TRANSLATIONAL CONTROL MECHANISMS ANALYZED

IN *NEUROSPORA CRASSA*

A Dissertation

by

JIAJIE WEI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Matthew S. Sachs
Committee Members,	Tatyana Igumenova
	Xiaorong Lin
	Keith A. Maggert
Department Head,	U.J. McMahan

August 2013

Major Subject: Biology

ABSTRACT

The *Neurospora crassa arg-2* gene encodes the small subunit of carbamoyl phosphate synthetase, the first enzyme in fungal arginine (Arg) biosynthesis. The arginine attenuator peptide (AAP), specified by an upstream open reading frame (uORF), stalls ribosomes at its termination codon in response Arg to control the translation of *arg-2*. In project 1, the effect of AAP and Arg on ribosome peptidyl transferase center (PTC) activity was analyzed in *N. crassa* and wheat germ cell-free translation extracts using the transfer of nascent AAP to puromycin as an assay. The results show that inhibition of PTC activity by the AAP and Arg is the basis for the AAP's function. The mode of PTC inhibition appears unusual because neither a specific amino acid nor a specific nascent peptide chain length was required for AAP to function.

In eukaryotic translation initiation, the stringency of start codon selection impacts initiation efficiencies at AUG codons in different contexts and at near-cognate codons (NCCs) that differ from AUG by a single nucleotide. In project 2, a codon-optimized firefly luciferase reporter was used to examine the stringency of start codon selection in *N. crassa. In vivo* and *in vitro* results indicated that the hierarchy of initiation in *N. crassa* is similar to that in human cells. The preferred context was more important for efficient initiation from NCCs than from AUG.

In project 3, the use of NCCs was also specifically examined for the *N. crassa cpc-1* gene. *cpc-1* and *Saccharomyces cerevisiae GCN4* are homologs specifying a transcription activator, which drives the primary transcriptional response to amino acid

starvation. *In vitro* studies showed that uORF1 and uORF2 in *cpc-1* are functionally analogous to uORF1 and uORF4 in *GCN4*. uORF1 promotes reinitiation at downstream start codons. uORF2 inhibits translation from the main *cpc-1* start codon. Four NCCs in the CPC1 reading frame and upstream of uORF2 can also be used for translation initiation.

In summary, we explored uORF-mediated translational regulation and the use of NCCs as initiation codons. Taken together, these studies establish *N. crassa* as a model system to examine mechanisms contributing to translational control including initiation and termination.

DEDICATION

To my mom, Jing Shi

To my dad, Jian Wei

To my husband, Yan Zhou

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude towards my advisor Dr. Matthew Sachs, for his inspiration, motivation, invaluable advice and consistent encouragement. Without his guidance, this dissertation would not have been possible.

Special thanks go to my committee members Drs, Tatyana Igumenova and Keith Maggert for their insights and suggestions. I am indebted to Dr. Xiaorong Lin for her support and encouragement throughout my entire Ph.D. studies. I would also like to thank Dr. Deborah Bell-Pedersen for her critical discussions and her lab members for generously sharing equipment and reagents.

My sincere thanks to Dr. Ivaylo Ivanov for his help with my research project and for his unconditional friendship. Thanks to Drs. Cheng Wu and Ying Zhang for their invaluable guidance, assistance and kindness thoughtout my graduate studies, and for their friendship that made my graduate school a wonderful experience. Thanks to Allyson Martinez, Fei Yang and Randolph Addison for their support.

I am forever indebted to my mother and father for their continuous love and encouragement, which without, I would be nowhere. I would like to give special thanks to my husband Yan Zhou for his love, patience and understanding for all these years. I dedicate this dissertation to my parents and my husband.

NOMENCLATURE

AAP	arginine attenuator peptide
ABC	ATP-binding cassette
AdoMetDC	*S*-Adenosylmethionine decarboxylase
ATF4	activating transcription factor 4
3-AT	3-aminotriazole
CPE	cytoplasmic polyadenylation element
CPEB	cytoplasmic polyadenylation element binding protein
CPS	carbamoyl phosphate synthetase
CrPV	cricket paralysis virus
CTT	C-terminal tail
CYH	cycloheximide
4E-BP	4E-biding protein
eEF	eukaryotic elongation factor
eIF	eukaryotic initiation factor
EMCV	encephalomyocarditis virus
ER	endoplasmic reticulum
eRF	eukaryotic release factor
FMDV	foot-and-mouth disease virus
FRET	fluorescence resonance energy transfer
GAAC	general amino acid conrol

GAP	GTPase activating protein
GCN2	general control non-derepressible-2
GCRE	general control response element
GEF	guanine nucleotide exchange factor
HCV	hepatitis C virus
HRI	heme-regulated inhibitor
ITAF	IRES trans-acting factor
IRES	internal ribosome entry site
LUC	luciferase
MARK	mitogen-activated protein kinase
MFC	multifactor complex
miRNA	microRNA
mRNP	messenger ribonucleoprotein
MSL	male specific lethal
mTOR	mammalian target of rapamycin
NCC	near-cognate codon
NMD	nonsense-mediated mRNA decay
NTT	N-terminal tail
ORF	open reading frame
PABP	poly(A) biding protein
Pak2	p21-activated kinase
PERK	PKP-like endoplasmic reticulum kinase

PIC	preinitiation complex
PKR	protein kinase RNA-dependent
PTC	peptidyl-transferase center
qPCR	quantitative PCR
RAP	ribosome-arrest peptide
RLU	relative light unit
RRF	ribosome recycling factor
SXL	sex-lethal
TC	ternary complex
TDE	translational derepression element
TOP	oligopyrimide
4-TU	4-thiouracil
uORF	upstream open reading frame
UTR	untranslated region

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Regulation of eukaryotic gene expression occurs at multiple steps including transcription, post-transcriptional processing, messenger RNA (mRNA) stability, translation, post-translational modification, and protein degradation.

Whereas the regulation of transcription controls the level of mRNA that can be translated into proteins, the regulation of translation has unique features, making it crucial for processes including cell cycle progression, cell proliferation, development and differentiation, aging and apoptosis (*1*). Translational control can rapidly modulate protein levels in cells exposed to various extracellular and intracellular signals. Most translational controls are through the reversible phosphorylation of translation factors, making it energetically economical. Regulation of translation provides a fine control compared to usually more pronounced changes in transcription rates. Systems that lack the transcriptional control, such as reticulocytes and platelets, use the translational control to modulate gene expression. Translation also supports spatial control, which is critical for polarity.

Translation includes three major stages: initiation, elongation, and termination. Multiple factors and steps are involved in translation to provide various means to exquisitely control gene expression including rates, accuracy, direction, and localization. This chapter summarizes the processes and mechanisms of translation and translational control in eukaryotic cells.

The regulatory role of upstream open reading frames (uORFs) and nascent peptides in translation is also reviewed in this chapter. uORFs are common regulatory elements that are found within many eukaryotic transcripts. Bioinformatic studies find that 49% of human and 44% of mouse transcripts contain at least one uORF (*2*). Two examples of genes containing uORFs will be discussed in the following chapters: the *Neurospora crassa arg-2* and *cpc-1* genes. The *arg-2* uORF encodes an evolutionarily conserved arginine attenuator peptide (AAP) that regulates translation in response to arginine (Arg). The AAP is shown to regulate translation as a uORF in fungal, plant, and animal cell-free systems. The AAP also retains its function when placed as a domain near the N-terminus or internally within a large polypeptide. Our detailed studies on the AAP are documented in CHAPTER 2. *cpc-1*, the homolog of *Saccharomyces cerevisiae GCN4*, is the major cross-pathway regulatory gene in *N. crassa*. It is required for *N. crassa* to induce expression of amino acid biosynthetic genes under amino acid starvation. Effects of two uORFs in the *cpc-1* transcript on its translation are discussed in CHAPTER 4.

This chapter also reviews translation initiation from non-AUG start codons. Many codons differing by a single nucleotide from AUG can be used as an initiation codon, albeit at significantly reduced levels. CHAPTER 3 describes the detailed studies on the usage of the non-AUG initiation codons in *N. crassa*. Effects of non-AUG initiation codons on the translation of *N. crassa cpc-1* are discussed in CHAPTER 4.

1.1. The molecular mechanism of eukaryotic translation

The ribosome is a multiprotein machine that plays a central role in protein translation. The ribosome consists of two subunits (30S and 50S in prokaryotes and 40S and 60S in eukaryotes) that combine to assemble three binding sites for tRNA, designated A (aminoacyl), P (peptidyl) and E (exit) sites (*3*). In eukaryotic cells, sequences within the 5'-untranslated region (5'-UTR), the open reading frame (ORF) region(s), and the 3'-untranslated region (3'-UTR) of mRNAs serve as *cis* elements that regulate various aspects of translation (Figure 1.1) (*4, 5*). Most eukaryotic mRNAs are modified at their 5' ends with a m^7G(5')ppp(5')N cap structure and at their 3' ends with a poly(A) tail. The cap structure is essential for multiple cellular processes, including translation initiation (*6, 7*), mRNA stability (*8, 9*), and mRNA transport and localization (*10, 11*). Although an internal ribosome entry site (IRES) can also serve as a translation initiation site in some cases, the majority of translation in eukaryotes is "cap-dependent." With respect to the poly(A) tail, the average length is ~70 nucleotides in yeast and ~200-250 nucleotides in mammals (*12*). The cap structure and the poly(A) tail can interact with each other via interactions between poly(A) binding protein (PABP), eIF4G (a translation initiation factor) and eIF4E (a mRNA 5' cap-binding protein). This results in the circularization of eukaryotic mRNAs and the enhancement of translation (*13-18*).

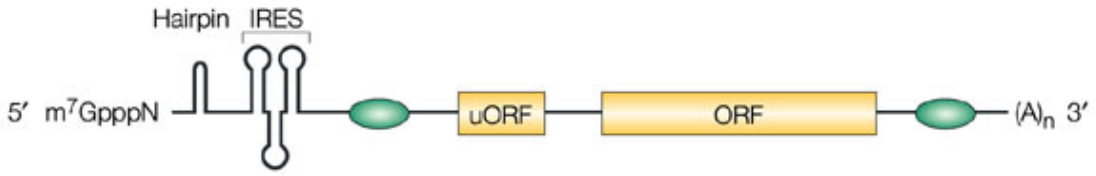


Figure 1.1. *cis*-acting elements on mRNA that affect gene expression. The cap structure at the 5' end of the mRNA and the poly(A) tail at the 3' end strongly stimulate translation initiation. Secondary structures, such as hairpins, block translation. IRESs mediate cap-independent translation. uORFs normally function as negative regulators by reducing translation from the main ORF. Adapted by permission from MacMillan Publishers Ltd (*5*).

1.1.1. Initiation

Translation initiation is considered to be the most highly regulated phase in translation in eukaryotic cells (*19*). It is a multi-step process that recruits initiator tRNA, translation initiation factors (eIFs), and ribosomal subunits to the 5' ends of capped mRNAs to form an initiation complex in the cytoplasm (Figure 1.2) (*20, 21*). A ternary complex (TC) is first formed by the GTP-bound eukaryotic initiation factor 2 (eIF2) and Met-tRNA$_i^{Met}$. TC joins with eIF1, eIF5, and eIF3 to form a multifactor complex that binds to the small (40S) ribosomal subunit to form a 43S pre-initiation complex (PIC). The 43S PIC joins with capped mRNA whose 5' end is bound to eIF4F (composed of eIF4E/eIF4G/eIF4A) and whose 3' end is bound to PABP to form a 48S PIC. This complex scans the 5' untranslated region of the mRNA until the start codon is

recognized by Met-tRNA$_i^{Met}$. After recognition of the start codon, the ejection of eIF1 from the PIC triggers the hydrolysis of eIF2 bound GTP, a process requires eIF5 (a GTPase activating protein), followed by the release of Pi and the dissociation of eIF2 and eIF5. eIF5B, a second GTPase, facilities the recruitment of the large (60S) ribosomal subunit to form the 80S elongating ribosome, which represents the conclusion of translation initiation (*20-23*). Specific molecular mechanisms involved in translation initiation are discussed below.

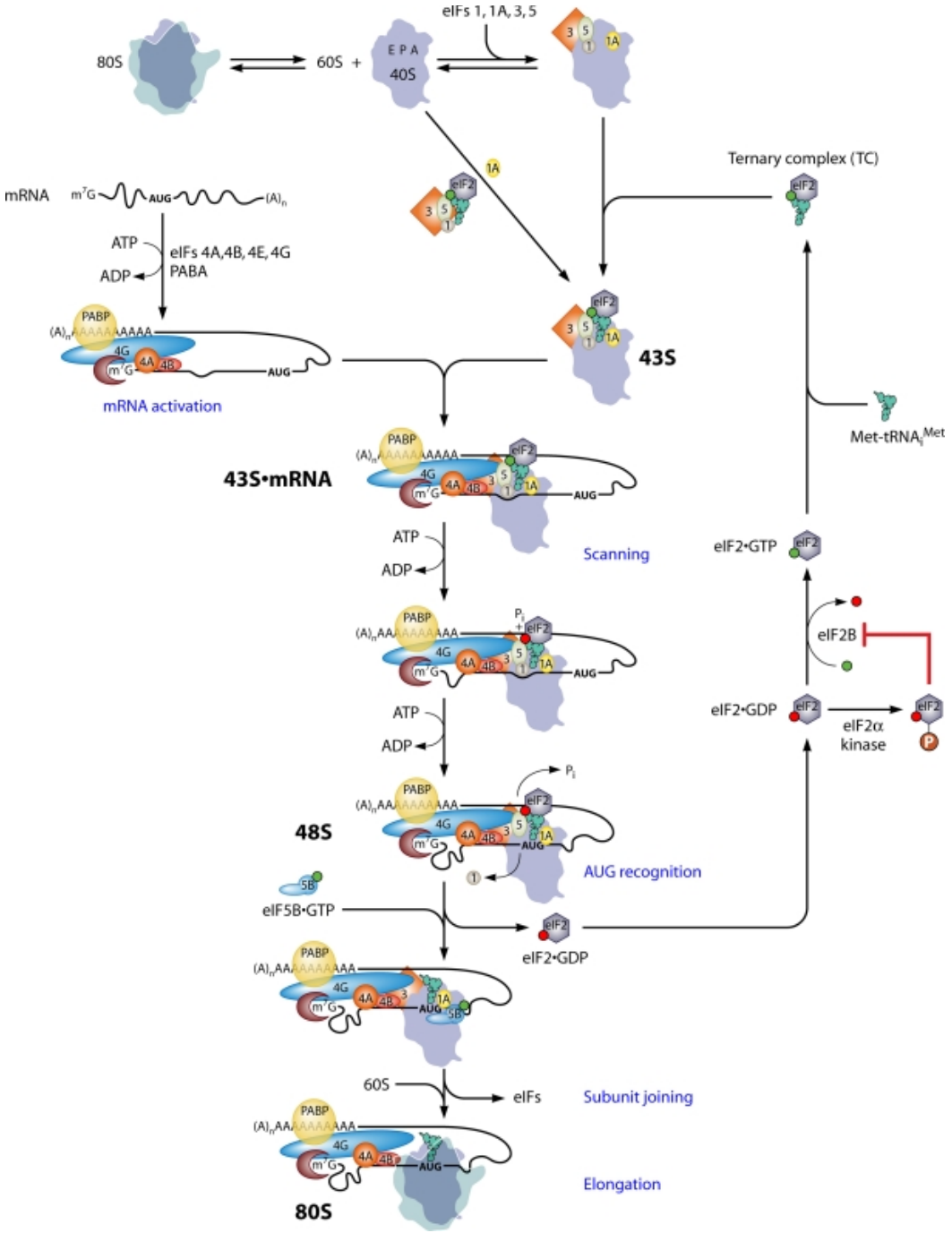


Figure 1.2. Simplified model for cap-dependent initiation of protein synthesis in eukaryotes. See the text for further details. Adapted by permission from American Society for Microbiology (*21*).

1.1.1.1. Formation of the TC

eIF2, consisting of α, β, and γ subunits, brings the initiator tRNA to the 40S subunit. The N-terminal region of eIF2β contains three lysine-rich segments (K-boxes), each consistent of 6-8 consecutive lysine residues. These K-boxes facilitate the binding of eIF2 to eIF5, eIF2B, and RNA (*24-26*). The γ subunit of eIF2 is implicated in GTP and Met-tRNA$_i$ binding and shows sequence and structural similarity to EF-Tu/eEF1A, which is a GTPase that delivers aminoacyl-tRNAs to their binding sites on ribosomes (*27-33*). Footprinting studies reveal direct interactions between the aminoacyl-tRNAs and γ and β subunits, but not the α subunit, suggesting that eIF2α may affect tRNA binding allosterically (*34*). The specific binding of eIF2 with the initiator tRNA and not elongator tRNAs results from the recognition of the methionyl residue and an A1:U72 pair at the end of the receptor stem of the initiator tRNA (*6, 22*). Direct contacts between eIF2γ and 18S rRNA helix 44, which spans the body of the 40S subunit, are also detected (*34*).

eIF2 binds Met-tRNA$_i$ 15 fold more tightly in its active GTP-bound form than in its inactive GDP-bound form (*22*). After the initiation codon is recognized, eIF2 leaves the ribosome in the inactive GDP-bound form. GDP is then exchanges for GTP by the guanine nucleotide exchange factor eIF2B to promote the binding of another Met-tRNA$_i$ and the reassembly of the TC (*6, 35-37*). The phosporylation of a conserved serine residue at position 51 on the eIF2α subunit results in a stable interaction between eIF2α and eIF2B. The sequestration of eIF2B reduces the rate of GDP-GTP exchange and the formation of TC, which decreases general translation initiation. This pathway provides a

way of translational control in response to multiple stress conditions by activating kinases that phosphorylate eIF2α (*38, 39*).

