding of any secondary structures resulting in an ideal binding site for the 43S complex (Lamphear et al., 1995). An RNA-dependent ATPase, eIF4A unwinds RNA duplexes in vitro (Ray et al., 1985; Rozen et al., 1990) and its ATPase activity is required for duplex unwinding in vitro (Pause and Sonenberg, 1992) and translation initiation in vivo (Blum et al., 1992). It does this by using ATP hydrolysis to disrupt RNA structure. In vitro, eIF4 is ineffective in unwinding RNA duplexes longer than ~10 base pairs and its activity decreases for these short duplexes if the stability of the duplex is increased by high G:C content (Rogers et al., 2001). The general model currently is that factors eIF4B and 4H captures regions of single-stranded RNA produced by eIF4A. The eIFG subunit of eIF4F perhaps performs a similar function via its three RNA-binding sites, which are necessary for successful translation initiation (Berset et al., 2003).
The large multisubunit factor, eIF3 may be involved in mRNA loading (Benne et al., 1978b; Phan et al., 2001; Trachsel et al., 1977). This factor binds the 40S ribosomal subunit and the ternary complex and its many subunits binds the RNA (Naranda et al., 1994; Verlhac et al., 1997) which leads to the possibility that it may play a role as a scaffold to alter the conformation of the 40S subunit, leading to easier access for the mRNA and/or ternary complex.

The 3' end of the mRNA also is important in the loading of the mRNA into the 40S subunit. The poly(A) tail is involved in the recruitment of the mRNA to the 43S complex. Previous studies showed that the poly(A) tail can facilitate binding of the 43S

complex to the mRNA, however, it cannot direct it to the 5'-end in the absence of the eIF4F-5'-cap complex (Preiss et al., 1998). This result is evidence that the 5'-cap promotes the usage of the 5'-end of the mRNA as the binding site for the 40S subunit. Experiments in mammalian cell extracts suggested that the poly(A) tail perhaps stimulates the association of the subunits rather than mRNA binding to the 43S complex (Jacobson, 1996; Munroe and Jacobson, 1990).

1.1.1.4 Scanning and AUG codon recognition. The "scanning" hypothesis proposes that once the 43S complex loads on the 5'-end of an mRNA, it moves towards the 3'-end until it encounters an initiation codon (Kozak, 2002). This scanning process requires the hydrolysis of ATP (Kozak, 1980). The 5' proximal AUG is utilized in ~90% of mRNAs that use this scanning mechanism (Kozak, 1980). The scanning 40S initiation complex recognizes the initiation codon by the dominant codon/anticodon interaction (Cigan et al., 1988). To recognize the initiation codon, Met-tRNAi should be bound to 40S ribosomes soon after the scanning process begins. There are three important players that are central to identifying the initiation codon: eIF1, eIF2 (any of its subunits), and eIF5, which promotes the hydrolysis of eIF2-GTP. By genetic analysis, these factors have been found to influence the fidelity of initiation codon selection (Castilho-Valavicius et al., 1990; Cui et al., 1998; Hashimoto et al., 2002). A model suggests that eIF1, the smallest initiation factor, changes the conformation of the P site of the 40S subunit which allows the decoding process to proceed (Pestova and Kolupaeva, 2002). eIF1 may detect the formation of the codon-anticodon interaction indirectly, perhaps by interacting with the initiator tRNA and responding to the conformational changes in it when the initiation codon is encountered. These structured changes in the tRNA may be significant for sensing the cognate codon-anticodon base-pairing during the A-site (the site which accepts the incoming aminoacylated tRNA) decoding process (Piepenburg et al., 2000). eIF5 is possibly regulated in response to a signal that the AUG codon has been reached. This signal could be direct, in which eIF5 could monitor the codon-anticodon interactions or it could be indirect mediated through conformational changes in the eIF1 or the 40S subunit (Huang et al., 1997).

Sequences surrounding the AUG codons are important in specifying which one is used for the initiation process (Kozak, 1997). These sequences play an important role since AUG codons in a poor context are bypassed. Initiation can be enhanced by presence of RNA secondary structure downstream from the codon (Kozak, 1991). In vitro experiments have suggested that eIF1 may play a role in discriminating between AUGs in favorable and unfavorable contexts by destabilization of preinitiation complexes on the least favorable AUGs (Pestova and Kolupaeva, 2002). Yeast and plants have different consensus sequences surrounding their initiation sites (AAAAA**AUG**UCU and AA(A/C)A**AUG**GC, respectively) (Cigan and Donahue, 1987; Hamilton et al., 1987; Joshi et al., 1997), the only common nucleotide being at the A postion –3 (with the A of AUG as 0). These sequences are important for accurate identification of the favored initiation site.

The final step in the translation initiation process in eukaryotes is the joining of the two subunits, 40S and 60S, following the recognition of the initiation codon and deposition of the initiator tRNA into the P site (the site which holds the tRNA with the nascent peptide chain) of the 40S subunit. In this process, two GTP hydrolysis events are required. The first is catalyzed by eIF2 upon recognition of the initiation codon and the other is after the 80S complex formation (Lorsch and Herschlag, 1999). eIF5B's GTPase activity is not necessarily required in the association of the two ribosomal subunits. Instead, the hydrolysis of GTP leads to the release of the factor from the 80S complex once the association of the subunits is complete (Lee et al., 2002; Pestova et al., 2000; Shin et al., 2002). In its GTP-bound form but not in its GDP-bound form, eIF5B has a high affinity for the ribosome. When initiation factors bound to 40S initiation complex dissociate, the large (60S) subunit can bind. EIF5B with GTP converts the preformed 40S complexes into the 80S initiation complexes.

1.1.2 Elongation

The sequence of codons on the mRNA directs the synthesis of the polypeptide chain. This complicated process occurs in the ribosome with high speed and high accuracy with an error rate between 10^{-3} to 10^{-4}. Translation elongation mechanisms

seem to be highly conserved, therefore, extensive studies in the prokaryotic system helped elucidate elongation mechanisms in eukaryotes (Ramakrishnan, 2002).

The ribosome consists of two subunits in all species and each subunit has three binding sites for tRNA, designated the A- (aminoacyl) site, P-(peptidyl) site, and the E-(exit), which holds the deacylated tRNA prior to leaving the ribosome. Both subunits are involved in translocation in which the tRNAs and mRNA move precisely through the ribosome, one codon at a time (Green and Noller, 1997). Translation involves other factors, many of which are GTPases activated by the ribosome. The ribosome plays a major role in the selectivity of tRNA by direct recognition of the geometry of codon-anticodon base pairing.

The peptide chain elongation cycle begins with a peptidyl RNA in the P site of the ribosome with an empty A site. The aminoacyl tRNA is carried to the A site as part of the ternary complex with GTP and eEF1A. The eEF1A·GTP·aa·tRNA complexes with either the cognate or noncognate aminoacyl tRNAs can bind to the A site of the ribosome. The cognate tRNA is selection for the next elongation step is based on several factors: codon-anticodon basepairing between the mRNA and the tRNA, conformational changes in the decoding center of the small ribosomal subunit, and GTP hydrolysis by eEF1A/EF-Tu. (Rodnina and Wintermeyer, 2001).

1.1.2.1 Decoding. The ribosome contains a "decoding site" that recognizes the geometry of codon-anticodon base pairings and sterically discriminates against mismatches. In a view termed "kinetic proofreading" the tRNA has two chances to dissociate: during initial selection and after GTP hydrolysis. Pre-steady state kinetic experiments show that cognate tRNA has lower dissociation rates and has much faster forward rates of GTPase activation and accommodation than near cognate tRNA. Kinetic experiments show that paromomycin increased both the affinity and GTPase activation rate of near cognate tRNA (Pape et al., 2000). It was proposed that cognate tRNA induces a conformational change, which is similar to a previous allosteric three site model (Nierhaus, 1990; Pape et al., 1999). In that model, the affinities of A- and E- site tRNAs are coupled. In the presence of E site tRNA, only cognate ternary complex has enough affinity for the A site to induce a conformational change in the ribosome on

binding, leading to the release of E site tRNA. Therefore, codon-anticodon complex formation is energetically important.

1.1.2.2 Role of EF-Tu/eEF1A. eEF1A and aminoacyl-tRNA synthase (ARS) are the proteins that advance the translation elongation cycle. Mammalian eEF1A evolved from prokaryotic EF-Tu (EF1A) by insertion of ~70 amino acids into 16 different sites, mostly at loop regions based on crystal structure for EF1A (Cavallius and Merrick, 1998). Its main function is in the formation of the ternary complex (eEF1A·GTP·aa·tRNA), which is then bound to the ribosomal A-site in a codon-dependent manner. eIF1A and ARS are supposed to play a main role in tRNA sequestering during mammalian translation (Negrutskii and El'skaya, 1998). The selection of tRNA begins with the binding of eIF1A ternary complex which presents the tRNA to the decoding site at a different angle (Stark et al., 1997b). The codon-anticodon recognition triggers the hydrolysis of GTP by EF-Tu through the transmission of a signal from the decoding site on the small ribosomal subunit (where the codon-anticodon interactions are recognized) to the large ribosomal subunit (where the GTPase domain of EF-Tu binds). This signal could be transmitted by the tRNA itself, since an intact tRNA is required for this process (Piepenburg et al., 2000). It has been shown that eEF1A was able to promote the renaturation of ARS and protect them against denaturation by dilution. In vitro renaturation occurs at the molar ratio of eIF1A to ARS equivalent to that found in the cytoplasm of higher eukaryotic cells. The chaperone-like activity of eEF1A might be important for maintaining the enzymes activity in mammalian cells (Lukash et al., 2004). eEF1A has been linked to the onset of cell transformation, is upregulated in cell-death, and also involved in the regulation of ubiquitin-mediated protein degradation. In addition, eEF1A undergoes phosphorylation and methylation that influence the activity of the protein (Lamberti et al., 2004).

1.1.2.3 Peptidyl-transferase and Translocation. The formation of the peptide bond between the entering amino acid and the peptidyl tRNA is catalyzed by the ribosomal peptidyl transferase center (Moore and Steitz, 2003). The formation of the peptide bond occurs rapidly and spontaneously when the aminoacyl end of the A- site tRNA enters the peptidyl transferase center (Pape et al., 1998). Following peptidyl

transferase, the P site tRNA is deacylated in an intermediate state with its acceptor end in the exit (E) site of the 60S subunit and its anticodon in the P site of the 40S subunit (Green and Noller, 1997). This complex is translocated with the deacylated tRNA in the E site, the peptidyl tRNA in the P site, and the mRNA shifted by three nucleotides to prepare the next codon of the mRNA to be positioned in the A site. This translocation process is facilitated by eEF2 upon its hydrolysis of GTP (Wintermeyer et al., 2001). Previous studies suggest that translocation can occur even in the absence of mRNA (Belitsina et al., 1981), therefore, tRNAs play a primary role in the translocation, possibly "dragging" the mRNA with them. Based on structural studies there is a kink between the A- and P site codons (Ogle et al., 2001; Yusupov et al., 2001) opening the possibility that this plays some role in the maintenance of the reading frame. This process is repeated until a termination codon is encountered.

1.1.2.4 Role of EF-G/eEF2 and eEF3. Eukaryotic elongation factor (eEF2), a ribosome activated GTPase, occupies an essential role in protein synthesis in which it catalyzes the translocation of the two tRNAs and the mRNA after peptidyl transfer on the 80S ribosome. Single rounds of translocation can occur even with nonhydrolyzable analogs of GTP, suggesting that GTP hydolysis is not a requirement for translocation but this may be required for the release of EF-G. However, other kinetic experiments showed that GTP hydrolysis occurs prior translocation suggesting that the energy of hydrolysis is used to drive translocation (Rodnina et al., 1997). Therefore, hydrolysis itself is not the driving force of translocation but rather the energy of the process is required for the release of EF-G. Therefore, this process contains a switch to drive the reaction. This type of switch is of physiological relevance since it has been shown that molecular motor proteins, such as myosin, have regions that act as switches. These regions have structural homology to G-protein elements that change conformation during nucleotide binding or exchange (Kull and Endow, 2004).

The mapping of EF-G on the ribosomes was done using two techniques: hydroxyl radical probing from tethered Fe on engineered cysteines (Wilson and Noller, 1998) and direct visualization by cryo-electron microscopy (cryo EM) (Agrawal et al., 1998). These techniques supported the view that EF-G bound to the ribosome in a similar way as

the ternary complex observed by cryo EM (Stark et al., 1997a). The cryo EM structures provide for a better understanding of the role of EF-G. This simplistic view of the translocation process is that it is slow without other factors, and EF-G binding to the A site lowers the kinetic barriers.

There is an exception to the evolutionary conservation in translation elongation and that is the existence of eEF3 which is exclusively present in fungi. eEF3 is an ATP-binding protein that promotes the release of deacylated tRNA from the E-site which increases the affinity of aa-tRNA for the A-site (Triana-Alonso et al., 1995). eEF3 was required in yeast ribosomes for poly-Phe synthesis assays in vitro in contrast with mammalian ribosomes (Skogerson and Engelhardt, 1977). The gene that encodes for eEF3 is necessary for yeast viability (Qin et al., 1990). eEF3 is found primarily associated with translating cystolic ribosomes and the majority of eEF3 is located in polysome fractions, which are consistent with the requirement for peptide bond formation. eEF3 has been shown to physically and genetically interact with eEF1A (Anand et al., 2003; Kovalchuke et al., 1998). The full-length histidine affinity-tagged form of eEF3 is functional in yeast (Andersen et al., 2004).

1.1.3 Termination

Translational termination occurs after a stop codon is translocated into the A site of the ribosome by EF-G/eEF2 and decoded at the small ribosomal subunit. The reaction triggered by a stop signal, the cleavage of the ester bond between the peptidyl and tRNA moieties of the peptidyl-tRNA occurs within the large ribosomal subunit at the peptidyl transferase center (PTC) of the ribosome. Decoding the stop signals could be implemented by the ribosome itself, external factors that bind to the ribosome in response to a stop signal and dissociate after completion and the combination of external factors and the ribosome. The final result is the release of a complete polypeptide following the hydrolysis of the ester bond that linked the polypeptide chain to the P site tRNA.

1.1.3.1 Role of eERF1 and eERF3. In eukaryotes, class I release factor protein, eRF1, recognizes all three stop codons: UAA, UAG, and UGA. Class I release factors associate transiently with the ribosomal A site. There is a direct interaction between stop codons and class I RFs (Moffat and Tate, 1994; Nakamura et al., 1996). Class I RFs has

been shown to cross-link specifically with stop signals in mRNA, although very low yield has been observed (Bulygin et al., 2002; Chavatte et al., 2002; Poole and Tate, 2000). eRF1-dependent stop codon is sensitive to both the position and the nature of the substituting amino acid. For stop codon recognition by eRF1, there are two proposed models: a protein-anticodon model (Nakamura et al., 2000) and a cavity model (Bertram et al., 2000; Inagaki et al., 2002). In the protein-anticodon model, the linear sequence of the amino acid decodes a stop codon, while in the cavity model, a combination of the amino acid from different parts of the polypeptide chain are clustered in space around a termination codon are responsible for decoding. eRF3 (Frolova et al., 1996; Zhouravleva et al., 1995), may be responsible for triggering the release of eRF1 from the ribosome following peptidyl-tRNA hydrolysis, which makes it an essential protein in eukaryotes. eRF3 binds to eRF1 in the absence of the ribosome and it has also been shown that it also binds to the poly(A) binding protein (Cosson et al., 2002; Hoshino et al., 1999).

1.1.3.2 The GGQ motif. A single sequence motif, GGQ is universally conserved in all eubacterial, archaebacterial and eukaryotic class I release factors and may have a similar function with the CCA end of tRNA. This tripeptide is positioned in a loop at the end of a stem region that interacts with the ribosomal peptidyl transferase center. This sequence seems to be a requirement for the activation of peptidyl-tRNA hydrolysis. The binding of eRF1 to the A site is presumably stabilized by interactions with both of the ribosomal subunits. A mutation in the GGQ domain of eRF1, which is located in both subunits, reduce the binding of eRF1 to the ribosome (Frolova et al., 2002; Seit-Nebi et al., 2002; Song et al., 2000a). The glycine residues of this GGQ in both eukaryotic and prokaryotic factors are indispensable for RF activity when tested both in vivo and in vitro (Frolova et al., 1999) (Mora et al., 2003). It has been proposed that the glutamine coordinates a water molecule and mediates a nucleophilic attack on the ester bond of the peptidyl-tRNA in the P-site (Song et al., 2000b). GAQ mutants of RF1 and RF2 are 4-5 fold less efficient in termination reaction versus wild-type, however, ability to bind to the ribosome is retained (Zavialov et al., 2002). The glutamine side chain of this motif is specifically methylated in both prokaryotes and eukaryotes. Methylation in *E. coli* is due PrmC; methylation results in strong stimulation of peptide chain release (Graille et al.,

2005). In *S. cerevisiae*, the product of the YDR140w gene is required for the methylation of eRF1 in vivo and for optimal yeast cell growth. The product of this gene has significant homology to PrmC (Heurgue-Hamard et al., 2005).

1.1.3.3 Mechanism of peptidyl-tRNA hydrolysis. The termination reaction is distinct from the elongation step since it relies on the inclusion of water in the peptidyl transferase center; water must be excluded during amide bond formation as seen in elongation. eRF1 may provide a channel for water but it may also undergo a conformational change in the ribosome that permits the entry of water into the PTC, as well as the activation of water and of the ester bond, which results in the hydrolysis of the peptidyl-tRNA (Frolova et al., 1999; Kapp and Lorsch, 2004b). Studies have shown that there might be some communication of the eRF1 and the ribosome. Some mutations of base 1054 in the decoding center of the yeast 18S rRNA may facilitate stop codon read-through, whereas different changes to the same base compensate for the defects that are associated with mutant eRF1 and eRF3 proteins that display omnipotent suppressor phenotypes (Chernoff et al., 1996). rRNA plays an significant role in the maintenance of translation termination efficiency along with its essential role in catalyzing peptide bond formation(Ogle et al., 2002).

1.1.4 Recycling of subunits

The fourth stage of eukaryotic translation is the recycling of ribosomal subunits for the preparation of the next round of translation initiation. Factors that increase the rate of dissociation of the subunits, mRNA and deacylated tRNA are required. eIF6 has dissociation activity and is believed to be involved in the recycling of ribosomes. Recent work has indicated that phosphorylation of eIF6 is involved in regulation of 60S subunit levels in vivo (Ceci et al., 2003).

eIF3 has the most intriguing role in ribosome recycling. It binds to the side of the 40S subunit opposite of the interface (Bommer et al., 1991; Srivastava et al., 1992; Valasek et al., 2003). Its dissociation activity is due to its induction of a conformational change in the 40S subunit rather than prevention of the binding of the 60S subunit. Induction of the conformational change could potentially increase the rate of subunit dissociation and a decrease in the rate of association.

The closed-loop model of eukaryotic mRNAs suggests that translation termination can be followed by the recycling of ribosomes over the poly(A) tail back to the 5'-end of the mRNA. This possibility was bolstered by the finding that there is an interaction between eRF3 and PAB, thus connecting the termination apparatus to the poly(A) tail (Hoshino et al., 1999). Class I RFs, in complex with the ribosome, may act as a GTPase-activator protein (GAP) or a GEF towards class II RFs. The GAP hypothesis states that the C domain of class I eRF1s contains a short sequence (GIRLY) that is crucial for their binding to eRF3 and for the subsequent induction of its GTPase (Frolova et al., 2000; Merkulova et al., 1999). This pentapeptide contains an arginine residue that is flanked by highly conserved hydrophobic residues. This element acts as an arginine finger that is typical for GAPs where the arginine residues play a catalytic role (Scheffzek et al., 1998).

1.2 TRANSLATIONAL CONTROL IN EUKARYOTES

In gene expression, the final step is the translation of mRNA to protein. In prokaryotes, there are specific structures in the mRNA that regulate gene expression. These structures are called riboswitches and they can cause premature termination of transcription or inhibition of translation initiation (Vitreschak et al., 2004). In eukaryotes, the regulation of the translation mechanism is extremely significant since this regulation is used to modulate gene expression in a wide range of biological situations from early embryonic development to cell differentiation and metabolism (Gebauer and Hentze, 2004). Translational control provides a means to control protein synthesis in the absence of transcriptional controls (Matthews et al., 2000). There are two general modes of translational control.

In global control, the translation of most mRNAs in the cell is regulated. This regulation mainly occurs by the modification, specifically phosphorylation of translation initiation factors. In mRNA-specific control, the translation of a specific group of mRNAs is modulated without affecting general protein biosynthesis. It occurs when *cis*-acting elements in the mRNA provide regulatory functions, such as regulatory protein complexes that recognize specific elements present in the 5' and or 3' untranslated

regions (UTRs) of the target mRNA. mRNA translation can also be regulated by small micro RNAs (miRNAs) that hybridize to mRNA sequences that are frequently in the 3' UTR. The specific molecular mechanisms of translational control will be discussed in this section by focusing on examples of both global and mRNA-specific translational control.

1.2.1 Global control

1.2.1.1 eIF2α and kinases. eIF2 is part of the ternary complex that associates with the small ribosomal subunit in a GTP-bound form. The hydrolysis of GTP occurs upon recognition of the initiator AUG, resulting in the eIF2 in the GDP-bound state. This GDP to GTP exchange on eIF2 is catalyzed by eIF2B which is necessary for the reconstitution of a functional ternary complex, for a subsequent round of translation initiation. As discussed previously, eIF2 consists of 3 subunits: alpha, beta, gamma; phosphorylation of the alpha-subunit at residue Ser-51 blocks this GTP-exchange reaction by reducing the dissociation rate of eIF2 from eIF2B.

Numerous kinases can phosphorylate the eIF2alpha at Ser-51 (Dever, 2002). These are: the heme-regulated inhibitor (HR1), which is stimulated by heme-depletion; GCN2 (general control non-derepressible-2), activated by amino acid starvation; PKR a protein kinase activated by double-stranded RNA and stimulated by viral infection and PERK, which is activated by stress in the endplasmic reticulum (ER). Abberant regulation of eIF2α kinase PKR can result in malignant transformation (Clemens, 2004).

1.2.1.2 eIF4E/4E-BPs. The eIF4E protein is crucial for translation initiation by virtue of its role in binding to the 5' cap. Cap-binding is the rate-limiting step and is, therefore, a focal point for global translational control. Regulation of the availability of the cap-binding protein eIF4E is used to control the general rate of translation. eIF4E can be phosphorylated through the p38 mitogen activated protein kinase (MAPK) pathway, with associated increases in translation (Wang et al., 2003). eIF4E interacts with the scaffold protein eIF4G for the 43S recruitment to the mRNA for translation initiation. The interaction between eIF4E and eIF4G requires a small domain in eIF4G that are

shared by a family of proteins called 4E-binding proteins (4E-BPs). Hypophosphorylated 4E-BPs bind to eIF4E and competitively displaces eIF4G, resulting in the inhibition of the 43S association with the mRNA, leading to translational repression. Extracellular cues such as insulin, activates a signaling cascade that triggers hyperphosphorylation of 4E-BP releasing eIF4E making it available to bind to eIF4G, promoting translation initiation. In contrast, nutrient deprivation by contrast will decrease phosphorylation of 4E-BP with an associated increase in binding between eIF4E and 4E-BP with a subsequent decrease in cap-dependent translation (Gingras et al., 1999). Proteolytic cleavage of translation factors also inhibits protein synthesis globally. The protein involved in the process of apoptosis, caspase-3, cleaves eIF4G and PABP (Bushell et al., 2000; Marissen et al., 2004). Cleavage by viral proteases interferes with the translation of cellular mRNAs (Schneider and Mohr, 2003).

1.2.2 Gene-specific control

1.2.2.1 Regulation by 3' UTRs. Translational regulators that function during embryonic development target the formation of the eIF4F complex. The cytoplasmic polyadenylation element binding protein (CPEB) regulates maternal mRNA during translation vertebrate oocyte maturation and early development. CPEB binds to the cytoplasmic polyadenylation element (CPE), which is located in the 3' UTR of target mRNAs, and promotes both silencing of mRNA prior to oocyte maturation and maturation-triggered subsequent cytoplasmic polyadenylation and translational activation (Mendez and Richter, 2001). Translational expression occurs when CPEB binds a protein, Maskin, that contains an eIF4E-binding domain which resembles the eIF4E-binding domain of eIF4G. Either Maskin or an expressed recombinant eIF4E-binding domain can inhibit translation in vivo, suggesting that Maskin competes with eIF4G for eIF4E association (Stebbins-Boaz et al., 1999).

Other regulators function as message-specific 4E-BPs. In *Drosophila melanogaster* embryogenesis, during the anteroposterior axis function, the mRNA that encodes the establishment of the posterior determinant, Nanos is specifically translated. The protein Smaug binds to the 3' UTR of unlocalized Nanos mRNA and represses its

translation by recruiting the eIF4E-binding repressor, Cup. Cup is recruited to the mRNA that encodes the posterior determinant, Oskar, by the RNA binding protein, Bruno, preventing Oskar synthesis (Nakamura et al., 2004; Wilhelm et al., 2003). In summary, Maskin and Cup are examples of regulatory proteins that indirectly associate with specific mRNAs by interaction with specific RNA-binding proteins, such as CPEB and Smaug, blocking eIF4E recognition by eIF4G.

1.2.2.2 Regulation via 5' leaders. Transcript specific regulation is mediated directly or indirectly by sequences found in the 5' and 3' UTRs. In general, the efficient 5' UTRs are relatively short and have a low GC content. Profound secondary structure or length approaching 200 nucleotides will decrease translational efficiency. Examples of specific 5' sequences that exert regulatory function are 5' terminal oligopyrimidine tract (TOP) sequences, internal ribosome entry sites (IRESes) and upstream open reading frames (uORFs). The functions of uORFs are discussed in Section 1.3.

Translation of mammalian mRNAs of ribosomal proteins is regulated in a growth-dependent manner. These mRNAs have short, unstructured 5' UTRs, ~40 nucleotides in length, and contain a 5' TOP consisting of a stretch of 5-14 consecutive pyrimidine residues. 5' TOPs inhibit translation in the absence of activation by $p70^{S6K}$ kinase, a downstream effector of mTOR (mammalian target of Rapamycin) signaling (Jefferies et al., 1997). Translational control of TOP mRNAs enable cells to express rapidly the biosynthesis of the translational machinery upon shortage of amino acids or growth arrest, thus blocking unnecessary energy output. The 5' TOP confers growth and nutrient-dependent expression on TOP mRNAs and thereby provides a mechanism for the coordinated expression of proteins that are required for ribosome biogenesis (Meyuhas, 2000).

Internal initiation was originally identified in picornovirus RNAs bearing very long and structured 5' UTRs (Le and Maizel, 1998). IRESes usually comprise highly structured areas within the 5' UTRs and often feature a polypyrimidine tract near the 3' terminal end (Jackson and Kaminski, 1995). This translation initiation mechanism is independent of recognition of the 5' mRNA and involves the direct recruitment of the 40S ribosomes to the vicinity of the initiation codon. It has been suggested that increased

expression of several proteins occur because of IRES control. These proteins include oncogenes, growth factors and proteins involved in regulation of programmed cell death, whose expression influences the development of numerous pathological conditions such as diabetes (Teshima-Kondo et al., 2004) cardiovascular diseases(Martin et al., 2003), and the development and progression of cancer (Holcik, 2004).

1.2.2.3 RNA-protein interactions. Sex-lethal, Sxl, is a regulator that inhibits the stable association of the 43S ribosomal complex with mRNA in a cap-independent manner. Sx1 prevents X-chromosome dosage compensation in *D. melanogaster* females by translational repression of *msl-2* mRNA, which is a crucial component of the dosage compensation complex. Sx1 binds to specific sites in both the 5' and 3' UTRs of *msl-2* mRNA, which results in recruitment of co-repressors to the 3' UTR (Gebauer et al., 2003; Grskovic et al., 2003). Although both the cap and poly(A) tail contribute to translation of *msl-2* mRNA, regulation is independent of these structures. The stable association of the small ribosomal subunit with the mRNA is affected by the translational repression of Sx1 due to the inhibition of the 48S complexes in the presence of Sxl. This translational mechanism is a molecular basis of a critical step in dosage compensation and a two step principle of translational control via multiple regulatory sites within an mRNA (Beckmann et al., 2005).

RNA-binding protein hnRNPK (heterogeneous nuclear ribonucleoprotein K) and hn RNP E1 inhibit the translation of 15-lipoxygenase (LOX) mRNA during early erythroid differentiation. They bind to a CU-rich repeat which is known as the differentiation-control element (DICE) in the LOX 3' UTR. Translation repression is independent of poly(A) tail; its translation is cap-independent and can occur during encephalomyocarditis virus (EMCV) or classical swine fever virus (CSVF) infection, which reduces translation of most cellular mRNAs. This IRES-regulation of LOX translation targets a late step in initiation. Sucrose-gradient analysis showed that the 80S ribosome was inhibited by hnRNP-K-hnRNP-E1 complex. These regulators prevent the binding of 60S subunit to the 40S subunit at the initiation codon. (Ostareck et al., 2001).

1.2.2.4 Translational control by miRNAs. Regulation of translation is controlled not only by protein factors but also by small RNA molecules of ~22 nts in

length called micro RNAs (miRNAs). Several hundred miRNAs have been described in plants and animals that regulate a broad spectrum of biological processes such as cell metabolism, cell differentiation, cell growth and apoptosis (Ambros, 2003; Enright et al., 2003; Stark et al., 2003). The first miRNAs discovered are *lin-4* and *let-7*, which are critical for regulation of developmental timing in *Caenorhabditis elegans* (Carrington and Ambros, 2003). miRNAs hybridize by incomplete base-pairing to numerous sites in the 3' UTR of target mRNAs. Since the mRNAs are still intact, miRNAs appear to repress translation rather than prevent it by the process of mRNA degradation.

1.3 TRANSLATIONAL CONTROL MEDIATED BY UPSTREAM OPEN READING FRAMES (uORFs)

Earlier studies showed that fewer than 10% of eukaryotic mRNAs contain AUG codons within their transcript leader regions, also referred to as the 5' untranslated region. 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; Suzuki et al., 2000). Presence of upstream AUGs (uAUGs) are the most conserved in the 5' UTR compared to orthologous genes of human and mouse or rat and mouse; their presence is linked to the weaker initiation context of the main AUG or principal AUG (pAUG) (Churbanov et al., 2005; Rogozin et al., 2001).

Upstream AUGs are present in two-thirds of cancer-causing genes or oncogenes and in other genes involved in cellular growth and differentiation (Kozak, 1991; Morris, 1995). The efficiency of translation is controlled by specific elements such as the GC content in the 5' untranslated region (5'UTR), the size of the 5' UTR and the location of the potentially active upstream initiation codons (uAUGs). Upstream AUGs (uAUGs) in the 5' UTR generally decreases the efficiency of initiation at the downstream open reading frame's (ORF) initiation codon. The average 5' UTR length varies from 90 to 210 nt for vertebrate mRNAs. Eleven to 42% of vertebrate mRNAs and 20-48% of human mRNAs contain at least one uAUG (Kozak, 1987; Pesole et al., 2000; Suzuki et al., 2000). The frequent presence of a uAUG within a good initiation context suggests that uAUGs are often involved in translational control. There are two modifications of

the most general ribosome "scanning" model to explain how the main ORFs downstream of uAUGs are translated. The first model called the "leaky" scanning model (Kozak, 1978) occurs when uAUGs are not efficiently recognized by the scanning ribosomes resulting in initiation on the downstream AUG. In the second model, downstream translation occurs when the small ribsosomal subunit (40S) remains bound to the mRNA after encountering the termination codon by the uORF, continues to scan, and then reinitiates. This is an inefficient mechanism and is only possible after translation of a short uORF. Reinitiation can be stimulated by the sequence context surrounding the termination codon (Miller and Hinnebusch, 1989).

1.3.1 Genes regulated by uORFs

A number of uORFs have been identified and investigated. The *Saccharomyces cerevisiae GCN4* mRNA is the best example of translational control via reinitiation (Hinnebusch, 1996; Hinnebusch, 2005). The 590-nucleotide 5' UTR of *GCN4* contains 4 uORFs. Without these uORFs, the GCN4 start codon is used efficiently, despite the long 5'-UTR. uORF1 decreases translation two-fold and uORF4 has shown the strongest effect, inhibiting overall translation. The effects of the other two uORFs are less important since the 5' UTR containing only uORF1 and uORF4 shows similar regulation as wild-type. When the cell is stressed due to amino acid or glucose starvation, general protein synthesis is downregulated due to phosphorylation of the alpha subunit of eIF2 by GCN2 kinase. This phosphorylation inhibits the GDP to GTP exchange, resulting in a halt in protein synthesis.

GCN4 has a reinitiation-dependent mechanism that stimulates *GCN4* translation when eIF2α is phosphorylated. The majority of scanning ribosomes initiate at uAUG1 and the sequences surrounding the termination codon of uORF1 enable ~50% of ribosomes to resume scanning after termination. In contrast, the rare proline codon and the 10 nt 3' to uORF4 are quite efficient in preventing reinitiation after translation of the uORF4. Upon eIF2α phosphorylation, only half of the ribosomal complexes can initiate and terminate at uORF4 (Miller and Hinnebusch, 1989). The first *in vitro* experiments supporting the mechanism of reinitiation were performed by the Sachs lab in collaboration with the Hinnebusch lab that mapped the positions of ribosomes using the

toeprint assay and by using cycloheximide to identify the first initiation events occuring on the transcript as well as initiation events during steady-state translation. It was shown that uORF1 enabled reinitiation and increases in eIF2 levels increased reinitiation at downstream uORFs (Gaba et al., 2001).

Activating transcription factor 4 (ATF4) is a mammalian gene that shows a similar form of uORF-regulation to GCN4. The mouse ATF4 5' UTR (272 nt) contains 2 uORFs, the second one overlapping with the ATF4 ORF (Harding et al., 2000). Translation of ATF4 increases after stress upon eIF2 phosphorylation by a kinase, PERK, and general translation is inhibited. Under this stress condition, translation of ATF4 increases 3-fold. When both of the ATF4 uORFs are present, stress induces a threefold increase in ATF4 translation. Modulation of uAUG1 of GCN4 leads to severe inhibition of translation. In contrast, uAUG1 of ATF4 mRNA has a stimulatory effect, therefore, translational control of ATF4 is not identical with control of GCN4 although has a similar response to eIF2 phosphorylation (Vattem and Wek, 2004).

The human cytomegalovirus early glycoprotein (gpUL4) repression of translation is dependent on the amino acid sequence of the second uORF (Degnin et al., 1993). Several amino acids in the 22-residue peptide are essential for repression (Cao and Geballe, 1996). uORF2 is conserved among several clinical strains of CMV. After translation of uORF2, the ultimate tRNA is not released and the ribosomal complex creates a barrier for upstream scanning 40S subunits. Translational repression by uORF2 is critical for UL4 expression levels; however, the uORF is not essential for viral replication (Alderete et al., 2001). Even though uAUG is recognized inefficiently, once the uORF is translated, the scanning subunits are stalled and gpUL4 ORF is strongly inhibited.

Mammalian S-Adenosylmethionine decarboxylase (AdoMetDC) is a key enzyme in the pathway of polyamine biosynthesis. The uORF in the 5' UTR encodes a hexapeptide with the sequence MAGDIS. The amino acid sequence, especially Asp-4 and Ile-5, are essential for translational control (Mize et al., 1998). The uORF functions as a negative regulatory element in the presence of polyamines and inhibition of translation is relieved by low levels of polyamines (Ruan et al., 1996). Polyamines stabilize an intermediate upon termination much like that with the peptidyl-tRNA-

ribosome complex found for gpUL4 which causes ribosome stalling (Raney et al., 2000). Primer extension inhibition assays confirm that 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). Translational expression of AdoMetDC by the uORF is dependent on cell-type (Hill and Morris, 1992) because in some cell types, such as non-lymphoid cells, initiation at the uAUG located 14 nt from the cap is insufficient.

These previous examples are significant due to their comparison to the *arg-2* AAP which I have extensively studied in this thesis. The *Neurospora crassa arg-2,* contains an evolutionarily conserved uORF encoding a 24 amino acid peptide called the arginine attenuator peptide (AAP). Unlike cytomegalovirus *UL4* uORF2 and in the AdoMetDC uORF, the AAP stalls ribosomes involved either in elongation or termination (Wang et al., 1998) (Fang et al., 2000). Stalling on the uORF is reversible and inhibition of *arg-2* translation is released when Arg is depleted (Wang et al., 1998) which is similar to AdoMet DC's uORF in which translation is relieved by low concentrations of polyamines. Critical amino acids residues in the uORF have been identified by in vivo genetic analysis (Freitag et al., 1996). Changing Asp-12 to Asn (D12N) eliminates arginine-specific negative control in vivo and in vitro. Toeprint assays show that ribosomes stalled around the termination codon but neither the intercistronic region nor a specific termination codon is necessary for *arg-2* regulation (Wang et al., 1998). Regulation is dependent on the peptide sequence and not the RNA coding sequence similar to MAGDIS peptide of AdoMetDC and *UL4* of CMV. Scanning mutagenesis revealed the key residues and the length of the AAP are important for regulation. Comparative analyses with its other fungal homologs and their associated uORF-encoded AAP identified conserved residues that are essential for regulation. In this thesis, Chapter 2 will discuss the details of how the AAP functions as an internal domain. Comparative analyses of the evolutionarily conserved features of the AAP with the other fungal AAP homologs will be discussed in Chapter 3. Finally, the specific requirements for Arg-mediated regulation by the AAP will be discussed in Chapter 4.

1.4 TRANSLATIONAL CONTROL BY NASCENT PEPTIDES

Arabidopsis CGS1 gene codes for cystathionine γ-synthase, an enzyme that catalyzes the first committed step of methionine biosynthesis in higher plants. Expression of the *CGS1* gene is regulated by a negative feedback mechanism at the step of mRNA stability in response to methionine or S-adenosyl-L-methionine (AdoMet). A short stretch stretch of amino acids called the MTO1 region, which is encoded within the first exon of *CGS1* is essential for regulation. This nascent peptide stalls ribosomes in response to AdoMet. The MTO1 region in the nascent polypeptide acts within the ribosomal exit tunnel to cause translation elongation arrest (Onouchi et al., 2005a). The translation of CGS1 mRNA is arrested temporarily at the Ser-94 codon in response to AdoMet after the MTO1 region is translated (Onouchi et al., 2005b). There is a strong correlation between the translation arrest and mRNA degradation. The MTO1 region is the *cis*-element in the post-transcriptional regulation of *CGS1*, and the amino acid changes in the MTO1 region abolish this regulation (Chiba et al., 1999; Ominato et al., 2002).