1.1.1.2. Formation of the 43S preinitiation complex (43S PIC)

The TC, formed by the GTP-bound eIF2 and Met-tRNA$_i^{Met}$, binds to the 40S ribosomal subunit to form the preinitiation complex (43S PIC), which can be purified by sucrose gradient centrifugation (*37*). The assembly and function of the PIC appears to be facilitated by the eIF1, eIF1A, eIF5, and eIF3 complex (*40-42*). Both N-terminal and C-terminal domains of eIF5 interact directly with eIF2 (*43*). eIF1, eIF1A, and eIF3 bind directly and cooperatively to the 40S subunit. eIF1 and eIF1A locate near the ribosomal P site and A site, respectively (*44, 45*). C-terminal and N-terminal tails (CTT and NTT) of eIF1A project into the P site in the PIC (*45*). The NTT interacts with eIF2 and eIF3 to promote TC recruitment *in vivo* (*46*). eIF1 and eIF1A induce an "open" conformation of the PIC to enable the binding of PIC to capped mRNA and the scanning of PIC on the mRNA to locate the start codon. In mammals, eIF3 is the largest and most complex initiation factor. It binds to the solvent face of the 40S subunit and interacts with the TC, eIFs 1, 1A and 5 (*47*). It is proposed to coordinate the assembly of PIC by serving as a bridge to connect components bound near the mRNA entry and exit channels (*48-50*). In the absence of the ribosome, eIF3 can also form a multifactor complex (MFC) with eIFs 1, 5 and the TC to prevent the diffusion of these initiation components (*40*).

1.1.1.3. Association of the preinitiation complex with mRNA

The association of the PIC with capped mRNA is a complex process that involves PABP and several initiation factors. The eIF4F complex is composed of eIF4A (an RNA helicase), eIF4E (a cap-binding protein), and eIF4G (*51*). eIF4G serves as a scaffold protein with multiple binding domains for other initiation factors and mRNA (*52, 53*). eIF4G increases eIF4E's affinity for the cap structure of mRNAs (*54-57*). eIF4A, an ATP-dependent RNA helicase, enhances the binding of the 43S PIC at the cap structure of mRNAs that contain secondary structures in their 5'UTR (*58, 59*). eIF4G interacts with mRNA sequences both upstream and downstream of the nucleotides in the initiation sites to position eIF4A at the mRNA entry channel (*60*). The binding of eIF4G to eIF4A helps to align the DEAD-box motif of eIF4A in the orientation required to unwind secondary structures in the mRNA to produce a single-stranded binding site for the 43S PIC near the 5' cap (*61-63*). eIF4G also interacts with the PABP (*64*). A stable, circular messenger ribonucleoprotein (mRNP), referred to as the "closed loop" structure, is created by linking the eIF4E bound cap structure and the eIF4G bound poly(A) tail. The formation of this "closed loop" promotes reinitiation by post-termination ribosomes and is proposed to provides a guard against the synthesis of truncated proteins that could be harmful to cells (*16, 65*).

In one model of mRNA recruitment, the mRNA is first processed to form an "activated mRNP", which then binds to the 43S PIC. eIF4G facilitates 43S PIC association with this "activated mRNP" by interacting simultaneously with mRNA and initiation factors bound to the PIC (eIF3 in mammals and eIF1 and eIF5 in yeast) (*21*).

As a result, the 43S PIC is recruited to the mRNA and is ready to scan downstream (*55*). Another model proposes that eIF4F factors bind directly to the PIC to form a "holoPIC" to directly recruit an mRNA. Distinguishing between these two models will require the development of new experimental approaches (*66*).

1.1.1.4. Recognition of the start codon

After binding to capped mRNA, the 43S PIC scans the mRNA for an initiation site (typically an AUG codon) in a suitable sequence context. During scanning, ATP hydrolysis enables eIF4A to remove any secondary structure (*67*). When an AUG enters the P site of the 40S subunit, a codon-anticodon interaction with Met-$tRNA_i^{Met}$ signals the start of translation. Sequences surrounding the AUG codon are also important for initiation codon recognition (*68, 69*). An AUG codon with an A at the -3 position and a G at the +4 position is favored for efficient initiation (*70*). An AUG codon in a poor context may be bypassed by the PIC. The scanning process and the recognition of the AUG codon require several initiation factors, including eIF1, eIF1A, eIF2, eIF3 and eIF5 (*6, 40, 71-74*). eIF1 and eIF1A promotes scanning by stabilizing an open conformation of the 43S PIC when non-AUG codons are in the P site (*74*). After eIF1 dissociation and Pi release in response to codon-anticodon base pairing, PIC closing is accompanied by movements of domains of eIF1A and eIF5 (*75-77*).

Pi release, which is markedly accelerated by start codon recognition, is more important to AUG recognition than GTP hydrolysis itself (*78*). Release of Pi, making GTP hydrolysis irreversible, is controlled by the dissociation of eIF1 from the PIC. eIF1A, which occupies the A-site, can regulate start codon selection in two ways. Its C-

terminal tail (CTT) supports continued scanning at non-AUG codons and must be evicted from the P site upon $tRNA_i$ accommondation, whereas its N-terminal tail has the ability to promote eIF1 release at AUG codons and arrest scanning (*45, 79*). eIF5 itself can enhance the dissociation of eIF1 from the PIC (*80*). Mutations in $tRNA_i$ and eIF3 that impair start codon recognition suggest that $tRNA_i$ itself and eIF3 also contribute to this molecular mechanism (*81*).

1.1.1.5. Assembly of the 80S ribosome

The translation initiation stage concludes when a second GTPase initiation factor eIF5B interacts with the CTT of eIF1A to stimulate the dissociation of eIF5 and eIF2·GDP and stimulates the recruitment of the 60S ribosomal subunit (*82*). After subunit joining, eIF5B hydrolyzes its bound GTP to trigger a conformational rearrangement in the 80S initiation complex (*82-84*). After the release of eIF5B, eIF1A leaves the interface region of the ribosome. eIF1A is thus a factor that functions from the beginning of translation initiation to its end (*85*). As a result, the 80S initiation complex contains Met-$tRNA_i^{Met}$ in the P-site and an empty A-site for the binding of the first aminoacyl-tRNA to start elongation.

1.1.1.6. Internal initiation

Although the majority of eukaryotic translation initiation is cap-dependent, some mRNAs can bypass the conventional scanning mechanism by using an internal ribosome entry site (IRES) to directly recruit ribosomes and initiate translation (*86*). Internal initiation was first discovered in picornaviruses (*87, 88*). Initiation from IRESs needs

several noncanonical IRES trans-acting factors (ITAFs), which can stabilize the active conformations formed by highly structured sequence elements within the IRES (*89, 90*). Translation initiation from IRESs provides a unique advantage for viral mRNAs. The IRESs of hepatitis C virus (HCV) bind directly to the 40S subunit without eIF4F (*91, 92*) and the IRESs of cricket paralysis virus (CrPV) can initiate translation without any eIFs or even Met-tRNA$_i$ (*93, 94*). This allows viral IRESs to recruit host ribosomes without the need to compete for eIFs in infected cells. Viral proteins can cleave host cell proteins or change the phosphorylation state of host proteins to shut down cellular cap-dependent translation, while allowing viral cap-independent translation (*95, 96*).

Several IRESs were discovered in cellular mRNAs that are active during mitosis or apoptosis. These IRESs facilitate the expression of essential regulators when cap-dependent translation is impaired (*97, 98*).

1.1.2. Elongation

Translation elongation mechanisms are relatively conserved between prokaryotic and eukaryotic systems (*99, 100*). Elongation can be divided into three major steps: the decoding of an mRNA codon by the cognate aminoacyl-tRNA, the formation of peptide bond through peptidyl transferase activity, and the translocation of the tRNA-mRNA complex. These steps are facilitated by several elongation factors.

At the beginning of elongation, a peptidyl-tRNA is in the ribosomal P site and the A site is empty to accept the aminoacyl-tRNA. The eukaryotic elongation factor eEF1A, homologous to prokaryotic EF-Tu, binds either the cognate or noncognate aminoacyl tRNA in a GTP-dependent manner and carries the tRNA to the A site. The

selection of the cognate tRNA relies on several factors, including the codon-anticodon interaction between the mRNA and the tRNA, conformational changes in the ribosome and GTP hydrolysis by eEF1A/EF-Tu (*101*). After the formation of the peptide bond between the peptidyl-tRNA in the P site and the aminoacyl-tRNA in the A site, translocation occurs to move the newly formed peptidyl-tRNA from the A site to the P site and the deacylated tRNA from the P site to the E site, thus resulting in an empty A site to accept another aminoacyl-tRNA.

1.1.2.1. Role of eEF1A

The eukaryotic elongation factor 1 (eEF1) consists of eEF1A and eEF1B (*102*). eEF1A delivers the aminoacyl-tRNA to the empty A site and hydrolyzes eEF1A-bound GTP by its GTPase activity. eEF1B, a guanine nucleotide exchange factor (GEF), functions similarly to prokaryotic EF-Ts but through a different mechanism. eEF1B accelerates the dissociation of GDP from eEF1A to recycle the inactive GDP-bound factor to its active GTP-bound state (*102*).

The fidelity of elongation is high with the rate of misincorporated tRNAs between 10^{-3} and 10^{-4} (*103*). Since both noncognate tRNA and cognate tRNA are carried to the A site of the ribosome, aminoacyl-tRNA must be carefully selected. During the initial selection, eEF1A·GTP·aa·tRNA presents the noncognate and cognate tRNAs to the decoding site of the mRNA-programmed 80S ribosome at different angles (*104*). For the noncognate tRNA, eEF1A·GTP·aa·tRNA dissociates from the ribosome more rapidly and the GTPase activation of eEF1A/EF-Tu occurs more slowly (*105*). An accepted codon-anticodon recognition triggers a signal transferred from the small

ribosomal subunit (decoding site) to the large ribosomal subunit (eEF1A GTPase domain binding site) to hydrolyze GTP (*101, 106, 107*). After this irreversible step of GTP hydrolysis, the rate of tRNA accommodation is faster for cognate tRNAs than the rate for noncognate tRNAs. This difference further ensures a high accuracy of tRNA selection (*101, 108-111*).

1.1.2.2. Peptidyl transferase and translocation

A peptide bond is formed between the aminoacyl-tRNA in the A site and the peptidyl-tRNA in the P site. The peptidyl transferase center (PTC), built up of highly conserved rRNA elements on the large ribosomal subunit, promotes peptide bond formation (*112*). The peptidyl transferase reaction starts when the α-amino group of the A site aminoacyl-tRNA nucleophilically attacks the carboxyl carbon of the C-terminal amino acid in the P site (Figure 1.3). The ribosomal large subunit does not chemically catalyze this reaction (*113, 114*). The ribosome accelerates the peptide bond formation by positioning the two substrates and changing the rate-limiting transition state (*115*).

Immediately after peptide bond formation, conformational changes in the ribosome move the acceptor ends of the tRNAs in the P and A sites to the E and P sites, with the anticodon loops remaining in the P and A sites, respectively. As a result, a so-called hybrid P/E and A/P state is formed (*3*). This hybrid state is stabilized the binding of eukaryotic elongation factor 2 (eEF2) or prokaryotic EF-G. The GTP hydrolysis by eEF2/EF-G and the release of Pi promote a conformational change in the ribosome to "unlock" the ribosome, triggering the movement of tRNA-mRNA, leading to the formation of P/P and E/E states (*116-118*). Spontaneously, another conformational

change is induced by the release of Pi from eEF2/EF-G to bring the ribosome back to its "locked" state. Pi release is also coupled to the release of eEF/EF-G from the ribosome and the release of the deacylated tRNA from the E site.

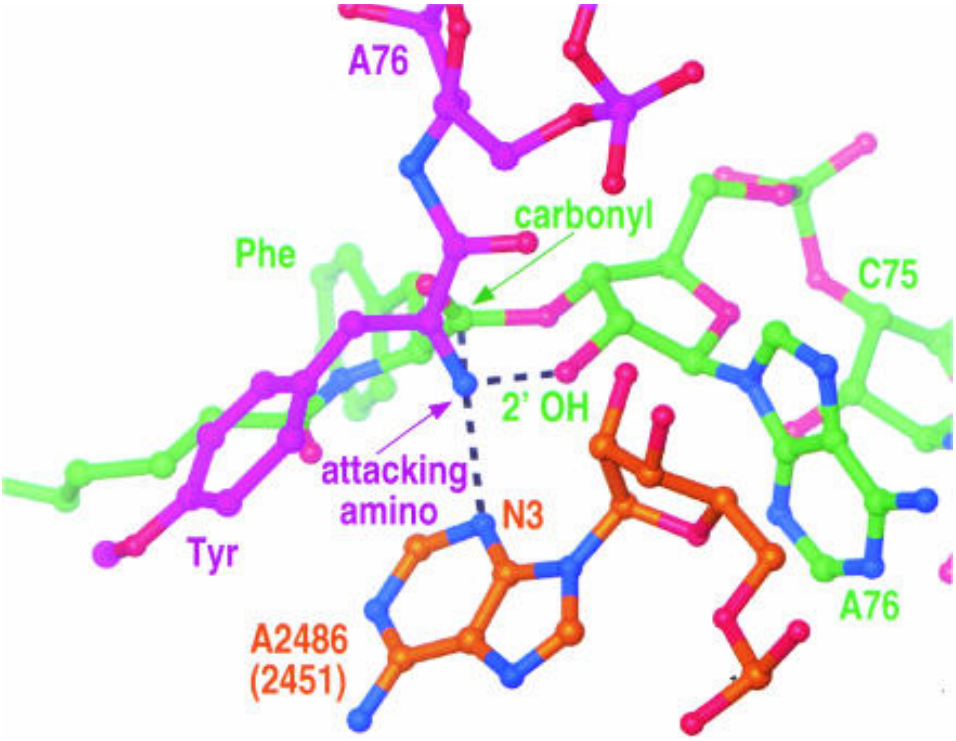


Figure 1.3. A model for the peptidyl transferase center from *Haloarcula marismortui* with substrates in both A site and P site. This model was obtained by superimposing an A-site substrate complex (PDB #1FGO, purple) on the structure of a P-site substrate complex (PDB #1M90, green). Possible hydrogen-bonding interactions involving the α-amino group and the N3 of A2486 (2451) and the 2'-OH of A76 are indicated. Adapted by permission from RNA society (*112*).

1.1.2.3. Role of eEF2 and eEF3

eEF2 binds to the 80S ribosome. It functions as a GTPase and catalyzes the translocation of the two tRNAs and the mRNA after peptide bond formation. The hydolysis of the eEF2/EF-G-bound GTP facilitates the translocation and the release of eEF2/EF-G from the ribosome (*116*). In the absence of eEF2/EF-G, tRNAs can move in the backward direction spontaneously. That is, tRNAs can move from the post- to the pre-translocation position (*119, 120*). The large-scale conformational changes of eEF2/EF-G during translocation may promote the directional movement of tRNAs and prevent the backward sliding of tRNAs (*121, 122*).

eEF3, found exclusively in fungi, is an exception to the evolutionary conservation of elongation (*123*). eEF3, containing two ATP-binding cassettes (ABCs), may facilitate the release of deacylated tRNA from the E site (*124-126*). The gene encoding eEF3 is essential for yeast viability (*127*). eEF3 associates with translating ribosomes, interacts with eEF1A, and is found primarily in polysome fractions, consistent with its requirement for peptide bond formation (*128, 129*). Further studies are needed to determine the function of eEF3 in translation elongation.

1.1.2.4. Role of eIF5A

eIF5A was originally shown to stimulate the transfer of methionine from Met-$tRNA_i$ in the 80S initiation complex to the aminoacyl-tRNA analog puromycin and was denoted as an initiation factor (*130*). Recent studies have implicated eIF5A in the process of translation elongation (*131*). Inactivation of an eIF5A temperature-sensitive mutant resulted in polysome retention in the absence of the elongation inhibitor

cycloheximide and increased ribosome transit times. The interaction between eIF5A and eEF2 further supports a role for eIF5A in translation elongation (*131*). eIF5A stimulates the reactivity of peptidyl-tRNA in the ribosomal P site with either aminoacyl-tRNA or protein ligands that enter the A site, suggesting its mechanism of action (*131*). eIF5A also promotes translation of polyproline motifs (Dever et al., unpublished data). The detail function of eIF5A in translation elongation needs to be determined by future studies.

1.1.3. Termination

Translation termination occurs when the ribosome reaches the end of the coding sequence with a stop codon entering the ribosomal A site. The sequence surrounding the termination codon has an effect on the efficiency of termination (*132-134*). The termination process includes the recognition of the stop codon and the hydrolysis of peptidyl-tRNA. Hydrolysis of peptidyl-tRNA is followed by the nucleophilic attack of water on the ester carbonyl group of the peptidyl-tRNA. The cleavage of an ester bond leads to the release of the nascent peptide. Prokaryotes have three release factors, RF1, RF2, and RF3. Eukaryotic cells have only two release factors, eRF1 and eRF3. The entry and the activation of water are facilitated by eRF1 through the induced conformational changes in the ribosome (*22*). Mutations in ribosomal RNA, eRF1 and eRF3 can increase stop codon read-through and lead to "nonsense suppression" phenotypes (*135, 136*).

1.1.3.1. eRF1 and eRF3

Class I release factors (RF1 and RF2 in prokaryotes and eRF1 in eukaryotes) directly recognize the stop codon and catalyze the hydrolysis of the peptidyl-tRNA (*137-141*). In prokaryotes, RF1 decodes UAA and UAG, while RF2 decodes UAA and UGA (*142*). eRF1 is the only class I release factor in eukaryotes and can recognize all three stop codons (UAA, UAG and UGA) (*143-145*). eRF1 is a tRNA-shaped protein composed of three domains: N-terminal (N), middle (M) and C-terminal (C) domains (*146*). Conserved residues of eRF1 from different parts of the polypeptide chain may cluster in space to form a three-dimensional network for decoding. A "cavity" model for the stop-codon recognition is suggested based on the importance of highly conserved NIKS and YxCxxxF motifs at the tip of eRF1 domain N. The highly conserved GGQ motif in the M domain of eRF1 triggers peptide-tRNA hydrolysis when correctly positioned in the PTC of the ribosome (*146*). The C domain is involved in the interaction with the class II release factor eRF3 (*146*).

Class II release factors (RF3 in prokaryotes and eRF3 in eukaryotes) are ribosome-dependent GTPases (*147*). RF3 promotes the release of RF1/RF2 from the post-termination complexes (*148*). eRF3 binds to eRF1 with high affinity to form a stable complex and to enter the ribosome together. eRF1 stabilizes the binding of GTP (but not GDP) to eRF3 (*149*). Both the 80S ribosome and eRF1 are required for the activation of eRF3's GTPase activity. The GTP hydrolysis by eRF3 couples the stop codon recognition and the peptidyl-tRNA hydrolysis by eRF1 (*150-153*). Peptide

release is very inefficient with eRF1 alone. eRF3 ensures rapid and efficient peptide release (*150, 151*).

1.1.3.2. The eRF1 GGQ motif

The GGQ motif is universally conserved in all class I release factors and positioned in a loop at the end of a stem region that interacts with the PTC (*154-156*). The binding of eRF1 to the A site is stabilized by eRF1's interactions with both of the ribosomal subunits. The GTP hydrolysis by eRF3 induces conformational changes in eRF1, triggering the GGQ loop to enter the PTC (*157, 158*). The binding of the GGQ motif to the PTC causes conformational changes in the ribosome, exposing the ester bond to nucleophilic attack by a water molecule. As a result, the peptidyl-tRNA is hydrolyzed (*154-156, 159*). A mutation in the GGQ motif of eRF1 reduces the binding of eRF1 to the ribosome (*146, 160, 161*). The two Gly residues in the GGQ motif adapt a unique backbone conformation, which facilitates the binding of the GGQ motif to the PTC. The Gln may function to coordinate the water molecule in the PTC and stabilize the product of peptidyl-tRNA hydrolysis or the transition state (*140, 146, 154-156, 162*).

1.1.4. Recycling of ribosomal subunits

After termination, 40S and 60S subunits must be dissociated from the 80S ribosome and the mRNA and deacylated tRNA released for the next round of translation. Recycling of ribosomal subunits is different in prokaryotes and eukaryotes. In bacteria, the ribosome recycling factor (RRF) functions with EF-G to split the ribosome into subunits, following the dissociation of class I release factors by RF3. IF3 promotes the

dissociation of tRNAs and the destabilization of mRNA binding (*163, 164*). In eukaryotes, ABCE1, a conserved ATP-binding cassette (ABC) protein, has the ability to promote recycling of post-termination ribosomes in the presence of eRF1 (*165-167*). After eRF1/ABCE1-mediated splitting of post-termination ribosomes, both tRNA and mRNA remain bound to 40S subunits. The departure of deacylated tRNA and mRNA is enhanced by Ligatin (also known as eIF2D) or MCT-1/DENR (*168*). These proteins may function by stabilizing the open state of the 40S subunit from which tRNA would be expected to dissociate more rapidly.

The dissociation of ribosomes is also accelerated by eIF1, eIF1A, eIF3, eIF3j (eIF3's loosely associated subunit) and eIF6 (*167, 169*). The opening of the mRNA entry channel is promoted in the ribosome in response to the binding of eIF1 and eIF1A (*170*). The binding of eIF3 to the 40S subunit leads to an increased rate of dissociation of the 80S ribosome to its 40S and 60S components and a decreased rate of association of these two ribosomal subunits (*171, 172*). The efficient dissociation of mRNA from the ribosome in the presence of eIF3 requires its 3j subunit (*169*). eIF3j and mRNAs bind to the 40S subunit with a negative cooperativity (*173*). eIF6 binds to the 60S subunit, preventing it from re-associating with the 40S subunit (*174, 175*). The phosporylation of eIF6 releases eIF6 from the 60S subunit and allows the 60S subunit to join the 40S subunit (*176*). eIF6 is also shown to be involved in ribosome biogenesis; its depletion impairs the biogenesis of the 60S subunit (*177, 178*). Based on the closed-loop model, the 40S subunit is shuttled back to the 5'-end of mRNA to facilitate translation reinitiation.

1.2. Translational control

The control of translation can be global or local. Global control regulates the translation of most mRNAs in the cell. Global translational control typically targets eukaryotic initiation factors and accessory proteins to regulate initiation, the first and rate-limiting step in translation. mRNA-specific control regulates the translation of mRNAs involved in a common function without disturbing general protein synthesis. This allows two distinct transcripts that are expressed at similar levels to produce different amounts of protein (*1*). Sequences in the 5' and/or 3' UTRs function as *cis*-acting elements that recruit regulatory protein complexes that can increase or decrease the efficiency of translation. Some of these sequences are recognized by small microRNAs (miRNA) that are components of regulatory complexes that modulate the efficiency of translation. The molecular mechanisms of both global and mRNA-specific translational control will be discussed.

1.2.1. Global control

Global control is achieved by regulating the activity of key components in the translation machinery. Many proteins involved in translation are phosphoproteins whose activity can be turned on and off by their phosphorylation status. The control of those proteins' phosphorylation and dephosphorylation is the major identified general mechanism of global translational control in eukaryotic cells.

1.2.1.1. Phosphorylation of eIF2

As discussed in section 1.1.1, eIF2 is a part of the ternary complex that brings Met-tRNA$_i^{Met}$ to the small ribosomal subunit. The hydrolysis of eIF2 bound GTP occurs during translation initiation. eIF2 is released from the ribosome in the inactive GDP bound form, which must be converted to the GTP bound form to participate in a subsequent round of translation initiation. The conversion from GDP to GTP is performed by eIF2B. Phosphorylation of serine 51 of eIF2α increases the binding affinity of eIF2B for eIF2 (*35*). The eIF2B·eIF2 complex is inactive and blocks the GTP-exchange reaction required for eIF2 activation. Since eIF2B is present in limited amounts compared to eIF2, a small change in the level of phosphorylated eIF2α has a significant impact on the overall translation rate (*39*).

A family of four kinases are known to phosphorylate eIF2α at serine 51 in response to different physiological conditions (*179*), including heme-regulated inhibitor (HRI), which is stimulated by heme-depletion (*180*); general control non-derepressible-2 (GCN2), which responds to amino acid starvation (*35*); double-stranded-RNA-dependent protein kinase (PKR), which is activated by viral infection (*181*); and the PKR-like endoplasmic reticulum kinase (PERK), which is triggered by endoplasmic reticulum (ER) stress (*182, 183*). eIF2α is dephosphorylated when a mitogen is added to quiescent cells (*184*). However, limited phosphorylation of eIF2α can promote the translation of bZIP transcription factors, such as yeast GCN4 (*185*), the details of which will be discussed later.

1.2.1.2. The interaction of eIF4E with 4E-BPs

During translation initiation, eIF4E (a cap-binding protein) interacts with eIF4G and eIF4A to form the eIF4F complex. The eIF4F complex recruits the 43S PIC to mRNAs. Interactions between eIF4E and the 5'-cap structure are enhanced when eIF4E is phosphorylated by the p38-mitogen-activated-protien kinase (MAPK) (*186, 187*). Thus, phosphorylation of eIF4E can increase translation when cells are treated with hormones, growth factors, cytokines, and mitogens (*186, 188*).

The binding domain of eIF4G for eIF4E is shared by 4E-binding proteins (4E-BPs). When hypophosphorylated 4E-BPs bind to eIF4E, they competitively reduce the binding of eIF4G, thereby inhibiting the association of PIC with mRNAs to repress cap-dependent translation (*188, 189*). 4E-BPs dissociate from eIF4E when they are in a hyperphosphorylated state. The phosphorylation of 4E-BPs is triggered by growth factors and nutrient status (*188, 190, 191*). In mammals, phosphorylation of 4E-BP1 is controlled by Akt/PKB and the mammalian target of rapamycin (mTOR) signaling pathways to regulate cell proliferation (*192-194*). In addition, during apoptosis, the caspase-dependent cleavage of 4E-BP1 produces a truncated polypeptide. This polypeptide retains the eIF4E-binding capacity. However, it cannot be sufficiently phosphorylated. It thus inhibits cap-dependent translation by failing to be released from eIF4E (*195*). Certain viruses, such as encephalomyocarditis virus (EMCV), also decreases the phosphorylation of 4E-BP1 to shut-off the host protein synthesis without affecting viral cap-independent translation (*96*).

1.2.1.3. eIF4G

eIF4G increases eIF4E's affinity for the 5'-cap structure of eukaryotic mRNAs. eIF4G itself can be phosphorylated by the cytostatic p21-activated kinase (Pak2) in response to a variety of stress conditions that inhibit cap-dependent translation (*196*). Pak2 phosphorylates several initiation factors including eIF4G, eIF3, and eIF4B, as well as the MP-kinase-interacting kinase: Mnk1 (*197, 198*). Phosphorylation of Mnk1 by Pak2 can reduce its binding to eIF4G and decrease the phosphorylation of eIF4G (*197*). eIF4G can be cleaved into two fragments by the leader proteinase of the foot-and-mouth disease virus (FMDV) (*199, 200*). This cleavage inhibits the cap-dependent translation of host proteins, whereas IRES-dependent translation used by the virus is not affected (*201*).

1.2.1.4. eEF2

eEF2 can be phosphorylated at its N-terminal GTP-binding domain to inhibit its binding to the ribosome and repress translation elongation. eEF2 kinase is a Ca^{2+}/calmodulin-dependent protein kinase that is phosphorylated by the mTOR and stress-activated protein kinase SAPK4 pathways (*202-204*). The phosphorylation of eEF2 kinase inhibits its activity, leading to an increased level of dephosphorylated eEF2 and more efficient translation elongation. Recent studies suggest that, under acute nutrient depletion conditions, eEF2 kinase blocks translation elongation to confer cell survival. eEF2 kinase can protect cells from nutrient deprivation and help tumor cell to adapt to metabolic stress (*205*). However, in response to eEF2 phosphorylation, the translation of certain neuronal mRNAs increases. The repression of general translation

may increase the availability of initiation factors and thus reduce the negative effect of eEF2 phosphorylation (*206*).

1.2.2. mRNA-specific translational control

mRNA-specific regulation confers rapid and coordinated control over one or more mRNAs, for example, over a set of transcripts encoding proteins with related functions, without affecting overall rates of translation.

1.2.2.1. Regulation by the 5'-untranslated region (5'-UTR)

Sequences in the 5'-UTR and 3'UTR can confer transcript-specific regulation. The length of 5'-UTRs is typically short with a low GC content. Long 5'-UTRs (>200 nucleotides) tend to decrease the efficiency of translation initiation. Translation can be controlled by certain elements in the 5'-UTR, including secondary structure, 5'-terminal oligopyrimide (TOP) tracts, IRESs and uORFs. The regulatory function of uORFs is discussed in Section 1.3.

5'-terminal oligopyrimide (TOP) tracts are found in 5'-UTRs of mammalian ribosomal proteins and other proteins involved in translation. TOP contains a stretch of 5-14 consecutive pyrimidine residues and provides regulation of TOP mRNAs in a growth-dependent manner. The regulation of TOP mRNAs is related to both mTOR and PI3K signaling pathways though the detailed mechanism is controversial (*207, 208*). Recent studies suggest that RNA binding proteins TIA-1 and TIAR, two general translational repressors in stressed cells, are required for the translational control of TOP mRNAs (*209*). Translational control of TOP mRNAs enables cells to coordinately

express proteins that are required for ribosome biogenesis in response to amino acid starvation or growth arrest (*210*).

As discussed in section 1.1.1.6, IRES can facilitate cap-independent translation. The IRES elements typically found in viral transcripts assume highly ordered structures within the 5'-UTR (*211*). IRES can recruit the 43S PIC to the mRNA without the 5' cap structure and even without initiation factors (*212-214*). HIV-2 genomic RNA can initiate translation by using IRES elements located downstream of the AUG start codon (*215*). IRES elements in cellular mRNAs increase the expression of several proteins, including oncogenes, growth factors and proteins involved in programmed cell death (*214, 216-221*).

1.2.2.2. Regulation by 3'-untranslated region (3'-UTR)

A cytoplasmic polyadenylation element (CPE) in the 3'-UTR of *c-mos* mRNA interacts with the cytoplasmic polyadenylation element binding protein (CPEB) to regulate translation during vertebrate oocyte maturation and early development (*222*). When CPEB binds to Maskin, a 4E-BP that regulates translation of target mRNAs, formation of CPE·CPEB·Maskin·eIF4E complexs repress translation. Following maturation, CPEB is phosphorylated, triggering elongation of the poly(A) tail, binding of PABP and the dissociation between Maskin and eIF4E. As a result, translation is activated (*223, 224*).

In *Drosophila melanogaster* embryogenesis, CUP functions similarly to Maskin in vertebrte oocyte development to bind with eIF4E and block the recognition of eIF4E by eIF4G. CUP is recruited to the 3'-UTR of *nanos* and *oskar* mRNAs by RNA binding

proteins Smaug and Bruno, respectively. *nanos* and *oskar* mRNAs encode the posterior determinant during the specification of the anteroposterior axis. This provides an example in which a group of functionally related genes is coordinately regulated during development (*225, 226*).

In *D. melanogaster*, male specific lethal 2 (MSL-2) protein is a crucial component of the dosage compensation complex. *msl-2* mRNA can be found in both females and males, while its expression is lethal in females. The regulation of MSL-2 is achieved by sex-lethal (SXL) protein. In females, SXL binds to poly(U) tracts in both the 5'- and 3'-UTRs of *msl-2* mRNA (*227, 228*). The mechanisms of repression are different for the 5'- and 3'-UTRs. In the 5'-UTR, SXL provides a physical block to interfere with ribosome scanning and the recognition of AUG start codons. SXL binding to the 3'-UTR recruits co-repressors(s) and inhibits translation by preventing the stable association of 43S PIC with mRNA (*229*).

1.2.2.3. Regulation by miRNAs

microRNAs (miRNAs) are 20-22 nucleotide small noncoding RNAs that can regulate gene expression by binding complementary sequences in mRNAs. Multiple biological processes are affected by miRNAs, such as cell metabolism, cell differentiation, cell growth and apoptosis (*230-232*). In plants, the complementary sites of miRNAs reside in both coding regions and non-coding regions. A limited number of mRNAs are regulated by miRNAs, which show a higher degree of complementarity to target mRNAs than animal miRNAs (*233, 234*).

In animals, it is estimated that 30% of protein-coding genes are regulated by miRNAs (*235, 236*). Most target sites reside in the 3'-UTRs of target mRNAs and multiple sites exist to improve the strength of regulation (*237-241*). miRNA can alter expression by promoting mRNA decay and inhibiting translation (*234, 242-247*). Ribosome profiling studies in mammals indicate that mRNA decay is the major force driving miRNA-mediated gene silencing (*248*). However, kinetic analyses suggest that translational repression occurs first, followed by mRNA deadenylation and decay (*249*).

Alternative polyadenylation can increase gene expression by generating a truncated 3'UTR lacking certain miRNA target sites (*250*). Enhanced mRNA stability and translatability usually results from deletion of these elements. Increased oncogene expression is related to 3'UTR shortening in cancer cells that is mediated by alternative polyadenylation (*251*).

Recent studies revealed a correlation between the presence of secondary structures in the 5'UTR and miRNA target sites. Unstructured 5' UTRs are refactory to miRNA repression. These data support that during miRNA-mediated gene regulation, translational repression via eIF4A2 is required first, followed by mRNA destabilization (*252*).

1.3. Translational control mediated by upstream open reading frames (uORFs)

In eukaryotes, cap-dependent translation initiation in eukaryotes requires ribosomes to scan from the 5'-cap structure to the 3'-end of mRNA to locate the start site (typically an AUG codon) (*69*). Some eukaryotic mRNAs contain AUG codons

upstream of the main AUG codon. These codons can serve as translation start sites in some circumstances (*253*). uAUGs are present in two-thirds of genes related in cancer and are common in genes involved in cellular growth and differentiation (*254, 255*). uAUGs in the 5'-UTR generally decrease the efficiency of initiation from the start codon of the main (genic) open reading frame (ORF). An upstream open reading frame (uORF) is defined as an upstream start codon (typical uAUG), at least one sense codon, and a termination codon. An evolutionary selection of functional uORFs is suggested by their higher conservation and lower frequency of appearance than expected by chance (*256, 257*). Bioinformatics studies indicate 49% of human and 44% of mouse mRNAs contain at least one uORF. An negative correlation between the prevalence of one or multiple uORF(s) in a transcript and the abundance of the respective protein has been established (*2*), consistent with uORFs generally reducing downstream translation.

uAUG and the main AUG are recognized to start translation using the same mechanism. The efficiency of initiation can be influenced by the initiation context, which is defined by the specific nucleotides flanking translation initiation codons (*68, 253, 258*). There are four major ways for uORFs to affect downstream ORF initiation. Firstly, a uORF could initiate translation and therefore reduce the possibility of initiation at a downstream start codon because ribosomes scan from the mRNA cap. Some 43S PICs will pass the uAUG codon, especially if it is in a poor initiation context, and reach the downstream AUG codon by leak scanning. Secondly, uORFs can control ribosome reinitiation. In the reinitiation model, the small ribosomal subunit remains associated with the mRNA after termination at the uORF stop codon and continues to scan. The

efficiency of reinitiation decreases as the length of the uORF increases, the rate of translation elongation decreases, and the distance between the uORF termination codon and the downstream start codon increases (*259, 260*). Thirdly, some uORF-encoded nascent peptides can arrest translating ribosomes at their termination codons. The stalled ribosome acts as a blockade to the scanning ribosome to prevent ribosomes from reaching the downstream initiation codon (*261-264*). In addition, translating ribosomes can stall within the uORF coding regions. The arrest stalls uORF translation and blocks ribosome scanning, thus reducing the translation of the main ORF (*265-267*).