Tryptophanase is a catabolic enzyme that degrades tryptophan to indole, pyruvate and ammonia, allowing tryptophan to serve as a carbon or nitrogen source (Kazarinoff and Snell, 1977). The *tna* operon of *E. coli* consists of a 319-base pair transcribed leader regulatory region that contains a coding region, TnaC, for the 24-residue leader peptide. Expression of the tryptophanase operon of *E. coli* is regulated by catabolite repression and tryptophan-induced transcription antitermination (Cruz-Vera et al., 2006; Gong and Yanofsky, 2002). Tryptophan induction requires TnaC synthesis with crucial tryptophan residue at position 12 (Gollnick and Yanofsky, 1990; Stewart and Yanofsky, 1985). The narrowest part of the exit tunnel formed with ribosomal proteins L4 and L22 responds to the segment of TnaC containing the crucial residue W12 by altering features of the peptidyltransferase center, creating the tryptophan induction site (Gong and Yanofsky, 2002). Current in vitro findings have established that W12 of TnaC-tRNA(Pro), the peptidyl-tRNA precursor of the leader peptide of this operon, is required for introducing specific changes in the peptidyl transferase center of the ribosome that activate free tryptophan binding. This resulted in peptidyl transferase inhibition. The free tryptophan

exerted its effect at or near the binding sites of several antibiotics in the peptidyl transferase center (Cruz-Vera et al., 2006).

There is an internal polypeptide domain I the prokaryotic regulatory protein, SecM, that can stall elongating ribosomes. In *E. coli*, the Sec protein translocase is responsible for translocation of extracytoplasmic proteins across the plasma membrane. The core machinery of the sec protein translocase is composed of (i) an ATPase, SecA, which is responsible for driving the movement of a preprotein and internal membrane complex, (ii) Sec YEG, which possibly functions as a polypeptide-conducting channel across the membrane (Mori and Ito, 2001). The expression of SecA is subject to specific regulation in response to the protein secretion status of the cell (Oliver and Beckwith, 1982). The regulation of SecA expression is significant and SecM, a product of the first open reading frame of the secA operon, has a crucial role in this regulated mechanism (Schmidt et al., 1988). Detailed mutagenesis studies showed that a sequence FXXXXXWIXXXXGIRAGP, in which X can be any amino acid and the final proline is the position of elongation arrest, is required to cause the arrest in SecM translation elongation. Other arrest-suppressing mutations in other cellular components were isolated (Nakatogawa and Ito, 2002). All the mutations characterized had an alteration in one of the selected 23S rRNA or L22 protein residues that are located close to the narrowest constriction of the exit tunnel. The arrest segment involved may assume a specific configuration within the tunnel, preventing movement along the tunnel or the PTC reaction. SecM embodies new concepts in regulation and gene expression because it is a protein that only functions in the ribosome. The examples of the different nascent peptides mechanisms of translational control helped in the understanding of how the AAP might be regulating translation. The *Neurospora crassa arg-2* uORF-encoded AAP causes ribosomes to stall after translation of the AAP with high concentrations of Arg (Wang et al., 1998). Arginine-specific translational regulation by the AAP is dependent on the amino acid sequence and not the mRNA sequence (Fang et al., 2000). Similar to the MTO1 region of Arabidopsis CGS1, the amino acid of the AAP is essential for regulation. The nascent AAP stalls ribosomes in response to a small molecule (amino acid) Arg, similar to *E. coli* TnaC that regulates translation in response to high concentrations of Trp. Extension of the uORF encoded AAP and its C-terminus resulted

in Arg-mediated ribosomal stalling during translational elongation within the extended region during termination. It is also similar to the MTO1 region of CGS1 in which this nascent peptide stalls in response to AdoMet. This translation elongation arrest is shown to be coupled with mRNA degradation (Chiba et al., 1999; Ominato et al., 2002). The *Saccharomyces cerevisiae* CPA1 AAP, the homolog of *arg-2* stalls ribosomes at the uORF termination codon in response to Arg that resulted in decreased levels of ribosomes at the downstream reporter start codon. The further analyses of CPA1 uORF using an in vivo reporter system with wild-type and nonsense mediated decay (NMD)-deficient strains indicated that uORF-mediated translational control regulated CPA1 mRNA levels by affecting transcript susceptibility to NMD (Gaba et al., 2005).

The AAP can be extended as its N-terminus without affecting its function; this led to the investigation whether the AAP can function when placed as an internal domain in a polypeptide, similar to that of SecM. The AAP exerted regulatory function whether placed near the N-terminus or internally within a large polypeptide. Pulse-chase analysis of the polypeptide products obtained during cell-free translation clearly indicated that the AAP causes regulated stalling of fungal, plant and animal ribosomes in response to Arg and that the peptide-sensing features are conserved in these systems (Fang et al., 2004). Details will be discussed further in Chapter 2.

1.5 TRANSLATIONAL CONTROL AND CANCER

There is growing evidence that suggests a correlation between the regulation of protein synthesis and the disruption of cell behavior that is characteristic of cancer. There are numerous ways in which translational control is relevant in cancer. The expression efficiency of major proteins involved in cell growth regulation, proliferation, or senescence may be controlled at the translational level by alterations in the activity of the protein synthesis machinery. Mutations leading to structural changes of individual mRNAs may alter the rates of protein production. Disruptions in regulation of signaling pathways that result in impairment of growth inhibitory or proapoptotic proteins may alter the balance of production of major components of the cell. Finally, tumor-

associated viruses that infect cells can result in interference with normal cellular controls on translation contributing to transformation.

Proliferative agents affect phosphorylation states of initiation factors. Phosphorylation of of the α subunit of eIF2 leads to strong inhibition of translation, through competitive binding to eIF2B. The increase in phosphorylation of eIF4E and availability of eIF4E results in a decrease in association with 4E-BPs. This change in initiation factor phosphorylation is due to various aspects of the signal transduction pathway (Kleijn et al., 1998). The phosphorylation of eIF2α alters protein synthesis in variety of stress conditions (Dever, 2002; Sonenberg et al., 2000). Both growth factor-activated MAP kinase pathway and stress-activated p38 pathway can result in Mnk activation and eIF4E phosphorylation (Waskiewicz et al., 1997). The PI3-kinase/mTOR cascade is responsible for phosphorylation of the 4E-BPs.

In many cancers, general protein synthesis rates and expression of several translation components are elevated, supporting the potential significance of translational control in tumor progression (Holland et al., 2004). Increased protein synthesis as a result of influences from mitogens, hormones, and growth factors is linked to increased eIF4F formation for highly structured (GC-rich) 5' UTRs (Koromilas et al., 1992). An increase in eIF4F formation could contribute to the unwinding of these specific mRNAs, facilitating ribosome binding. The phosphoinositide 3-kinase (PI3-kinase) signaling pathway (PI3K-Akt-mTOR) links the eIF4F complex formation and the 4E-BPs to cancer (Hidalgo and Rowinsky, 2000). This pathway regulates cell decisions such as growth, division, survival, and migration (Stokoe, 2005). Overexpression of eIF4E promotes malignant transformation in rodent fibroblasts (Lazaris-Karatzas et al., 1990) as well as in vivo in transgenic mice (Ruggero et al., 2004). Overexpression of eIF4E (Lazaris-Karatzas et al., 1990) and eIF4G (Fukuchi-Shimogori et al., 1997) has been shown to result in transformation of NIH-3T3 cells. Other studies showed that dominant negative forms of eIF2α protein kinase PKR or non-phosphorylatable mutant forms of eIF2α are observed in transformation (Donzé et al., 1995; Koromilas et al., 1992; Meurs et al., 1993). The assumption of which the mechanism by which transformation occurs is by elevated expression of key growth-stimulatory proteins, rather than just increased rates of overall protein synthesis. eIF4E levels are elevated in tumor cell lines with the most

pronounced increases observed in breast cancer(Kerekatte et al., 1995), and head and neck squamous cell carcinomas (Nathan et al., 1997). eIF4E is a suitable tumor marker in breast cancer. The gene amplification of eIF4G is seen in squamous cell lung carcinoma (Bauer et al., 2002).

1.5.1 Functional uORFs in proto-oncogenes

The 5' UTRs often contain GC-rich (70-90%) regions, indicating a high degree of secondary structure. There are elements within this 5' UTR that mediate translational control. Upstream open reading frames (uORFs) and internal ribosome entry sites IRESes are among these elements. Translation of mRNAs containing multiple uORFs is enhanced in response to the phosphorylation of the alpha-subunit of eIF2. These uORFs are often inhibitory to the downstream initiation codon (Shantz and Pegg, 1999). mRNAs with long leader sequences may contain additional upstream initiation codons that are not AUG codons. CUG is the most common alternative initiation codon (Hann, 1994). Presence of internal ribosome entry sites (IRESes), originally identified in picornavirus RNAs, comprise highly structured areas within the 5' UTRs and often have a polypyrimidine tract near the 3'-terminal end (Jackson and Kaminski, 1995). An IRES also allows preferential translation upon phosphorylation of eIF2α. The cellular growth-promoting genes have recently been found to contain IRESes (Willis, 1999). These translational elements are commonly seen in proto-oncogenes which are under specific control.

The uORF is significant due to its inhibitory effect on the translation of the downstream coding region. Single and multiple uORFs are found in mammalian genes, particularly those that encode oncoproteins, growth factors and growth receptors. An example is the MDM2 protein, which plays an important role in controlling the tumor suppressor p53 and other proteins in cell growth (Daujat et al., 2001; Michael and Oren, 2002; Momand et al., 2000). The oncogene *mdm2* is highly regulated and the protein *mdm2* mRNA contains two uORFs that occur in the long form in both human and mouse. These two uORFs together have a major impact on translation, however uORF2 has the weaker effect and the coding sequence of uORF1 contributes to its inhibitory influence. In human uORF2, changing residue 9 to valine reduced the inhibitory activity of human

uORF1 (Jin et al., 2003). The presence of an active IRES would circumvent the inhibitory activities of the two uORFs. The occurrence of the two uORFs with differing activities in both species suggests that these two elements play a fundamental role in regulating expression of the *mdm2* gene.

BCL-2 is a key inhibitor of intrinsic apoptotic signaling. It seems to regulate the cell cycle (Greider et al., 2002) and prevents a constitutive pro-apoptotic signal from p53 by direct binding to the transcription factor (Jiang and Milner, 2003). Its primary transcript contains a 1.45 kb 5' UTR including 10 upstream AUGs that may possibly restrict translation initiation via cap-dependent ribosome scanning. In eukaryotes, IRES-mediated translational initiation has been frequently observed for mRNAs that possess long and thermodynamically stable 5' UTRs with multiple potential uORFs, features that can dramatically inhibit scanning-dependent translation initiation. The 5' UTR of BCL-2 mRNA promotes the internal entry of ribosomes both in vitro (using RRL) and in vivo (Sherrill et al., 2004).

Tie2 is an endothelium-specific receptor tyrosine kinase required for normal blood vessel maturation. Angiogenesis is an essential step in allowing tumors to grow beyond 1-2mm in diameter (Risau, 1997). Angiopoietin-1 (Ang-1) is an agonist of endothelial cell tyrosine kinase receptor Tie1/Tek (Suri et al., 1996). Studies with Tie2 null mice indicate that the angiopoietin/Tie2 signaling system plays a role in the later steps of angiogenesis (Dumont et al., 1994). Tie2 expression is elevated in human cancers (breast, ovarian, hepatocellular, glioblastoma) with expression being localized in "vascular hot spots" at the leading edge of invasive tumors (Peters et al., 2004). Studies showed that the presence of an IRES within the Tie2 5'-UTR explains how this mRNA is able to escape the reduction in translation imposed by hypoxia. The Tie2 mRNA 5'UTR contains 5 uORFs and when both DNA and RNA transfections were performed, the expression of the reporter firefly LUC is increased when all AUGs are deleted. The interpretation was that some or all of the uORFs reside between the IRES and initiator AUG and are encountered by the ribosome (Park et al., 2005).

The *her-2* (neu, ErbB2) oncogene encodes a 185-kDa transmembrane receptor tyrosine kinase. The overexpression of the protein HER2 occurs in many primary human tumors and contributes to 25-30% of breast and ovarian carcinomas. The synthesis of

Her2 is controlled by a uORF present on its transcript. Previous studies had revealed that the well-conserved uORF in *her-2* controlled its expression at the level of translation (Child et al., 1999a). When fused to a reporter gene, the uORF inhibits downstream translation in multiple cell types (Child et al., 1999a). The uORF reduced translation of downstream cistron in in vitro cell-free translation systems, in vivo in HeLa cells, and was shown to affect translation start site selection. I will be discussing details of my work with *her-2* in Chapter 5. A recent finding showed that the 3' UTR relieves inhibition in specific tumor cells. The possibility that the stimulatory effect of *her-2* 3'-UTR is clinically relevant in tumors expressing high levels of HER2. (Mehta et al., 2006). These examples of proto-oncogenes are significant in showing the significance of 5' UTR elements such as uORFs and IRESes and their specific involvement in translational control.

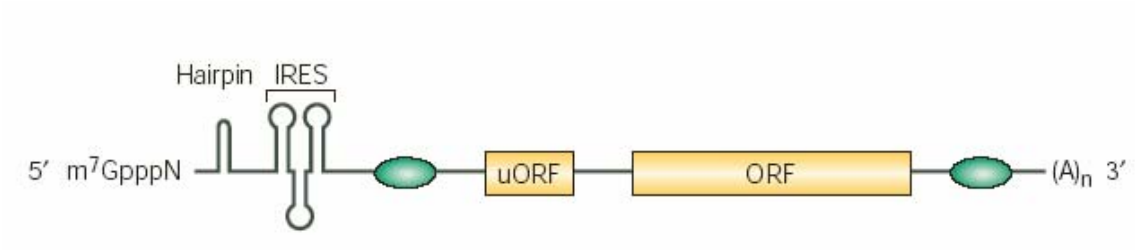

Figure 1.1 The cap structure at the 5' end of the mRNA, and the poly(A) tail at the 3' end, are canonical motifs that strongly promote translation initiation. Secondary structures, such as hairpins, block translation. Internal ribosome entry sequences (IRESs) mediate cap-dependent translation. Upstream open reading frames (uORFs) normally function as negative regulators by reducing translation from the main ORF. Adapted by permission from MacMillan Publishers Ltd (Gebauer and Hentze, 2004).

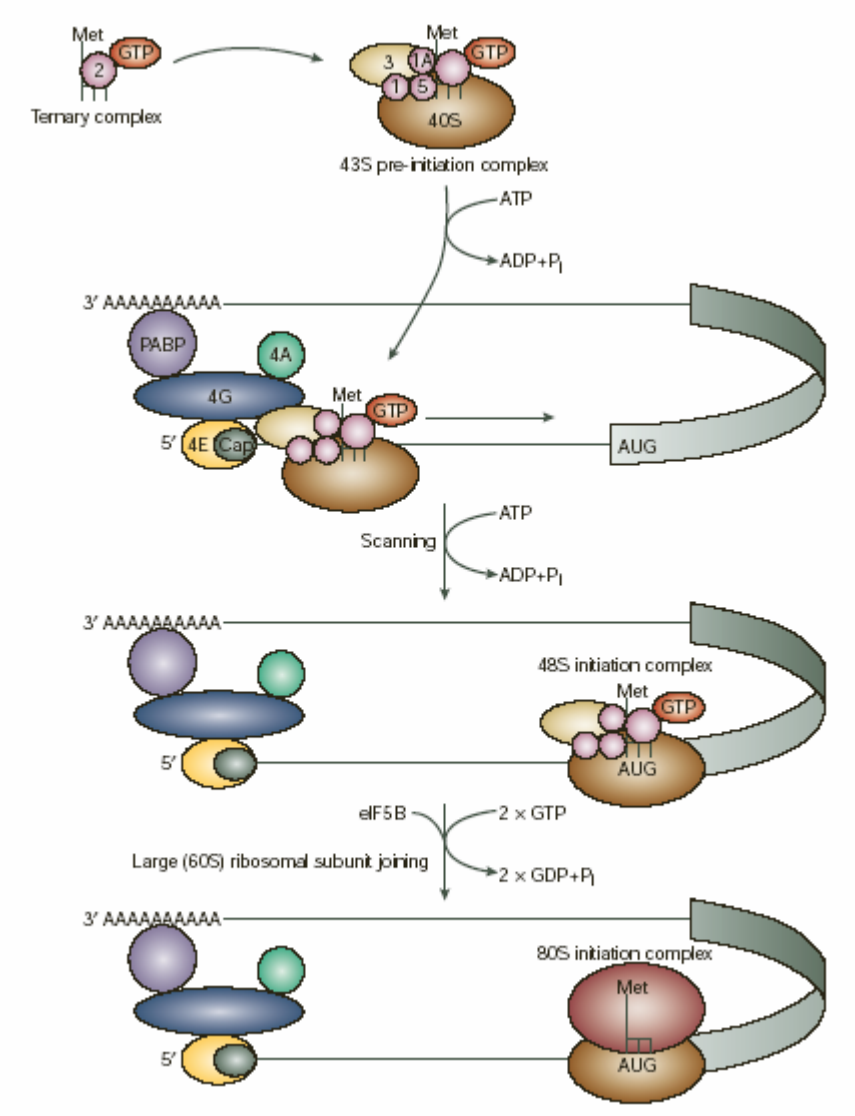

Figure 1.2 Cap-mediated translation initiation. Adapted by persmission from MacMillan Publishers Ltd (Gebauer and Hentze, 2004).

CHAPTER 2
A NASCENT POLYPEPTIDE DOMAIN THAT CAN REGULATE TRANSLATION ELONGATION[1]

2.1 INTRODUCTION

Nascent polypeptides can control translation. Signal peptides that direct polypeptides to the ER associate with the signal recognition particle to halt translation elongation until the nascent peptide docks with the ER (Keenan et al., 2001; Nagai et al., 2003). In addition, a variety of peptides specified by upstream open reading frames (uORFs) in eukaryotic and prokaryotic mRNAs can stall ribosomes involved in translation termination (Lovett and Rogers, 1996; Morris and Geballe, 2000; Raney et al., 2002; Sachs and Geballe, 2002; Tenson and Ehrenberg, 2002; Vilela and McCarthy, 2003). Expression of the small subunit of the arginine-specific carbamoyl phosphate synthetase, a fungal arginine (Arg) biosynthetic enzyme, is negatively regulated at the translational level. This is accomplished through the synthesis and/or action of the evolutionarily conserved, uORF-encoded Arg attenuator peptide (AAP). The nascent AAP normally causes ribosomes to stall at the uORF termination codon in response to Arg, thereby blocking the translating ribosome from reaching the initiation codon used for synthesis of the downstream enzyme (Gaba et al., 2001). Mutations that eliminate Arg-specific regulation of *Neurospora crassa arg-2* and *Saccharomyces cerevisiae CPA1* change a conserved Asp-residue, at positions 12 and 13 in each AAP, respectively, to Asn (Freitag et al., 1996). These mutations also abolish each AAP's capacity to stall ribosomes in fungal cell-free translation systems (Wang et al., 1999; Wang and Sachs,

---

[1] This material has been published in this or similar form in 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.*. **101**, 4059-4064.

1997b) . Unlike other uORF-encoded peptides that affect only translation termination, the AAP amino acid sequence allows Arg-regulated ribosome stalling when placed within a polypeptide, at its N-terminus (Wang et al., 1998). Stalling occurs during elongation, immediately downstream of the AAP coding region, and is independent of the sequence in the downstream region (Fang et al., 2000).

Most known nascent peptides which regulate translation are found encoded as uORFs or as N-terminal leader peptides. However, an internal polypeptide domain in the prokaryotic regulatory protein SecM can stall elongation. Studies on prokaryotic ribosomes synthesizing SecM indicate that the exit tunnel acts as a discriminating gate that permits regulation of polypeptide chain elongation as a consequence of the sequence of the nascent SecM chain (Berisio et al., 2003; Nakatogawa and Ito, 2001; Nakatogawa and Ito, 2002; Sarker and Oliver, 2002).

Could a general mechanism governing translation elongation enable internal domains within a nascent polypeptide chain to regulate completion of translation in response to a small molecule, and could such a mechanism regulate eukaryotic protein synthesis? The characteristics of the AAP suggested that it may provide such functions. To test this possibility we created large polypeptide coding sequences with the AAP coding sequence near the coding sequence for the N-terminus, or internally within the coding sequence. To facilitate protein detection, we removed Met-residues from our proteins except at the extreme N-terminus. Synthetic transcripts specifying these polypeptides were used to program fungal, plant, and mammalian cell-free translation systems. Polypeptide synthesis was monitored by pulse-chase analyses; the appearance of stalled peptidyl-tRNA intermediates was monitored by their ability to be precipitated with cetyltrimethylammonium bromide. The positions of ribosomes stalled on transcripts during translation were monitored by primer extension inhibition assay. The results of these studies indicated that an internally localized AAP domain does cause a translating ribosome to stall in response to Arg. This establishes that an internal nascent polypeptide domain can function as a *cis*-acting regulator of polypeptide elongation by modulating ribosome translation in response to changes in the concentration of a small molecule.

2.2 RESULTS

2.2.1 Pulse-chase analyses of radiolabeled polypeptide synthesis

We examined whether the AAP caused stalling as an internal polypeptide domain by performing pulse-chase experiments with mRNAs encoding fused polypeptide sequences in cell-free translation systems. Synthetic mRNAs were prepared that specified polypeptides containing a string of nine methionine residues at their N-termini and no internally positioned methionine residues (Figures 2.1A and 5). These polypeptides contained a domain from rabbit α-globin and a domain from firefly luciferase (LUC), or contained duplicated LUC domains. AAP domains were placed at two positions in each polypeptide: near the N-terminus, following the nine Met-residues (residues 12-33), and internally, between the globin and LUC domains (residues 101-123) or between the duplicated LUC domains (residues 131-153). We constructed all possible combinations containing the wild-type AAP (designated AAP_w) sequence, or a non-functional AAP sequence (AAP_m, in which the critical residue corresponding to *arg-2* AAP Asp-12 was substituted by Asn), at N-termini and internal positions.

Synthetic mRNAs were used to program cell-free translation extracts containing [^{35}S]Met, Arg at low (10 μM) or high (2000 μM) concentrations , and the other amino acids at fixed concentrations. Since Met-codons were only located at the N-termini coding regions, radiolabeled translation products would contain isotope only at their N termini. In our pulse-chase experiments, edeine was added to extracts after 2 min of incubation to block subsequent rounds of translation initiation during prolonged incubation. Nascent polypeptide N-termini were thus labeled to high specific activity with [^{35}S]Met during incubation prior to edeine addition. Polypeptide chain elongation progress was monitored by SDS-PAGE; this revealed the length of radiolabeled translation products as a function of time.

Programming *N. crassa* extracts with mRNA specifying polypeptides containing AAP_m domains at two locations showed that neither stalled polypeptide synthesis (Figures 2.1B and C). In either low or high Arg, full-length polypeptide synthesis was essentially complete by 6 min and no major intermediate length polypeptide was detected

(the tRNA was removed from the nascent peptide as a consequence of the analysis procedure). In contrast, with mRNA containing two AAP_w encoding regions, two intermediates in polypeptide synthesis were observed in the presence of high Arg (Figure 2.1E). The smaller (labeled N) migrated with a size consistent with it being a product of stalling after synthesis of the N-terminal AAP; the larger (labeled I) migrated with a size corresponding to a product formed by translational stalling after synthesis of the internal AAP. Synthesis of the full-length polypeptide (labeled F) was substantially delayed in high Arg, also indicative of translational stalling. 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 (Figure 2.6). The AAP_w near the N-terminus did not stall polypeptide synthesis in low Arg; the internal AAP_w caused some stalling in low Arg, but substantially more stalling in high Arg (Figure 2.1D).

2.2.2 Polypeptide synthesis appears to resume after stalling

If the short, radiolabeled products were intermediates resulting from stalled translation, they should be present as peptidyl-tRNAs. Also, if they were intermediates, then the radiolabeled methionine they contain should be quantitatively recovered in full-length polypeptide at later time-points in pulse-chase experiments. Therefore, we performed experiments with longer incubation periods and removed samples at different time-points for direct analysis by SDS-PAGE (Figure 2.2A) and following precipitation with cetyltrimethylammonium bromide (CTAB) (Figure 2.2B). CTAB selectively precipitates peptidyl-tRNA (Gilmore and Blobel, 1985). The radiolabeled polypeptides identified as intermediates of stalled translation (N and I) were selectively enriched by precipitation with CTAB compared to the levels of full-length polypeptide product (F) detected (compare Figure 2.2 A, B). This indicates that the intermediates were peptidyl-tRNAs, confirming that the shorter translation products are intermediates resulting from stalled polypeptide synthesis. Quantitative analysis of radiolabel in the major translation products observed between 3 and 30 min incubation (data from the experiment in Figure 2.2A and an independent replicate) showed radiolabel conservation consistent with the conversion of the small N-terminal product to a full-length polypeptide, with transient accumulation of a labeled intermediate resulting from stalled translation following the

internal AAP coding region (Figure 2.2C). These experiments indicate that the stalled products are true intermediates and that ribosomes resume chain elongation after stalling.

2.2.3 AAP-mediated ribosome stalling is observed using a toeprint assay

A primer extension inhibition (toeprint) assay (Sachs et al., 2002) was used to directly map the positions of the stalled ribosomes. The purpose was to verify that the translation intermediates observed by [^{35}S]Met-labeling arose from ribosome stalling as a consequence of synthesizing each AAP_w sequence. Previous studies showed that the ribosome involved in elongation that had synthesized an N-terminal AAP_w stalled on the mRNA when the final codon of the AAP coding sequence occupied the P-site of the translating ribosome (Fang et al., 2000; Wang et al., 1998). Therefore, ribosomes stalled by the N-terminal AAP_w or by the internal AAP_w would be expected to generate toeprint signals in a region starting ≈16-nt from the AAP's coding region's final GCG (Ala)-codon (Fang et al., 2000; Wang et al., 1998).

The results of toeprint analyses of translation reactions that were programmed with transcripts containing N-terminal and internal AAP_w domains (Figure 2.3) showed that ribosomes underwent Arg-dependent stalling at either position but did not do so in reactions programmed with AAP_m-encoding mRNAs. The toeprint data indicate that stalling caused by translation of the AAP coding region increases as the Arg-concentration is increased. Pulse-chase experiments of radiolabeled polypeptides and toeprint analyses of extracts treated with edeine indicate that increasing the Arg concentration increases the intermediate's half-life (data not shown), consistent with the increased signal observed in these experiments.

Mapping the positions of the toeprint bands arising as a consequence of Arg-mediated regulation by AAP_w domains revealed that the ribosomes stalled in a region beginning approximately ≈16-nt from each AAP's final GCG codon. Stalling of ribosomes mediated by the internal AAP appeared to occur within a narrower region downstream of the final AAP-codon than observed for the AAP near the N-terminus. While the reason for this difference is not known, these data are consistent with a model that stalling begins when the final amino acid-specifying codon of the AAP is in the ribosome P-site, as is observed when the AAP is encoded as a uORF. Irrespective of how

precisely the position of the AAP coding region can be placed with regard to the translating ribosome, it is clear from the toeprint data that translation of the AAP sequence causes ribosomes to stall on the mRNA at sites consistent with those expected from the sizes of the nascent polypeptide intermediates observed accumulating in high Arg in [^{35}S]Met pulse-chase experiments (Figure 2.1 and 2.2).

2.2.4 AAP can act in different contexts and can regulate polypeptide synthesis in heterologous systems

Might the nascent globin domain, chosen for the experiments described above because it naturally lacks in-frame Met-codons(Drabkin et al., 1993), coincidentally provide a special sequence context that allows internal stalling (Tenson and Ehrenberg, 2002)? To test this possibility, we replaced the globin domain with a LUC domain and repeated the above experiments Figure 2.4A). Pulse-labeling experiments using *N. crassa* extracts showed that AAP_w but not AAP_m functioned to stall ribosomes in response to Arg when placed at the N-terminus or internally in the LUC polypeptide (data not shown). Thus, the internal AAP sequence appears to function independently of the proximal nascent chain sequence.

Is the ability of AAP_w to cause stalling limited to fungal ribosomes? To test this we programmed wheat germ extracts and rabbit reticulocyte lysates with transcripts specifying dual LUC domains and performed pulse-chase analyses of the [^{35}S]Met products obtained in reaction mixtures containing low or high Arg (Figure 2.4). AAP_w caused Arg-regulated stalling in both systems, whether at the N-terminal or internal position. AAP_m did not cause stalling at either position and stalling at the internal AAP_w in low Arg was not detectable. The nascent wild-type AAP thus can function within ribosomes from fungi, plants and animals, and cause them to stall in response to Arg.

2.3 DISCUSSION

Several lines of evidence support the conclusion that the AAPcan function as an internal nascent peptide domain that causes regulated stalling of eukaryotic ribosomes in response to Arg. Pulse-chase analysis of the polypeptide products obtained during cell-

free translation directly indicated that the AAP causes a regulated pause in polypeptide synthesis (Figures 2.1 and 2.4). The shorter polypeptides which transiently accumulated as a consequence of AAP-mediated stalling appeared to be intermediates in polypeptide synthesis. The shorter polypeptides were present in the reaction mixture in the form of peptidyl-tRNA (Figure 2.2), which is appropriate, if they are translation intermediates. Furthermore, these shorter peptidyl-tRNAs appeared to be authentic intermediates because their radiolabeling was quantitatively chased into full-length polypeptides. This would not occur if the shorter peptidyl-tRNA intermediates were released from the ribosome, as happens in another nascent peptide-mediated stalling event, at a termination codon (Cao and Geballe, 1998).

We previously established that the amino acid sequence of the AAP, not the sequence of the its coding region, was responsible for regulation, and that regulation was effected by a high concentration of Arg, not aminoacylated Arg-tRNA$_{Arg}$ (Gaba et al., 2001) (Fang et al., 2000; Wang et al., 1998; Wang et al., 1999; Wang and Sachs, 1997b). Here we provided evidence that the AAP was a portable signal that caused stalling when placed either upstream or downstream of two different domains (derived from rabbit α-globin or firefly luciferase). Importantly, the AAP sequence functioned in each position to pause polypeptide synthesis in response to Arg in cell free translation systems derived from *N. crassa*, wheat germ, and rabbit reticulocytes. Mutation of the critical Asp12 residue of the AAP to Asn eliminated stalling during synthesis of the nascent polypeptide in each system tested, indicating that the nascent polypeptide and Arg were acting similarly in each system to cause stalling.

At least two models are suggested by these data that could account for how the AAP causes stalled polypeptide synthesis. The AAP and Arg might act in concert to constrain the movement of the nascent peptide in the ribosome tunnel. Elongation would slow or pause because the inability of the peptide to move in the tunnel would slow peptidyl transferase activity or translocation of the peptide in the translating ribosome. Normal translation would resume when the constraints on AAP movement in the tunnel ceased. Alternatively, AAP and Arg might directly affect peptidyl transferase activity by interfering with this domain of the ribosome.

The toeprinting data obtained with a variety of AAP sequences in different contexts, including the internal domain context examined here, indicate that stalling occurs when the AAP's carboxyl-terminal residue is at or near the ribosomal P-site. The critical Asp-12 and Asp-13 residues of the *N. crassa* and *S. cerevisiae* AAPs, established from both *in vivo* and *in vitro* studies to be critical for regulation, would each be located in the ribosome tunnel approximately 12 amino acids from the residue at the P-site. Interestingly, amino acids in the polypeptide chain which are critical for stalling prokaryotic ribosomes during the synthesis of SecM and TnaC are also located in the ribosome tunnel approximately 12 residues from the P-site. This region of the nascent polypeptide appears to interact with ribosomal protein L22 in the tunnel at a constriction point which has been called a "discriminating gate" (Berisio et al., 2003; Jenni and Ban, 2003; Nakatogawa and Ito, 2002). These data suggest that the eukaryotic ribosome has a similar "constriction gate" in its tunnel.

An observation that still requires explanation is that toeprints appear in a relatively wide region when the AAP is near the N-terminus (Figure 2.3), while the stalled radiolabeled polypeptide product appears to be a relatively discrete species (Figures 2.1 and 2.2). Direct comparisons of different AAP-containing mRNAs translated in *N. crassa* and *S. cerevisiae* cell-free systems showed that the most 5'-proximal toeprint could correspond to ribosomes with the final codon of the AAP in the ribosome P-site, as observed here, and that the distribution of 3'-distal toeprints was extract-dependent (Wang et al., 1999). One possible explanation that could account for these data is that additional factors are recruited to the stalled ribosome and that these factors result in additional toeprints 3'-distal to the ribosome.

How Arg interacts with the nascent peptide and/or the translational machinery to cause stalling remains unclear. Studies on the *Escherichia coli* leader peptide TnaC, which causes prokaryotic ribosomes to stall in response to tryptophan, indicate that free tryptophan exerts its stalling action by occupying the ribosomal A-site (Gong and Yanofsky, 2002). Replacement of the stop-codon (at which TnaC normally causes regulated stalling in response to Trp) with a Trp-codon causes constitutive stalling. The aminoacylated tRNA appears to place Trp in the proper spot to exert its regulatory effect. Therefore we placed an Arg-codon directly after the AAP and tested whether the direct

placement of Arg into the A-site by aminoacylated arginyl-tRNA caused stalling when the AAP was in the P-site (data not shown). We observed no effect on stalling in the *N. crassa* system (*i. e.*, stalling still required high Arg). Thus, Arg and AAP do not appear to act in a manner analogous to Trp and TnaC, in causing stalling.

The demonstration that the AAP functions as an internal domain to regulate elongation in response to Arg, establishes that such domains can provide a means of controlling translational elongation. There are no previous examples of the regulation of elongation by internal nascent polypeptide domains acting in concert with small molecules in either eukaryotes or prokaryotes, although such regulation by uORF-encoded nascent peptides is well established (Geballe and Sachs, 2000; Lovett and Rogers, 1996; Morris and Geballe, 2000; Raney et al., 2002; Sachs and Geballe, 2002; Tenson and Ehrenberg, 2002; Vilela and McCarthy, 2003).

How could the regulation of pausing by an internal domain and a small molecule contribute to the control of gene expression? Constitutive pausing during polypeptide synthesis has been proposed to lead to the formation of distinct intermediates that could contribute to the proper cotranslational binding of cofactors to the nascent polypeptide (Kim et al., 1991). The nascent polypeptide sequence at the N-terminus of rhodanese may affect release of nascent chain from the ribosome and thereby affect folding (Kudlicki et al., 1995); pause sites are also observed during chloramphenicol acetyl transferase synthesis, and these can be altered by modifying the N-terminus of the nascent polypeptide(Tsalkova et al., 1999). The regulated instability of the *Arabidopsis thaliana CGS1* transcript in response to the availability of S-adenosyl methionine requires the synthesis of a specific internal polypeptide domain. Translational stalling as a consequence of synthesis of this domain is a possible explanation of this regulatory effect (Chiba et al., 2003; Lambein et al., 2003). It has long been known that the occurrence of rare codons in polypeptide coding sequences slows elongation (Varenne et al., 1984). 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). 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 could potentially influence protein folding, mRNA stability, or frameshifting.

The data obtained concerning the regulation of translation by the AAP and Arg are consistent with the hypothesis that Arg acts by binding to the nascent peptide in the ribosomal exit tunnel to cause a conformational change that controls polypeptide synthesis. Alternatively, Arg might interact with the ribosome or associated factor to render the ribosome sensitive to the nascent peptide sequence. As noted above, tryptophan may directly occupy the ribosome A-site to act with the *E. coli* TnaC peptide as a regulatory element. Other small molecules are capable of acting in concert with nascent leader peptides to control translational events (Geballe and Sachs, 2000; Lovett and Rogers, 1996; Morris and Geballe, 2000; Raney et al., 2002; Sachs and Geballe, 2002; Tenson and Ehrenberg, 2002; Vilela and McCarthy, 2003). The leader peptides specified in bacterial chloramphenicol acetyl transferase transcripts can bind to and alter the conformation of 23S rRNA; in the presence of chloramphenicol, their synthesis stalls ribosomes (Harrod and Lovett, 1997). Polyamines act in concert with a uORF-encoded peptide in the 5'-leader of the transcript specifying mammalian S-adenosyl methionine decarboxylase to stall ribosomes at the uORF termination codon, negatively regulating gene expression (Law et al., 2001). Small molecules, including amino acids, have also been shown to interact directly with mRNA to exert a regulatory role. Lysine binds to the leader region of the *Bacillus subtilis LysC* transcript to regulate translation (Grundy et al., 2003) (Sudarsan et al., 2003). S-Adenosyl-methionine similarly affects regulation by binding to mRNA (McDaniel et al., 2003). Studies on the control of protein synthesis by small molecules are thus revealing that they can act in unanticipated ways to control gene expression. The control of fungal, plant and animal translation by AAP and Arg would seem to provide such an example of how a polypeptide and an amino acid can affect eukaryotic gene expression.

2.4 EXPERIMENTAL PROCEDURES

2.4.1 Constucts

The plasmids, used are listed in Table 2.1. They were derived from previous constructs using described procedures (Fang et al., 2002). Site-specific mutagenesis was used to remove every ATG codon (except for the 9 at the N-termini) in the three forward reading frames (Figure 2.5). The rabbit α-globin domain used was obtained by PCR from plasmid pSPα (Jobling and Gehrke, 1987) (from U.L. RajBhandary, MIT). Plasmid DNA templates were purified by equilibrium centrifugation (Wang and Sachs, 1997a) or by Promega Wizard midi-prep; templates were linearized with *Eco*RI.

2.4.2 Preparation of RNA and cell-free translation

Capped, polyadenylated RNA was synthesized with T7 RNA polymerase from linearized plasmid DNA templates and the yield of RNA was quantified as described (Wang and Sachs, 1997a). The reaction conditions for *in vitro* translation using *N. crassa* extracts were as described (Fang et al., 2000). Micrococcal-nuclease treated wheat germ extract and reticulocyte lysate were obtained from Promega and used according to the supplier's directions except as noted. Translation reaction mixtures were programmed at a final concentration of 6 ng/μl RNA; [^{35}S]Met was used at a final concentration of 0.5 μCi/μl.