Accumulating evidence indicates that eukaryotic uORFs can also influence translation by affecting mRNA stability (*261*). Nonsense-mediated mRNA decay (NMD) pathway is triggered by the uORF termination codon to decay mRNAs containing uORFs. In plants, a 50 amino acid long uORF can induce NMD efficiently while shorter uORFs (31 and 15 aa long) cannot, suggesting that uORFs trigger NMD in a size-dependent manner in plants (*268*). Ribosome stalling at the *N. crassa arg-2* uORF termination codon activates NMD to decay the *arg-2* transcript (*269*). Depletion of a factor essential for NMD increases the abundance of transcripts with uORFs in mammalian cells (*270*). In *Caenorhabditis elegans*, *gna-2* mRNA, containing two uORFs, is targeted by NMD (*271*).

1.3.1. Genes regulated by uORFs

A number of uORFs have been identified and their functions investigated. Several well characterized examples will be discussed here to demonstrate the molecular mechanisms of uORF function.

1.3.1.1. Mammalian *AdoMetDC*

Mammalian *S*-Adenosylmethionine decarboxylase (AdoMetDC) catalyzes a rate-limiting step in the synthesis of spermidine and spermine (both are polyamine compounds) (*272*). A uORF that starts only 14 nucleotides downstream of the 5'-cap is responsible for regulating the translation of AdoMetDC. This uORF encodes a short peptide with the sequence MAGDIS (*273, 274*). The amino acid sequence, especially Asp-4 and Ile-5, is crucial to translational control (*274*). The expression of AdoMetDC changes in response to the level of polyamine (*275*). When the concentration of polyamine is low, the scanning ribosome bypasses uORF initiation codon and initiates at the AdoMetDC ORF to induce the production of polyamine. When the concentration of polyamine rises, polyamine promotes the initiation of the uORF and causes ribosome stalling at uORF termination codon by impairing the hydrolysis of the peptidyl-tRNA upon termination (*276*). The stalled ribosome further prevents the initiation from AdoMetDC start codon to downregulate the polyamine synthesis (*273, 275, 277*).

1.3.1.2. Mammalian *her-2*

Oncogene *her-2* encodes a transmembrane receptor tyrosine kinase (*278, 279*). Overexpression of Her-2 protein occurs in many primary human tumors and contributes to approximately 30% of breast and ovarian carcinomas. A uORF is present in *her-2* mRNA and represses downstream HER-2 expression in multiple cell types (*280, 281*). This uORF affects the selection of downstream initiation site by interfering with the interaction between the ribosome and the primary *her-2* AUG codon (*282*). The translational derepression element (TDE) in the 3'-UTR in specific tumor cells abolishes

the repression effect of *her-2* uORF by allowing efficient translation reinitiation from the downstream ORF start codon (*283*).

1.3.1.3. Human cytomegalovirus gp48

The human cytomegalovirus gp48 gene encodes a structural glycoprotein (*284*). Three uORFs are found in the 5'-UTR of gp48 and the 22-codon uORF2 is responsible for regulating the expression of gp48 (*285*). The AUG codon of uORF2 is poorly recognized by the scanning ribosomes. Most ribosomes will pass it and initiate translation at the AUG codon of gp48 (*286, 287*). However, once the uORF2 is translated, it cause ribosomes to stall and strongly inhibit the translation of gp48 (*286*). The uORF2-encoded peptide inhibits translation at its own termination codon, resulting in a ribosome-peptidyl-tRNA complex (*288*). This complex forms a blockade to scanning ribosomes from reaching the gp48 initiation codon, thus obstructing the downstream translation (*289*). The inhibition of uORF2 translation is amino acid sequence-dependent and results from the impaired hydrolysis of peptidyl-tRNA (*289*).

1.3.1.4. *S. cerevisiae GCN4, N.crassa cpc-1, Aspergillus nidulans cpcA* and mammalian *ATF4*

Yeast *GCN4* is the best-studied example of translational control mediated by reinitiation (*185, 290, 291*). GCN4p, a bZIP transcription factor, mainly functions to activate the transcription of the majority of amino acid biosynthetic genes in response to amino acid starvation. Expression of GCN4 is low in standard laboratory media. The translation of *GCN4* mRNA is activiated during amino acid limitation.

The induction of GCN4 translation in response to amino acid limitation is performed by four small uORFs in the 590-nt 5'-UTR of *GCN4* mRNA. Without these four uORFs, the efficiency of initiation at *GCN4* start codon is low as a result of the long 5'-UTR. uORF1 and uORF4 together are sufficient to control regulation (*185, 291*). uORF1 is efficiently translated under both nutrient-replete and starvation conditions (*292*). Translation of uORF1 promotes downstream initiation via reinitiation either at another uORF start codon (typically uORF4) or at the *GCN4* start codon (*293, 294*). Sequences immediately downstream of the uORF1 termination codon are crucial for reinitiation (*295*). uORF4 decreases the translation of *GCN4*. Without uORF1, the translation of uORF4 by itself inhibits the initiation at the *GCN4* start codon by approximately 99% (*185*). eIF3b functions between 48S assembly and subunit joining to influence reinitiation on *GCN4* mRNA(*296*).

Through the general amino acid control pathway (GAAC), amino acid starvation results in an increased level of uncharged tRNAs, which activates GCN2, a kinase that phosphorylates the α subunit of eIF2 (*297*) (*298*). The phosphorylated eIF2α is a competitive inhibitor of eIF2B, which is required in the recycle of eIF2·GDP to eIF2·GTP. Thus amino acids starvation reduced the level of eIF2·GTP, decreases the formation of TC and slows the rate of overall protein synthesis.

The mechanism of GCN4 translational control was elucidated by genetic methods and *in vitro* studies (*299*). Ribosomes that scan from the 5' cap will translate uORF1 and in approximately 50% of the 40S ribosome resume the scanning process after translation termination of uORF1. Normally, the level of eIF2α phosphorylation is

low and the active TC is available soon enough to bind 40S ribosome to reinitiate translation at downstream inhibitory uORFs. The expression of *GCN4* is then repressed. Under amino acid starvation conditions, the TC level is reduced in response to eIF2α phosphorylation via GCN2. Some 40S subunits that resume scanning do not bind TC soon enough to translate uORFs. Instead, ribosomes reinitiate at the start codon of *GCN4* (*290, 300*). Consistent with this model, extending the distance between uORF1 and uORF4 reduces the expression of *GCN4* under starvation conditions (*301*). uORF-mediated control of *GCN4* expression can also be regulated by mechanism that are independent of eIF2α phosphorylation under conditions such as nitrogen starvation (*302*).

The *N. crassa* and *Aspergillus nidulans* homologs of *S.cerevisiae GCN4* are *cpc-1* and *cpcA*, respectively. Both *cpc-1* and *cpcA* have roles similar to *GCN4* in controlling the amino acid biosynthesis pathway, which is called cross-pathway control in filamentous fungi (*303*). In addition to amino acid biosynthesis genes, microarray studies in *N. crassa* and *S. cerevisiae* also identify other genes in other pathways that are affected by *cpc-1* and *GCN4,* suggesting a broader role of *cpc-1* and *GCN4* (*304, 305*). Both *cpc-1* and *cpcA* transcripts contain two uORFs, with the second uORF considerably longer than uORF2-4 found in yeast *GCN4*. Translation of *cpc-1* and *cpcA* are controlled by uORFs in their 5'UTR (*303, 306, 307*). Translational regulation performed by *cpcA* is supported by the evidence that the removal of uORFs leads to an increased expression under normal condition and a reduced regulation in response to stress (*308*). In response to amino acid limitation, native *cpc-1* mRNA moves to the polysomes fraction, indicating the existence of translational regulation (*309*). In this thesis,

CHAPTER 4 will discuss the translational control mediated by uORFs in *N. crassa cpc-1* transcript.

In mammals, *ATF4* (activating transcription factor 4) is a stress response gene with multiple downstream targets. There are two uORFs in the 5'-UTR of the *ATF4* mRNA, with the second uORF overlapping with the *ATF4* coding sequence (*310, 311*). Translation of *ATF4* is activated by eIF2α phosphorylation (*312*). Under normal cell conditions, after translation of the first uORF, TCs are most often recruited to the initiation codon of the second uORF. Since the second uORF overlaps the *ATF4* coding sequence, the expression of *ATF4* is strongly inhibited. Under stress conditions, eIF2α is phosphorylated, TC availability is reduced, and there is less chance of recruiting TCs to start initiation at the second uORF. The start codon of *ATF4* is used instead and translation is thus activated (*311*). The *ATF4* uORFs thus function similarly to the yeast *GCN4* uORFs: uORF1 controls downstream translation by reinitiation and *ATF4* uORF2 and *GCN4* uORF4 strongly inhibit the translation of the main ORFs (*310*).

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

In *N. crassa*, the *arg-2* gene specifies the small subunit of arginine-specific carbomoyl phosphate synthetase (CPS-A), which is the first protein in the Arg-specific biosynthesis pathway (*313*). This small subunit functions as a glutamine amidotransferase to transfer the glutamine amide nitrogen to the large subunit, where carbamoyl phosphate is synthesized. The ARG-2 polypeptide contains a mitochondrial signal for its subcellular location (*314*). *arg-2* is the only gene that is known to be negatively regulated by Arg in *N. crassa*. The expression of *arg-2* is controlled by a

uORF in the 5'-UTR of *arg-2* mRNA. This uORF encoded a 24-amino acid peptide, named the arginine attenuator peptide (AAP). The AAP is highly evolutionarily conserved among many fungal species (*314-317*).

The regulation of *arg-2* expression occurs at both the transcriptional and translational level. Short exposure of *N. crassa* to high levels of Arg decreases the synthesis of ARG-2 to 50% without changing the level of the *arg-2* transcript, while long exposure will reduce the level of the *arg-2* transcript (*309*). Polysome profile studies reveal that a high concentration of Arg in the media shifts *arg-2* mRNA to the monosome fraction (*318*). *In vivo* analyses also indicate that the removal of uORF start codon abolishes Arg-specific regulation (*309*). Forward genetic selection indentifies Asp-12 as a crucial residue in the AAP to negatively regulate *arg-2* expression. The Asp to Asn mutation (D12N) eliminates Arg-specific regulation (*319*).

The uORF-mediated translational control of ARG-2 synthesis has been studied using an *N. crassa* cell-free translation system. Synthetic transcript containing wild type AAP coding sequence upstream of the firefly luciferase reporter gene shows negative regulation of luciferase (LUC) synthesis in response to Arg. This regulation is not observed for transcripts in which the uORF start codon is eliminated (ΔAUG) or for the D12N uORF (*320*). By toeprint assay, ribosomes are observed at the wild-type uORF termination codon but not the ΔAUG or D12N uORF termination codon in response to a high concentration of Arg (*320*). Under the same conditions, the signal that corresponds to ribosomes associated with the downstream LUC initiation codon decreases for transcripts containing the wild-type AAP (*315*). These results indicate that wild-type

uORF has the ability to stall ribosomes at its termination codon in response to Arg and the stalled ribosome prevents scanning ribosomes from reaching the downstream initiation codon. Direct in-frame fusion of the AAP coding sequence to the LUC coding sequence retains AAP's capacity of stalling ribosome, indicating that the AAP can stall either terminating or elongating ribosomes to regulate translation (*321*). The amino acid sequence of the AAP determines its regulatory activity regardless the mRNA sequence and the charging status of arginyl-tRNAs (*322, 323*). The core region within the AAP amino acid sequence that is important for Arg-specific regulation is identified by *in vitro* assays and a comparative bioinformatic analysis (*317*). The AAP stalls plant, fungal and animal ribosomes, suggesting that AAP's function requires highly conserved feature of translation (*317, 321*).

Photo-crosslinking experiments show that probes place at Val-7 of the AAP crosslink more to ribosomal protein L4 and less to L17 in response to a high concentration of Arg, suggesting that Arg changes the conformation of AAP relative to the ribosomal tunnel (*324*). Cryo-EM structural studies of the AAP in the absence of Arg indicate that the C-terminal end of the AAP could exist as an alpha-helix structure in the ribosome exit tunnel and the nascent AAP interacts with residues of ribosomal proteins L4 and L17 at the tunnel constriction (*325*). The CCA end of peptidyl-tRNA may make direct contact with the ribosome peptidyl-transferase center (PTC) to silence its function (*325*). In this thesis, CHAPTER 3 will discuss the detailed study of Arg and AAP's effects on the PTC.

In *S. cerevisiae*, the AAP is 25 amino acids in length and is encoded by a uORF in the *CPA1* mRNA (*326*). It controls the Arg-specific regulation of *CPA1 in vivo* and *in vitro*. The D13N mutation abolishes AAP's regulatory function, similarly to the D12N mutation in the *N. crassa* AAP (*269, 327, 328*). AAP mediated regulation of *CPA1* is dependent on the peptide sequence but not the nucleotide sequence, with the active domain spanning from codon 6 to codon 23 (*329, 330*).

Yeast grown in Arg-supplemented medium shows a shorter half-life of the *CPA1* mRNA, suggesting that the level of *CPA1* transcripts is also regulated (*331*). The uORF-mediated nonsense mediated mRNA decay (NMD) pathway confers the regulation of *CPA1* at the mRNA level. NMD, including regulatory factors UPF1, NMD2 (UPF2) and UPF3 in yeast, targets mRNAs containing premature translation termination codons (*332*). In the wild-type yeast, the half-life of *CPA1* mRNA is 3 min. However, in an isogenic *upf1* Δ strain, the half-life increases to approximately 18 min, indicating that *CPA1* is a direct NMD target (*333*). UPF1 is shown to bind *CPA1* mRNA, suggesting that *CPA1* mRNA is a direct substrate of NMD (*334*).

The ability of *CPA1* to trigger NMD pathway is related to AAP's efficient translation (*269*). The wild-type AAP can trigger NMD pathway to degrade the *CPA1* transcript. The D13N AAP, which abolished ribosome stalling, cannot induce NMD in response to Arg. However, when the initiation context is improved, D13N AAP can efficiently trigger NMD pathway, the mechanism of which cannot be attributed to ribosome stalling. These results suggest that NMD is triggered by increased ribosomal occupancy of the uORF termination codon (*269*). Thus, translation of the uORF-

encoded AAP is required for reducing *CPA1* expression by affecting both translation efficiency and mRNA stability (*269*).

1.4. Translational control mediated by nascent peptide

The Arg- and AAP-mediated translational control of *arg-2* and *CPA1* depends on AAP's peptide sequence but not its mRNA sequences. Eliminating the initiation codon of the AAP abolishes its regulatory effect (*320*). However, when the C-terminus is extended or AAP is directly fused in-frame with downstream reporter genes, the ribosome stalling activity after the AAP's coding sequence retains (*322, 323*). The AAP can also exert its regulatory function as an internal domain within a large polypeptide and stall ribosomes during elongation (*317, 321*). These results suggest that the termination codon of the AAP is not crucial for its regulatory effect and the AAP functions as a nascent peptide in the ribosome exit tunnel. Other nascent peptides that control gene expression in eukaryotes and prokaryotes can help a better understanding of AAP-mediated control of the ribosome movement. Several representative nascent peptides are discussed below.

1.4.1. *Arabidopsis thaliana CGS1*

Arabidopsis CGS1 encodes cystathionine γ-synthase, which catalyzes the first committed step of methionine biosynthesis (*335*). The expression of *CGS1* gene is negatively regulated by methionine or its direct metabolite *S*-adenosyl-L-methionine (AdoMet). When the concentration of AdoMet is high, due to the mRNA degradation from the 5' end, the level of full-length *CGS1* mRNA decreases, and the amount of 5'-

truncated mRNA increases. The change of mRNA stability is mediated by a short peptide stretch ^{77}RRNCSNIGVAQIVA^{90} designated MTO1, which is encoded within the first exon of *CGS1*. The translation of MTO1 is crucial for the regulation of *CGS1* gene at the post-transcriptional level. Mutations in the MTO1 amino acid sequence or chemicals inhibiting translation abolish the regulation (*266, 336, 337*). *In vitro* experiments indicate that, in response to AdoMet, the nascent peptide forms a compact conformation within the ribosome exit tunnel (*338*). MTO1 induces a ribosome elongation-arrest immediately downstream of its coding region. The stalling ribosome contains Trp93 and Ser94 in the P and A sites respectively (*267*). The translation arrest and the mRNA degradation are closely related. When the ribosome is stalled, the 5'-end of *CGS1* mRNA is accessible for the RNA decay machinery (*267*).

1.4.2. *Escherichia coli SecM*

The *E. coli secMA* operon encodes two proteins, SecM (secretion monitor) and SecA (secretion driving ATPase). The latter drives the movement of preproteins and internal membrane complexes (*339, 340*). Translation of *secM* is responsible for the translation regulation of the downstream *secA* gene (*341, 342*). Under normal conditions, ribosomes stall transiently at the C-terminus of SecM during elongation, inducing mRNA structural changes to expose the Shine-Dalgarno sequence preceding the *secA* ORF. Other ribosomes engage in translation initiation at the *secA* gene to produce a basal level of SecA (*339, 343*). Conditions that inhibit SecM targeting prolong translation arrest and induce an overexpression of SecA. Translocation of SecM into the periplasm relieves the secretion stress and abolishes ribosome stalling (*344*). During the

time window of translation arrest, SecM polypeptide is co-translationally targeted by signal recognition particle to the cellular membrane. Translation arrest thus ensures that SecA is synthesized near the cellular membrane. The signal peptide in SecM functions as a molecular timer to co-ordinate member targeting with translation arrest (*345*).

SecM is 170 residues in length. The amino acid sequence F_{150}XXXXXWIXXXXGIRAGP_{166} at the C-terminal region is crucial for SecM's ribosome stalling activity (*263, 340, 346, 347*). SecM-mediated stalling requires Pro-166, which inhibits the formation of peptide bond between the peptidyl-$tRNA^{Gly}$ in the P site and prolyl-$tRNA^{Pro}$ in the A site (*346*). When the A site is occupied by prolyl-$tRNA^{Pro}$, the arrested peptidyl-tRNA is resistant to the peptide release induced by puromycin (*346*). Mutations that decrease or abolish the regulatory function of SecM have been identified in SecM itself, ribosomal protein L4, L22 and 23S rRNA (*347, 348*). These mutations are located either close to the narrowest constriction of the exit tunnel or near the PTC.

Cyro-EM analysis reveals structural rearrangements in both the large and small subunits of ribosomes during SecM-mediated translation arrest (*349*). Crosslinking and fluorescence resonance energy transfer (FRET) studies suggest that, when the synthesis of the arrest motif is complete, SecM adopts a compact conformation induced by the ribosome (*350, 351*). Specific interactions between SecM residues and ribosomal components may be responsible for the structural changes in both and the silenced PTC function.

1.4.3. *E. coli TnaC*

The *tnaCAB* operon contains *tnaC* (a 24-residue leader peptide), *tnaA* (a tryptophanase), and *tnaB* (a tryptophan-specific permease) (*352*). The transcription of *tnaA* and *tnaB* depends on the level of Trp in the cell. When the concentration of Trp is low, ribsosomes are dissociated from the mRNA after translating *tnaC*, allowing Rho termination factor to access the *rut* site adjacent to the *tnaC* stop codon. As a result, the transcription terminates before the RNA polymerase reaches the downstream *tnaA* and *tnaB* genes. When the concentration of Trp is high, the translation termination of *tnaC* is inhibited (*353, 354*). The nascent TnaC-tRNA remains uncleaved within the ribosome and the stalled ribosome prevents Rho factor from accessing its binding sites, thus allowing the transcription and expression of *tnaA* and *tnaB* (*355*).

Residues in the TnaC and components of the ribosome exit tunnel and the PTC are identified to affect the binding of Trp and the regulatory function of TnaC (*356-361*). Cryo-EM studies indicate that, in contrast to SecM, TnaC is mainly in an extended conformation inside the tunnel, although some compactions in the vestibule region of the tunnel could not be excluded (*362*). The presence of TnaC-tRNA^{Pro} within the ribosome is critical for the stalling action of Trp (*355*). Interactions between Trp, TnaC and specific ribosomal components may induce a cascade of structural changes that relay to the PTC, creating a Trp binding and function site in or near the PTC (*352, 353, 360, 362*). Three transmitting mechanisms are proposed, including through the nascent peptide itself, through interactions at the side of exit tunnel where ribosomal protein L4 resides,

and through interactions on the side of ribosomal protein L22 (*362*). So far, the last route contains most mutations that affect Trp induction (*363*).

1.4.4. *erm*

Macrolide antibiotics can induce the expression of *erm* genes in erythromycin-resistant bacteria (*364*). The translation of *erm*-encoded methyltransferase is regulated by drug- and nascent peptide-mediated ribosome stalling. Macrolides bind near the ribosome PTC (*365, 366*). *ermC* is the best-understood example of *erm* genes. It is controlled by *ermCL*, an ORF upstream of *ermC* (*367, 368*). In the absence of macrolides, *ermCL* is constitutively translated. The translation of *ermC* is repressed because an mRNA secondary structure produced between *ermCL* and *ermC* sequesters the ribosome binding site required for the ErmC synthesis. When low levels of antibiotic are present, translational arrest occurs at the ninth codon of *ermCL*, resulting mRNA structural rearrangements to expose the translation initiation region of *ermC* (*348, 369*). Four C-terminal amino acids (IFVI) of the ermCL nascent peptide are essential for ermCL's ribosome stalling function (*369*). The drug and the nascent peptide are recognized as a complex structure by specific components in the ribosome exit tunnel. (*370*). As a result, the properties of the ribosomal A-site change and the PTC activity is inhibited for specific amino acids to stall ribosomes (*371*).

1.5. Translational control mediated by non-AUG initiation codon

In eukaryotic cells, the 43S pre-initiation complex binds to the 5'-cap of mRNA and scans downstream for the initiation codon. Therefore the AUG that is closest to the

5'-cap and in a preferred initiation context is usually selected as the translation initiation site. Components in the translation initiation machinery ensure the stringency of start codon selection. Strong bias exists for nucleotides in the immediate vicinity of the start codon. Multiple initiation factors contribute to the stringency of start codon selection. However, in certain cases, translation can initiate at non-AUG codons, especially those that differ from AUG by a single nucleotide (near-cognate codons). In most cases, non-AUG initiated proteins begin with a methionine (*372*). Initiation from these non-AUG codons may increase coding capacity and function in regulatory mechanisms. The stringency of start codon selection changes in response to different physiological conditions, such as amino acid starvation.

1.5.1. The initiation context and near-cognate codons

The initiation context, which comprises six nucleotides upstream and one nucleotide immediately downstream of a potential initiation codon (-6 to +4, with the A of AUG as +1), has a strong influence on the initiation efficiency at that codon (*373, 374*). In mammalian cells, the Kozak consensus [GCC(A/G)CC**AUG**G] is optimal for initiation (*253, 374*). A purine at position -3 and a 'G' at position +4 are most important for efficient initiation (*374*). Mutations at position -3 can reduce initiation by more than an order of magnitude in mammalian cells (*374*). Substitutions at positions +4, -1 and -2, have lesser effects on translation efficiency, while their effects are intensified when the -3 nucleotide is suboptimal (*374*). Therefore, when a 5'-proximal AUG is surrounded by an unfavorable context, the frequency of initiation at that site is reduced.

Some near-cognate codons (NCCs) support translation with efficiencies that are less than that observed with AUG. The presence of a good Kozak context is crucial for initiation from such non-AUG codons (*68, 375, 376*). The measured efficiencies of initiation from NCCs depend on the gene and the study. Most efficiencies range from ~1 to 20% compared to AUG in the same context, except for "near-cognate" AAG and AGG, which do not permit initiation. That is, changing the 'U' at the second position to a purine eliminates translation initiation (*372, 377, 378*). Studies in human cells reveal a hierarchy in initiation efficiencies at NCCs (*379*). The most potent are CUG and GUG (19.5% and 9.2% of AUG initiated translation), while some NCCs are used at lower efficiencies (ACG 6.6%, AUA 3.3%, AUU 3.2%, UUG 1.9% and AUC 1.7%) (*379*). In plant protoplasts, CUG is the most active (30%), then GUG and ACG (15%). AUA, AUU, UUG and AUC are less active (2-5%) (*380*). In yeast, a recent *in vivo* assay demonstrates the following hierarchy: CUG (7%), GUG (5%), UUG (5%), ACG (4%), AUA (4%), AUU (2%) and AUC (1%) (*377*). However, other studies yielded different results, presumably because of the different initiation contexts used (*377, 378, 381-383*).

1.5.2. The existence of non-AUG initiation sites

In mouse embryonic stem cells, unannotated AUG or near-cognate initiation sites have been identified. The majority of these start sites drive the translation of uORFs in the 5' UTRs, potentially with regulatory function (*384*). For example, *Myc* and *Nanog*, genes involved in pluripotency, have multiple uORFs initiated at both AUG and NCCs to produce alternative translation products (*384*). In yeast, a genome-wide analysis reveals that NCC initiated uORFs can be translated at levels comparable to AUG-

initiated uORFs (*385*). Interestingly, translation of these non-AUG uORFs generally increases in response to amino acid starvation (*385*).

In some cases, an upstream in-frame NCC supports the translation of a longer isoform of protein in addition to that provided by the standard AUG codon (*69*). Protein isoforms with different N-terminal sequences can confer different functions at different subcellular locations (*386*). In yeast, *ALA1* and *GRS1* use upstream in-frame UUG and ACG, respectively, to produce a longer isoform targeted to mitochondria (*375, 381*).

Proteins identified to initiate exclusively at non-AUG codons include the following examples. In yeast, the glycyl-tRNA synthetase gene initiates from a UUG codon (*381*). *Podospora anserine* IDI-4 initiates at CUG (*387*). In plants, the *AGAMOUS* gene in *Arabidopsis* uses an ACG codon to start translation and some *RpoT* genes in *Nicotiana spp.* use CUG codons (*388-390*). Mammalian *NAT1* mRNA, encoding eIF4G2, uses GUG to start translation (*391*). Interestingly, the capsid protein in an insect RNA virus is initiated with glutamine encoded by a CAA codon instead of methionine (*392*). Another polypeptide initiated with a non-Met amino acid is presented by major histocompatibility complex (MHC) class I molecules. This peptide uses leucine-tRNA to recognize a CUG codon to start translation (*393*).

1.5.3. The stringency of start codon selection

The stringency of start codon selection can be regulated by several parameters. A high concentration of Mg^{2+} increases initiation at non-AUG codons *in vitro* (*68*). A high-throughput screen conducted in *S. cerevisiae* identified two structurally related molecules isoquinoline-1-carboxylic acid and 7-amino-5-iodo-8-quinolinol that decrease

the stringency of start codon selection (*394*). These compounds can also stimulate initiation at natural uORFs with near cognate start codons. The effects of these compounds are suppressed by the overexpression of eIF1 (*394*).

Genetic studies in yeast reveal that eIF1 (*395*), subunits of eIF2 (*71, 396*), eIF5 (*31*), eIF1A (*72, 397*), eIF3 (*398*) and eIF4G (*399*) regulate start codon selection *in vivo* conferring either the Sui^- (Suppressor of initiation codon) or Ssu^- (Supressor of *sui*) phenotypes. Sui^- mutations in these initiation factors can restore the expression of a *his4* allele lacking the AUG initiation codon by enabling initiation from the third codon (UUG) of the *his4* ORF. The Sui^- phenotype indicates increased translation efficiency from NCCs. Ssu^- mutations have the opposite effect, suppressing UUG-initiated translation conferred by Sui^- mutations (*71*).

Among initiation factors that change the stringency of start codon selection, eIF1 and eIF5 are of great interest because of their auto- and cross- regulatory roles (*400*) (*379*). eIF1 binds near the ribosomal P site and promotes an open conformation of the PIC. This conformation favors scanning and prevents base pairing of Met-$tRNA_i^{Met}$ with non-AUG codons (*44, 74, 170, 401, 402*). The release of Pi and the GAP (GTPase activating protein) function of eIF5 are blocked by eIF1 at non-AUG codons (*78, 401, 403*). AUG recognition triggers the dissociation of eIF1 from the 40S subunit, which is a critical step in start codon selection (*395, 401*). The release of eIF1 induces the formation of a closed conformation of the PIC, which is scanning-incompatible (*78, 80*). Mutations that enhance eIF1 release at non-AUG codons decrease the stringency of start codon recognition (*404*). Overexpression of wild-type eIF1 reduces the usage of AUG

codon in poor context and the usage of non-AUG start codons (*379*). Overexpression of eIF1 can also suppress the increased initiation at non-AUG codons conferred by Sui$^-$ mutations in other initiation factors (*397, 398*). Interestingly, eIF1 from different organisms initiates from its AUG codon in a poor context, which is shown to be crucial for an autoregulatory negative feedback loop (*379*). When the level of eIF1 is high, the stringency of start codon selection increases, resulting in eIF1 inhibiting its own translation (*379*). The autoregulatory function of eIF1 demonstrates that the stringency of start codon selection can be changed to regulate gene expression.

eIF5 hydrolyzes eIF2-bound GTP in response to initiation codon recognition (*405-407*). eIF5 can competitively bind to a site in the small ribosomal subunit, resulting the release of eIF1 (*80*). In contrast to eIF1, over-expression of eIF5 leads to an increased level of initiation at AUG codons in poor contexts and at non-AUG start codons (*400*). Similarly to eIF1, eIF5 auto regulates its own synthesis (*400*). Many eIF5 mRNAs contain one or more uORFs that initiate with AUG codons in poor contexts (*400*). In human cells, high levels of eIF5 decrease the stringency of start codon selection, induce translation of these inhibitory uORFs and reduce initiation from the downstream eIF5 start codon. (*400*). The auto- and cross- regulation achieved by eIF1 and eIF5 establish a regulatory feedback loop, which provides an additional mean to stabilize the stringency of start codon selection (*400*).

A genetic analysis in yeast also shows that deletion of the C-terminus of eIF1A causes Sui$^-$, indicating eIF1A's role in maintaining the stringency of start codon selection (*72*). A Ssu$^-$ mutation is also identified in eIF1A to decrease the rate of eIF1

release from the PIC (*404*). Mutations in eIF2 and its GTPase-activating protein eIF5 can increase initiation at non-AUG codons by increasing the dissociation of eIF2·GTP from Met-RNA$_i^{Met}$ or by enhancing the rate of GTP hydrolysis (*31*)

1.6. Dissertation research

In my dissertation, the translational control is analyzed in *N. crassa*. Puromycin release assay was used to show that AAP interferes with ribosome PTC in response to Arg (CHAPTER 2). The stringency of start codon selection was examined systematically both *in vivo* and *in vitro* in *N. crassa* (CHAPTER 3). *In vitro* studies showed that *cpc-1* uORF1 and uORF2 function similarly as *GCN4* uORF1 and uORF4 to regulate translation through reinitiation (CHAPTER 4). Four non-AUG start codons in the 5'UTR of *cpc-1* were revealed. The utilization of these codons might provide additional routes for translation regulation (CHAPTER 4).

CHAPTER II

THE ARGININE ATTENUATOR PEPTIDE INTERFERES WITH THE RIBOSOME PEPTIDYL TRANSFERASE CENTER*

2.1. Introduction

Translational control mediated by nascent peptides is demonstrated in mammals, fungi, plants, bacteria and viruses (*338, 342, 364, 408-411*). Among regulatory nascent peptides that control gene expression are some that are encoded by upstream open reading frames (uORFs) in mRNA 5'-leaders. The significance of eukaryotic uORFs is increasingly appreciated (*261, 412-415*). Translation of uORFs can reduce translation of downstream ORFs and also decrease mRNA stability. Regulation by eukaryotic uORFs and prokaryotic leader peptides (the designation for prokaryotic uORFs) has consequences for a variety of physiological processes (*2, 256*).

Regulatory nascent peptides can control translation from within the ribosome tunnel by causing ribosomes to stall. In *Escherichia coli*, tryptophanase expression is controlled in response to tryptophan by the TnaC leader peptide which acts as a ribosome-arrest peptide (RAP) during translation termination (*359*). *E. coli* SecM and *Bacillus subtilis* MifM nascent polypeptides contain domains that interact with the ribosome to cause ribosome-arrest during elongation (*340, 348, 416, 417*). Bacterial *erm*

* This material has been published in this or similar form in *Molecular and Cellular Biology* and is used here with permission of American Society for Microbiology.
Wei, J., Wu, C., and Sachs, M.S. (2012). The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. Mol Cell Biol *32*, 2396-2406.

and *cml* operons, which confer resistance to macrolides and to chloramphenicol, respectively, are regulated by nascent leader peptides that function as RAPs when the antibiotics are present (*371, 418-420*). A nascent peptide designated MTO1 within the *Arabidopsis thaliana CGS1* coding region causes ribosomes to stall during elongation in response to *S*-adenosyl-L-methionine; the stall induces mRNA degradation (*266, 267*). In mammals, the uORF-encoded RAP "MAGDIS" regulates the synthesis of *S*-adenosylmethionine decarboxylase in response to polyamines by stalling ribosomes (*276, 421*). Expression of the human cytomegalovirus gp48 gene is reduced by translation of its uORF2 RAP (*422*), which causes ribosomes to stall at the uORF2 termination codon (*287, 288*). Ribosomes synthesizing the uORF-encoded fungal arginine attenuator peptide (AAP) stall at the uORF-termination codon in response to a high concentration of arginine (Arg) (*315, 316, 321, 323, 423-425*).

The regulatory AAP uORF is present in the 5'-leaders of fungal mRNAs specifying the glutamine amidotransferase subunit of Arg-specific carbamoyl phosphate synthetase (*316, 328*). AAP-mediated stalling in response to Arg results in the reduced synthesis of the first enzyme specific for Arg biosynthesis (*426*). AAP is the best-understood example of a eukaryotic RAP. *In vivo* studies of the *Neurospora crassa arg-2* mRNA, which encodes the AAP uORF, show that the rate of ARG-2 synthesis is reduced in Arg-supplemented medium (*309*). Polysome-profile analyses show that adding Arg to the growth medium shifts the *N. crassa arg-2* and *Saccharomyces cerevisiae CPA1* transcripts that specify the wild-type uORF-encoded AAP toward the monosome fraction (*269, 318*). Furthermore, in *S. cerevisiae*, ribosome stalling at the

uORF termination codon triggers degradation of the mRNA through the nonsense-mediated mRNA decay pathway (*269*). Thus, AAP-mediated ribosome stalling can regulate gene expression in *cis* by reducing translation from a downstream start codon in the mRNA and by reducing the stability of this mRNA.

In vitro experiments have contributed to understanding the mechanistic basis of AAP function. Toeprinting (primer-extension inhibition), which maps the positions of ribosomes on mRNA, shows that when the AAP functions as a uORF, ribosomes arrested at the AAP termination codon block scanning ribosomes from reaching the downstream initiation codon for the genic ORF (*293*). AAP can also function as an internal polypeptide domain to cause stalling of ribosomes during elongation (*321, 424*). AAP causes Arg-regulated stalling of fungal, plant, and animal ribosomes, establishing that the AAP and Arg exploit highly conserved ribosome functions to cause stalling (*321*). Ribosomal peptidyl transferase activity is a likely target, but this has not yet been directly demonstrated.