2.4.3 Primer extension inhibition (toeprint) assays

Toeprint assays were accomplished as described (Wang and Sachs, 1997b) using primers FP93 (CTGGC GACGT AATCC ACG) and FP94 (CTTGT CCAGG GAGGC GTG); 8 μl of sample instead of 4 μl was loaded onto each gel lane. The gels were dried and exposed to screens of a Molecular Dynamics PhosphorImager for approximately 24 hours. All toeprinting data shown were representative of multiple experiments.

Table 2.1 Constructs used

Construct	Structure
pGL201	Met_9-AAP_w-globin-AAP_w-LUC
pGL202	Met_9-AAP_m-globin-AAP_w-LUC
pGL203	Met_9-AAP_w-globin-AAP_m-LUC
pGL204	Met_9-AAP_m-globin-AAP_m-LUC
pLL301	Met_9-AAP_w-LUC-AAP_w-LUC
pLL302	Met_9-AAP_m-LUC-AAP_w-LUC
pLL303	Met_9-AAP_w-LUC-AAP_m-LUC
pLL304	Met_9-AAP_m-LUC-AAP_m-LUC

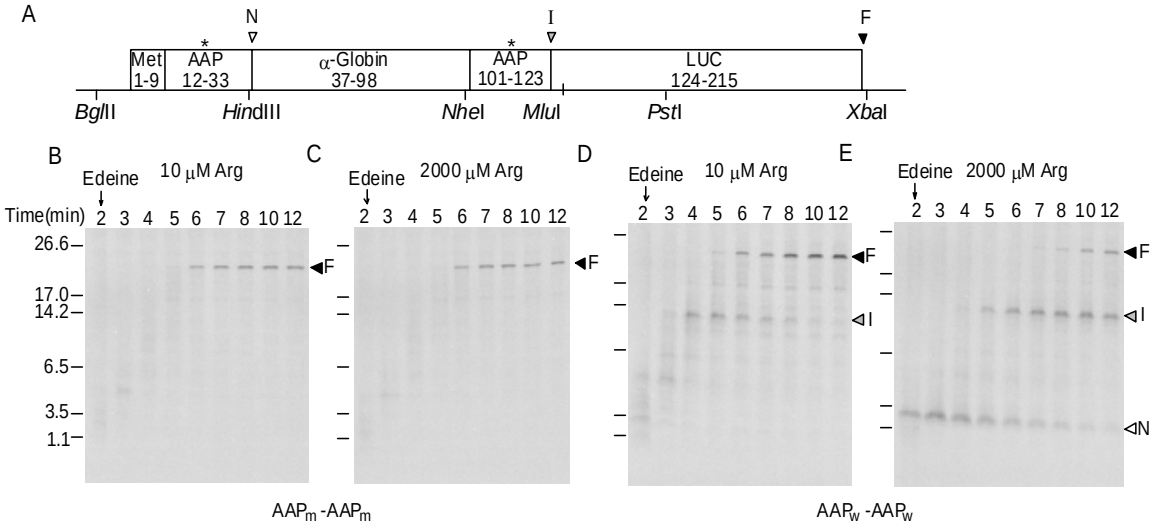

Figure 2.1 Polypeptide synthesis time-course in *N. crassa* cell-free extracts. (A) The Met_9-AAP-globin-AAP-LUC construct used (DNA sequence in Figure 2.5A. Asterisks indicate where wild-type AAP Asp codons were changed to Asn. Arrowhead *N* indicates the C-terminus of the Met_9-AAP polypeptide-intermediate; arrowhead *I*, the C-termini of the Met_9-AAP-globin-AAP intermediate; arrowhead *F*, the C-terminus of the completed polypeptide. Unique restriction enzyme sites are indicated. Transcripts specifying Met_9-AAP_m-LUC-AAP_m-LUC (B,C) or Met_9-AAP_w-LUC-AAP_w-LUC (D,E) were translated in extracts in low (B,D) or high Arg (C,E) (see text). Edeine was added at 2 min (arrow) and 10-µl aliquots of extracts were removed at the indicated time-points for analysis by SDS-PAGE (Fang et al., 2002). Arrowhead *N* indicates the intermediate Met_9-AAP; *I*, the polypeptide-intermediate Met_9-AAP-globin-AAP; *F*. the full-length polypeptide.

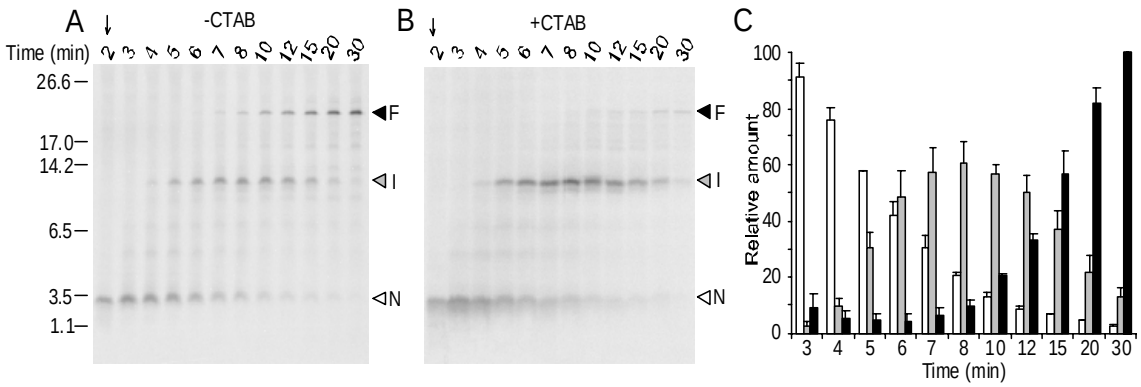

Figure 2.2 CTAB precipitation of peptidyl-tRNA from translation extracts and quantitative analysis of polypeptide intermediates and products. *N. crassa* extracts (150 µl) were programmed with Met_9-AAP_w-globin-AAP_w-LUC mRNA and incubated with high Arg as described in Figure 1.1. Edeine was added at 2 min and 10-µl aliquots removed at the indicated time points for analyses. (A) Total translation product analysis. (B) CTAB-precipitated translation product analysis. Arrowheads as in Figure 1.1. (C) Quantitative analysis of translation products obtained from data in panel A and an independent experimental replicate. The radiolabel in bands N, I, and F was determined using ImageQuant® 5.1 (Molecular Dynamics). The amount of radiolabel in band F at 30 min in each experiment was normalized to 100% and radiolabel in each band at each time-point calculated as a fraction of this value. White, radiolabel in intermediate N ; gray, radiolabel in intermediate I; black, radiolabel in full-length polypeptide F. The total (not normalized) radiolabel in bands N, I and F at 3-min was 88% of the amount at 30-min.

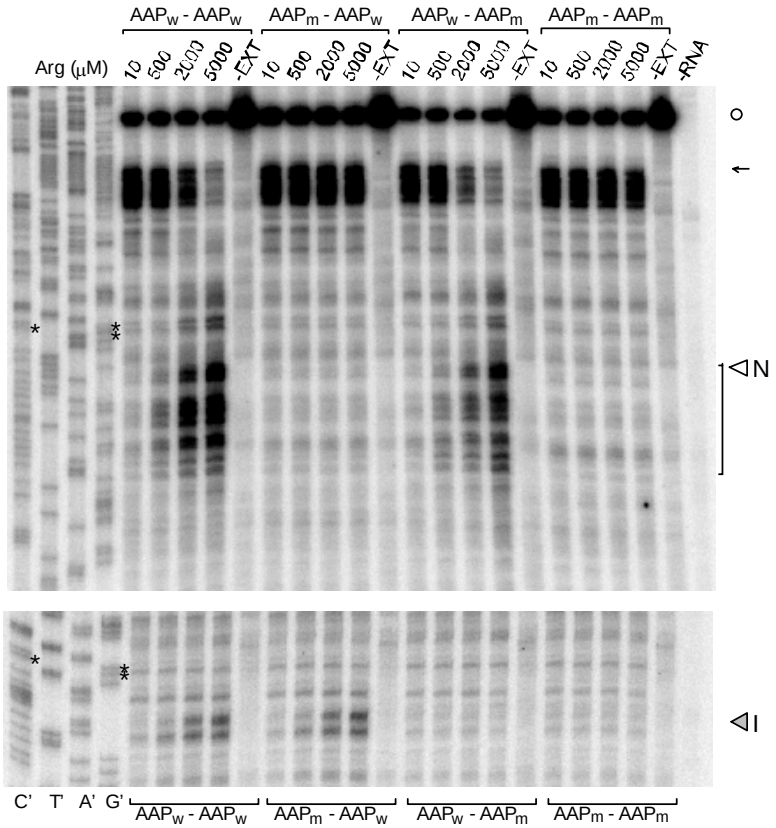

Figure 2.3 Toeprint analysis of ribosome stalling at AAP domains in Met_9-AAP-globin-AAP-LUC mRNA. Separate gels and primers -were used for analyses of stalling at the N-terminal (upper panel) and internal (lower panel) AAP domains for optimal resolution. N-terminal and internal AAP domains are indicated as wild-type (AAP_w) or mutated (AAP_m) above or below the corresponding lanes. Transcripts were translated in 20-μl reaction mixtures containing 10 μM, 500 μM, 2000 μM, or 5000 μM Arg as indicated and 10 μM of the other amino acids, and analyzed as described (Wang and Sachs, 1997b). Left: Sequencing reactions for the Met_9-AAP_w-globin-AAP_w-LUC template. The sequence can be directly read 5' to 3' from top to bottom. Controls: Products obtained from primer extension of RNA (18 ng) in the absence of extract (-EXT) and from an extract not programmed with RNA (-RNA). Primers FP94 and FP93 Figure 2.5A were used for the experiments shown in the upper and lower panels, respectively. The open circle indicates the mRNA 5'-end; the arrow, the position of ribosomes at the first Met (start) codon. Translation of the 9 contiguous Met codons was slow as evidenced by the toeprints of ribosomes in this region. Asterisks mark each AAP's final GCG-codons, which lie ≈16-nt upstream of the toeprints corresponding to the stall sites associated with the production of polypeptide intermediates N and I (indicated); toeprints that represent translational stalling after the Met_9-AAP coding region are indicated with a bracket.

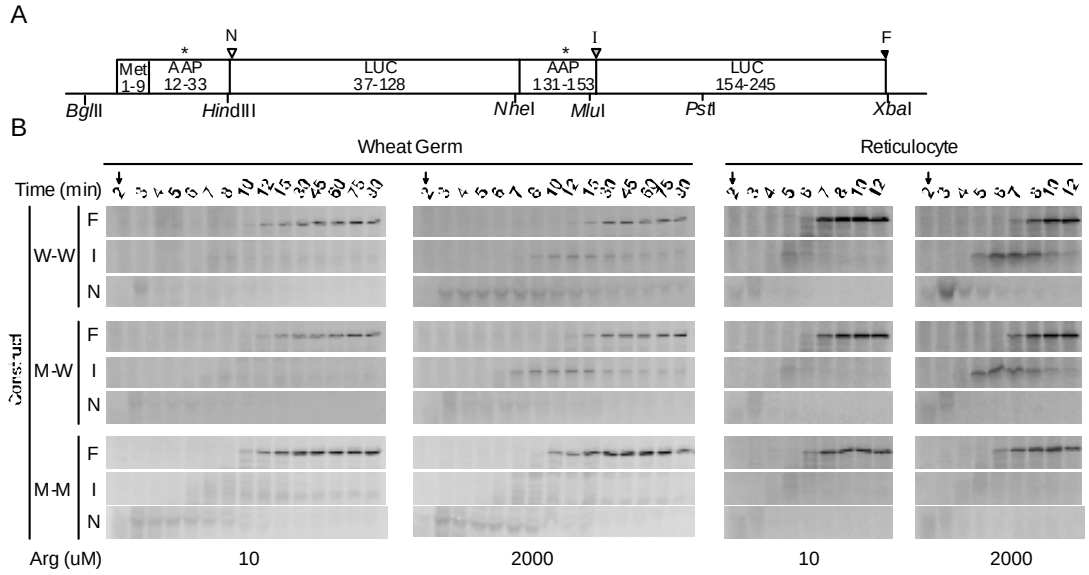

Figure 2.4 Time-course of polypeptide synthesis in wheat germ extracts and reticulocyte lysates. (A). The Met_9-AAP-LUC-AAP-LUC construct used (DNA sequence in Figure 2.5B. Designations are as in Figure 2.1 A except that arrowhead *I* is the C-terminus of the Met_9-AAP-LUC-AAP intermediate. (B) Wheat germ extracts (150 µl) or reticulocyte lysates (100 µl) were programmed with the indicated mRNA (W-W, Met_9-AAP_w-LUC-AAP_w-LUC, M-W, Met_9-AAP_m-LUC-AAP_w-LUC; and M-M, Met_9-AAP_m-LUC-AAP_m-LUC) and incubated at 25°C with either 10 µM or 2000 µM Arg and 10 µM each of the other amino acids. Edeine was added at 2 min and polypeptide products analyzed as described in Figure 2.1, except that reticulocyte samples were incubated in loading buffer at 25°C for 30 min prior to SDS-PAGE, and 3-µl of lysate was loaded per lane. *F*, full-length polypeptide; *I*, intermediate corresponding to Met_9-AAP-LUC-AAP; *N*, intermediate corresponding to Met_9-AAP.

A

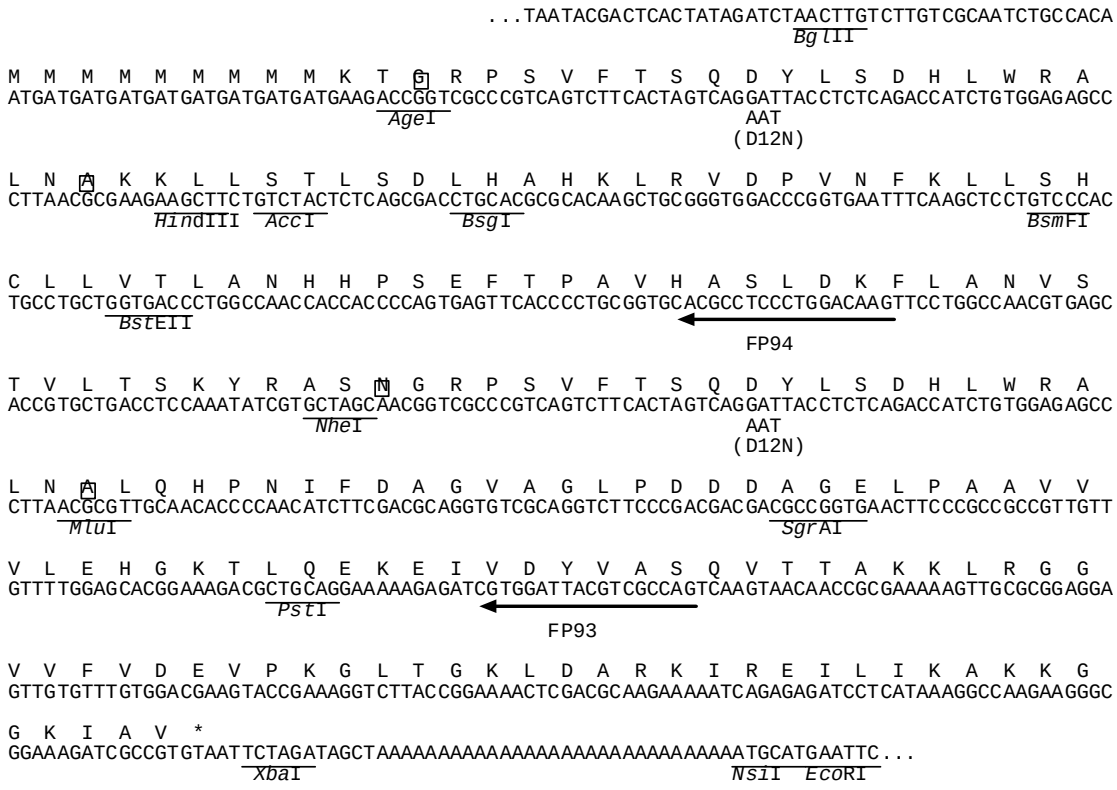

Figure 2.5A Sequences of constructs containing two Arg attenuator peptide (AAP) domains. Met_9-AAP-globin-AAP-LUC (LUC, luciferase). The DNA sequence shown begins with the T7 RNA polymerase-binding site and ends with the *Eco*RI site after the poly(A) tail. The Met_9-AAP-globin-LUC amino acid sequence is indicated. The coding region is a fusion of Met_9-AAP, a rabbit α-globin domain (residues 80-141), a second AAP coding region (residues 2-24), and a firefly LUC domain [residues 460-551 of LUC+NF (Promega)]. The AAP_m mutations corresponding to arg-2 AAP D12N mutations are shown below the wild-type sequence. Some unique restriction sites are underlined. The N- and C-terminal amino acid residues corresponding to each AAP domain are boxed. The sequences for which the reverse-complement DNA was synthesized for use as primers FP93 or FP94 for toeprint analyses are indicated by horizontal arrows.

B

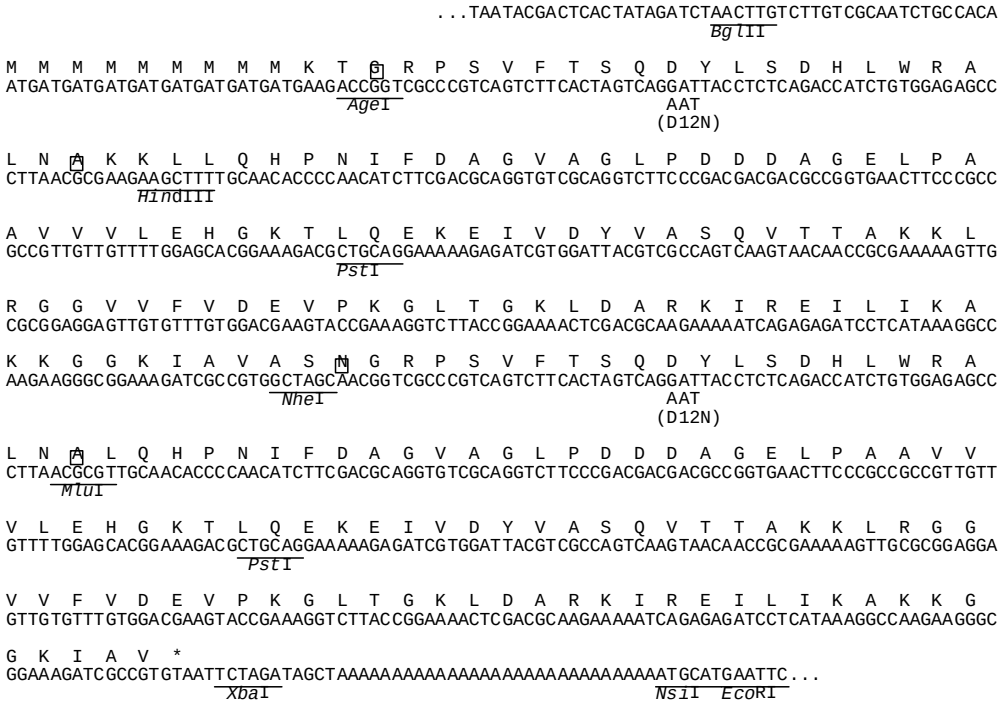

Figure 2.5B Met_9-AAP-LUC-AAP-LUC. Firefly LUC (residues 460-551) replaces globin. Designated features are as described for *A*. The *Eco*RI site was used to linearize templates for producing mRNA.

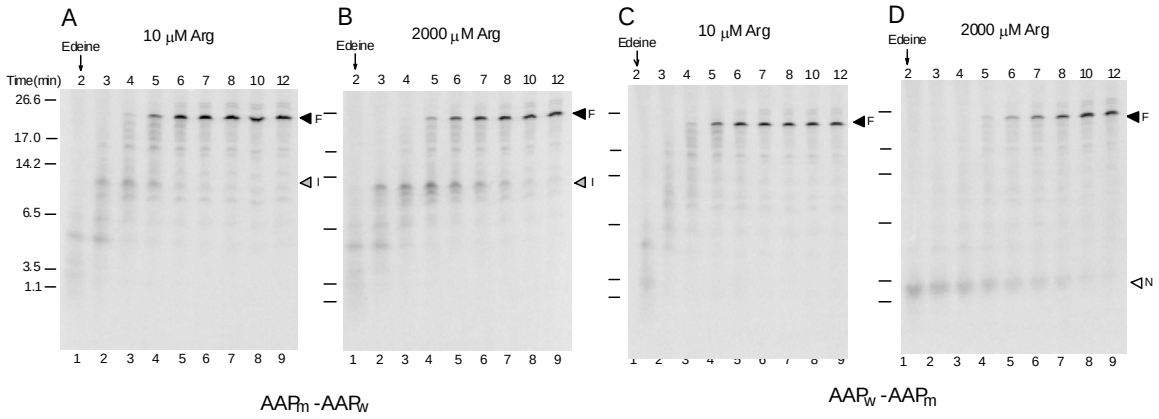

Figure 2.6 Polypeptide synthesis time course in *Neurospora crassa* cell-free extracts programmed with Met_9-AAP-globin-AAP-LUC transcripts. Transcripts specifying Met_9-AAP_m-globin-AAP_w-LUC (*A* and *B*) or Met_9-AAP_w-globin-AAP_m-LUC (*C* and *D*) were translated in extracts in low (*A* and *C*) or high Arg (*B* and *D*). Edeine was added at 2 min (arrow), and 10-μ l aliquots of extracts were removed at the indicated time points for analysis by SDS/PAGE (see text). Arrowhead N indicates the intermediate Met_9-AAP; I, the polypeptide-intermediate Met_9-AAP-globin-AAP; F. the full-length polypeptide.

CHAPTER 3
EVOLUTIONARY CHANGES IN FUNGAL CARBAMOYL-PHOSPHATE SYNTHETASE SMALL SUBUNIT GENE AND ITS ASSOCIATED UPSTREAM OPEN READING FRAME*

3.1 INTRODUCTION

The *Neurospora crassa arg-2* gene and the *Saccharomyces cerevisiae CPA1*, an arginine (Arg) biosynthetic enzyme (Davis, 1986; Radford, 2004). This enzyme is regulated both at the transcriptional and post-transcriptional levels (Davis, 1986; Gaba et al., 2005 and references therein). Under conditions of amino acid limitation, such as histidine starvation, the general amino acid control (GAAC) pathway (also called cross-pathway control in filamentous fungi) causes levels of the *arg-2* and *CPA1* transcripts to increase (Natarajan et al., 2001; Sachs and Yanofsky, 1991). *Magnaporthe grisea ARG2* mRNA also increases two-fold after histidine starvation (Shen and Ebbole, 1997). All three genes contain consensus binding sites (GCREs, general control response elements) for the GAAC transcriptional activators *N. crassa* CPC-1, *S. cerevisiae* Gcn4p, and *M. grisea* CPC1 respectively, which are induced during amino acid starvation (Hinnebusch, 2005; Orbach and Sachs, 1991; Shen and Ebbole, 1997), indicating the response has a transcriptional component. The *N. crassa* and *S. cerevisiae* genes specifying the CPS-A large subunit are also subject to GAAC (Kinney and Lusty, 1989; Natarajan et al., 2001). A separate enzyme, CPS-P, functions to provide carbamoyl phosphate for pyrimidine synthesis under normal growth conditions (Davis, 1986).

---

* This material has been published in this or similar form in *Fungal Genetics and Biology* and is used here with permission of Elsevier Science (USA).

Hood, H.M., Spevak, C.C., and Sachs, M.S. (2006). Evolutionary changes in the fungal carbamoyl-phosphate synthetase small subunit gene and its associated upstream open reading frame (In Press).

Arg metabolism in *S. cerevisiae* is further regulated by the ArgR/Mcm1p complex which is composed of Arg80p, Arg81p, Arg82p and Mcm1p (Messenguy and Dubois, 2003). The anabolic genes *ARG1*, *ARG3*, *ARG5*,*6*, and *ARG8* are repressed in the presence of Arg by the ArgR/Mcm1p complex while the catabolic genes *CAR1* and *CAR2* are induced. The ArgR/Mcm1p regulatory complex binds to arginine control (ARC) elements within the promoters of these genes (Crabeel et al., 1995; Messenguy and Dubois, 2003). Furthermore, an ARC element was identified in the *CPA1* promoter (Crabeel et al., 1995) which may contribute to altered transcription levels in the presence of Arg (Crabeel et al., 1990). The *CPA2* gene specifying the CPS-A large subunit lacks this element and has not been reported to be under control by ArgR/Mcm1p. The mechanism of control through ArgR/Mcm1p is absent in other yeasts such as *Schizosaccharomyces pombe* and filamentous fungi including *N. crassa* (Davis, 1986) (Van Huffel et al., 1994).

arg-2 and *CPA1* are negatively regulated at the level of translation in the presence of high concentrations of Arg (Davis, 1986; Davis and Ristow, 1987; Luo et al., 1995; Luo and Sachs, 1996; Orbach et al., 1990; Werner et al., 1987). This negative regulation is independent of GAAC (Kinney and Lusty, 1989; Sachs and Yanofsky, 1991), and is specific to the gene specifying the CPS-A small subunit (Davis, 1986). Sequencing and expression studies identified an upstream open reading frame (uORF) located within the 5' leader region of *arg-2* and *CPA1* (Nyunoya and Lusty, 1984; Orbach et al., 1990; Werner et al., 1987; Werner et al., 1985). In both *N. crassa* and *S. cerevisiae*, high concentrations of Arg negatively regulate translation in a *cis*-dependent manner via this uORF, which specifies the arginine attenuator peptide or AAP (Gaba et al., 2005; Luo and Sachs, 1996; Wang et al., 1998; Werner et al., 1987).

The AAP is a short peptide consisting of 24 residues in *N. crassa* and 25 residues in *S. cerevisiae* (Nyunoya and Lusty, 1984; Orbach et al., 1990; Werner et al., 1987; Werner et al., 1985). This peptide is evolutionarily conserved (Fang et al., 2000; Shen and Ebbole, 1997; Zhang and Dietrich, 2005). Genetic selection for mutations deficient in Arg-specific regulation in *N. crassa* and *S. cerevisiae* independently identified an aspartic acid residue of the AAP as critical for this negative *cis*-acting regulation (Freitag et al., 1996; Werner et al., 1987). Analyses of the distribution of mRNA on polysomes

demonstrated that translational regulation by the AAP requires this residue *in vivo* (Freitag et al., 1996; Gaba et al., 2005). Cell-free translation systems also showed that translational control by Arg was eliminated when there was a substitution of this aspartic acid residue with asparagine in the AAP (D12N in *N. crassa* and D13N in *S. cerevisiae*) (Wang et al., 1999). The mechanism of Arg-specific regulation was elucidated by *in vitro* studies in which Arg-specific regulation was preserved and in which ribosomes were shown to stall at the termination codon of the uORF specifying the wild-type *N. crassa*, *M. grisea*, *Aspergillus nidulans*, and *S. cerevisiae* AAPs in the presence of high Arg concentrations, resulting in reduced translation from the downstream CPS-A initiation codon (Fang et al., 2000; Wang et al., 1999). Arg-specific stalling, and thus regulation, is eliminated by the D12N or D13N mutations in the uORF-encoded AAP (Wang et al., 1999).

Over 30 fungal genomes have been sequenced (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi) and cDNAs are available in the form of ESTs for some of these fungi as well as for some fungi whose genomes have not been sequenced. Using these data, we performed a comparative analysis of both regulatory sequences and predicted polypeptide structures specified by identified *arg-2/CPA1* homologs across a wide range of fungi.

3.2 RESULTS

3.2.1 AAP sequence and structure

We identified 44 uORFs located upstream of the coding regions specifying the CPS-A small subunit in Ascomycota, Basidiomycota and Zygomycota phyla by searching genome and EST databases; among these were five that were previously described (Delbecq et al., 2000; Orbach et al., 1990; Shen and Ebbole, 1997; Werner et al., 1985). For all species examined, except for *K. lactis*, there was an intercistronic region ranging between 14 and 199 nucleotides in size that separated the AAP coding sequence from the CPS-A coding region (Table 3.1). The *K. lactis* AAP coding sequence overlapped that of CPS-A by 8 nucleotides. Because of this divergent genomic structure, we tested whether the *K. lactis* AAP retained functionality (see below). Although *S.*

pombe and *Candida spp*. posses homologous genes for the CPS-A small subunit, these lack uORFs.

The predicted uORF-encoded AAPs specified peptides between 22 and 25 amino acids in length. AAPs from Basidiomycetes, Hemiascomycetes and Zygomycetes did not vary in size across their respective species (22, 25 and 25 amino acids respectively) whereas AAPs from Euascomycetes ranged between 23 and 25 residues. Peptide alignment identified three conserved residues across all 44 species including an aspartic acid at position 12 (D12) of the *N. crassa* AAP and position 13 in *S. cerevisiae* which is known to be crucial for regulation by Arg (Figure 3.1; Freitag et al., 1996; Figure 3.1; Gaba et al., 2005; Wang et al., 1998; Werner et al., 1987). This peptide alignment also revealed conservation unique to each phylum. All Hemiascomycetes possessed a cysteine at residue 11, except for *Yarrowia lipolytica*. This fungus had a serine at this position and in this respect was more similar to the AAPs from Euascomycetes. A striking difference in amino acid composition between Ascomycota and Basidiomycota AAPs created a shift in the predicted peptides' distribution of electronic charge. For Ascomycetes, there was a conserved histidine (H17) not found in Basidiomycetes that was adjacent to the second absolutely conserved aspartic acid residue (D16). Moreover, there was an additional aspartic acid in Basidiomycetes (D9 in *C. neoformans*) not found in Ascomycetes. These variations resulted in different predicted peptide isoelectric points (*p*I) with the Basidiomycetes' AAPs *p*I more acidic than the Ascomycetes' (data not shown). Lastly, most species possessed a conserved DY(I/L)S motif except for *Yarrowia lipolytica* and *Rhizopus oryzae*, where a cysteine conservatively substitutes for the serine, and for *K. lactis*, where the serine is non-conservatively substituted by a glutamine. The potential affect of this glutamine substitution on regulation is discussed below.

The Neighbor-joining method was used to construct a phylogenetic tree of the AAP sequences Figure 3.2 (Saitou and Nei, 1987). The tree was rooted at the Basidiomycota clade. *U. maydis* had the most divergent AAP within this group and was assigned its own branch. The two Zygomycotes (*R. oryzae* and *Conidiobolus cornatus*) were grouped with the Basidiomycetes, probably because of under-representation of this phylum in this sequence collection. Subphyla divisions were apparent for the

Ascomycetes. The uppermost clade contained all of the Hemiascomycetes (Figure 3.2, in purple), with the *Y. lipolytica* AAP being the most divergent in this group consistent with it possessing a serine at residue 11 (see above). One branch solely consisted of yeasts in the *Saccharomyces* sensu stricto clade (*S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, and *S. cerevisiae*). Euascomycetes (Figure 3.2, in black) were mainly separated according to taxonomic class. All Sordariomycetes (starting with *Fusarium verticilloides*) were contained within a sub-branch of this tree. However, two species belonging to the Dothideomycetes (*Stagonospora nodorum* and *Cochliobolus heterostrophus*) class were placed with the class Eurotiomycetes, which includes *Aspergillus spp*. A phylogenetic tree was also constructed based on the predicted CPS-A coding sequences. Sequences *lipolytica*, which was placed as an outlier within the Euascomycetes and *R. oryzae*, which was placed with the Basidiomycetes (data not shown).

Previous analyses of the genomic structure of the *N. crassa arg-2* and *S. cerevisiae CPA1* uORF showed the AAP was contained within three exons in *N. crassa* (Figure 3.3A; Orbach et al., 1990) and in one exon in *S. cerevisiae* (Figure. 3.3C; Nyunoya and Lusty, 1984; Werner et al., 1987). We identified additional gene structures based on the 36 available genome sequences. The majority of AAPs from Euascomycetes were specified by three exons (Figure 3.3A), with the DY(I/L)S motif (Figure 3.1) always in the second exon. However, AAPs from *Aspergillus spp.* have one less intron, merging the DY(I/L)S containing exon with the next exon (Figure 3.3B). All of the yeast AAPs we found were encoded by one exon (Figure 3.3C), except for *Y. lipolytica*, which had two exons with the second exon containing DY(I/L)S, similar to *Aspergillus spp*.

The Basidiomycete and Zygomycete AAP gene structures were more varied (Figure 3.3D and E). A key change in the *C. neoformans* serotypes, *Coprinus cinereus*, and *Phanerochaete chrysosporium* was that the DY(I/L)S motif was split across the two exons (Figure 3.3D). The *Ustilago maydis* AAP lacked introns (Figure 3.3E). The Zygomycete *R. oryzae*'s AAP's exon/intron structure was unique among the fungi examined. This AAP was specified by two exons, with the intron at a unique position downstream of the DY(I/L)S motif.

3.2.2 Arginine-specific carbamoyl-phosphate synthetase gene structure and predicted subcellular location of the specific enzyme

We assessed 34 CPS-A coding regions for which there were genomic data for intron/exon arrangement. For all Hemiascomycetes, the CPS-A structural gene was contained within one exon. Predicted polypeptide lengths ranged from 399 to 448 amino acids, with *Y. lipolytica* having the longest coding sequence (Table 3.1). CPS-A was 38% identical and 80% similar across the nine Hemiascomycetes examined. Those from the *Saccharomyces* sensu stricto clade were the most related (data not shown). The *S. cerevisiae* CPS-A enzyme is located within the cytoplasm (Davis, 1986). We used the PSORT prediction program to evaluate the potential cellular compartment for the other predicted CPS-A small subunit polypeptides. For all Hemiascomycetes, CPS-A was predicted to be localized to the cytoplasm, with the exception of *Y. lipolytica*, which was predicted to be localized to the mitochondrion. The N-terminus of the *Y. lipolytica* peptide was enriched with arginine and serine residues, which is characteristic of mitochondrial signal sequences (von Heijne, 1986); moreover the first 30 residues of the *Y. lipolytica* N-terminus do not have a counterpart in the other yeasts.

Euascomycetes CPS-A gene structures were more varied. CPS-A was encoded in one exon (*e.g.*, *Aspergillus spp.*), two exons (*e.g.*, *Fusarium graminearum* and *Trichoderma virens*) or three exons for *M. grisea* (Shen and Ebbole, 1997) and *Botrytis cinerea*. Predicted polypeptide length ranged from 387 residues in *Uncinocarpus reesii* to 488 in *Coccidioides immitus* (Table 3.1). Although the peptide length varied across species, CPS-A was 33% identical and 76% similar across these 17 species. Subcellular prediction designated the mitochondrion as the cellular compartment for Euascomycetes CPS-A except for *Chaetomium globosum* (Table 3.1). The *N. crassa* enzyme is known to be mitochondrially localized (Davis, 1986). Further inspection of the automated annotation that predicted the *C. globosum*'s N-terminus indicated this prediction was incorrect based on comparisons to the other fungi. However, we could not firmly establish the correct N-terminus without cDNA evidence.

As was the case for the AAP coding regions, the arrangement of the Basidomycete CPS-A coding regions were the most diverse in this study. The *C.*

neoformans and *C. cinereus* CPS-A homologs contained four exons. In contrast, the homologs from *Ustilago maydis* and *P. chrysosporium* CPS-A were contained in one exon. All introns in Basidiomycetes genes were phase one (*i.e.*, occur after the first base of the codon), whereas all Ascomycete genes examined with introns and the one Zygomycete, for which we had genomic data, contained phase zero introns (*i.e.*, they were in-between codons).

Predicted Basidiomycete CPS-A polypeptide length ranged from 433 to 475 residues (Table 3.1). Sequence alignments showed 90% similarity across CPS-A from the Basidiomycetes and 45% identity among these sequences. All of these polypeptides were predicted to be localized to the mitochondrion (Table 3.1).

In fungi, a separate gene specifies pyrimidine-specific carbamoyl phosphate synthetase. Specifically, *N. crassa pyr-3* and *S. cerevisiae URA2* specify this pyrimidine-specific CPS. In each organism, the CPS domain is fused to another domain providing aspartate transcarbamoylase activity. We used BLAST to identify *pyr-3*/*URA2* homologs in the 34 species for which we had access to genomic data. Each of these fungal genomes is predicted to contain a separate CPS for pyrimidine (data not shown).

3.2.3 Regulatory Motifs

Both *S. cerevisiae CPA1* (Werner et al., 1987) and *N. crassa arg-2* (Orbach et al., 1990) genes contain multiple consensus sequences for the GCRE which enables the response to GAAC. In this study, we detected one GCRE (TGACTC) approximately 350 bp upstream of the AAP that aligned in all *Saccharomyces spp.*; in *S. castellii*, the GCRE was in the inverse orientation (Figure 3.4). In *S. paradoxus* there was a second GCRE ~170 bp upstream from the start of translation of the AAP; *K. waltii* also had two GCREs in similar positions. We did not detect any GCREs in *K. lactis CPA1*. Interestingly, there were six degenerate GCREs located upstream of the *K. lactis* AAP coding sequence, suggesting that CPS-A regulation by general control in this yeast species may be minimal or lost as appears to be the case for AAP-mediated regulation (see below). Previous studies detected four GCRE binding sites in *N. crassa arg-2* (Orbach et al., 1990) and two sites in *M. grisea ARG2* gene (Shen and Ebbole, 1997) between the AAP and *ARG2* coding sequences, including sites within introns. We detected an additional GCRE

approximately 300 bp upstream of the AAP coding sequence in *M. grisea*. The number of GCREs ranged from none to three for other Euascomycetes that had accessible genomic sequence (See Table 3.1). *A. fumigatus* was the only Euascomycete examined which lacked GCREs upstream of the CPS-A coding region although it possesses a GAAC regulatory system (Krappmann et al., 2004).

In *S. cerevisiae*, Arg metabolism is also regulated by the ArgR/Mcm1p complex that binds to arginine-responsive elements (ARC) within the promoters of several genes that are regulated by Arg (Crabeel et al., 1995; Messenguy and Dubois, 2003). The promoter region of *S. cerevisiae CPA1* contains an ARC element (Crabeel et al., 1995). Our analyses identified a conserved core ARC sequence (Chen et al., 2004). approximately 0.5 Kb upstream of the AAP translational start site only in species that belong to the *Saccharomyces* sensu stricto clade (Figure 3.4).