Structurally, site-specific photocrosslinking experiments indicate that Arg alters the conformation of the wild-type AAP relative to the ribosome tunnel (*324*). In high Arg, a crosslinker placed at AAP Val-7 reacted relatively less to ribosomal protein L17 and more to ribosomal protein L4. Consistent with these data, visualization of ribosome nascent chain complexes containing AAP in the absence of Arg by cryo-electron microscopy also indicates that the AAP interacts with ribosomal proteins L4 and L17 at the ribosome tunnel constriction (*325*).

A hypothesis to explain Arg-regulated ribosome stalling by AAP is that high Arg stabilizes a conformation of the nascent peptide relative to the ribosome that interferes with PTC activity, resulting in ribosome stalling. To test this, we used a puromycin-release assay to directly examine how the AAP and Arg affect PTC activity. Puromycin is an aminonucleoside antibiotic in which part of the molecule resembles the 3' end of tyrosyl-tRNA (*427*). During translation, puromycin enters the PTC A site and the peptidyl transferase reaction transfers the nascent peptide from tRNA to puromycin. The rate of nascent peptide chain transfer to puromycin thus can be used as an indicator of PTC activity (*113, 346, 350, 351, 353, 355, 428*).

Here we show that AAP functions with Arg to interfere with the PTC activity of *N. crassa* and wheat ribosomes. AAP containing the D12N mutation, which eliminates Arg-induced ribosome stalling, also eliminated Arg's effect on PTC activity. Importantly, the AAP interfered with the PTC before full-length AAP was synthesized, but full-length synthesis appeared important for most efficient stalling. These data support a model for AAP function in which the inhibition of PTC activity is the basis for the AAP's capacity to stall the ribosome. Unlike many other RAPs, specific features of the AAP near the ribosomal catalytic center appear relatively unimportant for stalling.

2.2. Metarials and methods

2.2.1. Plasmids

Plasmids containing the coding sequences for wild-type Met_9AAP (pJC102, Figure 2.1A) and D12N Met_9AAP (pJCS102) were constructed as described (*324*). The

wild-type AAP sequence was substituted with other sequences by replacing the small *Age*I-*Hin*dIII restriction fragments with annealed synthetic oligonucleotides by ligation (Table 2.1). To construct plasmid pJW201 (Figure 2.1B), which contains the uORF-encoded AAP and a downstream luciferase reporter, synthetic complementary oligonucleotides JW001+ and JW001– (Table 2.2) were annealed and ligated to gel-purified vector pR301 (*322*) that had been digested with *Mlu*I and *Nco*I. Mutations were introduced into the AAP-coding sequence by replacing the small *Spe*I-*Hin*dIII restriction fragment with annealed synthetic oligonucleotides by ligation (Table 2.2 and Figure 2.1B). To construct plasmid pJF401 (Figure 2.1C), which contained the AAP fused with in-frame luciferase reporter (AAP-LUC), synthetic complementary oligonucleotides (Table 2.3) were annealed and ligated to gel-purified vector pJW201 that had been digested with *Spe*I and *Bst*EII. Mutations were introduced into the AAP-coding sequence by replacing the *Spe*I-*Bst*EIII restriction fragment with annealed synthetic oligonucleotides by ligation (Tables 2.3 and Figure 2.1C). Plasmids specifying T7LUC and sea pansy luciferase were described (*315, 423*).

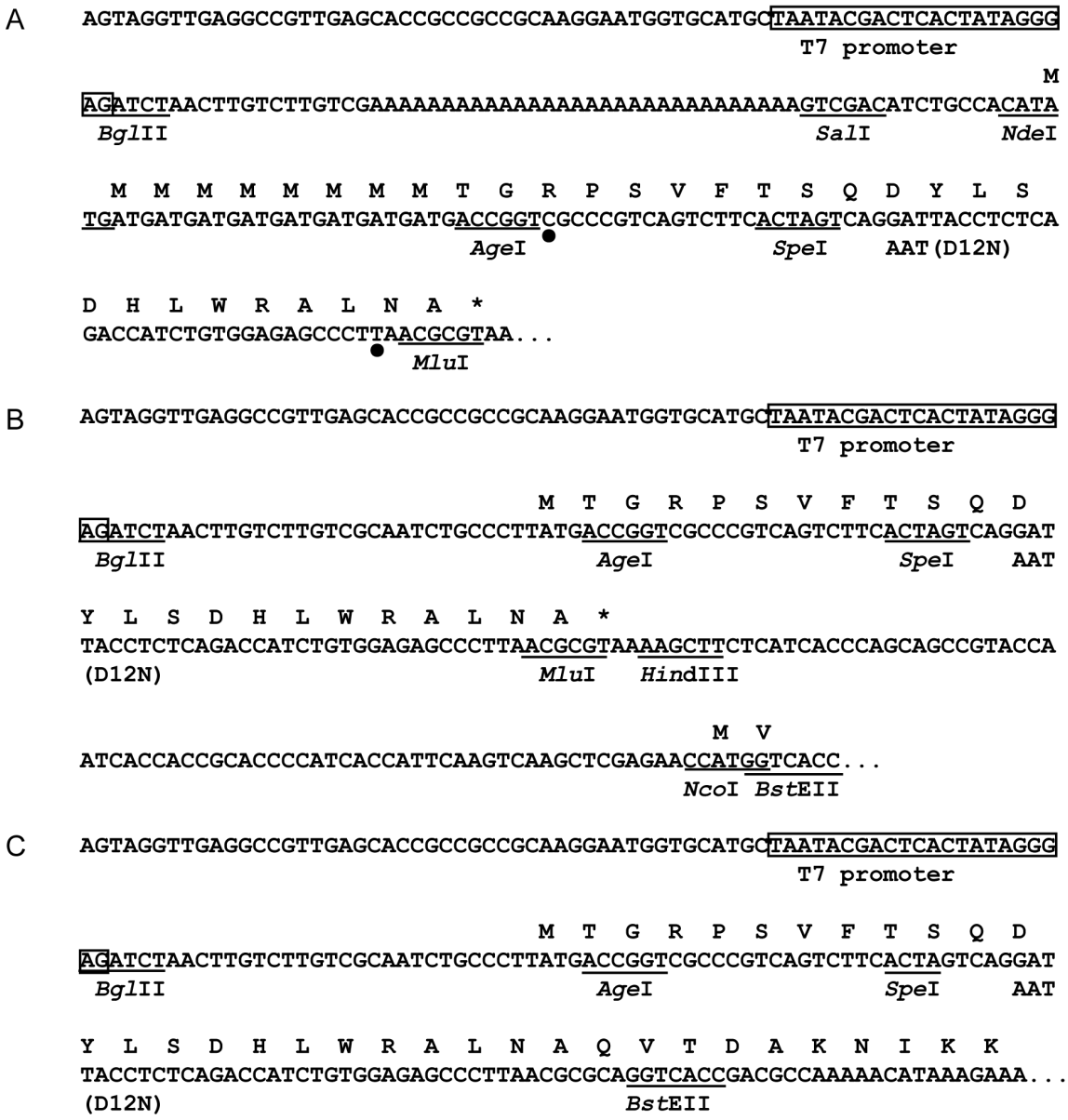


Figure 2.1. The sequences of AAP constructs used in this study. (A) The sequences of pJC102-based Met_9AAP constructs used for puromycin-release assays. Mutations and selected restriction enzyme sites are shown below the sequence. The cytosine marked with the first black dot was deleted and a thymine was inserted after the thymine marked with the second black dot to make FS AAP. (B) The sequences of pJW201-based AAP constructs used for toeprinting and luciferase reporter assays. (C) The sequences of pJF401-based AAP-LUC fusion constructs used for toeprinting and luciferase reporter assays.

Table 2.1. Oligonucleotides used to substitute wild-type AAP sequence with other peptide sequences in pJC102. Plus and minus strands were annealed and cloned into *Age*I/*Hin*dIII sites as shown in Figure 2.1A.

Oligo Name	Sequence (5′ - 3′)	Cloning site	Vector
FS+	CCGGTGCCCGTCAGTCTTCACTAGTCA GGATTACCTCTCAGACCATCTGTGGAG AGCCCTTTAACGCGTAAA	*Age*I/*Hin*dIII	pJC102
FS-	AGCTTTTACGCGTTAAAGGGCTCTCCA CAGATGGTCTGAGAGGTAATCCTGACT AGTGAAGACTGACGGGCA		
gp48 uORF2+	CCGGTCTGGTTCTCTCGGCGAAAAAAC TGTCGTCTTTGCTGACTTGCAAATACA TCCCGCCTTAAA	*Age*I/*Hin*dIII	pJC102
gp48 uORF2-	AGCTTTTAAGGCGGGATGTATTTGCAA GTCAGCAAAGACGACAGTTTTTTCGCC GAGAGAACCAGA		
ScAAP+	CCGGTTTATCGAACTCTCAATACACCT GCCAAGACTACATATCTGACCACATCT GGAAAACTAGCTCCCACTAAA	*Age*I/*Hin*dIII	pJC102
ScAAP-	AGCTTTTAGTGGGAGCTAGTTTTCCAG ATGTGGTCAGATATGTAGTCTTGGCAG GTGTATTGAGAGTTCGATAAA		

Table 2.2. Oligonucleotides used to construct pJW201 and introduce mutations into the AAP coding sequence in pJW201. Plus and minus strands were annealed and cloned into *Mlu*I/*Nco*I or *Spe*I/*Hin*dIII sites as shown in Figure 2.1B.

Oligo Name	Sequence (5′ - 3′)	Cloning site	Vector
JW001+	CGCGTAAAAGCTTCTCATCACCCAGCA GCCGTACCAATCACCACCGCACCCCAT CACCATTCAAGTCAAGCTCGAGAAC	*Mlu*I/*Nco*I	pR301
JW001–	CATGGTTCTCGAGCTTGACTTGAATGG TGATGGGGTGCGGTGGTGATTGGTACG GCTGCTGGGTGATGAGAAGCTTTTA		
D12N+	CTAGTCAG**AAT**TACCTCTCAGACCATC TGTGGAGAGCCCTTAACGCGTAAA	*Spe*I/*Hin*dIII	pJW201
D12N-	AGCTTTTACGCGTTAAGGGCTCTCCAC AGATGGTCTGAGAGGTA**ATT**CTGA		
AAP_{19}+	CTAGTCAGGATTACCTCTCAGACCATC TGTGG**TAG**GCCCTTAACGCGTAAA	*Spe*I/*Hin*dIII	pJW201
AAP_{19}-	AGCTTTTACGCGTTAAGGGC**CTA**CCAC AGATGGTCTGAGAGGTAATCCTGA		
AAP_{20}+	CTAGTCAGGATTACCTCTCAGACCATC TGTGGAGA**TAG**CTTAACGCGTAAA	*Spe*I/*Hin*dIII	pJW201
AAP_{20}-	AGCTTTTACGCGTTAAG**CTA**TCTCCAC AGATGGTCTGAGAGGTAATCCTGA		
AAP_{21}+	CTAGTCAGGATTACCTCTCAGACCATC TGTGGAGAGCC**TAG**AACGCGTAAA	*Spe*I/*Hin*dIII	pJW201
AAP_{21}-	AGCTTTTACGCGTT**CTA**GGCTCTCCAC AGATGGTCTGAGAGGTAATCCTGA		
AAP_{22}+	CTAGTCAGGATTACCTCTCAGACCATC TGTGGAGAGCCCTT**TAG**GCGTAAA	*Spe*I/*Hin*dIII	pJW201
AAP_{22}-	AGCTTTTACGC**CTA**AAGGGCTCTCCAC AGATGGTCTGAGAGGTAATCCTGA		
AAP_{23}+	CTAGTCAGGATTACCTCTCAGACCATC TGTGGAGAGCCCTTAAC**TAG**TAAA	*Spe*I/*Hin*dIII	pJW201
AAP_{23}-	AGCTTTTA**CTA**GTTAAGGGCTCTCCAC AGATGGTCTGAGAGGTAATCCTGA		

Table 2.3. PCR primers used to amplify linear DNA fragments as templates to generate mRNAs encoding AAPs and other peptides.

Oligo Name	Sequence (5′ - 3′)
T7 upstream+	AGTAGGTTGAGGCCGTTGA
AAP_{20} trunc-	TCTCCACAGATGGTCTGAG
AAP_{21} trunc-	GGCTCTCCACAGATGGTCT
AAP_{22} trunc-	AAGGGCTCTCCACAGATGG
AAP_{23} trunc-	GTTAAGGGCTCTCCACAGA
AAP_{24} trunc-	CGCGTTAAGGGCTCTCCAC
AAP_{25} trunc-	GACCGCGTTAAGGGCTCTC
A21R AAP_{21} trunc-	**ACG**TCTCCACAGATGGTCT
L22R AAP_{22} trunc-	**ACG**GGCTCTCCACAGATGG
N23R AAP_{23} trunc-	**ACG**AAGGGCTCTCCACAGA
A24R AAP_{24} trunc-	**ACG**GTTAAGGGCTCTCCAC
A24R AAP_{24} (CGG)trunc-	**CCG**GTTAAGGGCTCTCCAC
*25R AAP_{25} trunc-	**ACG**CGCGTTTAAGGGCTCT
A21R AAP_{24} trunc-	CGCGTTAAG**ACG**TCTCCAC
L22R AAP_{24} trunc-	CGCGTT**ACG**GGCTCTCCAC
N23R AAP_{24} trunc-	CGC**ACG**AAGGGCTCTCCAC
FS_F22R AAP_{24} trunc-	CGCGTT**ACG**GGGCTCTCCAC
FS_N23R AAP_{24} trunc-	CGC**ACG**AAAGGGCTCTCCAC
FS_A24R AAP_{24} trunc-	**ACG**GTTAAAGGGCTCTCCAC
Sc AAP trunc-	GTGGGAGCTAGTTTTCCA
gp48 uORF2 trunc-	AGGCGGGATGTATTTGCAAG
gp48 uORF2 P22A trunc-	**CGC**CGGGATGTATTTGCAAG

2.2.2. RNA synthesis

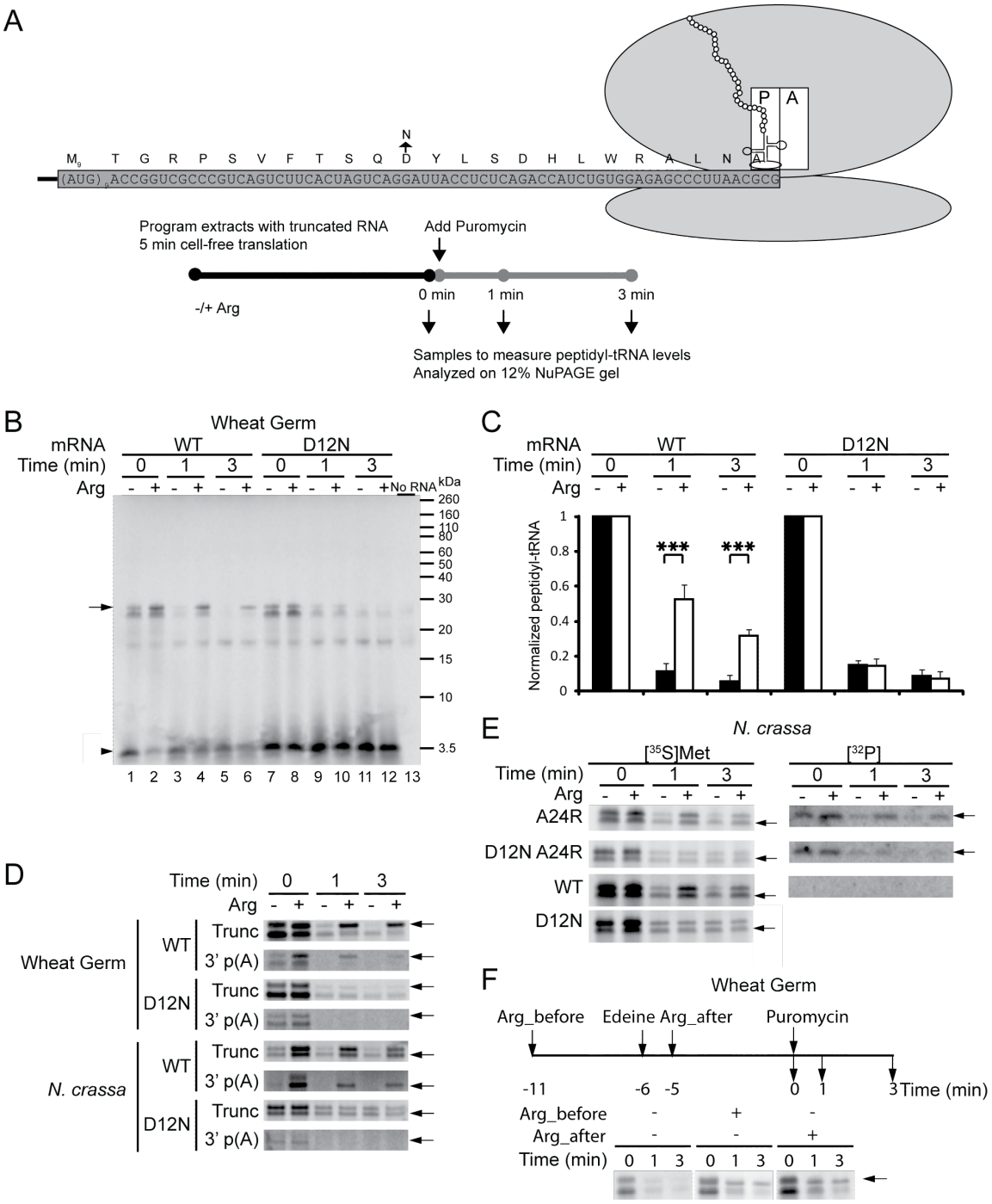
Capped and truncated mRNAs encoding nascent peptides were transcribed *in vitro* by T7 RNA polymerase (*424*) using PCR-generated DNA fragments as templates (primers are listed in Table 2.3). Capped and polyadenylated RNAs were transcribed *in vitro* by T7 RNA polymerase from plasmid DNA templates that were linearized with *Eco*RI (*322*). Aliquots of transcribed RNAs were electrophoresed in agarose gels

adjacent to nucleic acid standards of known quantity and stained with ethidium bromide. Gel images were obtained by a GE Typhoon Trio+ imager and analyzed by ImageQuantTL to determine the relative amounts of RNA.

2.2.3. Cell-free translation analyses

To visualize peptidyl-tRNA by [^{35}S]Met-labeling, *in vitro* translation reactions (20 µl) were programmed with 120 ng of RNA. Either micrococcal nuclease-treated *N. crassa* extracts or micrococcal nuclease-treated wheat germ extracts were used (*324*). [^{35}S]Met (>1000 Ci/mmol, MP Biomedicals) was added to reactions at a final concentration of 0.5 µCi/µl. Arg or Arg analogs were added to reactions as described in the text. For puromycin-release assays, translation reactions were incubated for 5 min at 26°C and then puromycin at concentrations indicated in the text was added to reactions. Samples (5 µl) were taken immediately before the addition of puromycin and at timepoints following addition as indicated (see for example Figure 2.2A). Samples were mixed with 5 µl 2X NuPAGE LDS Sample Buffer (Invitrogen) and put on ice to stop reactions, and then analyzed using 12% NuPAGE gels (Invitrogen) with MES running buffer. The gels were fixed and dried, exposed to phosphor screens overnight, and analyzed with a GE Typhoon Trio+ imager and ImageQuantTL software.

Figure 2.2. Puromycin-release assay to assess the effects of AAP and Arg on PTC activity. (A) Synthetic truncated mRNA specifying the *N. crassa* AAP is translated in cell-free systems. Because the mRNA lacks in-frame stop codons, the full-length nascent peptide should remain associated with ribosomes as peptidyl-tRNA with the last encoded amino acid in the ribosome P site. After translation at 26°C for 5 min, puromycin is added to the reaction. Samples are taken immediately before the addition of puromycin (Time 0), or 1 and 3 min following addition. Translation products are analyzed on 12% NuPAGE gels. Translation reactions contained either low Arg [20 μM (–)] or high Arg [2 mM (+)]. (B) Puromycin-release assay to examine the functions of WT and D12N AAPs in wheat germ extract. Translation reactions contain either low (–) or high (+) Arg. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin. Lane 13 is a control reaction to which no mRNA was added. Arrow: [^{35}S]Met-labeled AAP-tRNA that was resistant to puromycin-release in high Arg. Arrowhead: [^{35}S]Met-labeled free AAP and AAP-puromycin. (C) Quantification of the puromycin-release assay of WT and D12N AAP-tRNAs in wheat germ extract. The signals representing AAP-tRNAs at 1 and 3 min were normalized to the signal at time 0. Mean values and standard deviations from three independent experiments are given (***$p < 0.001$, Student's t-test). (D) Puromycin-release assay to compare the effect of translating truncated mRNA to mRNA with a 3' poly A tail in wheat germ and *N. crassa* extracts. mRNAs encoding WT and D12N AAPs were used. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin (wheat germ extract) or 0.1 mM puromycin (*N. crassa* extract). A representative result of triplicate experiments is shown. Trunc, truncated mRNA; 3' p(A), mRNA with an intact AAP stop codon and a 3' poly A tail. (E) Puromycin-release assay in *N. crassa* extract. Samples were taken at time 0, and 1 and 3 min after adding 0.1 mM puromycin. A24R, D12N A24R, WT and D12N AAPs were analyzed. Left panel: AAP-tRNA was labeled with [^{35}S]Met. Right panel: AAP-tRNA was detected with [^{32}P] labeled $tRNA^{Arg}$ (CGU) probe. A representative result of triplicate experiments is shown. (F) Adding Arg after AAP synthesis to assess the regulatory effect on PTC activity. The puromycin-release assay was performed as in the upper panel. After translation in wheat germ for 5 min, edeine was added to block translation initiation. 2 mM Arg was added either before translation was started or after the addition of edeine. The effects on puromycin induced release of nascent peptide were compared to a reaction that did not contain high Arg.


A
P
A
N
M9 T G R P S V F T S Q D Y L S D H L W R A L N
(AUG)9ACCGGUCGCCCGUCAGUCUUCACUAGUCAGGAUUACCUCUCAGACCAUCUGUGGAGAGCCCUUAACGCG
Program extracts with truncated RNA
5 min cell-free translation
Add Puromycin
-/+ Arg
0 min
1 min
3 min
Samples to measure peptidyl-tRNA levels
Analyzed on 12% NuPAGE gel
B
Wheat Germ
mRNA
WT
D12N
Time (min)
0 1 3 0 1 3
Arg
- + - + - + - + - + - +
No RNA
kDa
260
160
110
80
60
50
40
30
20
15
10
3.5
1 2 3 4 5 6 7 8 9 10 11 12 13
C
mRNA
WT
D12N
Time (min)
0 1 3 0 1 3
Arg
- + - + - + - + - + - +
Normalized peptidyl-tRNA
1
0.8
0.6
0.4
0.2
0
***
***
D
Time (min)
0 1 3
Arg
- + - + - +
Wheat Germ
WT
Trunc
3' p(A)
D12N
Trunc
3' p(A)
N. crassa
WT
Trunc
3' p(A)
D12N
Trunc
3' p(A)
E
N. crassa
[35S]Met
[32P]
Time (min)
0 1 3 0 1 3
Arg
- + - + - + - + - + - +
A24R
D12N A24R
WT
D12N
F
Wheat Germ
Arg_before
Edeine
Arg_after
Puromycin
-11
-6 -5
0 1 3
Time (min)
Arg_before
- + -
Arg_after
- - +
Time (min)
0 1 3 0 1 3 0 1 3

To visualize peptidyl-tRNA by [^{32}P]-probing, *in vitro* translation reactions (20 µl) were prepared as described above, except that reactions contained 20 µM Met and no [^{35}S]Met. Following gel electrophoresis, samples were transferred to Zeta-Probe nylon membranes (Bio-Rad). The procedures for electrophoretic transfer, nucleic acid fixation on the membrane, prehybridization, hybridization with [^{32}P]-labeled oligonucleotide and washing were as described (*429*), except that denatured salmon sperm DNA was not included in the prehybridization and hybridization solutions. After washing, membranes were exposed to phosphor screens overnight. The DNA probe (JW02: 5'-GATCCACCCAGGGGTCG-3') which is the reverse complement to 56-72 of *N. crassa* tRNA^{Arg} (CGU) (*430*) was [^{32}P]-labeled at its 5' end as described (*323*).

For luciferase reporter assays, the reaction conditions for *in vitro* translation using *N. crassa* extracts were as described (*320*). Translation reactions (10 µl) were incubated at 25 °C for 30 min, and translation was halted by adding 50 µl 1.2X Passive Lysis Buffer (Promega). For firefly luciferase activity measurements, equal amounts (12 ng) of mRNA encoding firefly luciferase were used to program extracts; 2.5 ng of synthetic mRNA encoding Renilla (sea pansy) luciferase was added to all reactions to serve as an internal control (*315*). Firefly and sea pansy luciferase enzyme activities were measured using the Dual-Luciferase Reporter Assay System (Promega) with a VICTOR 3 Multi-task Plate Reader (Perkin Elmer).

The primer extension inhibition (toeprint) assays were accomplished using [^{32}P]-labeled primer ZW4 as described (*320, 431*), except that when cycloheximide was added

to reaction after 10 min of translation, reactions were immediately processed to preserve signals from ribosomes stalled at termination codons.

2.3. Results

2.3.1. AAP interferes with PTC activity in response to Arg

We tested whether the AAP directly interfered with PTC activity using puromycin-release as an assay (Figure 2.2A). Cell-free translation extracts were programmed with truncated synthetic mRNA specifying wild-type or mutated AAPs. Ribosomes that translate the AAP coding sequence are expected to accumulate at the 3'-end of the mRNA with the last codon in the P site. The nascent AAP would thus be in peptidyl-tRNA form at the P site. The AAP is visualized by [^{35}S]Met-labeling; eight additional Met-residues are encoded at the AAP's N-terminus to increase labeling (*425*). After translation of the RNA for 5 minutes at 26°C in either low Arg [20 μM (-)] or high Arg [2 mM (+)], puromycin is added to the reactions. Samples are taken immediately before the addition of puromycin (Time 0), and at one-minute and three-minute time-points following puromycin addition (Figure 2.2A). The [^{35}S]Met-labeled reaction products are analyzed using 12% NuPAGE gels. AAP-tRNA and AAP released from tRNA can thus be resolved and the amount of [^{35}S]Met in AAP-tRNA form can be quantified.

The addition of Arg slowed the rate of puromycin-induced release of wild-type (WT) AAP-tRNA in wheat germ extract (Figure 2.2B). In low Arg, WT AAP-tRNA, which migrated with an apparent $M_r \approx 28$ kDa and which was RNase sensitive (Figure 2.3), disappeared rapidly following the addition of 1 mM puromycin (Figure 2.2B, lanes

3 and 5 versus lane 1; quantification is given in Figure 2.2C). The decrease in the AAP-tRNA species in response to puromycin was accompanied by an increase in the AAP (M_r ≈ 3.8 kDa, the AAP is expected to be linked to puromycin whose mass did not have a detectable impact on the AAPs' migration in this gel system). This indicated that, in low Arg, the PTC relatively efficiently transferred the nascent AAP from AAP-tRNA to puromycin. However, in high Arg, WT AAP-tRNA did not decrease as rapidly in response to puromycin, indicating that Arg interfered with PTC activity (Figure 2.2B, lanes 4 and 6 versus lane 2; quantification in Figure 2.2C). The difference in puromycin-release rates in low versus high Arg was significant ($p < 0.001$, Student's t-test). The WT AAP-tRNA band was stronger and the free AAP band was weaker in high Arg even before puromycin treatment (Figure 2.2B, lane 2 versus lane 1), which is also consistent with Arg having a stabilizing effect on the nascent WT AAP-tRNA in the ribosome. In other words, the increased level of WT AAP-tRNA in the presence of high Arg could reflect reduced spontaneous hydrolysis of the peptidyl-tRNA (*432*). Asp-12 of AAP is functionally important and the D12N mutation abolishes AAP-mediated stalling in response to Arg (*317 and reference therein*). For the nonfunctional D12N AAP, AAP-tRNA disappeared rapidly in response to puromycin in both low Arg (Figure 2.2B, lanes 9 and 11 versus lane 7, quantification in Figure 2.2C) and high Arg (Figure 2.2B, lanes 10 and 12 versus lane 8, quantification in Figure 2.2C), indicating that Arg did not affect PTC activity when ribosomes contained D12N AAP-tRNA. These data show that the D12N AAP did not inhibit PTC activity in response to Arg.

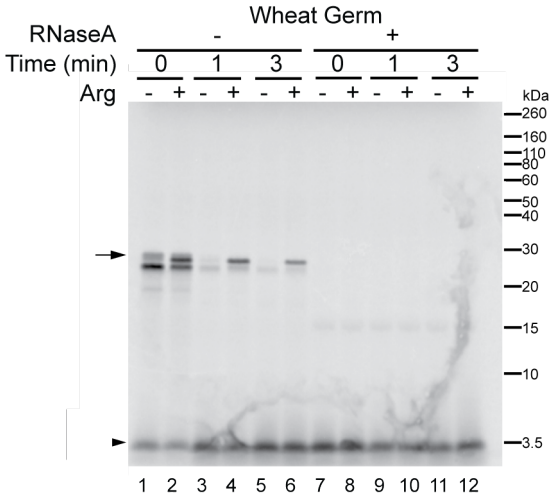


Figure 2.3. The species identified as AAP-tRNA are RNase-sensitive. Wheat germ extract was programmed with truncated mRNA encoding WT AAP. The puromycin-release assay was performed as in Fig. 2.2B. Half of the translation products were treated with 1 µg RNaseA and incubated at 37ºC for 10 min. Untreated and RNaseA-treated products were analyzed on 12% NuPAGE gels. Arrow: [^{35}S]Met-labeled AAP-tRNA that was resistant to puromycin-release in Arg. Arrowhead: [^{35}S]Met-labeled free AAP and AAP-puromycin.

The data in Figure 2.2B show AAP-tRNA as a doublet band. This doublet was RNase-sensitive (Figure 2.3), but only the upper band (arrow) was stabilized for WT AAP in high Arg. To assess whether the doublet occurred as a consequence of using truncated RNA as a template for translation and to test whether PTC activity of *N. crassa* ribosomes is also inhibited by AAP in high Arg, we compared wheat germ and *N. crassa* translation extracts programmed with truncated RNA or poly(A) mRNA (Figure 2.2D). The poly(A) mRNA contained an AAP with a termination codon and a 3'UTR. Overall, the truncated mRNAs encoding the WT AAP and the D12N AAP behaved similarly to the corresponding poly(A) mRNAs in both extracts (Figure 2.2D): the WT

AAP-tRNA species but not the D12N AAP-tRNA species showed resistance to puromycin-release in high Arg. In wheat germ extract, the upper band of the doublet was most resistant to puromycin; in *N. crassa* extract, the lower band was more resistant. The difference between wheat and *N. crassa* extracts could be expected to reflect differences in the tRNAs of these organisms. Furthermore, in extracts programmed with poly(A) RNA, the band that was not stabilized by Arg for WT AAP in response to puromycin was overall reduced. Based on these results and the data described below, this second band might represent ribosomes whose movement is blocked by the ribosome either at the 3' end of the truncated RNA or at the uORF termination codon. Regardless, as shown next, we confirmed that, in *N. crassa* extracts, the puromycin-resistant band (lower band) corresponded to full-length AAP-tRNA.

We examined Arg-specific regulation of AAP-tRNA release both by [^{35}S]Met labeling of the peptide and by [^{32}P]-probing of the tRNA in *N. crassa* extract. [^{35}S]Met-labeling (Figure 2.2E) showed that A24R AAP (AAP with a CGU-encoded Arg at its C-terminus) functioned similarly to WT AAP (AAP with a GCG-encoded Ala at its C-terminus) to inhibit puromycin-release in high Arg. D12N A24R AAP and D12N AAP did not respond to Arg (Figure 2.2E). Thus, consistent with observations that A24R AAP is functional for stalling (*317*), the A24R mutation at the C-terminus of the AAP did not affect the AAP's ability to regulate PTC activity. Next, samples that lacked [^{35}S]Met were probed with a [^{32}P]-labeled oligonucleotide, complementary to *N. crassa* tRNA^{Arg} (CGU) (*430*). The peptidyl-tRNA bands of A24R AAP and D12N A24R AAP but not WT AAP were detected (Figure 2.2E) and A24R AAP-tRNA was resistant to

puromycin-release in high Arg. These data showed that Arg stabilized the functional AAP that was fully synthesized from truncated RNA. AAP-tRNA appeared as a single band by tRNA-probing, not as a doublet, and based on its position relative to polypeptide size-markers, this band corresponded to the lower band of the doublet observed by [^{35}S]Met-labeling in *N. crassa* extract. We confirmed specificity of the CGU-probe by showing it did not recognize AAP-tRNA with a C-terminal CGC Arg-codon (Figure 2.4).

We next examined the effect on puromycin-release of adding high Arg to reactions in which full-length AAP was synthesized in low Arg (Figure 2.2F). Previous work indicates that the relative conformation of the AAP changes with respect to the ribosome under these conditions (*324*). We added edeine, which blocks translation initiation, after 5 min of translation to stop new synthesis of AAP. One minute later, 2 mM Arg was added, and five minutes after that, the puromycin-release assay was performed. The results of adding 2 mM Arg after edeine addition were similar to those of adding 2 mM Arg at the beginning of translation; in each case, PTC activity was inhibited relative to a reaction containing low Arg (Figure 2.2F). These data indicate that a high concentration of Arg can induce a change in PTC activity of ribosomes that have synthesized AAP in low Arg.

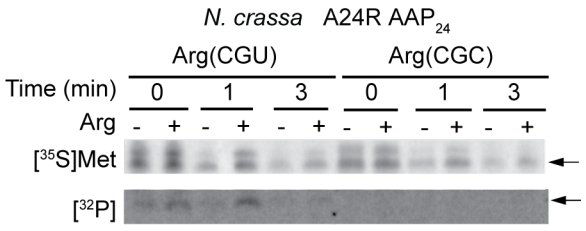


Figure 2.4. The specificity of Arg-tRNA detection with tRNA^{Arg} (CGU)-specific probe. Truncated mRNAs specifying A24R AAP with either an Arg (CGU) codon or an Arg (CGC) codon at position 24 were used to program *N. crassa* extract as in Figure 2.2E. AAP-tRNA was detected by [^{35}S]Met labeling (upper panel) or by [^{32}P] labeled tRNA^{Arg} (CGU)-specific probe (lower panel). When the AAP ends with Arg (CGU) but not Arg (CGC), the AAP-tRNA^{Arg} was detected by the probe.

2.3.2. Parameters that affect Arg- and AAP-mediated inhibition of PTC activity

We next examined effects of different parameters on the efficiency of puromycin-release in wheat germ and *N. crassa* extracts. The effects of different puromycin concentrations were tested (Figure 2.5A). In both extracts, 25 μM puromycin appeared sufficient in low Arg to effectively release the nascent AAP from the tRNA at the 3-min time point. In wheat germ extract, varying the concentration of puromycin from 0.025 mM to 1 mM had little impact on Arg's inhibitory effect on puromycin-release. However, in *N. crassa* extract, when the concentration of puromycin was greater than 0.1 mM, we did not observe an effect of Arg possibly because the release was too rapid relative to the time-points that were sampled. This difference in sensitivity of puromycin-release might be due to differences in the overall sensitivity of these translation extracts to puromycin. The *N. crassa* extract was more sensitive than the wheat germ extract to puromycin based on the ability to synthesize luciferase in the

presence of increasing amounts of puromycin (Figure 2.5B). We used 1 mM puromycin with wheat germ extract and 0.1 mM puromycin with *N. crassa* extract.