3.2.4 Are diverged AAPs functional?

We tested the functionality of the divergent AAPs from the Basidiomycete *C. neoformans* (serotype A), and the Hemiascomycete *K. lactis* using an *N. crassa* cell-free translation system. The predicted 5'-leader regions of these transcripts, including the uORF and CPS-A start codon, were placed upstream of the firefly luciferase coding region so that this luciferase reporter was fused in-frame with the CPS-A start codon. First we measured luciferase activity under low Arg (10 µM) or two higher Arg (500 µM and 2 mM) conditions. *N. crassa* uORF is fully activated by 150 µM Arg (Wang and Sachs, 1997a). For the *C. neoformans* AAP, luciferase activity was decreased by 2.8-fold in high Arg (2 mM) relative to low Arg (Figure 3.5). Parallel experiments using the *N. crassa* wild-type and D12N mutant AAPs replicated previous results showing reduced luciferase activity for the wild-type AAP in response to Arg and no Arg-dependent negative regulation for the D12N AAP. Regulation by wild-type *C. neoformans* and *N. crassa* AAPs were similar. In contrast, the *K. lactis* AAP showed a substantially reduced regulatory effect (1.5-fold) in response to Arg (Figure 3.5). Experiments using additional independent extracts and mRNA preparations also showed a strong Arg-dependent affect on luciferase activity for the *C. neoformans* and *N. crassa* wild-type AAPs, whereas the

effect of Arg on expression of constructs containing the *K. lactis* AAP was lower and the D12N AAP had no effect (data not shown).

High concentrations of Arg stall ribosomes at the *N. crassa*, *S. cerevisiae*, and *A. nidulans* termination codon of the uORF specifying the AAP (Fang et al., 2000; Sachs et al., 2002; Wang and Sachs, 1997b). Using primer extension assays, we examined ribosomal positions on the translated *C. neoformans* AAP- luciferase construct in response to increasing Arg concentrations (0.5 mM or 2 mM) (Figure 3.6). For both the *N. crassa* and *C. neoformans* uORFs, these increased concentrations of Arg caused increased ribosomal stalling at the uORF termination codon (asterisk; compare lanes 1, 2, and 3 and lanes 7, 8, 9; Figure 3.6). Control experiments using the *N. crassa* D12N mutant showed no ribosomal stalling (compare lanes 3 and 6) replicating previous results (Fang et al., 2000; Wang and Sachs, 1997b).

3.2.5 AAP structure prediction

We examined the AAP structure from *S. cerevisiae*, *N. crassa*, *C. neoformans*, and *K. lactis* by rendering them as α-helices and fitting them to each other using the Molecular Operating Environment program. We modeled them as α-helices because this could place the D12 and D16 of the *N. crassa* AAP in relative positions in which they could both interact with Arg and because preliminary NMR analyses of the synthetic peptide in solvent showed this region adopted an α-helical conformation. These particular species were selected because Arg-dependent negative regulation by the AAP has been extensively studied in *S. cerevisiae* and *N. crassa* and because our results here indicated that the *C. neoformans* AAP was functional, and that the *K. lactis* AAP was less functional. When we aligned and superimposed, D12, Y13 and D16 (numbering from *N. crassa* peptide) forms an unobstructed pocket in *N. crassa*, *S. cerevisiae* and *C. neoformans*. However, in the *K. lactis* AAP this region appeared to be occluded by the Gln15 substitution (S15Q) in the *K. lactis* AAP which was located in the highly conserved DY(I/L)S motif (data not shown). This change in predicted AAP in the highly conserved region of structure suggests a possible reason for why the *K. lactis* AAP is non-functional.

3.3 DISCUSSION

Analyses of the sequences of fungal CPS-A homologs and their associated genetic regulatory elements, including the uORF-encoded AAP, show both divergence and conservation among the 44 fungi examined. All CPS-A genes except *S. pombe* and *Candida* species encoded an upstream AAP. The number and positions of introns within the AAP and CPS-A coding regions varied across phyla and class. ARC elements were detected in the *Saccharomyces* sensu stricto clade but not in other yeasts or filamentous fungi. Varying levels of conservation were detected for GCREs. Although the predicted AAP sequences varied greatly at their N-termini, there are three residues within the AAP that are completely conserved, including the aspartic acid residue demonstrated to be critical for negative translational control in *N. crassa* and *S. cerevisiae*.

3.3.1 Regulatory Elements

We tested *C. neoformans* and *K. lactis* AAPs for function because these represented the most diverged examples of AAP sequences. The *C. neoformans* AAP efficiently reduced translation in response to Arg whereas the *K. lactis* AAP was comparatively poor as a regulator. The loss of features in the *K. lactis* gene that are conserved in all of the other fungal genes containing an AAP, including nonconservative changes in the peptide sequence and loss of intercistronic sequences to create overlapping reading frames, were consistent with the observed loss of function of the *K. lactis* AAP. While all species examined except *K. lactis* had an intercistronic region between the AAP and CPS-A coding regions, such an intercistronic region is not crucial for AAP function (Delbecq et al., 1994; Wang et al., 1998). It is possible, however, that a configuration in which the uORF overlaps with the CPS-A coding region could reduce occupancy of the uORF termination codon by stalled ribosomes.

Another key difference between the *K. lactis* AAP and all of the others examined was a nonconservative glutamine substitution (Figure 3.1) in a highly conserved region previously shown to be critical for AAP function. This change in sequence might also contribute to the AAP's reduced regulatory actions. Consistent with this, molecular modeling of the AAP as an α-helix, which would orient D12 and D16 on the same face of

the peptide surface, suggests that DY(I/L)S could form a potential interaction pocket for Arg with D12 and D16 lining the opening. Modeling predicts that the glutamine residue present in this region in the *K. lactis* AAP obstructed this pocket.

The AAP's N-terminus is not conserved, indicating that it is not important for function. Consistent with this, deletion of *N. crassa* AAP codons 2-4 does not affect function, whereas N-terminal deletions that include conserved residues abolish regulation (Fang et al., 2000). Similarly, studies in *S. cerevisiae* show that the AAP retains function following deletion of residues 2-5; but not residues 2-6 (Delbecq et al., 2000).

Three residues in the AAP were perfectly conserved across all species examined, including an aspartic acid residue known to be essential for Arg-dependent negative regulation (Delbecq et al., 2000; Freitag et al., 1996; Gaba et al., 2005; Wang and Sachs, 1997a; Werner et al., 1987). These residues occur in a highly conserved motif, DY(I/L)S. Two amino acid residues surrounding this conserved motif differed significantly between Ascomycetes and Basidiomycetes, but a Basidiomycete AAP still functioned in an Ascomycete cell-free system to confer Arg-specific negative translational regulation.

CPS-A regulation is subject to regulation in response to amino acid limitation or imbalance by the GAAC system in *N. crassa* and *S. cerevisiae* (Hinnebusch, 1992; Sachs, 1996). We detected GRCEs in all examined CPS-A homologs except for *A. fumigatus*, *K. lactis*, *S. pombe*, and *Candida spp*. *cpcA*, the *A. fumigatus GCN4* homolog, rescued the histidine starvation phenotype of a yeast *GCN4Δ* strain and deletion of *cpcA* from *A. fumigatus* caused slow growth under conditions of amino acid limitation, indicating that *A. fumigatus* has GAAC (Krappmann et al., 2004). The *HIS4*, *LEU1* and *TRP1* genes are under GAAC in *S. cerevisiae* (Hinnebusch, 1992); however their homologs in *K. lactis*, which do have GCREs, did not show transcriptional regulation following amino acid starvation suggesting that GAAC is minimal or non-existent in this yeast (Lamas-Maceiras et al., 1999). Our data suggest that arginine biosynthesis is not regulated by amino acid availability in the same way in *K. lactis* as in most of the other fungi.

S. cerevisiae has an additional level of Arg-dependent transcriptional regulation not found in other yeasts such as *S. pombe* or in other fungi (Davis, 1986; Van-Huffel et al., 1992). The ArgR/Mcm1p complex binds to ARC elements within the promoters of several Arg biosynthetic genes as well as to elements in Arg catabolic genes (Crabeel et

al., 1995; Messenguy and Dubois, 2003). A conserved ARC sequence is present in the promoter region of *CPA1*, the *S. cerevisiae* CPS-A (Crabeel et al., 1995). Our analyses detected a putative core ARC motif upstream of the AAP translational start site in yeasts within the *Saccharomyces* sensu stricto complex that we examined. This regulatory mechanism has not been extensively studied for *S. cerevisiae CPA1* but the conservation of the ARC motif within this clade suggests that it is functional.

3.3.2 Evolutionary Comparison

Phylogenetic analyses using AAP and CPS-A sequences recapitulated the Ascomycete/Basidiomycete distribution of fungi with the exception of *Y. lipolytica.* This fungus's predicted CPS-A polypeptide was grouped as an outlier on the Euascomycete branch instead of with the yeasts. Phylogenetic studies of 18S and 26S ribosomal DNA from ascomycetous yeasts also did not place *Y. lipolytica* sequences with the Hemiascomycetes(Keogh et al., 1998; Kurtzman and Robnett, 1998). Other aspects of *Y. lipolytica* CPS-A gene structure were more similar to Euascomycetes genes in that it contained an intron within the AAP coding region and a mitochondrial signal sequence for the predicted CPS-A small subunit. *Y. lipolytica* is a non-conventional yeast that is dimorphic, and under certain conditions it grows filamentously (Casaregola et al., 2000). Protein coding genes that contain introns are common in *Y. lipolytica* in contrast to *S. cerevisiae* (Barth and Gaillardin, 1997; Casaregola et al., 2000). Thus, the placement of *Y. lipolytica* with Euascomycetes based on CPS-A sequence, and its placement as an outlier among Hemiascomcyetes based on AAP sequence, provide additional evidence that its evolutionary history is incompletely understood.

Our analyses of genomic DNA suggest intron loss within the CPS-A gene and its uORF. For the AAP in Ascomycetes, there appears to be intron loss at the 3' end. The data suggest that *Aspergillus spp*. lost the second intron of the AAP and that Hemiascomycetes additionally lost the 5' intron. Moreover, the CPS-A homolog of *Aspergillus spp.* is encoded within one exon whereas this gene in other Euascomycetes is coded by two exons except for *M. grisea* CPS-A homolog which was contained in three exons (Shen and Ebbole, 1997). The evolutionary importance and mechanism of intron loss (or gain) continues to be debated (Rogozin et al., 2005). Recent literature examining

only fungal genomes or divergent eukaryotic genomes (including *S. pombe*) provide conflicting data on intron evolution. Nielson and colleagues (2004) examined orthologous genes from four Euascomycota genomes and show that introns tend to be lost more toward the middle of the gene compared to the 3' end (Nielsen et al., 2004). In contrast, Roy and Gilbert (2005a, b) show that introns across seven eukaryotic genomes were more likely to be lost near the 3' end supporting the model of intron loss by homologous recombination between intronless cDNA and its genomic counterpart (Fink, 1987; Mourier and Jeffares, 2003; Roy and Gilbert, 2005a; Roy and Gilbert, 2005b). The AAP data are consistent with the homologous recombination theory of intron loss. However, patterns of intron loss were not apparent within the Basidiomycete phylum, most likely due to the availability of less genomic information.

In conclusion, comparative analyses of fungal CPS-A homologs and their associated regulatory uORF specifying the AAP recapitulated the Ascomycete/Basidiomycete distribution of fungi. The AAP is the most highly conserved known uORF in the *Saccharomyces* sensu stricto clade (Zhang and Dietrich, 2005) and one of the most thoroughly studied uORFs in any eukaryotic system (Gaba et al., 2005 and references therein) giving it broad importance in elucidating regulatory mechanisms. Here, we identified three residues conserved across all AAPs, one of which was known through classical genetic studies to be crucial for Arg-dependent negative-regulation in *N. crassa* and *S. cerevisiae* (Freitag et al., 1996; Werner et al., 1987). Additional features, such as the organization of introns, the presence of a potential mitochondrial signal sequence for CPS-A, and the presence amino-acid responsive transcriptional regulatory elements, also generally followed the accepted phylogenies. However, there are interesting differences. While considered to be a yeast, *Y. lipolytica* appears to be more closely related to the Euascomycetes with regard to the mitochondrial localization of CPS-A and the sequence and organization of the AAP. Finally, since uORFs are more common than once thought, their presence must often add another layer to gene regulation (Galagan et al., 2005; Iacono et al., 2005; Vilela and McCarthy, 2003; Zhang and Dietrich, 2005). Comparisons of the wide range of fungal genomic sequences available should enable insight, based on conservation, into the potential importance of

other uORFs in controlling gene expression, as demonstrated for these studies of the uORF in the CPS-A small subunit gene.

3.4 EXPERIMENTAL PROCEDURES

3.4.1 Database queries

We searched databases using protein to nucleotide BLAST queries (Altschul et al., 1990) using AAP DNA and protein sequences from *N. crassa* (GenBank accession number J05512; Orbach et al., 1990) and *S. cerevisiae* (GenBank accession number X01764; Werner et al., 1985). The following databases were queried for fungal DNA sequences: NCBI BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi), The Broad Institute at MIT and Harvard Fungal Genome Initiative (http://www.broad.mit.edu/annotation/fgi/), Fungal genomes at the *Saccharomyces* Genome Database (http://seq.yeastgenome.org/cgi-bin/blast-fungal.pl), Genolevures (http://cbi.labri.fr/Genolevures/blast.php), the Institute for Genomic Research (http://tigrblast.tigr.org/er-blast/index.cgi?project=cna1), the Stanford Genome Technology Center (http://sequence-www.stanford.edu/group/C.neoformans/) and the Department of Energy's Joint Genome Institute (http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Phchr1&advanced=1). Because the AAP amino acid sequence is short (22-25 residues acids in length), we used the least stringent BLAST parameters. Additionally, we searched for CPS-A homologs using blastp when appropriate and manually inspected the upstream sequence, which enabled us to identify more divergent AAP sequences.

3.4.2 Alignments

Both genomic and predicted peptide sequences were aligned using CLUSTAL W (Thompson et al., 1994). AAP and CPS-A small subunit genomic sequences were manually annotated by evaluating all translational frames for previously identified conserved sequences. Exon-intron boundaries were determined by manual inspection for consensus splice donor and acceptor sites and were confirmed when cDNA sequence was

available. For some AAPs only cDNA sequences were available (see Table 1). A phylogenetic tree was constructed for the AAP amino acid sequences using NJPLOT which is part of the CLUSTAL W software package. This program uses the Neighbor-Joining method (Saitou and Nei, 1987). The tree was rooted at the Basidiomycete clade.

3.4.3 Subcellular localization prediction

A major distinction between the *N. crassa* and *S. cerevisiae* CPS-A enzymes is their subcellular localization. CPS-A in *N. crassa* is mitochondrial whereas the yeast enzyme is cytoplasmic (Davis, 1986). We used the PSORT program (http://psort.ims.u-tokyo.ac.jp/) to predict the subcellular localizations of CPS-A polypeptides and also scanned the N-termini of the predicted polypeptide manually for features characteristic of mitochondrial signal sequences.

3.4.4 Protein Modeling

In order to predict AAP structure we used the Molecular Operating Environment program (Chemical Computing Group, Quebec, CA). AAP amino acid sequences from *S. cerevisiae*, *Kluyveromyces lactis*, *N. crassa*, and *Cryptococcus neoformans* serotype A were imported into the program as α-helices. We aligned and superposed all three AAPs onto each other and examined the predicted conformations of the amino acid side chains in the highly conserved region of the AAP.

3.4.5 Motif Searches

Amino acid biosynthesis is subject to GAAC (Davis, 1986; Hinnebusch, 1992; Sachs, 1996). *N. crassa* Cpc-1, *S. cerevisiae* Gcn4p and *Aspergillus fumigatus* CpcA transcription factors bind to specific core sequences of the GCRE: TGA(G/C)TC found in many amino acid biosynthetic genes including those that code for CPS-A homologs studied to date (Hinnebusch, 1984; Hinnebusch, 2005; Paluh et al., 1988; Wanke et al., 1997). Additionally, in yeast the *trans*-acting ArgRp/Mcm1p complex regulates expression of genes involved in both arginine biosynthesis and catabolism. This complex binds to the arginine-responsive element, ARC (ARginine Control) that consists of a 23-nucleotide consensus sequence (Crabeel et al., 1995) with a highly conserved 7-

nucleotide core: CACTTAA (Chen et al., 2004). For sequences where genomic data was available, we searched the AAP and CPS-A coding sequences, intercistronic sequences, and introns (including 1kb upstream and 0.5 Kb downstream) for GCRE and ARC motifs.

3.4.6 DNA templates and RNA synthesis

Synthetic oligonucleotides (Sigma Genosys, The Woodlands, TX) containing the *C. neoformans* serotype A, MSS155 (5'-GATCTAACTTGTCTTGTCGCAATCTGCCACAATGCTGTGGACCGAGCGAGTTTCTGACGATTACCTTTCCGACCAAGTTTTCACGCTTTTGCATTTGTGAA-3') and MSS156 (5'-AGCTTTCACAAATGCAAAAGCGTGAAAACTTGGTCGGAAAGGTAATCGTCAGAAACTCGCTCGGTCCACAGCATTGTGGCAGATTGCGACAAGACAAGTTA-3') were designed to contain *Bgl*II and *Hind*III restriction sites placed in the 5' and 3' ends respectively. Synthetic oligonucleotides containing the *K. lactis* AAP sequence, CSY1 (5'-GATCTAACTTGTCTTGTCGCAATCTGCCACAATGTTTAGACTCTGCTTCTCTTCATACACTTGCCAAGACTACATTCAAGACCATATCTGGAAAAATAACAATGCCTAATTTCAAATCG-3') and CSY2 (5'-GTGACCGATTTGAAATTAGGCATTGTTATTTTTCCAGATATGGTCTTGAATGTAGTCTTGGCAAGTGTATGAAGAGAAGCAGAGTCTAAACATTGTGGCAGATTGCGACAAGACAAGTTA-3') were designed to contain *Bgl*II and *BstE*II restriction sites placed in the 5' and 3' ends respectively. The *C. neoformans* and *K. lactis* synthetic oligonucleotides were ligated to the gel purified vector pR701, which was cut with the same restriction enzymes that was specified in the respective synthetic oligonucleotides, and contains the T7 promoter and the firefly luciferase reporter coding region (Fang et al., 2002).

3.4.7 Cell-free translation of RNA and Primer Extension Inhibition Assays

Cell-free translation extracts were prepared from *N. crassa* mycelia as described (Fang et al., 2000; Wang and Sachs, 1997a). Reaction conditions for *in vitro* translation using *N. crassa* extracts were as described (Fang et al., 2000; Wang et al., 1998; Wang

and Sachs, 1997a). Briefly, 12 ng of synthetic mRNAs were used to program *N. crassa* extracts. Reactions (20 µl) were incubated at 25 ^{o}C for 30 minutes in the presence of high Arg (500 µM and 2 mM) or low Arg (10 µM). Translation was terminated by freezing in liquid nitrogen. Luciferase activity was measured by adding 5 µl of ice-thawed translation reactions to 50 µl of luciferase assay reagent (Promega, Madison, WI). Synthetic mRNA encoding the *Renilla* (sea pansy) luciferase mRNA was used as an internal control for translation (Wang et al., 1998). mRNA containing wild-type *N. crassa* AAP and mutated D12N AAP were used as positive and negative controls control for regulation (as described in Wang et al., 1998). All experiments were performed at least in duplicate with independent batches of RNA and extract. All data are represented as ratios of firefly luciferase activity to sea pansy luciferase activity.

The primer extension inhibition assays to map the positions of ribosomes at rate limiting steps in translation (*e.g.*, to determine whether Arg induced ribosome stalling by the different AAPs) were accomplished as described using primer ZW4 (Wang and Sachs, 1997b). Briefly, reactions programmed with mRNAs containing the *C. neoformans* AAP was added to 5.5 µl of reverse transcription buffer and was incubated at 50^{O}C for 2 minutes and then placed on ice. $^{32-}$P-labeled primer was added and annealed to the template by incubation at 37^{O}C for 5 minutes. Reverse transcription reaction mixtures (100 U of Superscript II RNase H-; Gibco BRL) were incubated for 30 minutes at 37^{O}C. Reactions were stopped by extraction with 10 µl of phenol-chloroform. The aqueous phase was added to an equal volume of DNA sequencing stop solution and then was heated at 85^{O}C for 5 minutes, cooled on ice, and loaded on 6% DNA denaturation sequencing polyacrylamide gels (7 µl/sample). The gels were dried and exposed to screens of a Molecular Dynamics PhosphorImager for approximately 24 h.

Table 3.1

Species	AAP intron #	AAP AA#	Theoretical pI	intercistronic	GCREs	ARC	CPS intron #	CPS AA#	PSORT[1]
A. gossypii	0	25	7.85	26	1	0	0	399	C
K. lactis	0	25	6.49	-8	0	0	0	409	C
S. castellii	0	25	4.75	119	1	0	0	399	C
S. kluyveri	0	25	6.69	31	1	0	0	truncated	C
S. paradoxus	0	25	5.94	74	2	1	0	411	C
S. mikatae	0	25	5.94	86	1	1	0	411	C
S. kudriavzeii	0	25	5.94	72	1	1	0	411	C
S. bayanus[2]	0	25	5.94	ND	ND	ND	ND	ND	ND
S. cerevisiae	0	25	5.94	56	2	1	0	411	C
K. waltii	0	25	5.96	72	2	0	0	399	C
K. thermotolerans[3]	ND	25	6.5	ND	ND	ND	ND	ND	ND
Y. lipolytica	1	23	6.8	64	2	0	0	448	M
A. flavus	1	24	6.5	57	2	0	0	450	M
A. nidulans	1	25	5.19	55	2	0	0	454	M
S. nodorum	2	23	6.5	53	2	0	1	476	M
A. fumagatus	1	25	5.36	57	0	1	0	453	M
A. terreus	1	25	5.9	83	3	0	0	453	M
A. niger[3]	ND	25	6.5	ND	ND	ND	ND	ND	ND
C. heterostrophus[5]	ND	24	4.22	ND	ND	ND	ND	ND	ND
H. capsulatum	2	24	6.5	48	1	0	0	388	M
P. brasiliensis[3]	ND	24	6.5	ND	ND	ND	ND	ND	ND
U. reesii	2	24	5.36	55	3	0	0	387	M
C. posadasii[3]	ND	24	5.36	ND	ND	ND	ND	ND	ND
C. immitis	2	24	5.36	60	2	1	0	488	M
S. sclerotiorum	2	24	6.5	57	3	0	1	457	M
B. cinerea	2	23	6.5	58	2	0	2	455	M
F. verticillioides	2	24	6.69	66	2	0	1	456	M
F. graminearum	2	24	6.5	72	2	0	1	456	M
C. globosum	2	24	8.36	199	4	0	1	417	C
P. anserina	2	24	5.36	79	1	0	1	453	M

Species	AAP intron #	AAP AA#	Theoretical pI	intercistronic	GCREs	ARC	CPS intron #	CPS AA#	PSORT[1]
N. crassa	2	24	6.5	63	4		1	453	M
T. reesei	2	24	5.36	80	3	0	1	453	M
T. virens	2	24	5.36	89	3	0	1	453	M
M. grisea	2	24	5.36	77	3	0	2	471	M
U. maydis	0	22	5.28	55	2	0	0	475	M
C. neoformans B	1	22	4.22	71	0	0	3	446	M
C. neoformans D	ND	22	4.22	ND	ND	ND	ND	446	M
C. neoformans A	1	22	4.22	75	2	0	3	446	M
C. neoforman D	1	22	4.22	74	2	0	3	446	M
C. cinereus	1	22	3.93	16	2	0	3	433	M
P.involutus[3]	ND	23	4.43	ND	ND	ND	ND	ND	ND
P. chrysosporium	1	22	4.41	14	2	0	2	439	M
R. oryzae	1	25	4.75	28	0	0	0	398	C
C.cornatus[3]	ND	25	4.83	ND	ND	ND	ND	ND	ND

1. For CPS-A subcellular predictions, C represents cytoplasmically located whereas M indicated that the peptide is predicted to be mitochondrial.
2. Whole-genome shot gun sequence contained a partial sequence for CPS-A homolog.
3. AAP sequence was obtained from cDNA; therefore there was no information on genomic structure.
4. ND = not determined due to lack of sequence data.
5. Unpublished data obtained from Dr. Scott Baker.

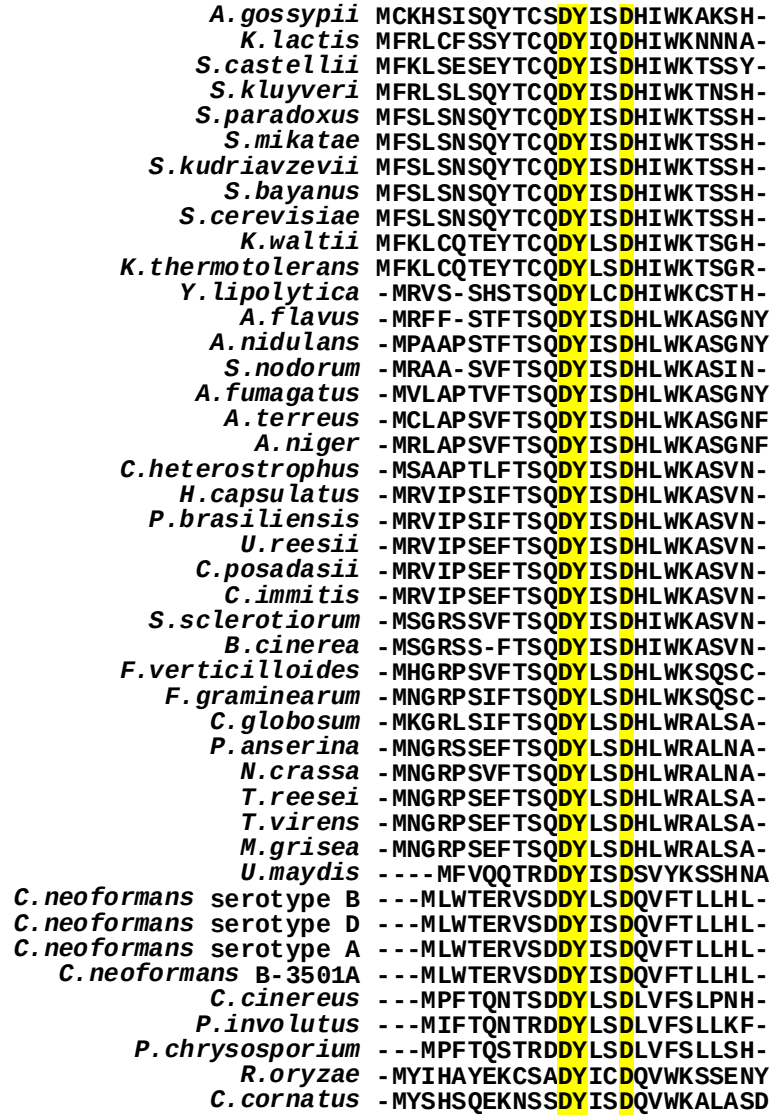

Figure 3.1 Alignment of the arginine attenuator peptide (AAP) upstream open reading frame identified residues critical for function. Using blastp (Altschul *et al.*, 1990), we queried several genome databases (see methods) with AAP from multiple species and retrieved putative AAP sequences from Ascomycota, Basidiomycota and Zygomycota phyla. ClustalW alignment (Thompson et al, 1994) of 44 AAP sequences identified three universally conserved residues (in yellow). Studies in *S. cerevisiae and N. crassa* identified the region of the AAP containing these residues as critical for function (Delbecq *et al.*, 2000; Fang *et al.*, 2000).

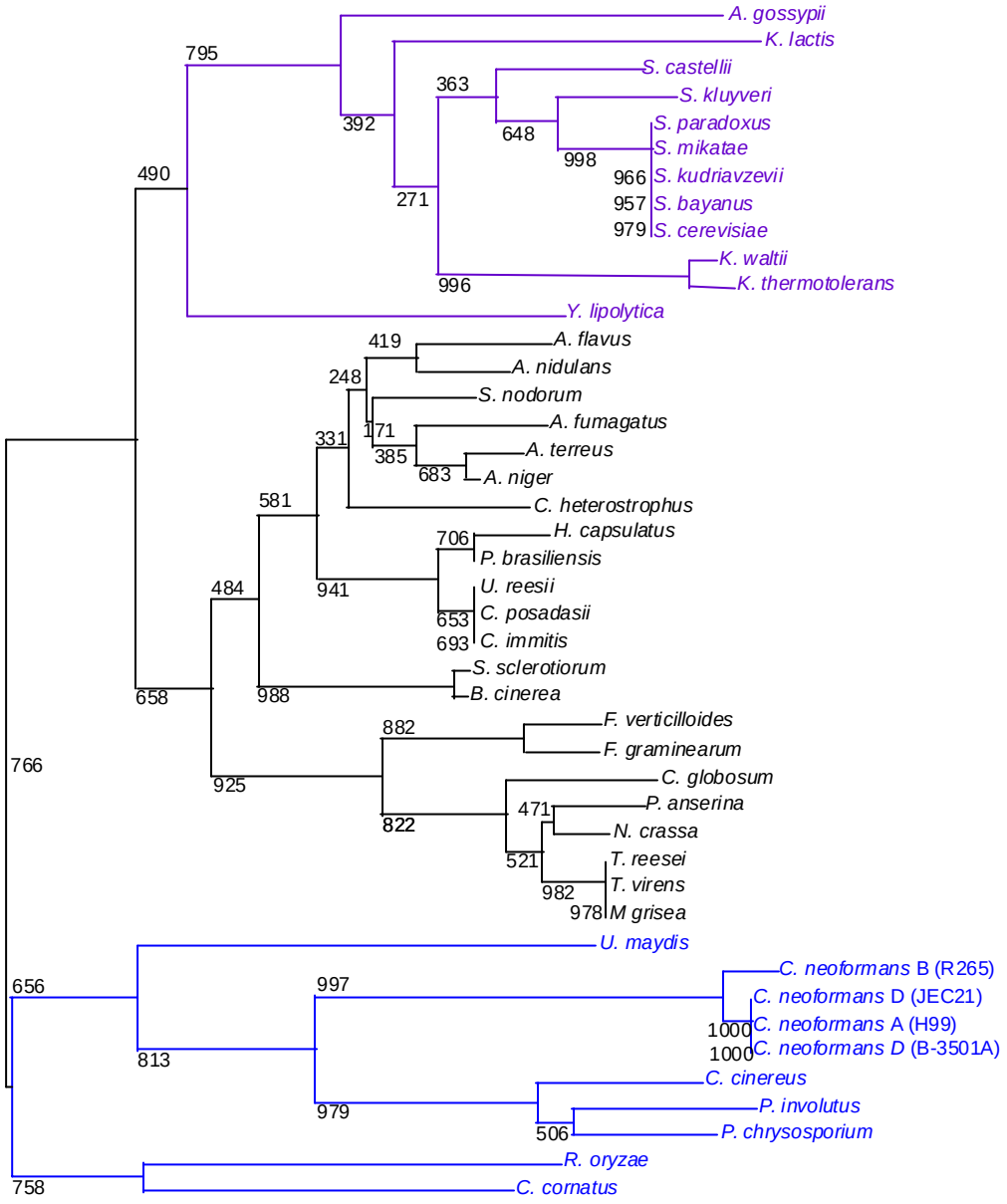

Figure 3.2 This tree based on AAP residues was created using the neighbor-joining method of Saitou and Nei (1987) and was rooted at the Basidiomycetes clade. Three major clades were apparent. (A) The uppermost clade contained species within the Saccharomyces family. (B) All of the known AAPs for Euascomycetes were contained within the middle clade. Furthermore, the Eurotiomycetes and Sordariomycetes classes being the major subdivisions. (C) AAPs from Basidiomycetes were contained in the last major branch with *Ustilago maydis* peptide sequence being the most diverged within this branch.

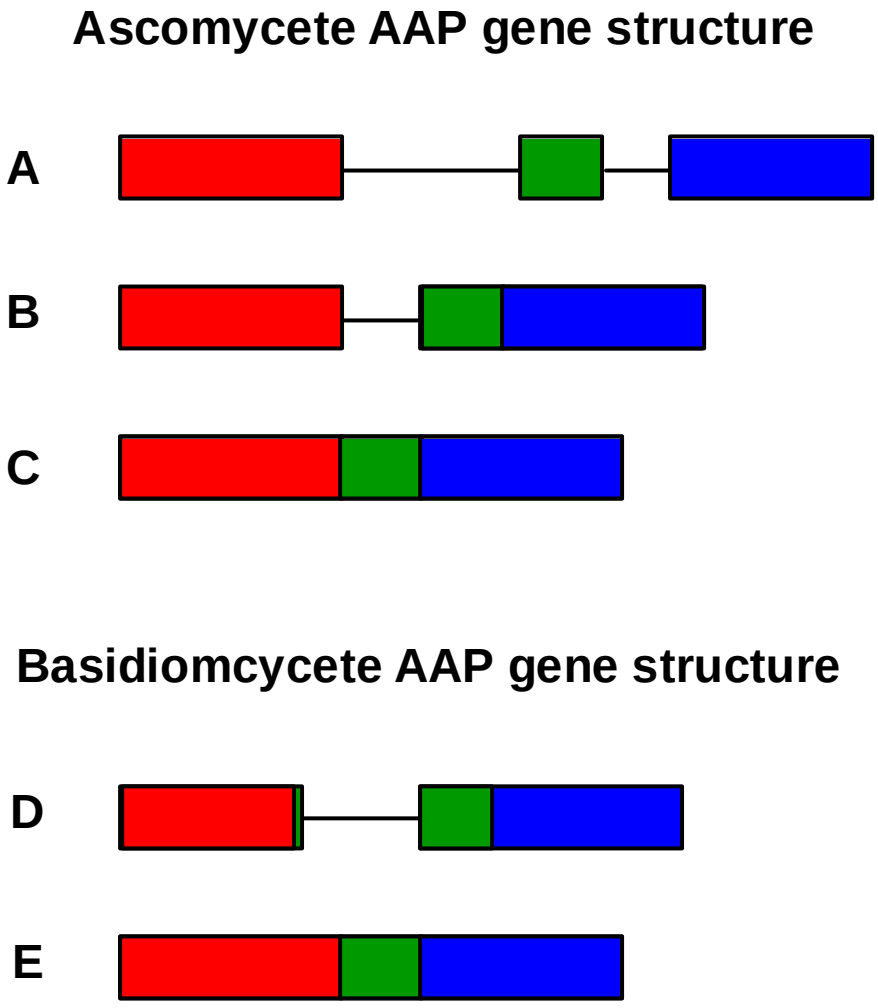

Figure 3.3 AAP genomic structure differed across fungal species. (A) Most Euascomycetes, except for *Aspergillus spp.* (see B), possessed similar exon/intron structure consisting of three exons (these three exons positions are color-coded for reference) and two introns (represented by lines). The second exon coded for the highly conserved motif, DY(I/L)S. (B) *Aspergillus spp*. lost an intron of the AAP, merging two exons. (C) AAP genomic structure for all yeast species, except for *Yarrowia lipolytica*, consisted of a single exon, which was expected considering the paucity of introns in yeast. The genomic structure of *Y. lipolytica* consisted of two exons in the same arrangement of as the *Aspergillus spp.* as shown in (B). In *Kluyveromyces lactis*, the AAP open-reading frame (ORF) and *CPA1* ORF overlap (data not shown). (D) In Basidiomycetes species we detected an alternate intron/exon placement. The gene structure for *Cryptococcus neoformans* strains, *Phanerochaete chrysoporium*, and *Coprinus cinereus* AAP consisted of two exons and one intron. The key difference was that the highly conserved DY(I/L)S motif was interrupted by an intron. (E) *Ustilago maydis* was the only Basidiomycete to date to have one exon coding for AAP.

Figure 3.4 Transcription factor binding elements conservation. Genomic alignments identified a conserved core Arginine control element (ARC) sequence (TTAAGTG) that binds the Arg80 transcription factor in *S. cerevisiae* (Chen *et al.*, 2004). This element was located approximately 500 bp upstream of the AAP translational start site in yeasts belonging to the *Saccharomyces* sensu stricto clade (numbering based in *S. cerevisiae* sequence; highlighted in green). However, this motif was not identified in other fungi. We also detected varying numbers of general control/cross-pathway control motif (TGACTCA; highlighted in blue) across all fungal species that contain an AAP suggesting differences in regulatory control efficiencies (data shown for only *Saccharomyces spp.*).

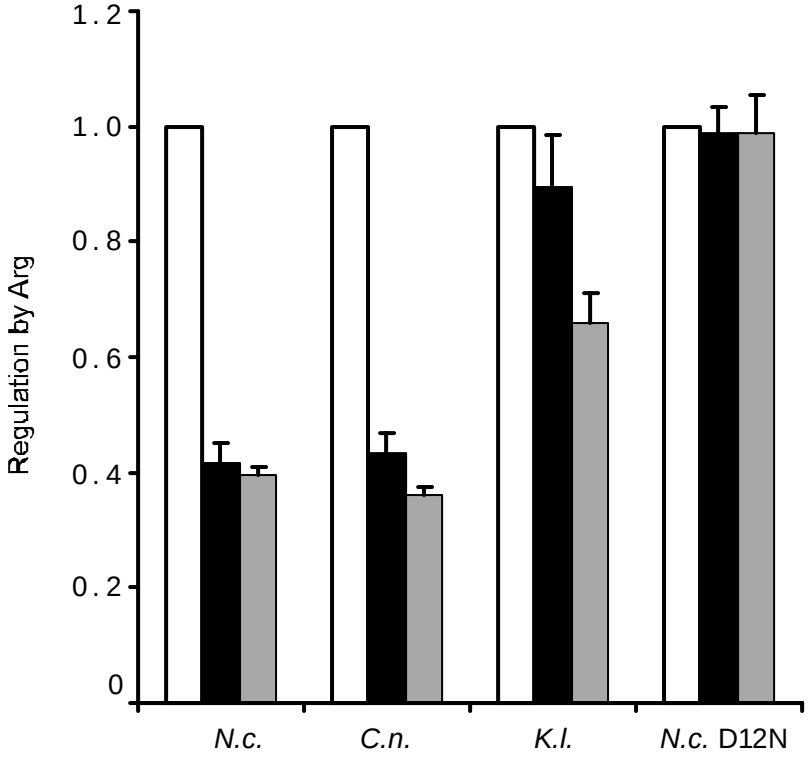

Figure 3.5 Arg-dependent regulation of translation by diverged AAPs. The effects of uORFs from *N. crassa, C. neoformans*, and *K. lactis* on the synthesis of firefly luciferase in *N. crassa* cell-free translation extracts programmed with two independent preparations of mRNA was assessed using three different concentrations of Arg (10 μM, white bars; 500 μM, black bars; 2 mM; gray bars). Firefly luciferase activity was normalized to *Renilla* luciferase activity (see Methods) and expressed as a ratio of normalized activity at a given Arg concentration relative to the normalized activity in 10 μM Arg (i.e., the value of 10 μM Arg is unity). The mutated *N. crassa* D12N AAP, which lacks the capacity to regulate translation in response to Arg, was used as a negative control. In response to higher Arg concentrations, the *C. neoformans* AAP reduced translation similarly to the *N. crassa* wild-type AAP, while the *K. lactis* AAP functioned less effectively to reduce translation in response to Arg.

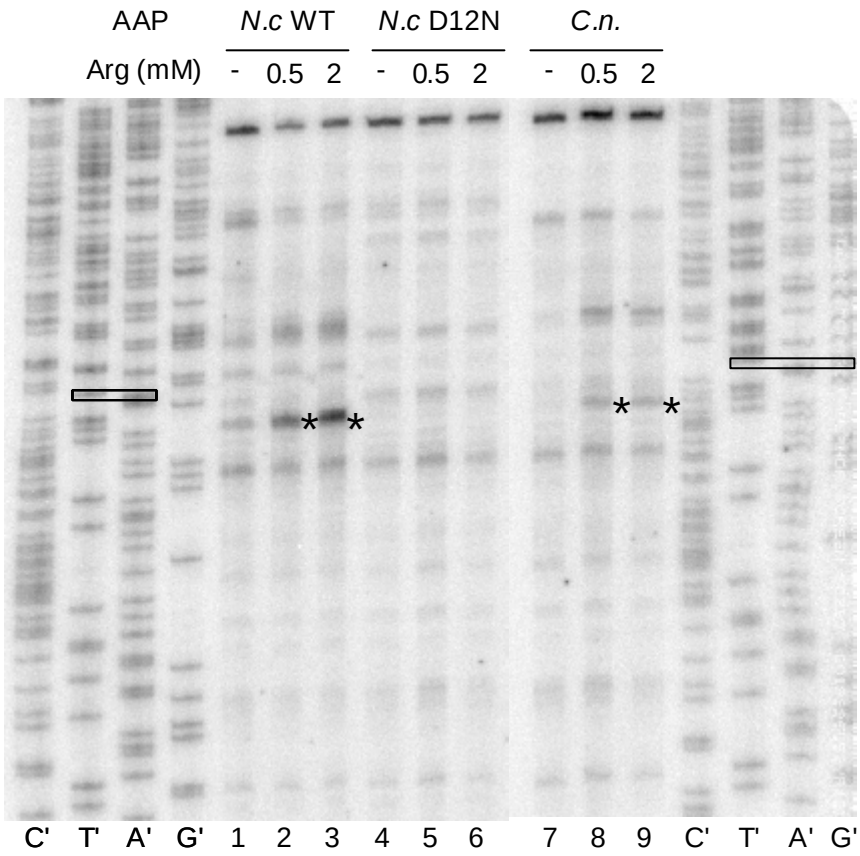

Figure 3.6 Toeprint analysis of Arg-regulated stalling by the wild-type *C. neoformans* and *N. crassa* AAPs. Arg-induced ribosomal stalling was assessed using a primer extension assay (see Methods). Sequencing reactions for the template mRNA are adjacent to the toeprints. The 5' to 3' sequence reads from top to bottom. The wild-type *N. crassa* AAP (lanes 1-3) showed Arg-dependent stalling at the AAP termination codon (asterisks) whereas stalling was not observed in the D12N mutant (lanes 4-6). The more diverged *C. neoformans* AAP also produced Arg-dependent ribosomal stalling (lanes 7-9). The site of the Arg-induced toeprint is ≈13 nt downstream of each uORF's termination codon (boxed).

CHAPTER 4
SPECIFIC REQUIREMENTS FOR A NASCENT PEPTIDE THAT STALLS RIBOSOMES IN RESPONSE TO ARGININE

4.1 INTRODUCTION

The *Neurospora crassa arg-2* gene encodes the small subunit of arginine specific carbamoyl phosphate synthetase, the first enzyme in fungal arginine (Arg) biosynthesis (Davis, 1986; Radford, 2004). The *arg-2* mRNA contains an upstream open reading frame (uORF) specifying a 24-residue peptide named the Arg attenuator peptide (AAP). The uORF encoded AAP is highly conserved in many fungi including Hemiascomycete, Ascomycete, and Basidiomycete species (Hood et al., In press).The nascent *N. crassa arg-2* AAP and homologous *Saccharomyces cerevisiae CPA1*-encoded AAP cause ribosomes to stall in response to high concentrations of L-arginine as demonstrated by *in vitro* studies (Wang et al., 1998; Wang et al., 1999). AAP-mediated Arg-specific regulation is independent of the level of aminoacylation of arginyl-tRNA (Wang et al., 1999). AAP-mediated ribosome stalling at the uORF termination codon in response to Arg blocks scanning ribosomes from reaching the downstream start codon for the biosynthetic enzyme, reducing gene expression at the level of translation (Gaba et al., 2001). Arg-regulated stalling of ribosomes mediated by the AAP is observed in translation systems from plant, animal and fungal sources, indicating the mechanism does not require fungal-specific factors (Fang et al., 2004). Finally, ribosome stalling at the termination codon for the uORF-encoded *CPA1* AAP triggers degradation of the *CPA1* mRNA by the nonsense-mediated mRNA decay (NMD) pathway *in vivo* (Gaba et al., 2005)

Previous studies have examined some of the sequence requirements for Arg-specific regulation mediated by the AAP. *In vivo*, it was demonstrated that mutations of

a conserved Asp residue to Asn abolish regulation (Freitag et al., 1996; Werner et al., 1987). Analyses of the distribution of ribosomes on reporter mRNAs *in vivo* show these mutations (D12N in the *N. crassa* AAP and D13N in the *S. cerevisiae* AAPs, respectively) act to abolish regulation at the translational level (Freitag et al., 1996; Gaba et al., 2005). Neither the intercistronic region (Delbecq et al., 1994; Fang et al., 2000) nor a specific termination codon is required for regulation (Wang et al., 1998). Arg-specific translational regulation by the AAP is dependent on the amino acid sequence and not the mRNA sequence (Delbecq et al., 1994; Fang et al., 2000). Studies in both *N. crassa* and *S. cerevisiae* identified the evolutionarily conserved region of the AAP as important (Delbecq et al., 2000; Delbecq et al., 1994; Fang et al., 2000; Wang et al., 1998), but the minimal functional domain required for AAP function has not been determined.

Other eukaryotic uORFs specify nascent peptides that cause ribosomes to stall at the uORF's termination codon (Geballe and Sachs, 2000; Morris and Geballe, 2000 and references therein), but the AAP, unlike these other nascent peptides, can cause regulated stalling of ribosomes involved in elongation as well as termination,. Thus, AAP-mediated stalling is observed even when it is placed as an internal domain within a large polypeptide (Fang et al., 2004).

In a formally similar system, the leader peptide in the *E. coli* tryptophanase (*tna*) operon, named TnaC, causes prokaryotic ribosomes to stall in response to tryptophan (Trp). The interactions between TnaC and the large subunit of the ribosome have been mapped near the large subunit polypeptides L22 and L7, and rRNA mutations in this region affect TnaC function. Recently, it was determined that the Trp-12 residue of TnaC-tRNA(Pro) is required to activate free tryptophan binding which results in the inhibition of peptidyl transferase (Cruz-Vera et al., 2006). This is relevant to understanding the AAP because the distance from Trp-12 to the ribosome peptidyl transferase center (PTC) when stalling occurs in response to Trp is similar to the distance between the essential Asp-12 of the AAP and the PTC when stalling occurs in response to Arg.

An internal domain within *E. coli* SecM causes ribosomes to stall. The arrest sequence, whose critical residues were identified as FXXXXWIXXXXGIRAGP, interacts with the ribosome tunnel close to the ribosomal protein L22 (Nakatogawa and

Ito, 2002). Studies on prokaryotic ribosomes synthesizing SecM indicate that the exit tunnel acts as a discriminating gate that permits regulation of polypeptide chain elongation as a consequence of the sequence of the nascent SecM chain (Nakatogawa et al., 2004). Recently, it has been shown by *in vitro* experiments and toeprinting that the last essential amino acid in the arrest sequence (Pro166) is not placed into the nascent chain when stalling occurs, but remains at the ribosome A-site (Muto et al., 2006).

The similarities between TnaC, SecM, and the AAP and interactions of prokaryotic peptides TnaC and SecM with the ribosomal protein L22, suggests that these peptides are important to consider when formulating a model to determine the ribosomal components that are interacting with the AAP.

Here, we determined by extensive analyses of amino acid substitutions in the AAP which residues are important for translational regulation. Scanning mutagenesis analysis of the AAP coding region by changing conserved residues to alanine and proline revealed the core residues, D12, Y13, L14 and W19 are essential for function. The minimal sequence required for AAP to function as an internal domain or as a uORF was determined by systematically substituting its natural N- and C-terminal regions with segments of other polypeptides. Residues T9-R20 are necessary for regulatory function. This minimal domain functions to stall ribosomes when placed as a uORF.
Arg-specific regulation appears independent of the extent of arginyl-tRNA charging status (Wang et al., 1999); tests on a variety of Arg analogs for their regulatory effects showed that the chiral center, guanidino group and primary amino group of Arg are most important for regulatory function. Determination that the carboxyl group of Arg was not important led us to test the effect of the short synthetic peptides including RGD and GRGD. Analyses of the kinetics of regulation and concentrations of Arg or RGD to translation extracts showed that both Arg and RGD acted similarly, but GRGD does not exert regulation, consistent with the idea that the primary amino group is important for regulatory function.

4.2 RESULTS

4.2.1 Systematic mutagenesis revealed key residues in the AAP

Synthesis of the 24-residue evolutionarily-conserved *N. crassa* AAP causes ribosomes to stall on the mRNA in the presence of high concentrations of Arg. The amino acid sequence of the AAP, not the nucleotide sequence of its coding region, is responsible for regulation (Fang et al., 2000). Therefore, we completed alanine-scanning mutagenesis to determine which residues of the AAP were important for Arg-specific. Proline-scanning was also performed to determine whether substitution of each residue with Pro, which should constrain peptide folding, had different effects on the nascent peptide's regulatory function. Since residues 2-5 are entirely dispensable for regulation (Fang et al., 2000), we started scanning mutagenesis on residue Ser-6. Scanning-mutations were placed into the *N. crassa* AAP coding region using PCR with mutagenic primers (Freitag et al., 1996; Luo and Sachs, 1996) and/or megaprimer PCR (Sarkar and Sommer, 1990) (Figure 4.1, Table 4.1). Regulation exerted by the uORF containing the individual alanine and proline mutagenic sequences (Table 4.1) was compared with regulation mediated by the wild-type uORF and by a uORF in which Asp-12 was mutated to Asn (D12N), a mutation that abolishes regulation *in vivo* (Freitag et al., 1996) and *in vitro* (Fang et al., 2000; Wang et al., 1998; Wang and Sachs, 1997b).

Equal amounts of capped and polyadenylated synthetic RNAs representing each construct (Table 4.1) were used to program *N. crassa* extracts. Previous *in vitro* experiments showed that, in translation reactions programmed with RNAs containing the wild-type uORF, 150 μM Arg is sufficient for regulation. However, since the mutated AAPs could be expected to have reduced function, higher concentrations of Arg (500 μM or 2 mM) were added to the translation reactions. Figure 4.2 summarized the mutations analyzed and their effects on regulation based on the expression of a firefly luciferase reporter. These data were obtained from at least two independent experiments using at least two independent RNAs and at least two independently prepared cell-free translation extracts.

The effects of Ala and Pro substitutions were overall similar, but if there were differences, the effect of proline was always more severe. Thus, substitution with Pro at some positions had more severe effects on regulation in experiments using 500 Arg; and for Ser-10, substitution with Ala did not abolish regulation, but substitution with Pro did. Examination of the relative amount of luciferase produced per unit amount of input mRNA was accomplished to determine whether any of the mutations had significant effects on the synthesis of luciferase that were independent of their effects on Arg-specific regulation. The amount of luciferase activity produced by translation of each transcript did not decrease or increase significantly in that, in all cases, luciferase activity was at least 80% of wild-type mRNA control in extracts containing supplemented with a low concentration (10 μM) of Arg. Thus none of these mutations appeared to increase stalling in low Arg (*i.e.*, cause constitutive stalling), a result that was borne out by the toeprint assays described below.

Examination of the luciferase reporter enzyme activity in Figure 4.2 showed that three fundamentally different effects of mutations could be observed. First, substitutions could reduce regulation. Thus, comparing Fig. 2A to 2B shows how the specific mutations: S6A, S6P, D16P, H17P, R20P, and A24P resulted in reduced regulation (Fig. 2A yellow dots) in response to 500 μM Arg. However, when the Arg concentration was raised to 2 mM, the regulatory effect of these uORFs increased (*i.e.*, residues marked in yellow in Fig. 2A are green in Fig. 2B). Another way regulation appeared reduced was that regulation was not apparent when 500 μM Arg was added, but was apparent when 2 mM Arg was added. Such effects were seen with mutations: T9A, T9P, R20A, R20P, and A21P. A second class of mutations eliminated regulation at each concentration of Arg tested; this occurred with any substitution tested at residues D12, Y13, and L14, which are highly conserved in fungi (Hood et al., In press). A final class of mutations showed no discernible regulatory consequences. These substitutions could be found distributed throughout the AAP. Surprisingly in this regard, the highly conserved residue D16 (Hood et al., In press), could be substituted with Ala, Pro or Glu without eliminating the AAP's function; however, the D16N mutation eliminates function (Wang and Sachs, 1997b and data not shown).

To verify that the differences in regulatory effects of these and additional mutations that affected regulation based on luciferase assays arose from differences in the capacity of the specific AAPs to stall ribosomes, the positions of ribosomes on each RNA were assayed by toeprinting. As expected, in extracts containing high Arg (500 μM) compared to low Arg (10 μM), an increased toeprint signal corresponding to ribosomes stalled at the wild-type uORF termination codon (Fang et al., 2000) was seen (Figure 4.3A, compare lane 2 to lane 1). Also as expected, the D12N mutation did not show an increased toeprint at the uORF stop codon in response to high Arg (Figure 4.3A, compare lanes 4 and 3). Mutation of residues, D12, Y13, L14 to alanine or proline abolished regulation, as seen by assay of LUC synthesis and by toeprinting (Figures 4.2 and 4.3B, and). The toeprint results for changes at residue D16 were also entirely consistent with the luciferase data (Figure 4.3C). Interestingly, substituting the conserved W19 residue with alanine or proline abolished regulation, but changing it to Tyr did not (Figure 4.2 and Figure 4.3D. The W19F mutation, also did not affect regulation as determined by LUC assay (data not shown). This result is consistent with the comparative analysis data across fungi showing that all AAPs contained an aromatic residue at this position, but not always Trp (Hood et al., In press). Additional toeprinting data were obtained for all of the Ala and Pro substitutions shown in Figure 4.2; these data (not shown) were consistent with the luciferase assays.

Since S15, D16, H17 are highly conserved in ascomycetes and hemiascomycetes (Hood et al., In press), and each can be individually changed to Ala without eliminating function (Figure 4.2A and Figure 4.2B), we substituted Ala for S15, D16, H17 in all possible combinations. Preliminary data indicated that all three residues can be converted to Ala simultaneously and regulation maintained, albeit at a reduced level. The effects of mutations were additive, in that the effects of single mutations were generally less than double, which were less than the triple mutation (data not shown).

Additional studies were performed to investigate whether altering the distance between the conserved residues S15, D16, H17 and the conserved aromatic residue W16 affected regulation. Insertion of a single Ala residue after L14 or H17, or deletion of residue L14 or L18 abolished regulation as determined by luciferase assay. Since

regulation was abolished in preliminary experiments, these studies were not pursued further.

4.2.2 Determination of the minimal AAP sequence required for regulation as an internal domain

What is the minimal functional domain of the AAP? Previous studies in which the AAP was specified as an independent uORF showed that the N-terminal segment (residues 2-5) of the AAP could be deleted without affecting regulation, but further deletions eliminated regulation (Fang et al., 2000). Shortening the C-terminus by even one residue eliminated regulation (Wang and Sachs, 1997a). It was not established from these studies whether the loss of regulation occurred because the length of the nascent peptide was crucial or specific residues were crucial. Because other studies showed that the entire AAP from residues 2-24 could be placed internally within a larger polypeptide and that the AAP would still function to stall ribosomes from fungal, plant and animal sources (Fang et al., 2004), this opened the possibility for studying the requirements for specific amino acid sequences to retain AAP function without affecting any absolute length requirement.

We examined the minimal sequences required for the AAP using constructs to systematically remove the N- and/or C-terminal residues of the AAP when it was positioned internally in a larger polypeptide. The effects of removing these residues were assessed by [^{35}S]Met pulse-chase experiments, in which a stall in translation is detected by the appearance of a polypeptide intermediate in SDS-PAGE analyses (Fang et al., 2004). Synthetic mRNAs were derived from plasmid pLL301 (Fang et al., 2004) and were prepared as described in the Experimental Procedures. These synthetic mRNAs were used to program cell-free translation extracts containing Arg at low (10 μM) or high (2 mM) concentrations, and the other amino acids and [^{35}S]Met at fixed concentrations. Radiolabel is incorporated into these polypeptides essentially only near the N-termini. For pulse-chase analyses, translation reactions are incubated for two minutes, and the edeine is added to block subsequent rounds of initiation, and polypeptide chain elongation progress was monitored by SDS-PAGE as previously described (Fang et al., 2004).

Programming *N. crassa* and wheat germ extracts with mRNAs specifying the various N- and C-termini systematic deletions of the AAP and examining the radiolabeled translation products using the pulse-chase protocol established the minimal sequence required for Arg-specific regulation as summarized in Figure 4.4. Systematic deletion from the N-terminus showed that deletion of residues 2-8 (ΔNGRPSVF) did not significantly affect regulation, but removal of residues 2-9 affected regulation (ΔNGRPSVFT) (Figure 4.5, compare panel B to panel C). Systematic deletion from the C-terminus showed that residues 21-24 (ΔALNA) (Figure 4.4), could be deleted without affecting regulation, however, residue R20 is critical since the deletion of codons 20-24 (ΔRALNA) eliminated regulation (data not shown).

After determining which segments of the AAP's C- and N-termini were essential for Arg-specific regulation, a construct was created that combined the largest functional deletions from each terminus to produce an AAP construct containing only 12 residues. To eliminate the possibility that residues from the adjacent luciferase domains had compensated or influenced the deleted AAP residues, alanine was substituted for all amino acid positions of the AAP that were shown to be dispensable for function (pCS914)(Figure 4.4). This transcript was used to program *N. crassa* and wheat germ extracts and polypeptide chain elongation was monitored by pulse-chase analysis. For the internal AAP wild-type construct (Figure 4.5A), an intermediate (I), consistent with stalling at the internal AAP was observed in the presence of high Arg. Synthesis of the full-length polypeptide (labeled F in Figure 4.5A) was also delayed in high Arg (full-length synthesis at 6 min rather than 5 min), consistent with translational stalling. Thus, it appears that the AAP region consisting of residues T9-R20 is sufficient to cause ribosome stalling. When the transcript specifying the determined minimal 12-residue internal AAP was programmed into *N. crassa,* extracts and wheat germ cell –free extracts, an intermediate (I) referring to stalling at the internal AAP was observed in the presence of high Arg (Figure. 4.5D). Thus, pulse-chase analysis in *N. crassa* (and wheat germ extracts) showed that the 12 core residues (T9-R20) bordered by Ala residues still conferred regulation (Figure 4.5D). While the Ala-replacement construct shares two residues within the full-length AAP outside the core, this concern was addressed by using

a construct in which the 12 core AAP residues are bounded by LUC residues eliminating any similar residues with the wild-type AAP (data not shown).

4.2.3 The minimal internal AAP exerts regulation when placed as a uORF

To determine if the minimal 12-residue sequence of the internal AAP domain was sufficient to stall ribosomes at the termination codon when placed as a uORF. We used the minimal AAP construct flanked by alanines (pCS914) to analyze its effect by assaying for enzyme activity activity after addition of high Arg concentrations (500 μM). Regulation was reduced below wild-type levels, but was detectable and above that observed for the D12N control (preliminary data). We proceeded to perform toeprinting experiments to map ribosomes on the transcript. Equal amounts of each RNA sample (120 ng) were translated in reaction mixtures containing low Arg (10 μM) and a higher Arg concentration (2 mM) to better assess the effect of Arg. For the wild-type uORF, increasing the concentration of Arg from 10 μM to 2 mM caused ribosomes to stall at the uORF termination codon (Fig. 6, lanes 1-3). The minimal uORF sequence containing the 12 essential residues acted similarly, although did not show a strong regulatory effect compared to the wild-type (Figure 4.6, lanes 7-9) with increasing Arg concentrations. However, in contrast to the Arg-specific regulation by ribosome stalling conferred by the wild-type uORF, and the core AAP uORF, the D12N uORF in good initiation context did not confer Arg-specific regulation (Figure 4.6, lanes 4-6). This reveals that the conserved region of the AAP containing the 12 essential residues (AAPmin) is sufficient to cause ribosomes to stall in high concentrations of Arg.

4.2.4 Arginine analogs and short peptides provide insights into the structural requirements for regulatory function

Since previous studies showed that the level of aminoacylated Arg-tRNA is not important for regulation (Wang et al., 1999), Arg or a metabolite of Arg must be important. Many Arg analogs have been tested for their effects on regulation (Table 4.3) on the wild-type AAP and on all the alanine, proline, and residue substitution mutations (4.1). Equal amounts of each RNA (12 ng) were translated in the reaction mixtures supplemented with 500 μM of the Arg analogs The preliminary results show that L-Arg

methyl-ester, L-Arg ethyl-ester and phopho-L-Arg are the only analogs that conferred similar regulatory effects as Arg. The conclusions from that study are that the chiral center, guanidino group and primary amino group are most important for regulatory function. With this result, we then further tested the effects of short peptides containing Arg such as RGD, RGDS, GRGD and GRGDS. This would enable future experiments, such as cross-linking, to determine the site of Arg regulatory function. RGD and RGDS showed similar effects as Arg to cause Arg-specific negative regulation through the wild-type *arg-2* AAP but not the D12N AAP in *N. crassa* extracts as determined by luciferase assays (data not shown). In the wild-type AAP construct (Figure 4.7A), increasing concentrations (10-500 μM) of Arg added to *N.crassa* cell-free translation reactions resulted in a toeprint at the uORF termination codon revealed Arg-specific regulation (Figure 4.7A, compare lane 1 to lanes 2-6). The tripeptide RGD showed a similar effect as Arg (Figure 4.7A, compare lanes 2-6 to lanes 7-11) to cause regulated ribosome stalling. Addition of GRGDS did not result in a toeprint at the termination codon when added to translation reactions containing the wild-type AAP (Figure 4.7A, compare lane 1 to lanes 12-16). In the D12N AAP construct (Figure 4.7B), increasing concentrations of Arg, RGD, and GRGDS (10-500μM) did not reveal a toeprint at the uORF termination codon (Figure 4.7B). Addition of Arg and these analogs did not result in regulation, consistent with the previous effects seen with this D12N mutant in previous experiments. Analyses of the kinetics of regulation and the effects of adding different concentrations of Arg or RGD to *N.crassa* extracts showed that both Arg and RGD acted at similar rates on the wild-type AAP-LUC fusion (data not shown). HPLC analyses confirmed that the RGD peptide added in these experiments was not contaminated with Arg.

In contrast to the results with RGD, GRGD, similar to the effect of GRGDS, did not exert regulation (data not shown), consistent with the importance of the primary amino group of Arg. We also addressed concerns that RGD was hydrolyzing to Arg in the extracts in the time-frame used for these assays by addition of numerous protease inhibitors and short incubation periods (data not shown). We performed pulse-chase experiments and detected stalling soon after initiating translation when comparing the effects of adding 300 μM Arg or RGD to extracts programmed with dual AAP wild-type pLL301 constructs (Fang et al., 2004). We used 300 instead of 2000 μM concentrations

to demonstrate the similar sensitivity of stalling to Arg and RGD and to minimize potential issues of hydrolysis (Figure 4.8). In conclusion, these data indicate that RGD can replace Arg in vitro to stall ribosomes. The fact that Arg-containing peptides exert regulation allows for future experiments aimed at identifying the site of Arg regulatory action.

4.3 DISCUSSION

The results indicate the specific sequence requirements of the AAP to enable Arg-mediated ribosome stalling. Alanine- and proline-scanning mutagenesis revealed the core residues that are essential for function. These residues are the most highly conserved (Chapter 3). The discovery that the AAP can function as an internal domain to stall fungal, plant, and animal ribosomes (Fang et al., 2004) along with the previous demonstration that the length of the N-terminus may be important since shortening the C-terminus eliminates function (Wang and Sachs, 1997b) led to the question of whether the length of the AAP is a requirement for regulation. Since the AAP functioned as an internal domain, systematic deletion of nonconserved and/or nonessential residues from both the N- and C- termini allowed the determination of the core residues necessary without affecting the length of the AAP. These results showed that the 12-residue AAP core, T9-R20, was necessary and sufficient to confer regulation and are also able to function when placed as a uORF. Even though this 12-residue AAP functions to stall ribosomes, the complete 24-residue peptide still exerts a stronger effect.

The functional similarities between the AAP and prokaryotic TnaC and SecM provide insights on the specific events that could be occurring in the ribosome tunnel. From these collective analyses, AAP Asp-12, which is critical for function, would be located in the ribosome tunnel ~12 amino acids from the residue at the P site. Interestingly, amino acids in the polypeptide chain that are critical for stalling prokaryotic ribosomes during the synthesis of SecM and the TnaC are also located in the ribosome tunnel ~12 residues from the P site. This region interacts with ribosomal protein L22 in the tunnel. The sequence similarities between TnaC, SecM, and the AAP and the demonstrated interactions of the prokaryotic peptides with the ribosome near L22 make these prokaryotic peptides relevant to consider as a model for the possible interactions of

proteins in the ribosomal tunnel with the AAP, such as L17 (the eukaryotic homolog of L22).

Previous analog studies provided insights into the structural requirements for the AAP's regulatory function. The Arg-specific mediated regulation is also dependent on the structure of Arg, since L-Arg, but not its stereoisomer D-Arg, confers regulation. The conclusions that the chiral center, guanidino group, and primary amino group are most important for regulatory function led to the testing of short peptides that contain Arg, such as RGD and GRGD. RGD worked like Arg to cause negative regulation through the wild-type but not the D12N AAP in *N. crassa*. GRGD does not exert regulation which is consistent with the importance of the primary amino group of Arg. The fact that short peptides containing Arg function in regulation provide an excellent opportunity for further investigation of the site of Arg function such as cross-linking experiments.

Another example of eukaryotic ribosome stalling mediated by a nascent peptide is the *Arabidopsis CGS1* gene, which codes for cystathionine γ-synthase. This enzyme catalyzes the first committed step of methionine biosynthesis in higher plants. A high level of methionine or S-adenosyl methionine (AdoMet) reduces mRNA levels. The MTO1 region, which is a short stetch of amino acids encoded by the first exon of CGS1, in the mRNA is essential for regulation by acting within the ribosomal exit tunnel to cause elongation arrest. Recent studies suggested that the ribosome stalled at the step of translocation (Onouchi et al., 2005a). This is a recent example of a nascent peptide-mediated translation elongation arrest that coupled with mRNA degradation. Perhaps there is a similar control mechanism with the AAP since it has been shown to control mRNA stability in vivo as seen in *S. cerevisiae* CPA1. The CPA1 uORF has a critical role in modulating NMD. The model states that the nascent AAP and Arg act together to decrease CPA1 expression by regulating translation and mRNA stability.

The interaction of Arg, AAP, and the translational machinery is still unclear. The previous demonstration that the AAP functions as an internal domain in a polypeptide provides a means of controlling translational elongation. The results of these studies provide insights for further characterization of the requirements for Arg-mediated regulation by the AAP either as an internal peptide or as a uORF. The regulation of

translational elongation by an internal nascent polypeptide domain that acts together with a small molecule is an interesting aspect of gene regulation.

Understanding the key residues, the length of the nascent peptide, and the structure of arginine itself is important in deciphering the mechanisms involved in ribosome stalling by the AAP. The results have further defined the specific sites in the AAP and specific structures of Arg that are necessary for regulation. The data presented here provide a very detailed study that a eukaryotic peptide can cause the regulated stalling of ribosomes involved in both termination and elongation and the features of the nascent peptide itself is primarily responsible for controlling ribosome movement.

4.4 EXPERIMENTAL PROCEDURES

4.4.1 DNA templates and RNA synthesis

For scanning mutagenesis to replace AAP residues with Ala, Pro and selected other residues, synthetic oligonucleotides (Sigma Genosys, The Woodlands, TX) were designed for use in megaprimer PCR (Sarkar and Sommer, 1990) to place specific mutations affecting the AAP in PCR-generated DNA fragments that contained *Bgl*II and *Nco*I restriction sites near their 5' and 3' ends respectively (Table 4.1). After digestion with these enzymes, the PCR products could be ligated to reporter vectors for further studies. The PCR template was pR301 (Fang et al., 2002). The vector contains the T7 RNA polymerase promoter and firefly luciferase coding region. Forward primers CS1 (5'GAGCACCGCCGCCGCAAG-3') or RF1 (5'-CCGCAAGGAATGGTGCAT-3') were used for first-round PCR and the various mutagenic primers listed in Table 4.1 were used as the reverse primers for first round PCR. The product from the first round PCR (megaprimer) was used as the forward primer and primers CS12 (5'-CCTTTCTTTATGTTTTTGG-3') or RF12 (5'-TGTTTTTGGCGTCGGTGA-3') were used as the reverse primers. The final megaprimer PCR products were digested with *Bgl*II and *Nco*I and ligated to similarly digested and gel-purified vector pR401, which differed from pR301 by its containing a unique diagnostic *Nde* I site.

For analyses of the minimal functional AAP domain, AAP sequences were ligated to a vector that was derived from plasmid pLL301 (Supplemental data from Ref. Fang et

al., 2004). This plasmid, pCS801, contains a T7 RNA polymerase promoter to enable *in vitro* synthesis of a polyadenylated transcript containing a reporter ORF consisting of (i) nine methionine codons at the N-terminus (ii) an HA-tag, (iii) a luciferase domain (iv) a modified AAP region containing strategically placed unique restriction sites to enable mutation or truncation of the AAP region and (v) another luciferase domain (Figure 4.4). Synthetic oligonucleotides were annealed together and ligated to appropriately digested pCS801 vector to produce plasmids(Figure 4.4) containing mutated and/or truncated AAP coding segments with the amino acid sequences depicted in Figure 4.4. Annealed oligonucleotides had 5' overhangs and were directly ligated to digested, gel-purified vector.

4.4.2 Cell-free translation and primer extension inhibition (toeprint) assays

Plasmid DNA templates were purified by using either the Qiagen Plasmid Midi Prep kit or the Wizard Plus Midi Prep kit (Promega). Capped, polyadenylated RNA was synthesized with T7 RNA polymerase from plasmid DNA templates that were linearized with *Eco*RI and the yield of RNA was quantified as described (Wang and Sachs, 1997b). The reaction conditions for *in-vitro* translation using *N. crassa* extracts were as described (Wang et al., 1998; Wang and Sachs, 1997b). Cell-free translation extracts were prepared from *N. crassa* as described (Fang et al., 2000; Wang et al., 1998; Wang and Sachs, 1997b). For firefly luciferase activity measurements, equal amounts (12 ng) of mRNA were used to program extracts; synthetic mRNA encoding Renilla (sea pansy) luciferase was added to all reactions to serve as an internal control (Wang et al., 1998). All translation reaction mixtures were incubated at 25°C for luciferase assays and translation was halted by freezing reaction mixtures in liquid nitrogen after 30 min.

For [^{35}S]Met pulse-chase experiments, micrococcal nuclease treated *N. crassa* extracts or micrococcal nuclease treated wheat germ extracts (Promega) were programmed with 60 ng of RNA and [^{35}S]Met was used at a final concentration of 0.5 uCi/ul. Edeine was added as previously described at 2 min after translation was initiated (Fang et al., 2004).

The primer extension inhibition (toeprint) assays were accomplished as described using primer ZW4 (Wang and Sachs, 1997b). The gels were dried and exposed to

screens of a Molecular Dynamics PhosphorImager for approximately 24 h. All toeprint data shown were representative of multiple experiments.

Table 4.1 Firefly LUC constructs made by PCR or megaprimer PCR (MP)

Construct	5'leader structure	Forward primer	Backward mutagenic primer
pCS201	S6P	CS1	RC1
pCS202	S6A	CS1	RC2
pCS203	V7P	RF1	RC3
pCS204	V7A	CS1	RC4
pCS205	F8P	CS1	RC5
pCS206	F8A	CS1	RC6
pCS207	T9P	CS1	RC7
pCS208	T9A	CS1	RC8
pCS209	S10P	CS1	RC9
pCS210	S10A	RF1	RC10
pCS301	Q10P	RF1	RC11
pCS302	Q11A	RF1	RC12
pCS303	D12P	CS1	RC13
pCS304	D12A	RF1	RC14
pCS305	Y13P	RF1	RC15
pCS306	Y13A	CS1	RC16
pCS307	L14P	CS1	RC17
pCS308	L14A	CS1	RC18
pCS309	S15P	CS1	RC19
pCS310	S15A	CS1	RC20
pCS401	D16P	CS1	RC21
pCS402	D16A	CS1	RC22
pCS403	D16E	CS1	RC23
pCS404	H17P	CS1	RC24
pCS405	H17A	CS1	RC25
pCS406	L18P	CS1	RC26
pCS407	L18A	CS1	RC27
pCS408	W19Y	CS1	RC28

Construct	5'leader structure	Forward primer	Backward mutagenic primer
pCS101	W19P	RF1	RF2
pCS102	W19A	RF1	RF3
pCS103	R20P	RF1	RF4
pCS104	R20A	RF1	RF5
pCS105	A21P	RF1	RF6
pCS106	L22P	RF1	RF7
pCS107	L22A	RF1	RF8
pCS108	N23P	RF1	RF9
pCS109	N23A	RF1	RF10
pCS110	A24P	RF1	RF11

1. The final background primer for megaprimer PCR was either RF1 or RF12
2. pR301 was used as the wild-type AAP; pS301 was used as the D12N AAP
3. The wild-type pR301 was used as the PCR template and pR401 (differing from pR301 only by a single NdeI site) was used as the ligation vector.

Table 4.2 LUC Assays

Construct	Mutation	500 µM Arg	2 mM Arg
pCS201	S6P	1.56	4.12
pCS202	S6A	1.38	3.46
pCS203	V7P	1.91	4.13
pCS204	V7A	2.98	4.28
pCS205	F8P	2.56	4.54
pCS206	F8A	2.39	5.03
pCS207	T9P	1.21	2.22
pCS208	T9A	1.29	1.97
pCS209	S10P	0.94	1.03
pCS210	S10A	2.93	4.62
pCS301	Q10P	1.03	0.99
pCS302	Q11A	2.35	2.92
pCS303	D12P	0.93	1.0
pCS304	D12A	1.06	0.94
pCS305	Y13P	1.0	0.8
pCS306	Y13A	0.93	1.16
pCS307	L14P	0.90	1.07
pCS308	L14A	1.03	1.16
pCS309	S15P	2.63	2.97
pCS310	S15A	3.33	3.45
pCS401	D16P	1.56	2.76
pCS402	D16A	1.91	5.5
pCS403	D16E	1.90	2.63
pCS404	H17P	1.23	2.04
pCS405	H17A	1.85	4.0
pCS406	L18P	1.11	1.14
pCS407	L18A	2.28	4.39
pCS408	W19Y	2.14	4.39

Construct	Mutation	500 µM Arg	2 mM Arg
pCS101	W19P	1.02	0.97
pCS102	W19A	1.09	1.14
pCS103	R20P	1.49	2.21
pCS104	R20A	1.27	2.58
pCS105	A21P	1.32	1.8
pCS106	L22P	2.39	4.08
pCS107	L22A	3.18	3.21
pCS108	N23P	2.54	2.82
pCS109	N23A	3.06	3.61
pCS110	A24P	2.25	2.28
pR301	Wild-type	2.80	3.92
pS301	D12N	1.02	1.05

1. The ratio of luciferase enzyme activity produced after 30 min in *N. crassa* translation extracts supplemented with 10 µM Arg to those supplemented with 500 µM or 2 mM Arg.
2. Values were normalized to the sea pansy LUC internal control. These values are representative of triplicate experiments (using at least two independent RNAs and two independent extracts).
3. A value >1.5 is representative of Arg-specific regulation and a value of 1.0 represents a loss in Arg-specific regulation.

Table 4.3 Arg analogs effect on AAP-mediated regulation

Arginine Analog	Regulator
D-Arginine	No
Agmatine	No
L-Argininamide	No
L-Arginine Methyl Ester	Yes
L-Arginine Ethyl Ester	Yes
L-Arginine Hydroxamate	No
L-Argininic Acid	No
L-N^G – Monomethyl Arginine	No
Phospho-L-Arginine	Yes
L-N^5 –Iminoethyl – Arginine	-
L-Canavanine	-
L-Homoarginine	No
L-Citrulline	No
L-Ornithine	No
L-Histidine	No
L-Lysine	No
Arg-Gly-Asp	Yes
Arg-Gly-Asp-Ser	Yes
Gly-Arg-Gly-Asp	No
Gly-Arg-Gly-Asp-Ser	No

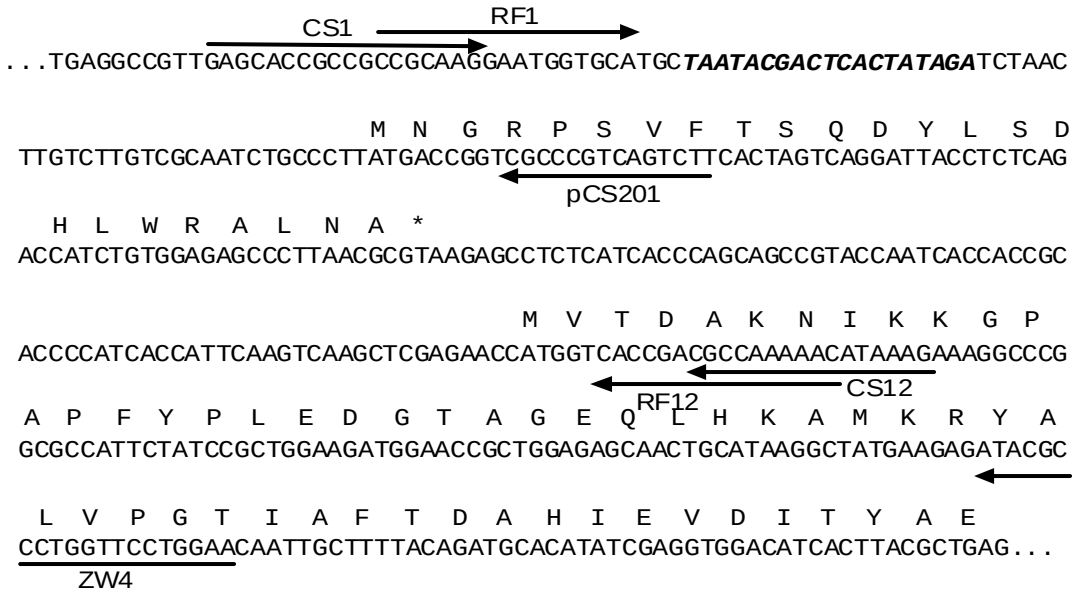

Figure 4.1 The 5' leader region of the *arg-2* LUC gene used in this study. The sequence begins with the T7 RNA polymerase-binding site and ends with the LUC coding region. The 24 amino acid sequences of the *arg-2* AAP and the N-terminus of LUC are indicated. The forward primers (RF1, CS1) and the final reverse primers (RF12, CS12) that were used for PCR are shown by horizontal arrows. One of the reverse primers used to introduce the individual mutations (pCS201) used for the first round of PCR is shown by a horizontal arrow (See Table 4.1 for all primers used for all individual mutations). The sequence for which the reverse complement was synthesized and used as primer ZW4 for toeprint analysis is indicated by a horizontal arrow below the sequence.

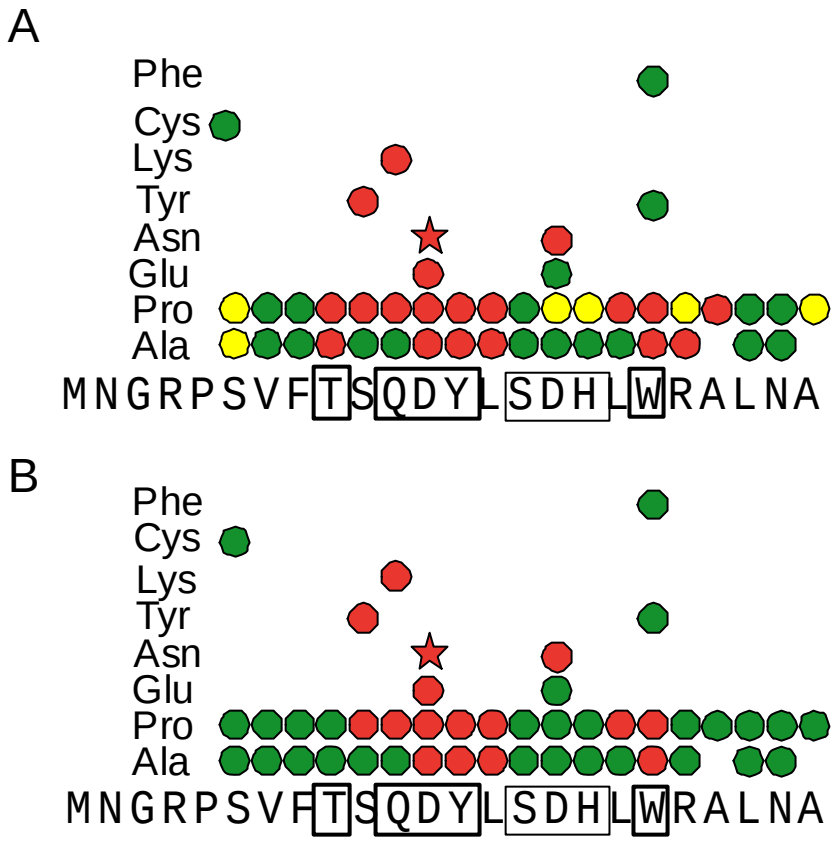

Figure 4.2. Summary of effects of changing individual amino acids on the function of the AAP. A. The wild-type *arg-2* AAP sequence is shown and the conserved residues are boxed. Residues 6-24 were changed to Ala and Pro individually or substituted with other residues as indicated and all AAPs were placed as uORFs in front of a luciferase reporter. In reaction mixtures containing low Arg (10μM) or high (500 μM) Arg; the effects of single amino acid changes in the region important for regulation were assessed by luciferase assays and toeprint assays. Red indicates regulation is lost by the specific changes, yellow indicates that regulation was reduced and green shows that regulation is retained. The red star indicates the loss in regulation in the D12N mutant which is the key mutant in these studies. B. Effects of single amino acid changes in reaction mixtures containing low Arg (10μM) or high (2 mM) Arg.

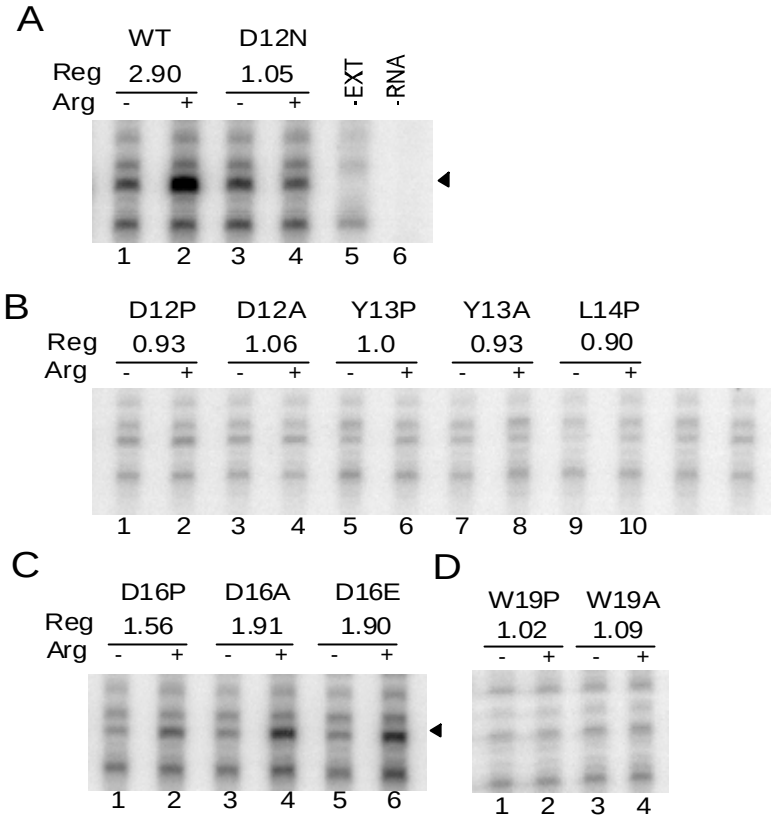

Figure 4.3 Toeprint analyses of the critical residues in Arg-specific regulation. The different constructs examined are indicated on the top. The ratio of LUC activity produced after 30 min in reaction mixtures containing 10 µM Arg vs 500 µM Arg are also indicated at the top. Equal amounts of synthetic RNA transcripts (120 ng) were translated in 20 µl reaction mixtures at 25 ºC. Reaction mixtures contained 10 µM (-) or 500 µM (+) Arg and 10 µM each of the other 19 amino acids. After 20 min of translation, 3 µl of translation mixtures were toeprinted with primer ZW4 and analyzed next to dideoxynucleotide sequencing of wild-type construct pR301 (Table 4.1).

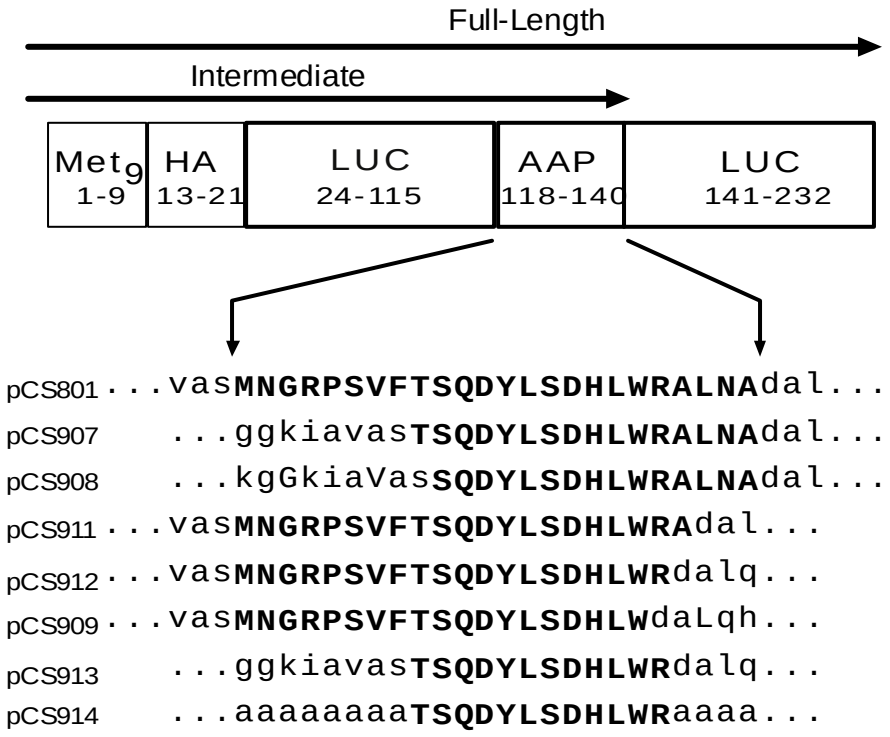

Figure 4.4 The different constructs that were used to analyze the minimal sequence required for an internal AAP to function. Deletions from the N- and C- terminal regions are compensated by the LUC sequences on both sides (in lower case). The LUC residues that are at the same positions as the original AAP residues are capitalized. Pulse-chase time course experiments and SDS-PAGE were performed to examine whether the modified internal AAP continued to cause ribosome stalling. The constructs corresponding to the respective mutations are labeled on the left.

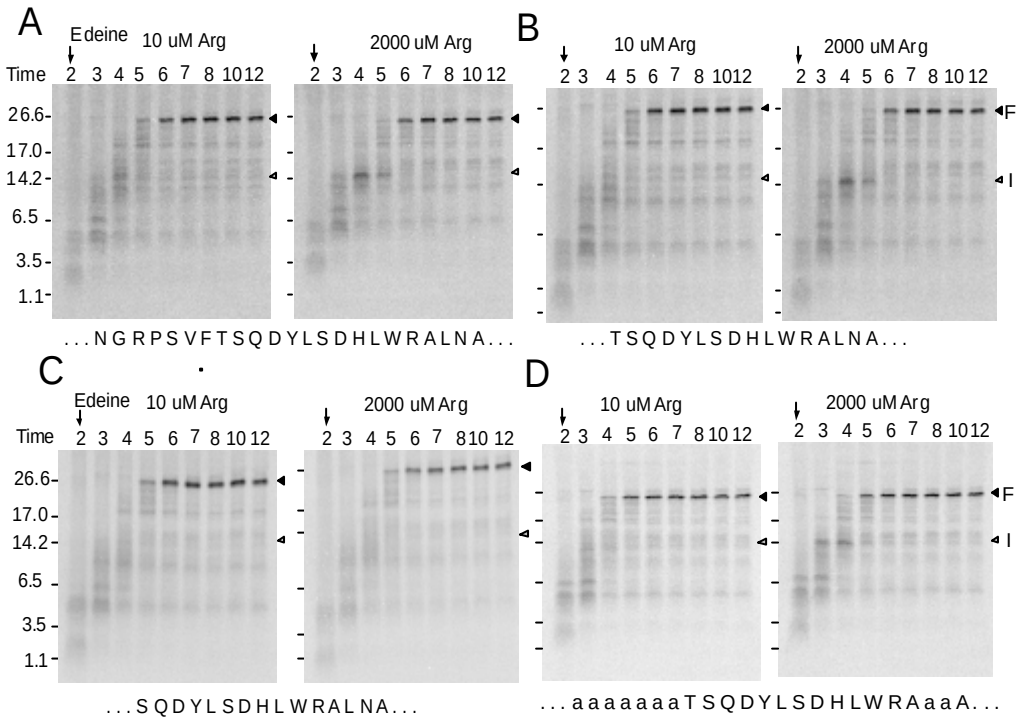

Figure 4.5 Polypeptide synthesis time course in *N.crassa* cell-free extracts. Transcripts specifying the full-length wild-type AAP as an internal domain starting from the second residue (panel A), specifying systematic N-terminal deletions T9-A24 (panel B) and S10-A24 (panel C) and containing the core AAP sequence consisting of residues T9-R20 (panel D), were examined to determine the minimal AAP sequence required for regulation. The mRNAs were translated in extracts supplemented with low (10 µM) or high (2 mM) Arg. Edeine was added at 2 min (arrow) and 10 µl aliquots of extracts were removed at the indicated time points for analysis by SDS/PAGE. Arrowhead I refers to the intermediate that accumulates when the AAP stalls ribosomes; arrowhead F refers to the full-length polypeptide. The full AAP and systematically deleted AAP sequences area shown below the corresponding panels in capital letters. In panel D, the flanking Ala residues that align with Ala-residues in the full AAP are indicated by capital letters; others are lower case.

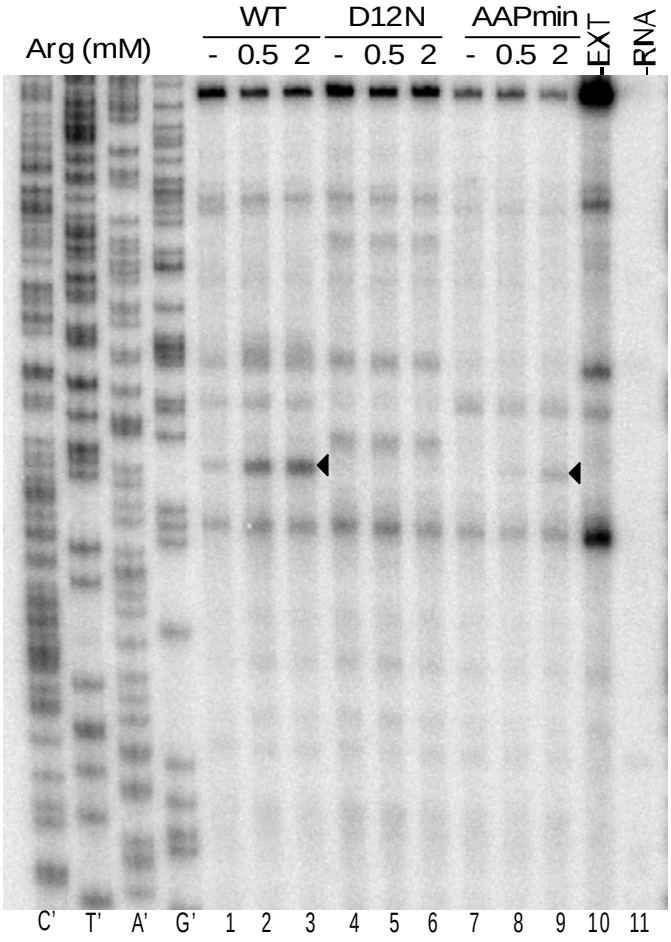

Figure 4.6 The minimized AAP functions as a uORF. The 12-residue core AAP was analyzed by toeprint assay and was compared to the wild-type and mutant (D12N) AAPs in reactions containing various Arg concentrations as indicated at the top. Sequencing reactions for the template encoding the wild-type AAP are on the left. The arrowheads indicate the toeprints corresponding to Arg-specific stalling of ribosomes at the uORF stop codon. The products of a reaction obtained from primer extension of wild-type RNA in the absence of translation mixture (-EXT), and extract in the absence of added RNA (-RNA) are shown as controls.

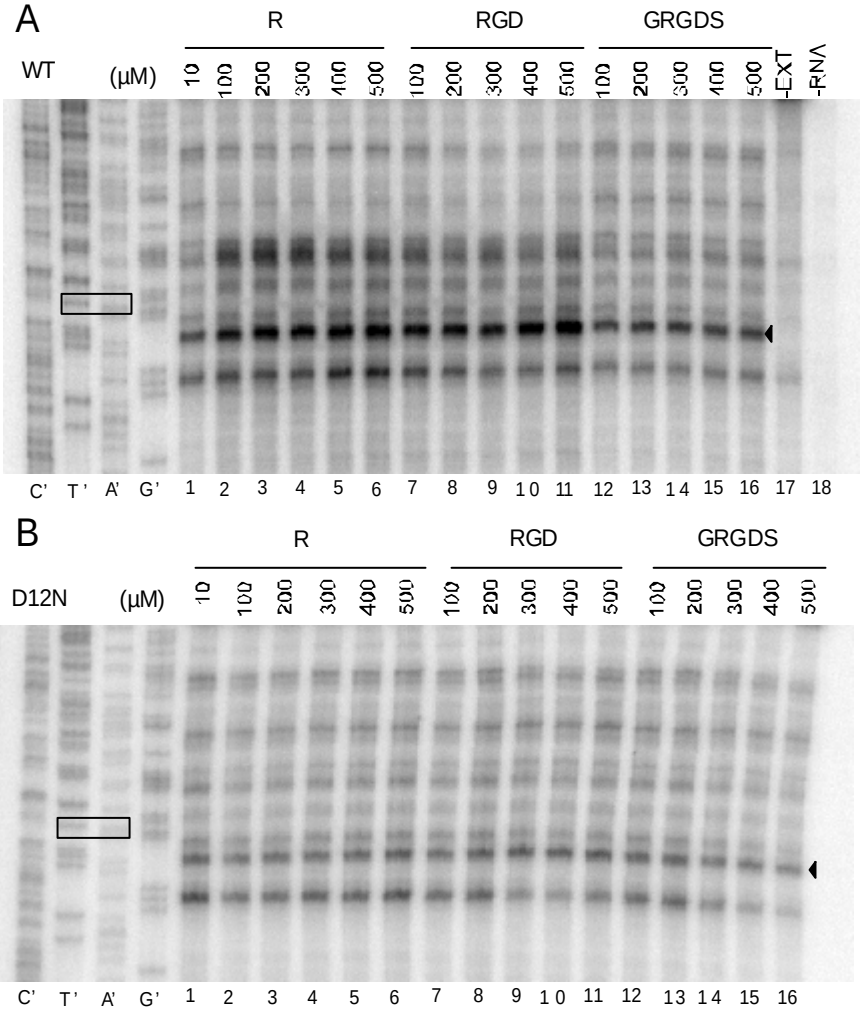

Figure 4.7 Effects of Arg, RGD and GRGDS on the AAP. Transcripts encoding either wild-type (WT) or D12N mutant uORF were examined in *N.crassa* extracts. Reactions were supplemented with different concentrations of Arg, RGD, or GRGDS as shown. The products obtained for primer extension of pure RNA in the absence of translation reaction mixture (-EXT) and from a translation mixture not programmed with RNA (-RNA) are shown for comparison.

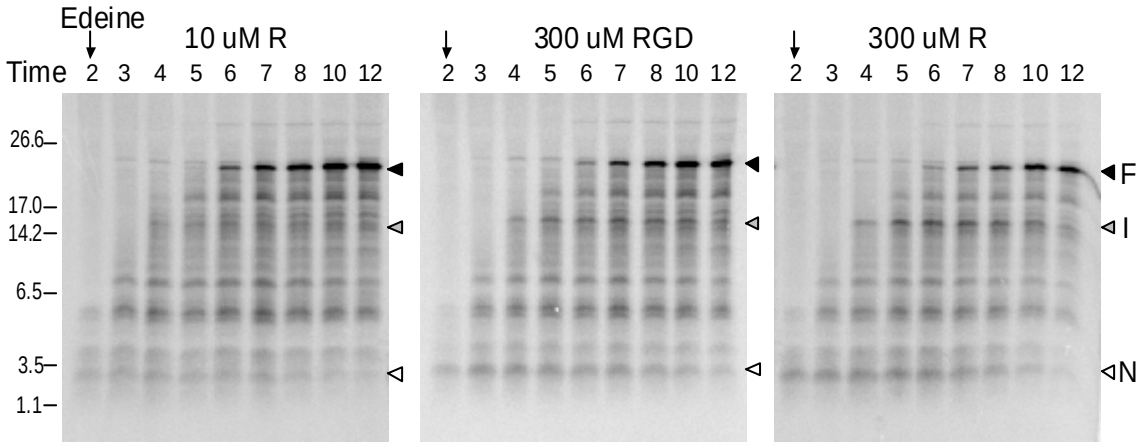

Figure 4.8 Time course of polypeptide synthesis with RGD. The RGD tripeptide confers regulation when added to translation reactions with *N.crassa* programmed with the Met_9-AAP-LUC-AAP-LUC mRNA and incubated at 25ºC with low Arg (10 μM), 300 μM RGD, and 300 μM Arg and 10 μM each of the other amino acids. Edeine was added at 2 min (arrow) and polypeptide products were analyzed by SDS/PAGE. Arrowhead N refers to intermediate Met_9-AAP, arrowhead I refers to intermediate corresponding to Met_9-AAP-LUC-AAP and arrowhead F referring to the full-length polypeptide.

CHAPTER 5
HER-2 UPSTREAM OPEN READING FRAME EFFECTS ON THE USE OF DOWNSTREAM INITIATION CODONS

5.1 INTRODUCTION

In eukaryotes, genes specifying products involved in growth control and development often contain short upstream open reading frames (uORFs) in their mRNA 5' leaders. One of these genes, *her-2* (*neu*, *erb*B-2), which encodes a 185-kDa transmembrane receptor tyrosine kinase that is a member of the epidermal growth factor receptor family, contains a single uORF in its mRNA. The HER2 protein is overexpressed in a large number of cancers including approximately 30% of breast cancers, and HER2 overexpression plays an important role in promoting metastasis (Slamon et al., 2001). The HER2 receptor level varies without changes in mRNA levels, indicating that post-transcriptional mechanisms participate in control of its expression (Buhring et al., 1995). Since the expression of many translation factors, such as eIF2α, is altered in different tumors (Clemens and Bommer, 1999), it is possible that the *her-2* uORF has a translational regulatory role during oncogenesis, as well as possible roles in growth and development.

Two distinct translational mechanisms are known to control HER2 protein synthesis (Child et al., 1999a). The first is a cell-type-dependent mechanism that causes translation of *her-2* mRNA to increase in transformed cells in comparison with primary cells. The second is a cell-type-independent repression of downstream translation mediated by the uORF. The work described here is aimed at understanding the roles in translation of the *her-2* uORF and the multiple in-frame initiation codons near the 5'end of the *her-2* coding region. Experiments with reporter constructs showed that the uORF was inhibitory to translation *in vitro* in mammalian, plant, and fungal systems and *in vivo* in mammalian cells. *In vitro* translation of *her-2* occurred primarily from the first in-

frame start codon when that start was the only one present. Primer extension inhibition (toeprint) mapping of ribosomes on the mRNA *in vitro* showed that ribosomes were clearly associated with the first *her-2* start codon in the absence but not the presence of the uORF. *In vitro* studies raised the possibility that alternative start codons might be used when the uORF was present and/or absent, and that the uORF could change the initiation event at the first *her-2* start codon so that it was not detected by toeprinting. Finally, additional features of the *her-2* leader rendered the translation of reporter transcripts exquisitely sensitive to ionic conditions in the heterologous fungal cell-free translation system.

5.2 RESULTS

5.2.1 The her-2 uORF inhibits downstream translation in cell-free systems based on luciferase assays

To explore the role of the *her-2* uORF in controlling translation, we used synthetic capped and polyadenylated mRNAs containing sequences derived from the 5' region of the *her-2* transcript, including the uORF and part of the HER2 coding region, fused to the firefly luciferase (LUC) reporter (Figure 5.1A). For constructs pCS604-609, 51 codons specifying the amino terminus of HER2, including three in-frame Met (initiation) codons were fused to LUC; for short constructs pCS701-704, the first two codons of *her-2* were fused to LUC (Figure 5.1B). Rabbit reticulocyte lysates, wheat germ extracts, and *Neurospora* extracts were programmed with mRNAs derived from these constructs.

The mRNAs specifying LUC fused to HER2 after the second codon (Figure 5.1B) produced active enzyme when *her-2* AUG1 was present (Figure 5.2A). LUC activity in the pCS701 (+/+) and pCS702 (-/+) constructs was derived primarily from initiation at *her-2* AUG1 because eliminating this codon (constructs pCS703 (+/-) and pCS704 (-/-)) substantially reduced LUC activity in each extract tested. When the wild-type construct containing the uORF and *her-2* AUG1 (pCS701, +/+) was compared with that lacking the uORF (pCS702, -/+), an approximate 2-fold inhibitory effect of the uORF was observed using rabbit reticulocyte lysates. Under optimized salt conditions in *Neurospora*, the

inhibitory effect of the uORF was comparable to that of reticulocyte lysates. An approximate 10-fold inhibitory effect was seen in the wheat germ system.

When *N. crassa* extracts were used to assay the translational function of the *her-2* uORF with the constructs in which LUC was fused to HER2 after its second codon, the standard reaction conditions used previously did not support translation of the HER2-LUC reporters (Figure 5.3). We found that, in *N. crassa* extracts, the synthesis of the fusion polypeptide from these transcripts was highly sensitive to K^+ and Mg^{2+} concentrations. While the K^+ and Mg^{2+} concentrations used previously (150 mM and 3.75 mM, respectively) did not support translation of these mRNAs, when they were adjusted to 100 mM and 1.75 mM, there was a two-order of magnitude increase in luciferase production from *her-2*-LUC transcripts containing HER2 AUG1. The reason why changing these ion concentrations in the fungal extract so dramatically affected the translation of mRNA containing the *her-2* sequences is not known, but presumably, these ions influence the structure of the mRNA (which is 80% GC in its 5'-leader) in a way that affects translation or affects the activity of factors involved in translation. The effect of these changes in ion concentrations on transcripts lacking *her-2* sequences was small (less than 2-fold) (data not shown).

To examine the effects of the uORF *in vivo*, transient RNA transfections into HeLa cells were performed using short pCS701-704 synthetic mRNAs and the luciferase they produced was measured (Figure 5.2B). Luciferase synthesis *in vivo* was strongly dependent on *her-2* AUG1. The uORF reduced LUC activity ~15-fold. These results using transfected mRNAs are consistent with previous experiments using a different reporter system and plasmid transfections assays, which showed the uORF had a ~5-fold effect in both primary and transformed cells (Child et al., 1999a).

Luciferase activity measurements showed that the long pCS604-609 constructs that contained 51 residues of *her-2* fused to LUC were complicated by the observation that constructs lacking *her-2* AUG1 produced more enzyme activity per unit mRNA than constructs containing AUG1, instead of substantially reducing it (data not shown). The simplest explanation for this is that the 51-residue N-terminal extension reduced specific activity of the reporter enzyme, and that shortening the N-terminus reduced the inhibitory effect. Thus, it is complex to assess enzyme activity data by direct comparison of

constructs, since the uORF appears to affect downstream start site selection (see below), but these results were consistent with the uORF having an inhibitory effect when matched constructs (either containing or lacking AUG1) were compared (data not shown).

5.2.2 The her-2 uORF both quantitatively and qualitatively affects translation of the luciferase reporter based on [^{35}S] –Met labeling

To assess the effect of the uORF and the potential use of alternative start sites, [^{35}S] Met labeling experiments were performed using equal amounts of mRNAs from both long and short constructs to program rabbit reticulocyte lysates (Figure 5.4A). Reaction mixtures programmed with firefly LUC encoding mRNA lacking *her-2* sequences were used as controls. In reticulocyte lysates, the long construct lacking the uORF (pCS604, -/+) produced a single major product whose migration was consistent with its predicted mass of 68-kDa (Figure 5.4A, lane 2). In contrast, the long construct which contained the uORF (pCS605, +/+) produced two major products in reticulocyte lysates (Figure 5.4A, lane 1), an upper band that could represent initiation at *her-2* AUG1 (~68 kDa) and a lower band (~63 kDa) presumably resulting from initiation at an in-frame downstream start codon (see Figure 5.1A). The absence of the lower band from lysates programmed with the pCS604 (-/+) mRNA argues against the lower band representing a degradation product or a C-terminally truncated product. Quantitation of radioactivity indicated that the larger product was reduced approximately 2-fold by the uORF (Figure 5.4A, compare lanes 1 and 2). These data are consistent with the uORF having both quantitative and qualitative effects on downstream *her-2* translation.

Surprisingly, in reticulocyte lysates, regardless of the presence (pCS608, +/-) or absence (pCS609, -/-) of the uORF, translation of mRNA from the long constructs that lacked *her-2* AUG1 generated polypeptide products of approximately the same size as those observed for the wild-type pCS605 (+/+) construct (Figure 5.4A, compare lanes 3 and 4 with lane 1). Similar results were found in a wheat germ cell-free system where two major products (an upper band at ~68 kD and a lower band at ~63 kD) were again observed after the translation of mRNA from the long construct pCS609 (-/-) (Figure 5.4B, lane 4). While elucidating which codon, in the absence of AUG1, was used to initiate the synthesis of the ~68 kD product has not yet been determined, it is evident that

this product was obtained in translation extracts from mammals and plants (Figure 5.4A and Figure 5.4B) from a template demonstrably lacking AUG1 (see sequencing reaction results for pCS609 in Figure 5.5).

For the short constructs in which LUC was fused directly after the second HER2 codon, (pCS701-704), the presence of the uORF also inhibited by two-fold protein synthesis from AUG1 (Figure 5.4A, compare lanes 6 and 7). For the short constructs lacking AUG1, full-length fusion polypeptides were not produced (Figure 5.4A, lanes 8 and 9), consistent with the loss of production of luciferase activity (Figure 5.2); indicating AUG1 was crucial for initiation in these constructs.

5.2.3 Toeprinting analyses of her-2 uORF control

A primer extension inhibition (toeprint) assay was used to directly map the positions of the ribosomes at *her-2* initiation codons. The purpose of these studies was to verify where ribosomes associated with the mRNA in the presence and absence of a functional uORF. All *her-2* constructs examined by LUC assay, [^{35}S]Met labeling, and RNA transfections were analyzed by toeprinting. Cycloheximide (CYH) was added to extracts prior to incubation of translation reactions (T_0) or added after the translation reaction was underway for 10 min. Adding CYH at T_0 should show where ribosomes first initiate translation, and adding it to extracts at T_5 or T_{10} can reveal primary initiation events or reinitiation events (Gaba et al., 2001).

The results of toeprinting mRNAs produced from the long constructs are shown in Figure 5.5A and 5B. The region downstream of AUG2 is not shown because there were no differences in signals among constructs. Regardless of when CYH was added, a toeprint corresponding to the uORF initiation codon were observed when the uORF was present (Figure 5.5A, lanes 1 and 2, Fig. 6B, lanes 1 and 2, and lanes 5 and 6) but not when it was absent (Figure 5.5A, lanes 3 and 4, Figure 5.5B, lanes 3 and 4, and lanes 7 and 8). These results indicate that the uORF initiation codon is recognized and likely used to start translation. In striking contrast, *her-2* AUG1 showed a strong signal when the uORF was absent (Figure 5.5A, lanes 3 and 4, Figure 5.5B, lanes 3 and 4) and did not yield a strong toeprint signal when the uORF was present (Figure 5.5A, lanes 1 and 2, Figure 5.5B, lanes 1 and 2, and lanes 5 and 6). The loss of this strong signal when both

the uORF and AUG1 were absent (Figure 5.5A, lanes 7 and 8; Figure 5.5B, lanes 7 and 8) verifies it corresponds to AUG1. There are some weak toeprint bands in lanes 7 and 8 that suggest the possibility that some ribosomes are recognizing the mutated AUG or alternatively could correspond to ribosomes recognizing the nearby CUG1. Possibly, some of the background bands in the region of gels in which toeprints corresponding to ribosomes at AUG1 are observed could arise from interactions between the RNA in this region and RNA-binding proteins. Importantly, when the start codons for both the uORF and AUG1 were mutated, a toeprint that corresponds to *her-2* AUG2 became stronger (Figure 5.5B, compare the signal in lanes 7 and 8 to the other lanes). However, there was no evidence of initiation at AUG3 after analyzing the toeprint gels. The shorter constructs showed similar results to the longer constructs (Figure 5.6) in that the majority of ribosomes associated with the first AUG encountered on the transcripts, but did not show significant fraction associated with AUG1 when the uORF was present.

5.3 DISCUSSION

HER2 is overexpressed in human cancers and its overexpression is associated with poor prognosis (Nicholson et al., 2001). In breast cancer, overexpression of HER2 occurs in ~30% of cases and is often attributed to gene amplification (Bofin et al., 2004). The bulk of the transforming events mediated by overexpression of HER2 are the result of enhanced signaling through the PI3K/Akt pathway. Cells sensitive to the cytotoxic effects of herceptin (which interferes with HER2 activity) become resistant upon expression of a constitutively activated Akt (Xia et al., 2004; Yakes et al., 2002). Previous studies have suggested that additional translational mechanisms modulate HER2 protein expression (Child et al., 1999b). The results described here strengthen and extend these studies by showing that synthesis of HER2 protein is quantitatively and qualitatively affected by its uORF. The uORF affects downstream translation in reticulocyte, wheat germ, and *Neurospora* cell-free systems as well as in transfected HeLa cells.

The [^{35}S]Met data (Figure 5.4) support the idea that the uORF is inhibitory to translation. Furthermore, they indicate the uORF causes some ribosomes to initiate from alternative in-frame downstream start codons. This would result in N-terminally

truncated products. A similar mechanism is seen in C/EBPα and C/EBPβ mRNAs containing a uORF (Calkhoven et al., 2000). In the case of *her-2*, this might generate a primary translation product that would lack its signal sequence and thus might not enter the ER as would be expected for full-length HER2. The presence of the uORF in *her-2* may, therefore, serve as a *cis*-regulatory element that controls the site of translation initiation and possibly reinitiation to AUG2 resulting in a truncated product (~62 kDa) as seen in Figure 5.4A, lanes 1 and 3. This truncated product may induce proliferation similar to the truncated products seen in C/EBPα and C/EBPβ (Calkhoven et al., 2000). The direct demonstration of the production of shorter polypeptides from the authentic *her-2* coding region is consistent with the results of reporter studies suggesting alternative downstream reinitiation at a heterologous ATG codon is stimulated by the uORF (Child et al., 1999b). Possibly, the uORF regulates reinitiation at different downstream start codons, as seen for *Saccharomyces cerevisiae GCN4* (Hinnebusch, 2005) and mammalian *ATF4* (Harding et al., 2000). That is, the uORF might enable ribosomes to access a more remote downstream initiation codon by reducing initiation at a more proximal one.

The toeprint data confirm that the uORF has a very strong effect on how ribosomes access the first *her-2* AUG codon, since no ribosome signal is detected there when the uORF is present. Since ribosomes are not observed there when the uORF is present, we cannot determine whether the ribosomes reach the *her-2* AUG codon by leaky scanning or reinitiation. Why ribosomes are not observed at the *her-2* AUG when the uORF is present is not clear. The experiments using the short constructs (pCS701-704) measuring LUC activity (Figure 5.2 and Figure 5.3) and [^{35}S]Met-incorporation (Figure 5.4) establish that *her-2* AUG1 is necessary for expression. Thus, the absence of a toeprint signal from this codon when the uORF is present (Figure 5.6) cannot be due to its not having a role in translation. One possibility for the absence of signal is therefore that ribosomes that arrive at the *her-2* AUG1 codon after translating the uORF do not bind as tightly and are lost during toeprinting compared to ones that arrive there when the uORF is absent.

N-terminally truncated HER2 receptors have been previously shown to have enhanced cell transformation activity compared with the full-length product (Bargmann

and Weinberg, 1988). These deletions within the extracellular domain (ECD) of HER2 have been shown to increase autokinase and transformation activities. Herceptin (trastuzumab), which binds to the ECD of HER2, was not effective against tumors expressing truncated products (Xia et al., 2004). Additionally, it did not block receptor activation and downstream signal transduction pathways that are involved in disease progression (Xia et al., 2004). The resistance to herceptin therapy may be due to increased levels of these N-terminally truncated proteins. The possibility that the uORF may promote ribosomal bypass of the first initiation codon to a downstream initiation codon signifies that further elucidation of how this uORF affects gene expression and protein synthesis will be valuable in the determination of therapeutic strategies.

5.4 EXPERIMENTAL PROCEDURES

5.4.1 Plasmids

Plasmid pEQ582 (Child et al., 1999b), containing the wild-type *her-2* 5'leader, uORF and coding region fused to β-galactosidase, was used as a template for PCR. Primer CCS3 (5'-CAAGAGATCTGCGCCCGGCCCCCACC-3') and CCS4 (5'-CCACCTGGTGACCTGGTAGAGGTGGCG-3') were used to amplify the region containing the *her-2* 5'leader, uORF and the coding region containing the first three *her-2* in-frame ATG codons. These primers also introduced a *Bgl*II restriction site upstream of the 5' leader and a *Bst*EII site downstream of *her-2* ATG3. Plasmid pEQ581 (Child et al., 1999b) differing from pEQ582 by a mutation of the uORF ATG codon to AAG, was also used as a PCR template. Due to the GC-rich 5'-leader, betaine (Sigma-Aldrich) was used in the PCR reactions at a 1 M final concentration to improve amplification (Henke et al., 1997).

The PCR products from pEQ582 and pEQ581 templates were digested with *Bgl*II and *BstE*II, gel purified and ligated to pR301 (Fang et al., 2000) which was digested with the same restriction enzymes which cut between the T7 promoter and the firefly luciferase reporter coding region. This resulted in constructs pCS604 and pCS605 that produced synthetic mRNA containing the *her-2* 5' leader followed by a HER2-LUC fusion reading frame. Additional constructs containing or lacking the uORF ATG and lacking *her-2*

ATG1 were made using mutagenic primers and megaprimer PCR to yield the set of plasmids pCS608-609 (Figures 5.1 and 5.2). Another matched set of constructs in the pR301 vector contained luciferase fused at codon two of HER2 (pCS701-704, Figures 5.1 and 2). These were obtained by PCR using primers CCS13 (5'-CTATAGATCTGCGCCCGGCCC-3') as the forward primer (containing restriction site *Bgl*II), and CCS17 (5'-AGGCGGTGACCTCCATGGTGCT-3'), CCS18 (5'-AGGCGGTGACCTCCTTGGTGCT-3') as the reverse primer (containing restriction site *Bst*EII).

5.4.2 Preparation of synthetic RNA

Plasmid templates were purified using the Wizard Plus Midi Prep kit (Promega). For *in vitro* studies, capped, polyadenylated RNA was synthesized with T7 RNA polymerase from *Eco*RI-linearized plasmid templates and yields of RNA were quantified as described (Wang and Sachs, 1997a). For *in vivo* studies, capped and polyadenylated mRNAs were prepared as previously described (Pelletier and Sonenberg, 1985); the yield of RNA transcripts were quantified by measuring UV-absorbance at 260nm and the quality of each RNA preparation assessed by SYBR gold (Molecular Probes; Eugene, OR) staining following fractionation on formaldehyde/1% agarose gels.

5.4.3 Cell-free translation and primer extension inhibition (toeprint) assays

Translation reaction conditions using nuclease-treated rabbit reticulocyte lysates and wheat germ extracts (Promega) were essentially those specified by the supplier, but with a final reaction volume of 10 μl instead of 50 μl. Preparation of translation extracts from *Neurospora crassa* and the reaction conditions for *in vitro* translation were as described (Wang et al., 1998; Wang et al., 1999), except that to achieve maximum activity for *her-2* containing reporters, K^+ and Mg^{2+} final concentrations were 100 mM and 1.75 mM respectively. For luciferase activity measurements, equal amounts (6 ng) of mRNA were used to program extracts; synthetic mRNA encoding Renilla (sea pansy) LUC was added to all reactions to serve as an internal control (Wang et al., 1998). All translation reaction mixtures were incubated at 25°C for luciferase assays; translation was

halted by freezing reaction mixtures in liquid nitrogen after 30 min (*N. crassa* and wheat germ) or 45 min (reticulocyte lysates).

For [^{35}S]Met labeling of polypeptides, synthetic RNA (60ng) were used to program reticulocyte lysates and wheat germ extracts and translation reactions were incubated for 45 min and 30 min respectively. [^{35}S]Met was used at a final concentration of 0.5 μCi/μl. The reaction mixtures for both reticulocyte lysates and wheat germ extracts were mixed with SDS-containing loading buffer but were not heated prior to SDS-PAGE. Radiolabeled products were examined by SDS-PAGE and analyzed by phosphorimaging.

The toeprint assays were accomplished as described (Wang and Sachs, 1997b) using primers ZW4 (5'-TCCAGGAACCAGGGCGTA-3') and primer CS3 (5'-CATGTCCAGGTGGGTCTC-3'). Cycloheximide (CYH) was added to a final concentration of 0.5 mg/ml as described (Gaba et al., 2001). The results shown are representative of at least three independent experiments. The radiolabeled products were examined using a Molecular Dynamics Phosphorimager and ImageQuant software.

5.4.4 Cell Culture and RNA Transfections

HeLa cells were cultured in Dulbecco's modified Eagle's medium (Sigma; St. Louis, MI) containing 10 % fetal bovine serum. Transient RNA transfections into HeLa cells were performed using DMRIE-C (Invitrogen; Carlsbad, CA), as specified by the manufacturer. Briefly, approximately 2 x 10^5 cells were seeded per well of a 6-well plate 18 hr prior to transfection. Cells were then transfected with 4 μg of capped and polyadenylated mRNAs from templates pCS701-704 and harvested ~ 16 hr post transfection. Transfection efficiencies were assessed by co-transfecting cells with Renilla luciferase RNA (1 μg/well) and luciferase activities were measured using the Dual-Luciferase assay. The results of three independent sets of transfection experiments were analyzed.

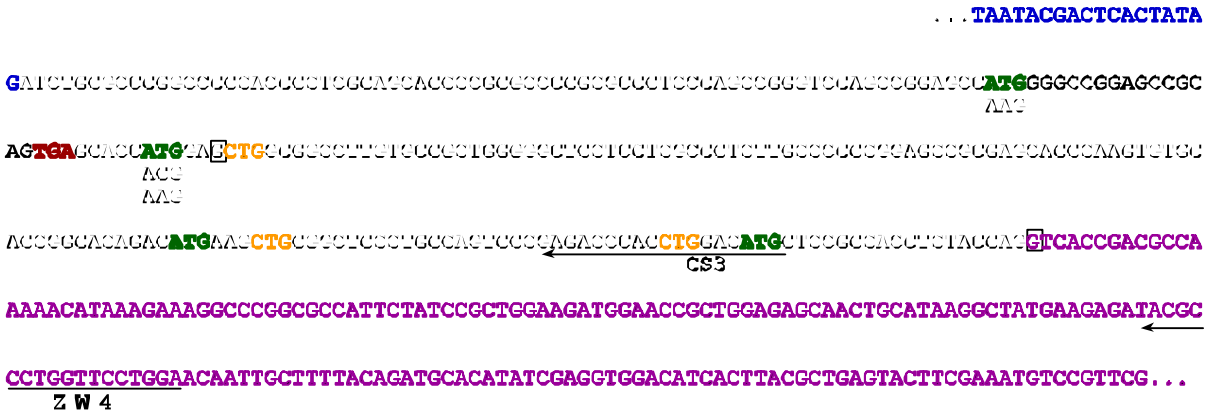

Figure 5.1A The 5' leader regions , *her-2* sequence, and constructs used in this study. The sequence shown begins with the T7 RNA polymerase-binding site (blue) and ends with the LUC coding region (purple). The upstream open reading frame ATG and the 3 in-frame ATGs (ATG1, ATG2, and ATG3) are in green. The in-frame CTGs are in yellow. The termination codon of the uORF region is in red and the mutations are shown directly below the sequence. The sequences whose reverse complements were synthesized and used as primers (CS3 and ZW4) for toeprint analyses are indicated by a horizontal arrow below the sequence. For the shorter constructs (pCS701-704), the luciferase reporter sequence was fused directly after the second codon after ATG1, the boxes in the sequence represent where the sequences were fused to make these shorter constructs.

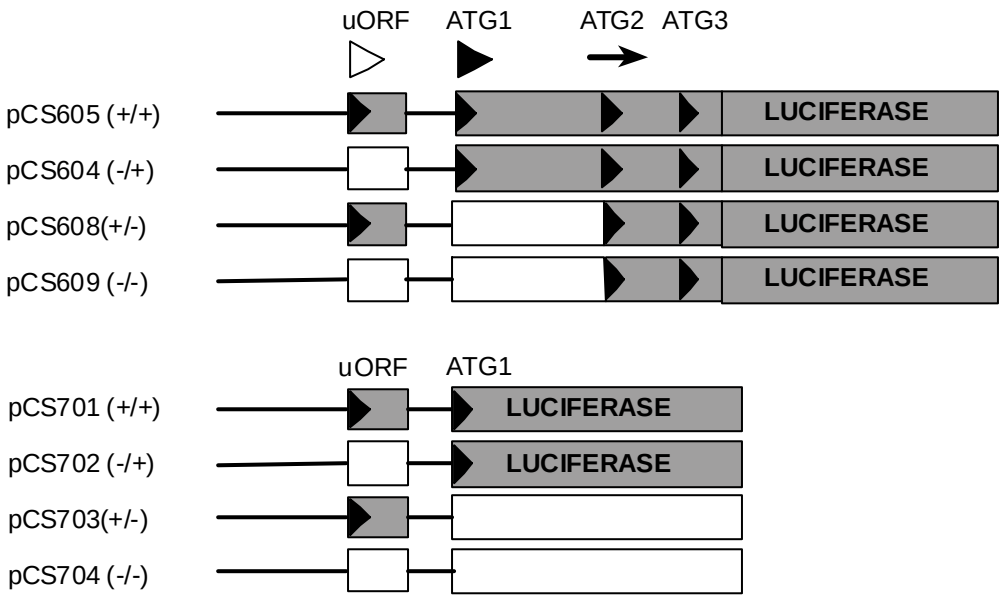

Figure 5.1B Constructs composed of the 97 nt *her-2* 5' leader including the uORF and intercistronic region, and the N-terminal coding sequence that includes 3 in-frame ATGs. Wild-type (pCS605) and mutated sequences (lacking the uORF ATG codon and/or the first ATG codon, pCS604, 608, and 609) were fused in-frame to a firefly luciferase reporter. The symbols in parentheses represent the presence and/or absence of an ATG (e.g. The -/+ indicates the absence of the uORF ATG and the presence of the first *her-2* ATG). These constructs were used to produce capped and polyadenylated synthetic RNAs. Shorter constructs composed of the 97 nt *her-2* 5' leader including the uORF region and the first 2 codons of *her-2* fused in-frame with firefly luciferase. The black triangles represent the ATG codons and the white boxes represent the coding regions that would not be translated if the ATG codons that were mutated were responsible for initiating translation. Additional symbols above the constructs correspond to the uORF ATG (open arrowhead), ATG1 (filled arrowhead) and ATG2 (arrow) respectively.

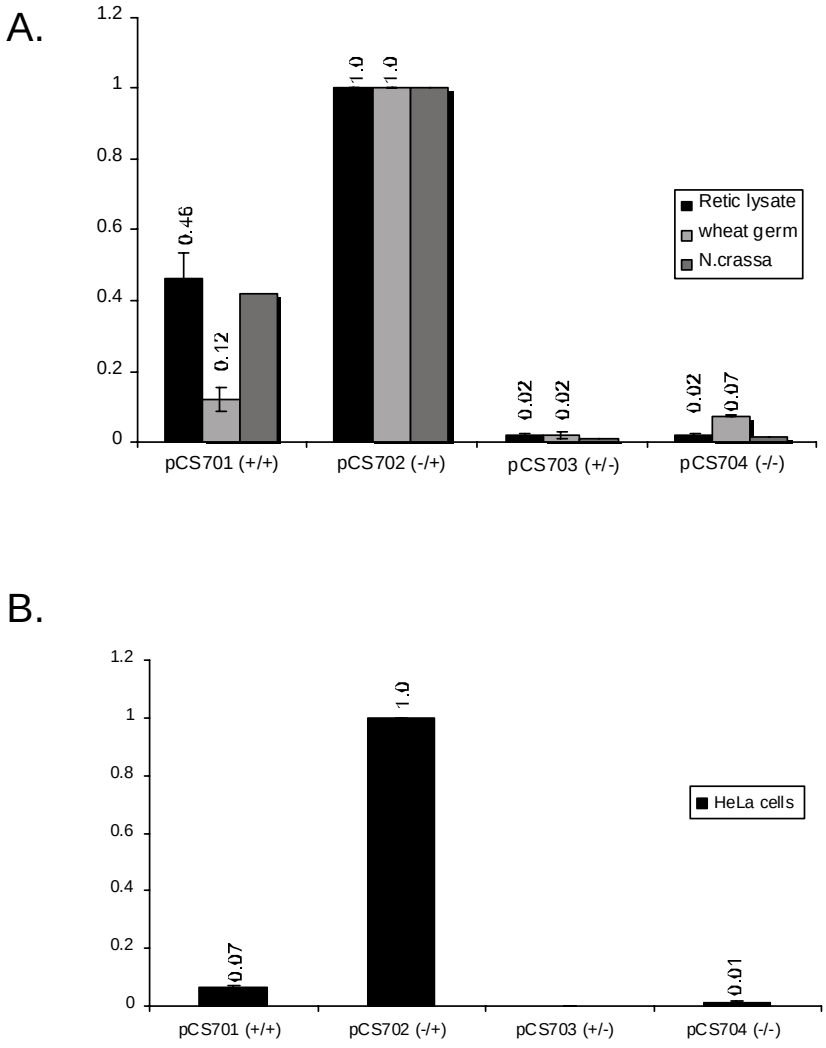

Figure 5.2 Luciferase enzyme activity assays. A. In-vitro translation reactions were performed for each extract and luciferase activity assays performed for rabbit reticulocyte lysates, wheat germ extracts, and *N. crassa* extracts. A dual luciferase (LUC) reporter assay system was used to measure firefly LUC and Renilla LUC (internal control) for constructs pCS701-704. All firefly LUC values were normalized to the internal control. The level of normalized firefly LUC expression from multiple independent experiments (two experiments using *N. crassa*, four experiments using wheat germ and reticulocyte lysates) with each construct were averaged and normalized to the expression from pCS702, and the standard deviations calculated for experiments using wheat germ and reticulocyte lysate (error bars). B. Luciferase enzyme activity obtained from RNA transfection into HeLa cells as describe in Experimental Procedures using RNA from constructs pCS701-704.

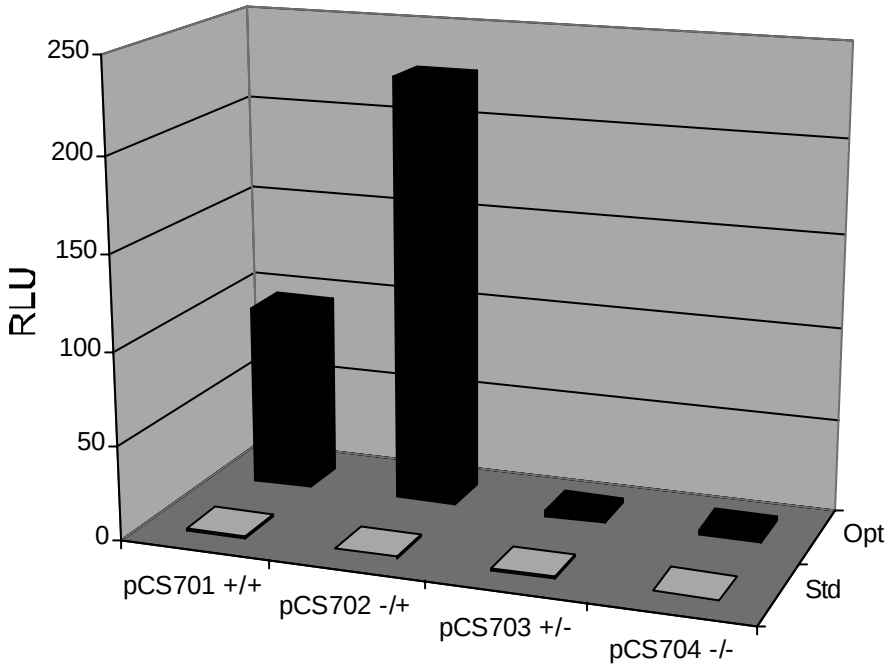

Figure 5.3 Luciferase activity assays show that translation of mRNAs from pCS701-704 containing the *her-2* 5'-UTR are highly sensitive to K^+ and Mg^{2+} concentrations in *N. crassa* extracts. *N. crassa in-vitro* translation reactions were performed in parallel using 150 mM K^+ and 3.75 mM Mg^{2+} salt concentrations (denoted Std, standard), or 100 mM K^+ and 1.75 mM Mg^{2+}, which yielded optimal synthesis of luciferase from mRNAs containing the *her-2* 5 leader (denoted Opt). After 30 min of incubation at 25ºC, LUC synthesis was determined by measuring enzymatic activity [expressed as RLU (relative light units)] using 5 µl aliquots from each reaction mixture. The *N. crassa* optimized salt concentrations showed less than a 2-fold effect on translation of the mRNAs lacking *her-2* sequences (T7-LUC) but a 200-fold effect on transcripts containing the *her-2* 5'-UTR (data not shown). Using optimized salt concentrations, the construct lacking the uORF (pCS702) showed a ~2-fold increase in reporter expression compared to wild-type (pCS701).

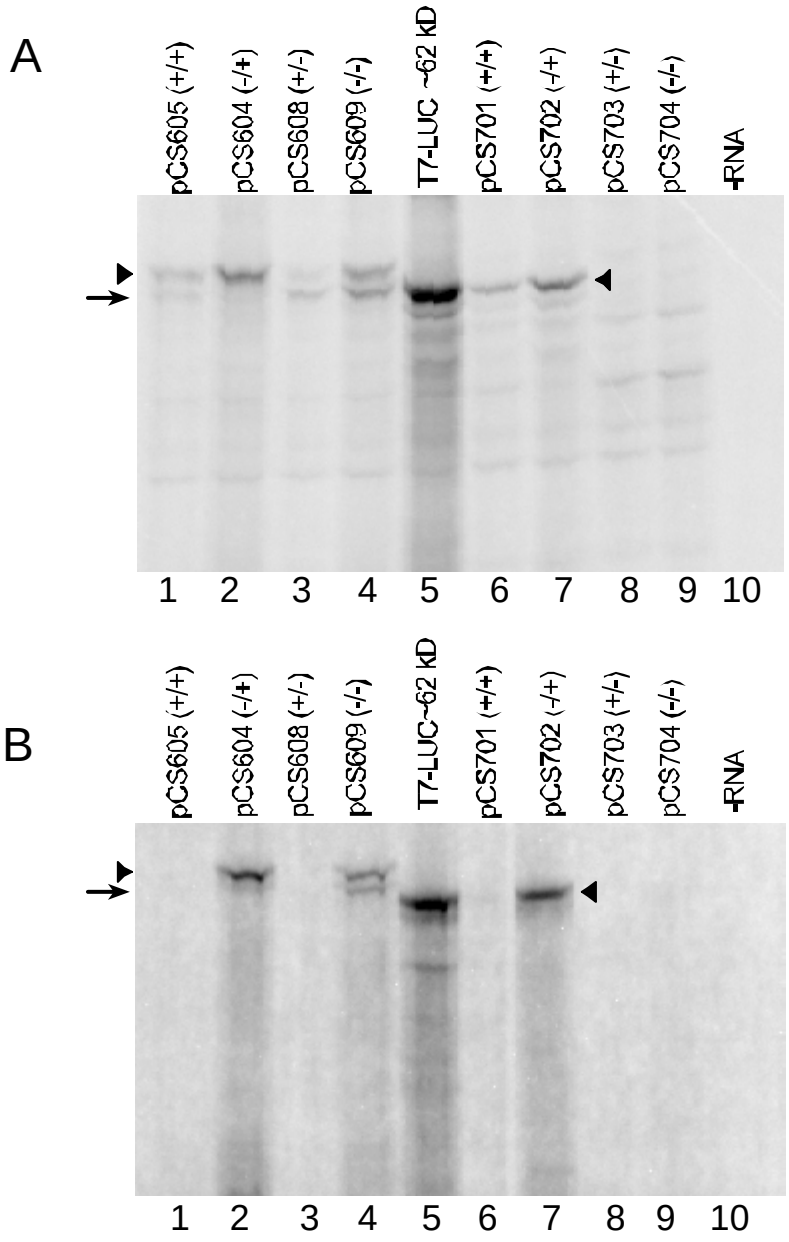

Figure 5.4 [^{35}S]Met labeling of products obtained from each of the constructs shown in Fig.1B. A. Synthetic RNA (60 ng) were used to program reticulocyte lysates (10 μl). Reactions were incubated for 45 min at 25ºC. Radiolabeled products were examined by SDS-PAGE and phosphorimaging. Controls included a reaction mixture programmed with a firefly LUC encoding mRNA lacking *her-2* sequences (T7-LUC) and a reaction mixture to which no RNA was added (-RNA). B. Synthetic RNA (60 ng) were used to program wheat germ extracts (10 μl). Reactions were incubated for 30 min at 25ºC. In each panel, filled arrowheads correspond to polypeptide products (where present) arising from (i) initiation at AUG1 (lanes 6-7); (ii) initiation at AUG1 or possibly at a nearby CUG codon (lanes 1-4). The arrow corresponds to polypeptide products (where present) arising from possible initiation at AUG2 (lanes 1-4).

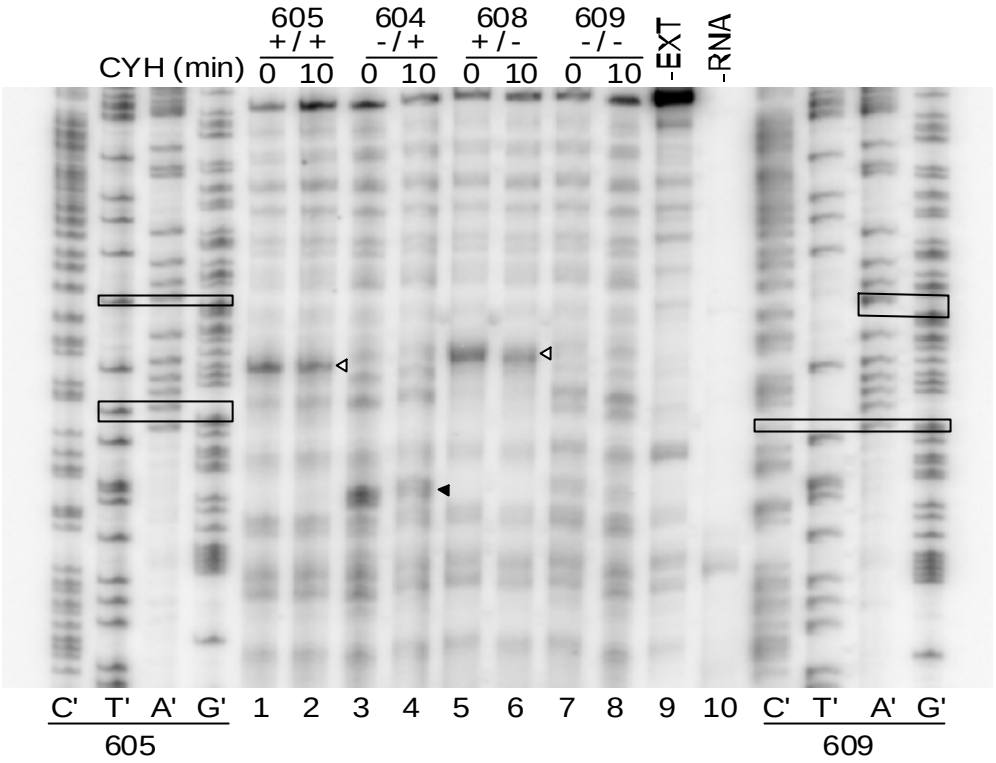

Figure 5.5A Primer extension inhibition (toeprint) assays examine initiation on transcripts containing or lacking the *her-2* uORF AUG. A. Rabbit reticulocyte lysates (10 μl) were programmed with mRNA (60 ng) from long constructs pCS604-609. Cycloheximide (CYH) was added at time 0 (prior to incubation of translation reactions) or 10 min (when translation was underway). The reactions were incubated for a total period of 15 min and toeprinted with primer CS3. Toeprints obtained when CYH is added at time 0 should show where ribosomes first initiate translation; toeprints obtained when CYH was added after 10 min should reveal where primary initiation events and reinitiation events are occurring. Open arrowheads correspond to the uORF and filled arrowheads correspond to AUG1.

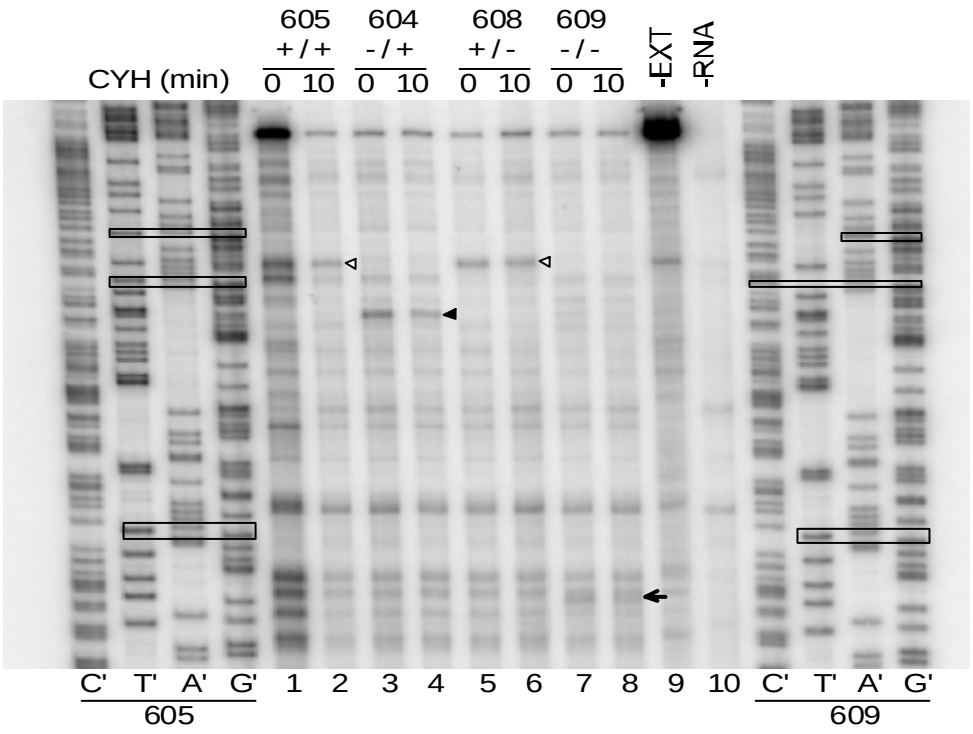

Figure 5.5B Toeprint assay using a primer (ZW4) that is further downstream for visualization of the second *her-2* AUG codon. When both the uORF AUG and HER2 AUG1 are mutated, ribosomes load on AUG2 (arrowhead). In these reactions, CYH was added at time 0 and after 10 min with a total reaction time of 15 min. These results indicate that most ribosomes initiate translation predominantly at the first AUG codon encountered in the transcript. Open arrowheads correspond to the uORF, filled arrowheads to *HER-2* AUG1, and the arrow to *HER-2* AUG2. Sequencing reactions of both 605 (+/+) and 609 (-/-) are shown on the left and right sides of the toeprint gel respectively. The sequencing reactions in panel A displayed some migration artifacts due to compression in the GC-rich regions. In the sequencing lanes, the uORF AUG, *her-2* AUG1, and mutations that eliminate each, are boxed in both panels; *her-2* AUG2 is boxed in panel B. The increased intensity of the specific signals in Lane 1 is due to an increase in the amount of cDNA sample loaded.

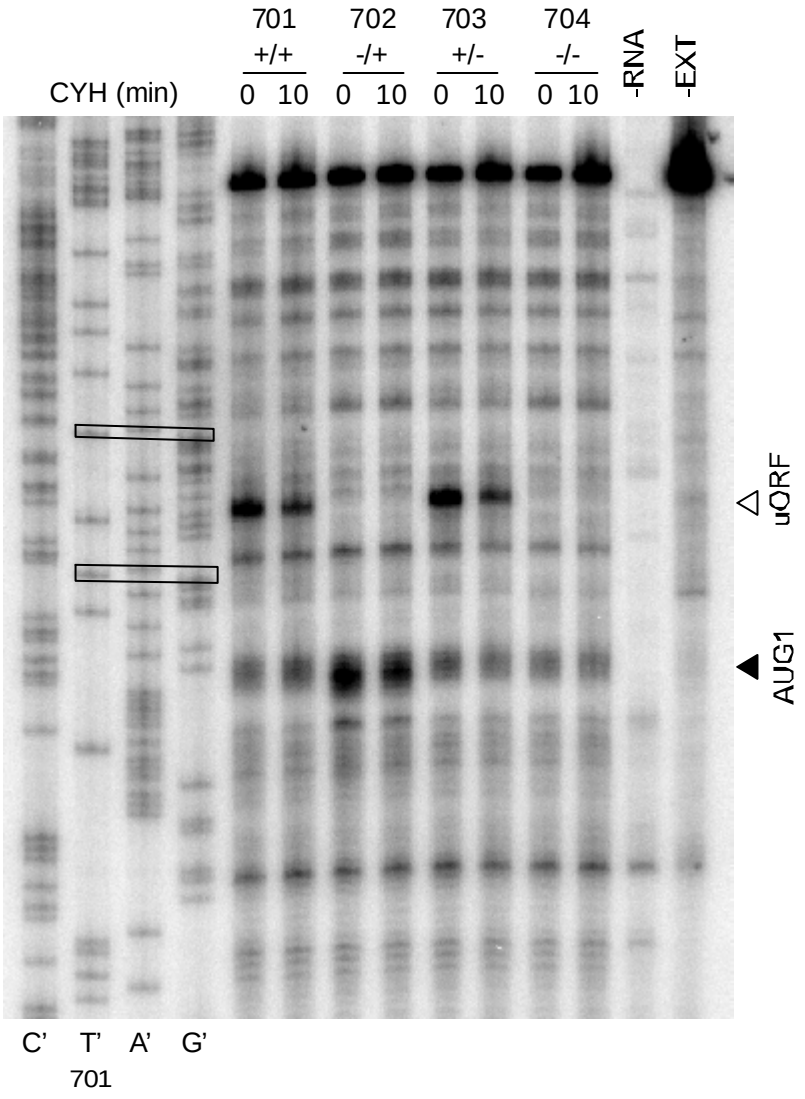

Figure 5.6 Toeprint assays to examine initiation on transcripts containing or lacking the *her-2* uORF AUG and/or *her-2* AUG1 using shorter constructs, pCS701-704. Rabbit reticulocyte lysates were programmed with mRNA and CYH was added at time 0 or after 10 min and reactions were incubated for a total period of 15 min. Ribosomes are preferentially associated with the uORF AUG when it is present. In the sequencing lanes, the uORF AUG and *her-2* AUG1 are boxed.

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

6.1 SUMMARY OF RESEARCH

6.1.1 Systematic mutagenesis reveals the importance of key residues in the AAP

Alanine and proline-scanning mutagenesis studies were completed to examine the effect of replacing residues with either Ala or Pro, leaving other residues unaffected. Analyses of the regions crucial for AAP function showed that relatively conserved residues can be mutated and yet the peptide responds to 2 mM Arginine. The absolutely conserved D12, Y13 and L14 cannot be changed to Ala or Pro but the D16 residue can be mutated to Ala, Glu, or Pro and function is maintained at 2 mM Arg; the D16N mutation eliminates function comparable to the D12N mutation. The D16, H17, and L18 can be individually changed to Ala while function is maintained. Ala was substituted for S15, D16, H17 (all possible combinations). All three residues can be converted to Ala simultaneously and regulation maintained, although changing two residues reduced regulation and changing at all three residues reduced regulation further when compared to wild-type and single residue changes. Single residue addition and deletion mutations surrounding the S15, D16, and H17 region abolishes Arg-specific regulation and contributes to the importance of the length of the AAP and conserved residues. The highly conserved S15 can be mutated to Ala or Pro and still function. The yeast *Kluyveromyces lactis* predicted AAP contains Gln at this position and this was tested in vitro and did not confer regulation due to that larger side chain of Gln which would have greater impact on the structure of this region. However, a diverged basidiomycete AAP was shown to retain function to stall ribosomes in an arginine-specific manner. Therefore, within the evolutionarily conserved region, some residues are more important than others for AAP to function which confirms the previous study concluding that the coding region is not what is responsible for regulation by Arg.

6.1.2 The AAP functions as an internal domain that stalls ribosomes in response to arginine

Our studies support the conclusion that the AAP can function as an internal nascent peptide domain that causes regulated stalling of eukaryotic ribosomes in response to Arg as seen by the pulse-chase protocol and toeprinting assays. This was examined by performing pulse-chase experiments with mRNAs encoding fused polypeptide sequences in cell-free translation systems. The wild-type AAP caused Arg-regulated stalling whether placed at the N- terminal or internal position in a large polypeptide. Stalling occured in fungal, plant, and mammalian in vitro cell-free systems. These findings demonstrate that an internal polypeptide domain in a nascent chain can regulate eukaryotic translation elongation in response to a small molecule. These results provide precedents for translational strategies that would allow domains located within nascent polypeptide chains to modulate gene expression.

6.1.3 Analyses of the evolutionarily conserved features of various fungal AAPs show importance of certain sequences required for regulation

We performed comparative analyses of the genomic structure and predicted peptide sequence of the AAPs across 51 fungi and observed divergence across phyla. Differences detected at the genomic level included variation in intron and position within the AAP and CPS-A coding regions and differences in known regulatory motifs. Although differences exist in AAP sequences, there were these universally conserved amino acid residues corresponding to *N. crassa* AAP residues D12, Y13, and N16. There was a notable difference and a change-shift in the Basidiomycete sequences (an additional Asp on one side of the core; and a loss of His on the other). However, a diverged Basidiomycete AAP from *Cryptococcus neoformans* still functioned in vitro.

6.1.4 The minimal sequence required is highly conserved and is sufficient for regulation as an internal domain and as a uORF

The discovery that the AAP functioned as an internal peptide that could be assayed by the pulse-chase protocol allowed systematic removal of AAP-specific residues from each end without concern that the absolute length of the nascent peptide would be altered with respect to the ribosome. We concluded that the 12-residue AAP core, residues T9-R20, is necessary and sufficient to confer regulation. A uORF consisting only of a Met codon, Ala residues at positions that are not critical for regulation and the T9-R20 AAP core residues confer regulation when placed as a uORF, as determined from luciferase and toeprint assays.

6.1.5 The study of Arg analogs provide insights into the structural requirements for regulatory function

Since the level of aminoacylated Arg tRNA is not important for regulation, Arg or its metabolite must be important. Many Arg analogs were tested for their effects on regulation. The results from these experiments showed that the chiral center, the guanidino group and primary amino group are most important for regulatory function. Based on these results, we then proceeded to test the effects of short peptides RGD and GRGD. We obtained evidence from LUC assays and toeprinting that the tripeptide RGD stalled ribosomes but GRGD did not, consistent with the importance of the primary amino group of Arg. Pulse-chase analyses also indicated that RGD can replace Arg in vitro to stall ribsomes. The fact Arg-containing peptides exert regulation allows for future experiments aimed at identifying the site of Arg regulatory action such as crosslinking experiments.

6.1.6 The effect of *her-2* uORF on the use of downstream initiation codons

The proto-oncogene, *her-2*, has a uORF that reduced translation of the downstream cistron in *N.crassa*, wheat germ, and reticulocyte in vitro cell-free systems and RNA-transfected cells in vivo. Protein labeling experiments of in vitro translation products obtained indicated that the uORF also affected downstream start site selection.

Toeprinting assays showed that ribosomes loaded at initiation codons in reticulocyte lysates and indicated that the uORF affected the interaction of ribosomes with the primary *her-2* AUG codon.

6.2 FUTURE DIRECTIONS

We completed alanine and proline scanning mutagenesis studies to determine the specific residues, in addition to length, required for Arg-specific regulation by the AAP. The AAP has sequence similarities with bacterial TnaC and SecM and these prokaryotic peptides also are involved in the stalling of elongating ribosomes. Since ribosomal protein L22 in bacterial systems seems to be involved near the site in the tunnel responsible for ribosome stalling, the eukaryotic homolog L17 may play an important role in this process in terms of the AAP. After residue comparisons, it was determined that crucial residues in nascent TnaC, SecM, and AAP all function at a similar distance from the peptidyl transferase center which is approximately 12 residues. Experimental approaches to determine whether there is an AAP structural change in response to Arg in the ribosomal tunnel is crucial in understanding the mechanisms involved.

One method that can address the AAP's structure in the ribosomal tunnel is to measure the length of the AAP and its derivatives near the peptidyl transferase center and then compare these to a nascent peptide with already established lengths. If there is a structural change in the AAP, an extended conformation occurs in compact α-helical conformation which would place the core AAP residues in different regions of the ribosome tunnel. A method to measure nascent peptide compaction has already been developed (Lu and Deutsch, 2005). This method uses a reference nascent peptide containing a reporter Cys as a "tape measure" used to "PEGylate" the reporter Cys. PEGylation increases the mass of the peptide and quantification of the rate and extent of PEGylation provides a measure of relative accessibility of the Cys residue. By placing the Cys at defined positions in the nascent chain relative to the peptidyl transferase center, it is possible to measure the length of the nascent chain by examining the accessibility of Cys to PEG-MAL. The compaction of the AAP in the tunnel can be determined by comparing Cys accessibility in a tape measure. An engineered and well-characterized N-

terminal segment of potassium channel Kv1.3 containing wild-type and mutated AAPs can be used to measure PEGylation of these AAP-substituted tape measures in the presence and absence of Arg relative to compacted and extended E64C controls. I have started to make constructs for these studies.

Most of the work described in this thesis has been done in an in vitro system. Due to the AAP's functional similarity to TnaC, the progress in characterizing the ribosome structures of bacteria, and the availability of selectively changed sequences of the proteins and rRNAs that comprise these ribosomes, the bacterial system is a good candidate system for studying AAP function. I have already made constructs containing the TnaC Shine-Dalgarno sequence proximal to the double AAP construct (Met_9-AAP-LUC-AAP-LUC) coding region containing either two wild-type or two D12N AAPs; these have been used for in vitro translation in *E.coli* S30 extracts. Preliminary results showed that peptidyl-tRNA accumulated at 2 positions for the wild-type construct which corresponded with the two predicted stalls that was previously seen in *N.crass*a extracts. Currently, Arg-dependence has not been observed in these extracts. However, after modifications to the assays, continued exploration in this bacterial system might be valuable in understanding AAP mechanisms.

After extensive analysis of the AAP and determining whether the AAP interacts with Arg, the previous experiments point to the residues in the conserved core, and possibly Asp-12 as the residue that interacts with Arg. In principle, such interactions might be tested by 2D-NMR. Analyses of the AAP's structure in the presence or absence of Arg indicated that the structure in the ribosome tunnel and in solution were similar.

Arg- regulated ribosome stalling by the nascent AAP is hypothesized to be stabilized by high Arg concentrations. The interaction of the AAP with the ribosome causes perturbation of the completion of the peptidyl-transferase reaction. This altered interaction of the nascent AAP and the ribosome should be detectable by incorporating photoreactive moieties into the nascent peptide and examining the components of the translational apparatus that crosslink to the nascent wild-type peptide but not the D12N peptide in high versus low Arg. I have already made constructs containing 9 Met codons at the N-terminus followed by an affinity tag for wild-type and D12N AAPs with amber codons and lysine codons at positions 7, 10, 17, and 22. Nitrene-generating cross-linker

N^{ε}-(5-azido-2-nitrobenzoyl) (εANB) and carbine-generating equivalent N^{ε}-4-(3-trifluoromethyldiazirino) benzoyl (εTDB) can be placed into the AAP at selected codons where Lys does not interfere with function. The photoactivation of the cross-linkers is expected to covalently attach to the nascent AAP to nearby components of the translational machinery which would be informative in explaining Arg-specific regulation by the AAP.

The proto-oncogene, *her-2*, encodes for the protein HER2 that is overexpressed in human cancers. The HER2 protein is quantitively and qualitively affected by its uORF; this uORF also affects downstream translation in cell-free systems. Our in vitro studies set a foundation for future in vivo studies. We have demonstrated the possible usage of alternative initiation codons, such as a CTG codon. The roles of the in-frame CTG codons and their influence on downstream translation can be investigated by synthesizing constructs that contain these CTGs in addition to constructs where they are eliminated. Start site selection can prove to be extremely significant, especially when targeting for therapeutic strategies, since truncated products may induce proliferation and result in poor prognosis.

Overall, my studies regarding the two genes, *arg-2* and *her-2*, both containing upstream open reading frames, prove to be significant in the understanding of translational control mechanisms. uORFs can not only influence or inhibit downstream translation of the ORF, but it can also influence start site selection, as seen in *her-2*. Determination of the specific *cis*-acting requirements for translational control in *arg-2* will benefit other studies with genes that contain a uORF under similar controls.

LITERATURE CITED

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986). mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol 6*, 2932-2943.

Agrawal, R. K., Penczek, P., Grassucci, R. A., and Frank, J. (1998). Visualization of elongation factor G on the Escherichia coli 70S ribosome: the mechanism of translocation. *Proc Natl Acad Sci U S A 95*, 6134-6138.

Alderete, J. P., Child, S. J., and Geballe, A. P. (2001). Abundant early expression of gpUL4 from a human cytomegalovirus mutant lacking a repressive upstream open reading frame. *J Virol 75*, 7188-7192.

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.*, et al.* (2002). Development and characterization of a reconstituted yeast translation initiation system. *RNA 8*, 382-397.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol 215*, 403-410.

Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell 113,* 673-676.

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. F., Anand, M., Boesen, T., Van, L. B., Kinzy, T. G., and Andersen, G. R. (2004). Purification and crystallization of the yeast translation elongation factor eEF3. *Acta Crystallogr D Biol Crystallogr 60,* 1304-1307.

Åström, S. U., von Pawel-Rammingen, U., and Byström, A. S. (1993). The yeast initiator tRNAMet can act as an elongator tRNA(Met) in vivo. *J Mol Biol 233*, 43-58.

Bargmann, C. I., and Weinberg, R. A. (1988). Oncogenic activation of the neu-encoded receptor protein by point mutation and deletion. *EMBO J 7*, 2043-2052.

Barth, G., and Gaillardin, C. (1997). Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. *FEMS Microbiol Rev 19*, 219-237.

Bauer, C., Brass, N., Diesinger, I., Kayser, K., Grasser, F. A., and Meese, E. (2002). Overexpression of the eukaryotic translation initiation factor 4G (eIF4G-1) in squamous cell lung carcinoma. *Int J Cancer 98*, 181-185.

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.

Belitsina, N. V., Tnalina, G. Z., and Spirin, A. S. (1981). Template-free ribosomal synthesis of polylysine from lysyl-tRNA. *FEBS Lett 131*, 289-292.

Benne, R., Brown-Luedi, M. L., and Hershey, J. W. (1978a). 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.

Benne, R., Edman, J., Traut, R. R., and Hershey, J. W. (1978b). Phosphorylation of eukaryotic protein synthesis initiation factors. *Proc Natl Acad Sci USA 75*, 108-112.

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.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from *Saccharomyces cerevisiae. Rna 9*, 871-880.

Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G., and Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *Rna 6,* 1236-1247.

Blum, S., Schmid, S. R., Pause, A., Buser, P., Linder, P., Sonenberg, N., and Trachsel, H. (1992). ATP hydrolysis by initiation factor 4A is required for translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A 89,* 7664-7668.

Bofin, A. M., Ytterhus, B., Martin, C., O'Leary, J. J., and Hagmar, B. M. (2004). Detection and quantitation of HER-2 gene amplification and protein expression in breast carcinoma. *Am J Clin Pathol 122*, 110-119.

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

Buhring, H. J., Sures, I., Jallal, B., Weiss, F. U., Busch, F. W., Ludwig, W. D., Handgretinger, R., Waller, H. D., and Ullrich, A. (1995). The receptor tyrosine kinase p185HER2 is expressed on a subset of B-lymphoid blasts from patients with acute lymphoblastic leukemia and chronic myelogenous leukemia. *Blood 86*, 1916-1923.

Bulygin, K. N., Repkova, M. N., Ven'yaminova, A. G., Graifer, D. M., Karpova, G. G., Frolova, L. Y., and Kisselev, L. L. (2002). Positioning of the mRNA stop signal with respect to polypeptide chain release factors and ribosomal proteins in 80S ribosomes. *FEBS Lett 514*, 96-101.

Bushell, M., Poncet, D., Marissen, W. E., Flotow, H., Lloyd, R. E., Clemens, M. J., and Morley, S. J. (2000). Cleavage of polypeptide chain initiation factor eIF4GI during apoptosis in lymphoma cells: characterisation of an internal fragment generated by caspase-3-mediated cleavage. *Cell Death Differ 7*, 628-636.

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

Cao, J., and Geballe, A. P. (1996). Coding sequence-dependent ribosomal arrest at termination of translation. *Mol Cell Biol 16*, 603-608.

Cao, J., and Geballe, A. P. (1998). Ribosomal release without peptidyl tRNA hydrolysis at translation termination in a eukaryotic system. RNA *4*, 181-188.

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

Casaregola, S., Neuveglise, C., Lepingle, A., Bon, E., Feynerol, C., Artiguenave, F., Wincker, P., and Gaillardin, C. (2000). Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia lipolytica. *FEBS Lett 487*, 95-100.

Castilho-Valavicius, B., Yoon, H., and Donahue, T. F. (1990). Genetic characterization of the *Saccharomyces cerevisiae* translational initiation suppressors *sui1*, *sui2* and *SUI3* and their effects on *HIS4* expression. *Genetics 124*, 483-495.

Cavallius, J., and Merrick, W. C. (1998). Site-directed mutagenesis of yeast eEF1A. Viable mutants with altered nucleotide specificity. *J Biol Chem 273*, 28752-28758.

Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L. A., Offenhauser, N., Marchisio, P. C., and Biffo, S. (2003). Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature 426*, 579-584.

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.

Chavatte, L., Seit-Nebi, A., Dubovaya, V., and Favre, A. (2002). The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *Embo J 21*, 5302-5311.

Chen, G., Hata, N., and Zhang, M. Q. (2004). Transcription factor binding element detection using functional clustering of mutant expression data. *Nucleic Acids Res 32*, 2362-2371.

Chernoff, Y. O., Newnam, G. P., and Liebman, S. W. (1996). The translational function of nucleotide C1054 in the small subunit rRNA is conserved throughout evolution: genetic evidence in yeast. *Proc Natl Acad Sci USA 93*, 2517-2522.

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 gamma-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 gamma-synthase gene in *Arabidopsis*. *Proc Natl Acad Sci USA 100*, 10225-10230.

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

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

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. *Nucleic Acids Res 33*, 5512-5520.

Cigan, A. M., and Donahue, T. F. (1987). Sequence and structural features associated with translational initiator regions in yeast– a review. *Gene 59*, 1-18.

Cigan, A. M., Feng, L., and Donahue, T. F. (1988). tRNAi(met) functions in directing the scanning ribosome to the start site of translation. *Science 242,* 93-97.

Clemens, M. J. (2004). Targets and mechanisms for the regulation of translation in malignant transformation. *Oncogene 23*, 3180-3188.

Clemens, M. J., and Bommer, U. A. (1999). Translational control: the cancer connection. *Int J Biochem Cell Biol 31,* 1-23.

Cosson, B., Couturier, A., Chabelskaya, S., Kiktev, D., Inge-Vechtomov, S., Philippe, M., and Zhouravleva, G. (2002). Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. *Mol Cell Biol 22,* 3301-3315.

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

Crabeel, M., De Rijcke, M., Seneca, S., Heimberg, H., Pfeiffer, I., and Matisova, A. (1995). Further definition of the sequence and position requirements of the arginine control element that mediates repression and induction by arginine in *Saccharomyces cerevisiae*. *Yeast 11*, 1367-1380.

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., 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.

Cui, Y., Dinman, J. D., Kinzy, T. G., and Peltz, S. W. (1998). The Mof2/Sui1 protein is a general monitor of translational accuracy. *Mol Cell Biol 18*, 1506-1516.

Daujat, S., Neel, H., and Piette, J. (2001). MDM2: life without p53. *Trends Genet 17*, 459-464.

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

Davis, R. H., and Ristow, J. L. (1987). Arginine-specific carbamoyl phosphate metabolism in mitochondria of *Neurospora crassa*. *J Biol Chem 262*, 7109-7117.

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., Calvo, O., Filipkowski, R. K., Pierard, 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.

Delbecq, P., Werner, M., Feller, A., Filipkowski, R. K., Messenguy, F., and Piérard, 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.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell 98*, 835-845.

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

Donzé, O., Damay, P., and Spahr, P. F. (1995). The first and third uORFs in RSV leader RNA are efficiently translated: implications for translational regulation and viral RNA packaging. *Nucl Acids Res 23*, 861-868.

Drabkin, H. J., Helk, B., and RajBhandary, U. L. (1993). The role of nucleotides conserved in eukaryotic initiator methionine tRNAs in initiation of protein synthesis. *J Biol Chem 268,* 25221-25228.

Dumont, D. J., Gradwohl, G., Fong, G. H., Puri, M. C., Gertsenstein, M., Auerbach, A., and Breitman, M. L. (1994). Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo. *Genes Dev 8*, 1897-1909.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. *Genome* Biol *5*, R1.

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

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 USA 101*, 4059-4064.

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.

Farruggio, D., Chaudhuri, J., Maitra, U., and RajBhandary, U. L. (1996). The A1 x U72 base pair conserved in eukaryotic initiator tRNAs is important specifically for binding to the eukaryotic translation initiation factor eIF2. *Mol Cell Biol 16*, 4248-4256.

Fink, G. R. (1987). Pseudogenes in yeast? *Cell 49*, 5-6.

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., 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., Seit-Nebi, A., and Kisselev, L. (2002). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA 8*, 129-136.

Frolova, L. Y., Merkulova, T. I., and Kisselev, L. L. (2000). Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. *RNA 6*, 381-390.

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.

Fukuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Ekimoto, H., and Igarashi, K. (1997). Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res 57*, 5041-5044.

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.

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.

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., Clutterbuck, J*., et al.* (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature 438*, 1105-1115.

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

Geballe, A. P., and Sachs, M. S. (2000). Translational control by upstream open reading frames, In Translational control of gene expression, N. Sonenberg, J. W. B. Hershey, and M. B. Mathews, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 595-614.

Gebauer, F., Grskovic, M., and Hentze, M. W. (2003). Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol Cell 11*, 1397-1404.

Gebauer, F., and Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol 5*, 827-835.

Gilmore, R., and Blobel, G. (1985). Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants. *Cell 42*, 497-505.

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.

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. (2002). Instruction of the translating ribosome by nascent peptide. *Science 297*, 1864-1867.

Graille, M., Heurgue-Hamard, V., Champ, S., Mora, L., Scrima, N., Ulryck, N., van Tilbeurgh, H., and Buckingham, R. H. (2005). Molecular basis for bacterial class I release factor methylation by PrmC. *Mol Cell 20*, 917-927.

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

Greider, C., Chattopadhyay, A., Parkhurst, C., and Yang, E. (2002). BCL-x(L) and BCL2 delay Myc-induced cell cycle entry through elevation of p27 and inhibition of G1 cyclin-dependent kinases. *Oncogene 21*, 7765-7775.

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.

Grundy, F. J., Lehman, S. C., and Henkin, T. M. (2003). The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci USA 100*, 12057-12062.

Hamilton, R., Watanabe, C. K., and de Boer, H. A. (1987). Compilation and comparison of the sequence context around the AUG startcodons in Saccharomyces cerevisiae mRNAs. *Nucleic Acids Res 15*, 3581-3593.

Hann, S. R. (1994). Regulation and function of non-AUG-initiated proto-oncogenes. *Biochimie 76*, 880-886.

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.

Harrod, R., and Lovett, P. S. (1997). Leader peptides of inducible chloramphenicol resistance genes from gram-positive and gram-negative bacteria bind to yeast and Archaea large subunit rRNA. *Nucl Acids Res 25*, 1720-1726.

Hashimoto, N. N., Carnevalli, L. S., and Castilho, B. A. (2002). Translation initiation at non-AUG codons mediated by weakened association of eukaryotic initiation factor (eIF) 2 subunits. *Biochem J 367*, 359-368.

He, H., von der Haar, T., Singh, C. R., Ii, M., Li, B., Hinnebusch, A. G., McCarthy, J. E., and Asano, K. (2003). The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection. *Mol Cell Biol 23*, 5431-5445.

Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S. A. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res 25*, 3957-3958.

Hershey, J. W. B., and Merrick, W. C. (2000). The Pathway and Mechanism of Initiation of Protein Synthesis, In Translational control of gene expression, N. Sonenberg, J. W. B. Hershey, and M. B. Mathews, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 33-88.

Heurgue-Hamard, V., Champ, S., Mora, L., Merkulova-Rainon, T., Kisselev, L. L., and Buckingham, R. H. (2005). The glutamine residue of the conserved GGQ motif in Saccharomyces cerevisiae release factor eRF1 is methylated by the product of the YDR140w gene. *J Biol Chem 280*, 2439-2445.

Hidalgo, M., and Rowinsky, E. K. (2000). The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. *Oncogene 19*, 6680-6686.

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.

Hinnebusch, A. G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc Natl Acad Sci USA 81*, 6442-6446.

Hinnebusch, A. G. (1992). General and pathway-specific regulatory mechanisms controlling the synthesis of amino acid biosynthetic enzymes in *Saccharomyces cerevisiae*, In The Molecular and Cellular Biology of the Yeast *Saccharomyces*, E. W. Jones, J. R. Pringle, and J. R. Broach, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 319-414.

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

Hinnebusch, A. G. (2005). Translational regulation of Gcn4 and the general amino acid control of yeast. *Annu Rev Microbiol 59*, 407-450.

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

Holland, E. C., Sonenberg, N., Pandolfi, P. P., and Thomas, G. (2004). Signaling control of mRNA translation in cancer pathogenesis. *Oncogene 23*, 3138-3144.

Hood, H. M., Spevak, C. C., and Sachs, M. S. (In press). Evolutionary changes in the carbamoyl-phosphate synthetase small subunit gene and associated upstream open reading frame.

Hood, H. M., Spevak, C. C., and Sachs, M. S. (In press). Evolutionary changes in the carbamoyl-phosphate synthetase small subunit gene and associated upstream open reading frame.

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.

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.

Inagaki, Y., Blouin, C., Doolittle, W. F., and Roger, A. J. (2002). Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. *Nucleic Acids Res 30*, 532-544.

Jackson, R. J., and Kaminski, A. (1995). Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA 1*, 985-1000.

Jacobson, A. (1996). Poly(A) metabolism and translation: the closed-loop model, In Translational control, J. W. B. Hershey, M. B. Mathews, and N. Sonenberg, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 451-480.

Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. (1997). Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. *Embo J 16*, 3693-3704.

Jenni, S., and Ban, N. (2003). The chemistry of protein synthesis and voyage through the ribosomal tunnel. *Curr Opin Struct Biol 13*, 212-219.

Jiang, M., and Milner, J. (2003). Bcl-2 constitutively suppresses p53-dependent apoptosis in colorectal cancer cells. *Genes Dev 17*, 832-837.

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.

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

Joshi, C. P., Zhou, H., Huang, X., and Chiang, V. L. (1997). Context sequences of translation initiation codon in plants. *Plant Mol Biol 35*, 993-1001.

Kapp, L. D., and Lorsch, J. R. (2004a). GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J Mol Biol 335*, 923-936.

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

Kazarinoff, M. N., and Snell, E. E. (1977). Essential arginine residues in tryptophanase from Escherichia coli. *J Biol Chem 252*, 7598-7602.

Keenan, R. J., Freymann, D. M., Stroud, R. M., and Walter, P. (2001). The signal recognition particle. *Annu Rev Biochem 70*, 755-775.

Keogh, R. S., Seoighe, C., and Wolfe, K. H. (1998). Evolution of gene order and chromosome number in Saccharomyces, Kluyveromyces and related fungi. Yeast *14*, 443-457.

Kerekatte, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. (1995). The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas. *Int J Cancer 64*, 27-31.

Kim, J., Klein, P. G., and Mullet, J. E. (1991). Ribosomes pause at specific sites during synthesis of membrane-bound chloroplast reaction center protein D1. *J Biol Chem 266*, 14931-14938.

Kinney, D. M., and Lusty, C. J. (1989). Arginine restriction induced by δ-N-(phosphonacetyl)-L-ornithine signals increased expression of *HIS3*, *TRP5*, *CPA1*, and *CPA2* in *Saccharomyces cerevisiae*. *Mol Cell Biol 9*, 4882-4888.

Kleijn, M., Scheper, G. C., Voorma, H. O., and Thomas, A. A. (1998). Regulation of translation initiation factors by signal transduction. *Eur J Biochem 253*, 531-544.

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.

Koromilas, A. E., Roy, S., Barber, G. N., Katze, M. G., and Sonenberg, N. (1992). Malignant transformation by a mutant of the IFN-inducible dsRNA-dependent protein kinase. *Science 257*, 1685-1689.

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. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem 266*, 19867-19870.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J 16*, 2482-2492.

Kozak, M. (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.

Krappmann, S., Bignell, E. M., Reichard, U., Rogers, T., Haynes, K., and Braus, G. H. (2004). The *Aspergillus fumigatus* transcriptional activator *CpcA* contributes significantly to the virulence of this fungal pathogen. *Mol Microbiol 52*, 785-799.

Kudlicki, W., Odom, O. W., Kramer, G., Hardesty, B., Merrill, G. A., and Horowitz, P. M. (1995). The importance of the N-terminal segment for DnaJ-mediated folding of rhodanese while bound to ribosomes as peptidyl-tRNA. *J Biol Chem 270*, 10650-10657.

Kull, F. J., and Endow, S. A. (2004). A new structural state of myosin. *Trends Biochem Sci 29*, 103-106.

Kurtzman, C. P., and Robnett, C. J. (1998). Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek *73*, 331-371.

Kurzchalia, T. V., Bommer, U. A., Babkina, G. T., and Karpova, G. G. (1984). GTP interacts with the gamma-subunit of eukaryotic initiation factor eIF-2. *FEBS Lett 175*, 313-316.

Kyrpides, N. C., and Woese, C. R. (1998). Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. *Proc Natl Acad Sci U S A 95*, 3726-3730.

Lamas-Maceiras, M., Cerdan, M. E., and Freire-Picos, M. A. (1999). Kluyveromyces lactis HIS4 transcriptional regulation: similarities and differences to Saccharomyces cerevisiae HIS4 gene. *FEBS Lett 458*, 72-76.

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

Lamberti, A., Caraglia, M., Longo, O., Marra, M., Abbruzzese, A., and Arcari, P. (2004). The translation elongation factor 1A in tumorigenesis, signal transduction and apoptosis: review article. *Amino Acids 26*, 443-448.

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. Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem 270*, 21975-21983.

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., Montine, K. S., and Sonenberg, N. (1990). Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature 345*, 544-547.

Le, S. Y., and Maizel, J. V., Jr. (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 USA 99*, 16689-16694.

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.

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.

Lorsch, J. R., and Herschlag, D. (1999). Kinetic dissection of fundamental processes of eukaryotic translation initiation in vitro. *EMBO J 18*, 6705-6717.

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

Lu, J., and Deutsch, C. (2005). Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol 12*, 1123-1129.

Lukash, T. O., Turkivska, H. V., Negrutskii, B. S., and El'skaya, A. V. (2004). Chaperone-like activity of mammalian elongation factor eEF1A: renaturation of aminoacyl-tRNA synthetases. *Int J Biochem Cell Biol 36*, 1341-1347.

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.

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.

Marissen, W. E., Triyoso, D., Younan, P., and Lloyd, R. E. (2004). Degradation of poly(A)-binding protein in apoptotic cells and linkage to translation regulation. *Apoptosis 9*, 67-75.

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

Matthews, M. B., Sonenberg, N., and Hershey, J. W. B. (2000). Origins and principles of translational control, In Translational control of gene expression, N. Sonenberg, J. W. B. Hershey, and M. B. Mathews, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 1-31.

McDaniel, B. A., Grundy, F. J., Artsimovitch, I., and Henkin, T. M. (2003). Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci USA 100*, 3083-3088.

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.

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

Merkulova, T. I., Frolova, L. Y., Lazar, M., Camonis, J., and Kisselev, L. L. (1999). C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their *in vivo* interaction. *FEBS Lett 443*, 41-47.

Messenguy, F., and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene 316*, 1-21.

Meurs, E. F., Galabru, J., Barber, G. N., Katze, M. G., and Hovanessian, A. G. (1993). Tumor suppressor function of the interferon-induced double-stranded RNA-activated protein kinase. *Proc Natl Acad Sci USA 90*, 232-236.

Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem 267*, 6321-6330.

Michael, D., and Oren, M. (2002). The p53 and Mdm2 families in cancer. Curr Opin Genet Dev *12*, 53-59.

Michel, Y. M., Poncet, D., Piron, M., Kean, K. M., and Borman, A. M. (2000). Cap-Poly(A) synergy in mammalian cell-free extracts. Investigation of the requirements for poly(A)-mediated stimulation of translation initiation. *J Biol Chem 275*, 32268-32276.

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.

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.

Moffat, J. G., and Tate, W. P. (1994). A single proteolytic cleavage in release factor 2 stabilizes ribosome binding and abolishes peptidyl-tRNA hydrolysis activity. *J Biol Chem 269*, 18899-18903.

Momand, J., Wu, H. H., and Dasgupta, G. (2000). MDM2--master regulator of the p53 tumor suppressor protein. Gene *242*, 15-29.
Moore, P. B., and Steitz, T. A. (2003). After the ribosome structures: how does peptidyl transferase work? *RNA 9*, 155-159.

Mora, L., Heurgue-Hamard, V., Champ, S., Ehrenberg, M., Kisselev, L. L., and Buckingham, R. H. (2003). The essential role of the invariant GGQ motif in the function and stability in vivo of bacterial release factors RF1 and RF2. *Mol Microbiol 47*, 267-275.
Mori, H., and Ito, K. (2001). The Sec protein-translocation pathway. *Trends Microbiol 9, 494-500.*

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.

Mourier, T., and Jeffares, D. C. (2003). Eukaryotic intron loss. *Science 300*, 1393.

Muckenthaler, M., Gray, N. K., and Hentze, M. W. (1998). IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol Cell 2*, 383-388.

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.

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., and Jovine, L. (2003). Structure, function and evolution of the signal recognition particle. *EMBO J 22*, 3479-3485.

Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. *Dev Cell 6*, 69-78.

Nakamura, Y., Ito, K., and Ehrenberg, M. (2000). Mimicry grasps reality in translation termination. *Cell 101*, 349-352.

Nakamura, Y., Ito, K., and Isaksson, L. A. (1996). Emerging understanding of translation termination. *Cell 87*, 147-150.

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.

Nakatogawa, H., Murakami, A., and Ito, K. (2004). Control of SecA and SecM translation by protein secretion. *Curr Opin Microbiol 7*, 145-150.

Naranda, T., MacMillan, S. E., and Hershey, J. W. (1994). Purified yeast translational initiation factor eIF-3 is an RNA-binding protein complex that contains the PRT1 protein. *J Biol Chem 269*, 32286-32292.

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

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., and Marton, M. J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol 21*, 4347-4368.

Nathan, C. A., Liu, L., Li, B. D., Abreo, F. W., Nandy, I., and De Benedetti, A. (1997). Detection of the proto-oncogene eIF4E in surgical margins may predict recurrence in head and neck cancer. *Oncogene 15*, 579-584.

Negrutskii, B. S., and El'skaya, A. V. (1998). Eukaryotic translation elongation factor 1 alpha: structure, expression, functions, and possible role in aminoacyl-tRNA channeling. *Prog Nucleic Acid Res Mol Biol 60*, 47-78.

Nicholson, R. I., Gee, J. M., and Harper, M. E. (2001). EGFR and cancer prognosis. Eur J *Cancer 37 Suppl 4*, S9-S15.

Nielsen, C. B., Friedman, B., Birren, B., Burge, C. B., and Galagan, J. E. (2004). Patterns of intron gain and loss in fungi. *PLoS Biol 2*, e422.

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

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.

Oliver, D. B., and Beckwith, J. (1982). Regulation of a membrane component required for protein secretion in Escherichia coli. *Cell 30*, 311-319.

Olsen, D. S., Savner, E. M., Mathew, A., Zhang, F., Krishnamoorthy, T., Phan, L., and Hinnebusch, A. G. (2003). Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. *Embo J 22*, 193-204.

Ominato, K., Akita, H., Suzuki, A., Kijima, F., Yoshino, T., Yoshino, M., Chiba, Y., Onouchi, H., and Naito, S. (2002). Identification of a short highly conserved amino acid sequence as the functional region required for posttranscriptional autoregulation of the cystathionine gamma-synthase gene in Arabidopsis. *J Biol Chem 277*, 36380-36386.

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

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

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

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.

Ostareck, D. H., Ostareck-Lederer, A., Shatsky, I. N., and Hentze, M. W. (2001). Lipoxygenase mRNA silencing in erythroid differentiation: The 3'UTR regulatory complex controls 60S ribosomal subunit joining. *Cell 104*, 281-290.

Paluh, J. L., Orbach, M. J., Legerton, T. L., and Yanofsky, C. (1988). The cross-pathway control gene of *Neurospora crassa*, *cpc-1*, encodes a protein similar to *GCN4* of yeast and the DNA-binding domain of the oncogene *v-jun*-encoded protein. *Proc Natl Acad Sci USA 85,* 3728-3732.

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.

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. V. (2000). Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nat Struct Biol 7*, 104-107.

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

Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *Embo J 11*, 2643-2654.

Pelletier, J., and Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell 40*, 515-526.

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.

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., 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.

Peters, K. G., Kontos, C. D., Lin, P. C., Wong, A. L., Rao, P., Huang, L., Dewhirst, M. W., and Sankar, S. (2004). Functional significance of Tie2 signaling in the adult vasculature. *Recent Prog Horm Res 59*, 51-71.

Phan, L., Schoenfeld, L. W., Valasek, L., Nielsen, K. H., and Hinnebusch, A. G. (2001). A subcomplex of three eIF3 subunits binds eIF1 and eIF5 and stimulates ribosome binding of mRNA and tRNA(i)Met. *Embo J 20*, 2954-2965.

Piepenburg, O., Pape, T., Pleiss, J. A., Wintermeyer, W., Uhlenbeck, O. C., and Rodnina, M. V. (2000). Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome. *Biochemistry 39*, 1734-1738.

Poole, E., and Tate, W. (2000). Release factors and their role as decoding proteins: specificity and fidelity for termination of protein synthesis. *Biochim Biophys Acta 1493*, 1-11.

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., Muckenthaler, M., and Hentze, M. W. (1998). Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA 4*, 1321-1131.

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.

Radford, A. (2004). Metabolic highways of Neurospora crassa revisited. *Adv Genet 52*, 165-207.

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

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.

Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C., and Thach, R. E. (1985). ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem 260*, 7651-7658.

Risau, W. (1997). Mechanisms of angiogenesis. *Nature 386*, 671-674.

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.

Rogozin, I. B., Kochetov, A. V., Kondrashov, F. A., Koonin, E. V., and Milanesi, L. (2001). Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics 17*, 890-900.

Rogozin, I. B., Sverdlov, A. V., Babenko, V. N., and Koonin, E. V. (2005). Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief Bioinform 6*, 118-134.

Roy, S. W., and Gilbert, W. (2005a). The pattern of intron loss. *Proc Natl Acad Sci U S A 102*, 713-718.

Roy, S. W., and Gilbert, W. (2005b). Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc Natl Acad Sci U S A 102*, 5773-5778.

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., 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.

Ruggero, D., Montanaro, L., Ma, L., Xu, W., Londei, P., Cordon-Cardo, C., and Pandolfi, P. P. (2004). The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nat Med 10*, 484-486.

Sachs, A. B., and Kornberg, R. D. (1985). Nuclear polyadenylate-binding protein. *Mol Cell Biol 5*, 1993-1996.

Sachs, M. S. (1996). General and cross-pathway controls of amino acid biosynthesis, In The Mycota: Biochemistry and Molecular Biology, R. Brambl, and G. A. Marzluf, eds. (Heidelberg: Springer-Verlag), pp. 315-345.

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 translational apparatus components at initiation and termination codons of fungal mRNAs. *Methods 26*, 105-114.

Sachs, M. S., and Yanofsky, C. (1991). Developmental expression of genes involved in conidiation and amino acid biosynthesis in *Neurospora crassa*. Dev Biol *148*, 117-128.
Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4*, 406-425.

Sarkar, G., and Sommer, S. S. (1990). The "megaprimer" method of site-directed mutagenesis. *BioTechniques 8*, 404-407.

Sarker, S., and Oliver, D. (2002). Critical regions of secM that control its translation and secretion and promote secretion-specific secA regulation. *J Bacteriol 184*, 2360-2369.

Scheffzek, K., Ahmadian, M. R., Wiesmuller, L., Kabsch, W., Stege, P., Schmitz, F., and Wittinghofer, A. (1998). Structural analysis of the GAP-related domain from neurofibromin and its implications. *Embo J 17*, 4313-4327.

Schmidt, M. G., Rollo, E. E., Grodberg, J., and Oliver, D. B. (1988). Nucleotide sequence of the secA gene and secA(Ts) mutations preventing protein export in Escherichia coli. *J Bacteriol 170*, 3404-3414.

Schmitt, E., Blanquet, S., and Mechulam, Y. (2002). The large subunit of initiation factor aIF2 is a close structural homologue of elongation factors. *Embo J 21*, 1821-1832.

Schneider, R. J., and Mohr, I. (2003). Translation initiation and viral tricks. Trends *Biochem Sci 28*, 130-136.

Seit-Nebi, A., Frolova, L., and Kisselev, L. (2002). Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep 3*, 881-886.

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. (1997). Cross-pathway and pathway-specific control of amino acid biosynthesis in *Magnaporthe grisea*. *Fungal Genet Biol 21*, 40-49.

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

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.

Skogerson, L., and Engelhardt, D. (1977). Dissimilarity in protein chain elongation factor requirements between yeast and rat liver ribosomes. *J Biol Chem 252*, 1471-1475.

Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M.*, et al.* (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med 344*, 783-792.

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., Hershey, J. W. B., and Mathews, M. B. (2000). Translational control of gene expression (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

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

Song, W., Raden, D., Mandon, E. C., and Gilmore, R. (2000b). Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell 100*, 333-343.

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*, 301-314.

Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of Drosophila MicroRNA targets. *PLoS Biol 1*, E60.

Stark, H., Orlova, E. V., Rinke-Appel, J., Jünke, N., Mueller, F., Rodnina, M., Wintermeyer, W., Brimacombe, R., and van Heel, M. (1997a). Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy. *Cell 88*, 19-28.

Stark, H., Rodnina, M. V., Rinke-Appel, J., Brimacombe, R., Wintermeyer, W., and van Heel, M. (1997b). Visualization of elongation factor Tu on the *Escherichia coli* ribosome. *Nature 389*, 403-406.

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.

Stewart, V., and Yanofsky, C. (1985). Role of leader peptide synthesis in tryptophanase operon expression in *Escherichia coli* K-12. *J Bacteriol 164*, 731-740.

Stokoe, D. (2005). The phosphoinositide 3-kinase pathway and cancer. *Expert Rev Mol Med 7*, 1-22.

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

Suri, C., Jones, P. F., Patan, S., Bartunkova, S., Maisonpierre, P. C., Davis, S., Sato, T. N., and Yancopoulos, G. D. (1996). Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell 87*, 1171-1180.

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

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.

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

Teshima-Kondo, S., Kondo, K., Prado-Lourenco, L., Gonzalez-Herrera, I. G., Rokutan, K., Bayard, F., Arnal, J. F., and Prats, A. C. (2004). Hyperglycemia upregulates translation of the fibroblast growth factor 2 mRNA in mouse aorta via internal ribosome entry site. *Faseb J 18*, 1583-1585.

Thomas, A., Goumans, H., Voorma, H. O., and Benne, R. (1980). The mechanism of action of eukaryotic initiation factor 4C in protein synthesis. *Eur J Biochem 107*, 39-45.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*, 4673-4680.

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

Triana-Alonso, F. J., Chakraburtty, K., and Nierhaus, K. H. (1995). The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. *J Biol Chem 270*, 20473-20478.

Tsalkova, T., Kramer, G., and Hardesty, B. (1999). The effect of a hydrophobic N-terminal probe on translational pausing of chloramphenicol acetyl transferase and rhodanese. *J Mol Biol 286*, 71-81.

Valasek, L., Mathew, A. A., Shin, B. S., Nielsen, K. H., Szamecz, B., and Hinnebusch, A. G. (2003). The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. *Genes Dev 17*, 786-799.

Valasek, L., Nielsen, K. H., and Hinnebusch, A. G. (2002). Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo. *Embo J 21*, 5886-5898.

Van-Huffel, C., Dubois, E., and Messenguy, F. (1992). Cloning and sequencing of arg3 and arg11 genes of Schizosaccharomyces pombe on a 10-kb DNA fragment. Heterologous expression and mitochondrial targeting of their translation products. *Eur J Biochem 205*, 33-43.

Van Huffel, C., Dubois, E., and Messenguy, F. (1994). Cloning and sequencing of Schizosaccharomyces pombe car1 gene encoding arginase. Expression of the arginine anabolic and catabolic genes in response to arginine and related metabolites. *Yeast 10*, 923-933.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol 180*, 549-576.

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

Verlhac, M. H., Chen, R. H., Hanachi, P., Hershey, J. W., and Derynck, R. (1997). Identification of partners of TIF34, a component of the yeast eIF3 complex, required for cell proliferation and translation initiation. *Embo J 16*, 6812-6822.

Vilela, C., and McCarthy, J. E. (2003). Regulation of fungal gene expression via short open reading frames in the mRNA 5'untranslated region. *Mol Microbiol 49*, 859-867.

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet 20*, 44-50.

von Heijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *Embo J 5*, 1335-1342.

von Pawel-Rammingen, U., Astrom, S., and Bystrom, A. S. (1992). Mutational analysis of conserved positions potentially important for initiator tRNA function in Saccharomyces cerevisiae. *Mol Cell Biol 12*, 1432-1442.

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., 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.

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.

Wanke, C., Eckert, S., Albrecht, G., van Hartingsveldt, W., Punt, P. J., van den Hondel, C. A., and Braus, G. H. (1997). The *Aspergillus niger GCN4* homologue, *cpcA*, is transcriptionally regulated and encodes an unusual leucine zipper. *Mol Microbiol 23*, 23-33.

Waskiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. (1997). Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *Embo J 16*, 1909-1920.

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

Werner, M., Feller, A., and Piérard, A. (1985). Nucleotide sequence of yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoyl-phosphate synthetase: homology of the deduced amino acid sequence to other glutamine amidotransferases. *Eur J Biochem 146*, 371-381.

Wilhelm, J. E., Hilton, M., Amos, Q., and Henzel, W. J. (2003). Cup is an eIF4E binding protein required for both the translational repression of oskar and the recruitment of Barentsz. *J Cell Biol 163*, 1197-1204.

Willis, A. E. (1999). Translational control of growth factor and proto-oncogene expression. *Int J Biochem Cell Biol 31*, 73-86.

Wilson, K. S., and Noller, H. F. (1998). Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell 92*, 131-139.

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.

Xia, W., Liu, L. H., Ho, P., and Spector, N. L. (2004). Truncated ErbB2 receptor (p95ErbB2) is regulated by heregulin through heterodimer formation with ErbB3 yet remains sensitive to the dual EGFR/ErbB2 kinase inhibitor GW572016. *Oncogene 23*, 646-653.

Yakes, F. M., Chinratanalab, W., Ritter, C. A., King, W., Seelig, S., and Arteaga, C. L. (2002). Herceptin-induced inhibition of phosphatidylinositol-3 kinase and Akt Is required for antibody-mediated effects on p27, cyclin D1, and antitumor action. *Cancer Res 62*, 4132-4141.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 A resolution. *Science 292*, 883-896.

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.

Zhang, Z., and Dietrich, F. S. (2005). Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in *Saccharomyces cerevisiae*. *Curr Genet 48*, 77-87.

Zhouravleva, G., Frolova, L., Le Goff, X., Le Geullec, 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.

BIOGRAPHICAL SKETCH

Christina C. Spevak was born in Manila, Philippines on January 4, 1976. In 1998, she received her B.A. in Biology from the University of San Diego. After being employed during the fall of 1998 at Scantibodies Laboratory, Inc. in San Diego, she decided to pursue her M.S. in Biochemistry and Molecular Biology at the Oregon Graduate Institute during the winter of 1999. She decided to continue with her Ph.D degree in Biochemistry and Molecular Biology at the Oregon Health and Science University under the supervision and research of Dr. Matthew S. Sachs. Following the successful defense of her dissertation, Christina will begin her postdoctoral research in the laboratory of Dr. David Stokoe at the Cancer Research Institute of the University of California, San Francisco.

Publications

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. 101*, 4059-4064.

Galagan, J.E, Calvo, S.E, Cuomo, C, Ma, L-J, Wortman, J., Batzoglou, S., Lee ,S-I, Baştürkmen, M.**,** Spevak, C.C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G.H., Draht, O., Busch, S., D'Enfert, C., Bouchier ,C., Goldman, G.H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J.H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E.U., Archer, D.B., Peñalva, M.A., Oakley, B.R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W.C., Denning, D.W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs,M.S., Osman, S.A., and Birren, B.W. (2005). Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *Nature 438*, 1105-1115.

Hood, H.M., Spevak, C.C., and Sachs, M. S. (2006). Evolutionary changes in the carbamoyl-phosphate synthetase small subunit gene and associated upstream open reading frame. (In press).

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. (Submitted).

Spevak, C.C., and Sachs, M.S. (2006). Specific requirements for a nascent peptide domain that stalls ribosomes in response to arginine (In preparation).