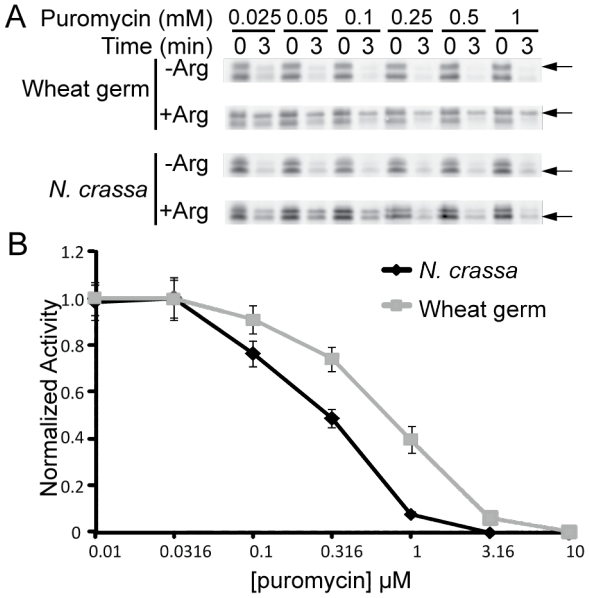


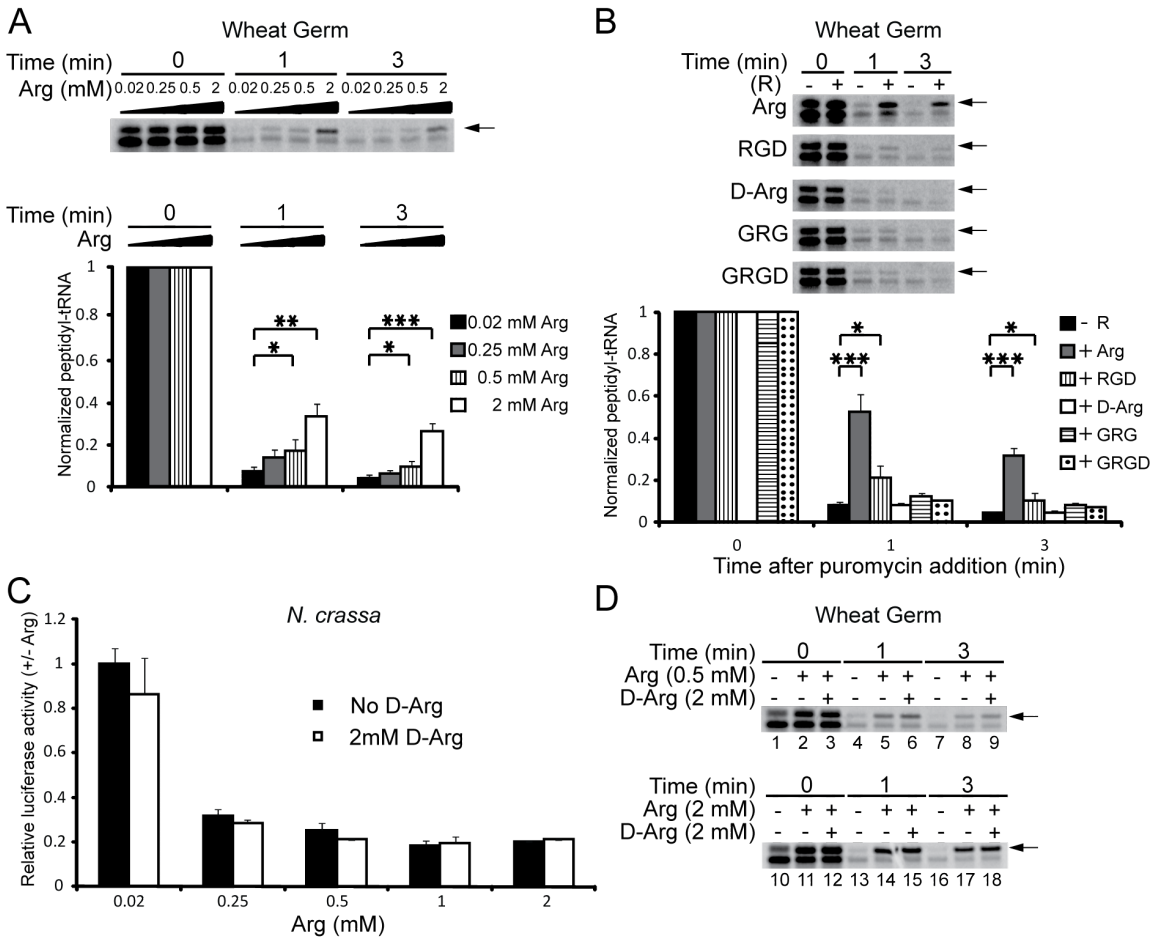
Figure 2.5. The effect of increasing puromycin concentration on translation in wheat germ and *N. crassa* extracts. (A) The effect of increasing puromycin concentration on the release of WT AAP-tRNA in wheat germ and *N. crassa* extracts. Reactions were performed in low (-) or high (+) Arg. Samples were taken at time 0 and 3 min after adding the indicated concentrations of puromycin. Arrows: [^{35}S]Met-labeled AAP-tRNA that was resistant to puromycin-release in the presence of high Arg. A representative result of triplicate experiments is shown. (B) The effect of increasing puromycin concentration on the synthesis of firefly luciferase reporter enzyme. Translation reactions were programmed with T7 LUC mRNA (mRNA encoding firefly LUC reporter only), and different concentrations of puromycin as indicated were added to reaction mixtures prior to the start of translation. Translation reactions were incubated for 30 min and the luciferase synthesis quantified by assay of luciferase activity. Luciferase synthesis in the presence of puromycin was normalized to the synthesis in the absence of puromycin and plotted versus puromycin concentration. The X axis is log-scale.

The effect of Arg concentration on puromycin-release was tested next (Figure 2.6A). We observed increased inhibition of WT AAP-tRNA release as the concentration of Arg increased from 0.25 mM to 2 mM in wheat germ extract. Was this due to nonspecific electrostatic effects of Arg? To test this, we compared the effects of adding either 2 mM Arg or 2 mM of the stereoisomer D-Arg on puromycin-release. D-Arg, which does not induce stalling (*317*), also had no discernible impact on puromycin-induced release (Figure 2.6B), showing that the effect of Arg on PTC activity required l-Arg.

Although D-Arg did not directly induce regulation by the AAP, it was possible that it could competitively inhibit Arg's capacity to induce regulation. We tested this in *N. crassa* extract by analyzing the effects of increasing concentrations of Arg on the synthesis of a firefly luciferase reporter in the presence or absence of 2 mM D-Arg (Figure 2.6C). The ability of the uORF-encoded AAP to down-regulate firefly luciferase reporter synthesis is a measurement of ribosome stalling at the uORF (*423*). As observed previously, Arg-regulation by AAP was relatively efficient, and addition of Arg at concentrations ranging from 0.25 mM to 2 mM significantly reduced luciferase synthesis ($p<0.001$, Student's t-test). Addition of 2 mM D-Arg as a competitive inhibitor at each of these Arg concentrations did not interfere with regulation. In addition, when D-Arg was added to translation reactions with 0.5 mM Arg or 2 mM Arg, puromycin-induced release of AAP-tRNA was not affected (Figure 2.6D, lanes 6, 9, 15, 18 versus lanes 5, 8, 14, 17). Therefore, the effects of Arg on AAP-specific regulation were stereo-specific, and the site(s) at which Arg functioned were not blocked by D-Arg.

Arg-regulated stalling by the AAP requires that Arg has a free N-terminus (*317*). Arg-Gly-Asp (RGD) tripeptide, which can induce AAP-mediated ribosome stalling (*317*) and induce a change in the AAP's relative conformation in the ribosome (*324*), also was inhibitory to puromycin-release, albeit more weakly that Arg (Figure 2.6B). Gly-Arg-Gly (GRG) tripeptide and Gly-Arg-Gly-Asp (GRGD) tetrapeptide did not inhibit puromycin-release (Figure 2.6B). These data indicate that the free amino group of Arg is important for the AAP-mediated inhibition of PTC activity.

Figure 2.6. The effects of Arg concentration and Arg-analogs on the puromycin-release assay. (A) Wheat germ extract supplemented with different concentrations of Arg (0, 0.25, 0.5 and 2 mM) was programmed with truncated mRNA specifying WT AAP. Samples were taken at time 0, and 1, and 3 min after adding 1 mM puromycin. Arrows: [^{35}S]Met-labeled AAP-tRNA that was resistant to puromycin-release in high Arg. In the low panel, the signals representing peptidyl-tRNAs at 1 and 3 min were normalized to the signal at time 0 in different concentrations of Arg. Mean values and standard deviations from three independent translation reactions are shown (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Student's t-test). (B) The effect of Arg-analogs on the puromycin-release in wheat germ extract programmed with truncated mRNA encoding WT AAP. (–): the translation reaction contained low Arg; (+) the translation reaction contained 2 mM of Arg or the following Arg-analogs as indicated: D-Arg, RGD, GRGD and GRG. In the lower panel, the signals representing AAP-tRNAs at 1 and 3 min were normalized to the signal at time 0. Mean values and standard deviations from three independent translation reactions are shown (*$p < 0.05$, ***$p < 0.001$, Student's t-test). (C) The effect of Arg concentration on AAP-mediated regulation of luciferase reporter synthesis. Equal amounts of AAP-LUC mRNA were translated in *N. crassa* extract supplemented with 0.02, 0.25, 0.5, 1 or 2 mM Arg in the absence or presence of 2 mM D-Arg. All reactions contained mRNA for sea pansy luciferase. Firefly luciferase synthesis was first normalized to sea pansy luciferase synthesis. The relative levels of synthesis are shown relative to the extract containing 0.02 mM Arg and no D-Arg. Mean values and standard deviations from three independent experiments, each performed in triplicate, are given. (D) The effect of D-Arg on the release of peptidyl-tRNA by puromycin. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin to wheat germ extract. D-Arg and Arg were present as indicated. A representative result of triplicate experiments is shown.



2.3.3. The regulatory functions of extended and shortened AAPs

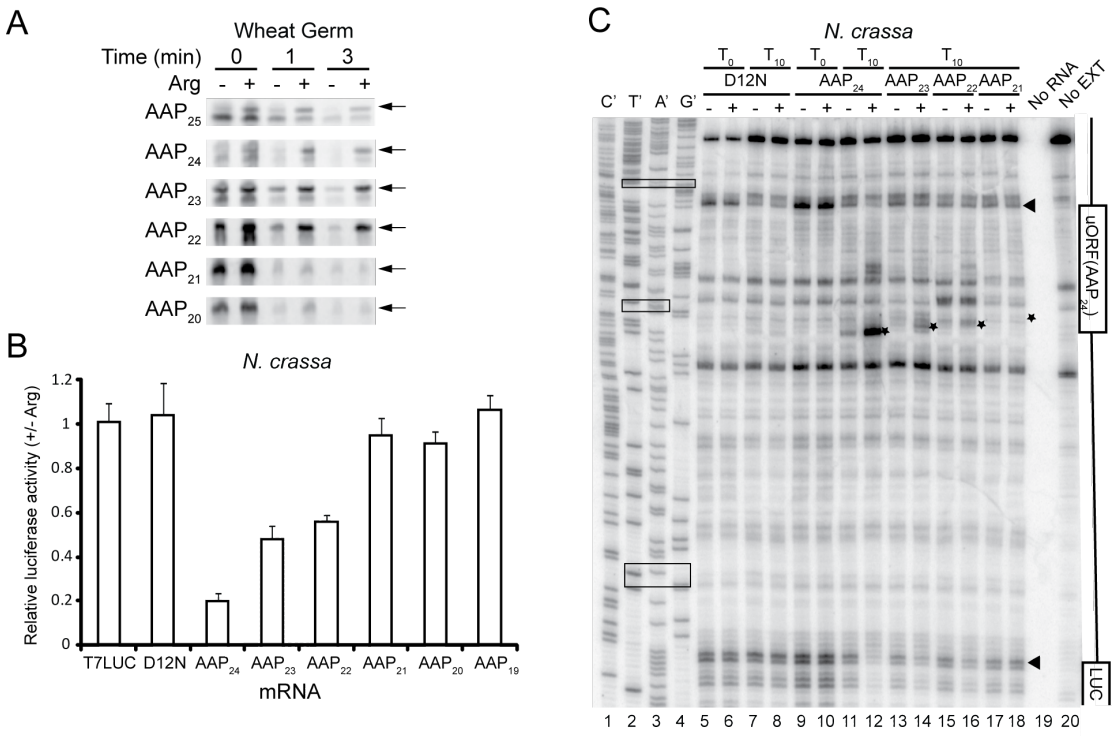
Toeprinting analyses indicated that Arg-specific ribosome-stalling was considerably diminished when the *N. crassa* AAP was shortened by a single residue at its C-terminus (*322*). However, C-terminally shortened AAPs still undergo a change in conformation relative to the ribosome in response to Arg (*324*). Since puromycin-

induced release is a direct assay of PTC activity, we used this assay to examine the function of C-terminally extended and shortened AAPs.

Programming wheat germ extracts with truncated mRNAs specifying AAPs ending at Val-25 (AAP_{25}, which the AAP stop codon is changed to a valine codon), Ala-24 (AAP_{24}, the WT AAP), Asn-23 (AAP_{23}), and Leu-22 (AAP_{22}) showed that each AAP inhibited puromycin-induced release of the nascent chain in high Arg (Figure 2.7A). However, AAPs ending at Ala-21 (AAP_{21}) or Arg-20 (AAP_{20}) did not inhibit puromycin-release (Figure 2.7A). These results indicate that AAPs extended by one residue or shortened at the C-terminus by one or two residues regulated PTC activity in response to Arg, but further shortening eliminated this regulatory function.

We tested whether C-terminally shortened AAPs were capable of regulating translation in response to Arg when encoded as uORFs in the 5'UTRs of capped and polyadenylated luciferase reporter mRNAs (Figure 2.7B). In *N. crassa* extract, the full-length wild-type AAP (AAP_{24}) conferred approximately 5-fold regulation; AAP_{23} and AAP_{22} conferred approximately 2-fold regulation. AAPs truncated further (AAP_{21} and AAP_{20}) showed no regulatory function in response to Arg. Similarly, mRNAs that lacked the uORF-encoded AAP (T7 LUC) or that contained the D12N AAP did not show Arg-regulated synthesis of luciferase (Figure 2.7B). These results are consistent with those obtained from the puromycin-release assay (Figure 2.7A) and indicate that AAPs shortened by one or two residues still confer regulation in response to Arg.

Figure 2.7. The effects of truncating AAP on Arg-specific translational control. (A) The effect of AAP truncation or extension on puromycin-release in wheat germ extract. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin. Translation reactions were performed in low (–) or high (+) Arg. For the truncated AAP mRNAs used in these experiments, the numbers in subscripts indicate the length of the AAP (*e.g.*, AAP_{24} indicates AAP was truncated after position 24). AAP_{25} changes the wild-type stop codon to a valine codon. A representative result of triplicate experiments is shown. Arrows: [^{35}S]Met-labeled AAP-tRNA that was resistant to puromycin-release in high Arg. (B) The regulatory effects of wild-type and C-terminally truncated AAPs positioned as uORFs in a capped and polyadenylated luciferase reporter. uORFs were created with stop codons at AAP positions 20-25 as indicated (*e.g.*, AAP_{24} contains a stop codon at position 25). Equal amounts of each mRNA were translated in *N. crassa* extract containing low or high Arg. As controls, equal amounts of mRNA encoding firefly LUC reporter only (T7 LUC) or mRNA containing the D12N AAP uORF were translated in parallel. All reactions contained mRNA for sea pansy luciferase to serve as an internal control for translation. Firefly luciferase synthesis was normalized to sea pansy luciferase synthesis and then luciferase synthesis in high Arg was calculated relative to synthesis in low Arg. Mean values and standard deviations from three independent experiments, each performed in triplicate, are given. (C) Toeprint analysis to assess Arg-regulation by WT, D12N and truncated AAPs. *N. crassa* extract was programmed with equal amounts (60 ng) of mRNA specifying D12N, AAP_{24} (WT), AAP_{23}, AAP_{22} or AAP_{21} in the 5'-leader. mRNAs were those used for luciferase measurements in (B). Cycloheximide was added either prior to incubation (T_0) or after 10 min of incubation (T_{10}). Radiolabeled primer ZW4 was used for primer extension analysis and for sequencing the AAP_{24} template (lanes 1-4). The nucleotide complementary to the dideoxynucleotide added to each sequencing reaction for the AAP_{24} template is indicated above the corresponding lane so that the sequence of the template can be directly deduced. Asterisks: toeprint products corresponding to ribosomes stalled at the termination codon of AAP. Arrowheads: toeprint products corresponding to ribosomes bound at the initiation codons. Boxes (top to bottom): AAP initiation codon, AAP termination codon and LUC initiation codon. No EXT, RNA without extract. No RNA, extract without RNA.



We next directly examined the capacity of these uORF-encoded AAPs to stall ribosomes in *N. crassa* extracts using the toeprinting assay, which shows the positions of ribosomes engaged in the translation of mRNA (Figure 2.7C). Cycloheximide (CYH) was added at 0 min (T_0) or 10 min (T_{10}) to increase the signals from ribosomes at translation initiation sites (*269*). When CYH was added at T_0, both D12N and WT (AAP_{24}) mRNAs showed similar levels of ribosomes at the uORF and luciferase (LUC) initiation codons, either in low or high Arg (Figure 2.7C, lanes 5 and 6, and lanes 9 and 10). This is consistent with leaky-scanning of ribosomes past the uORF initiation codon (*269*). When CYH was added after 10 min, an increased toeprint signal that corresponded to ribosomes stalled at the AAP_{24} uORF termination codon and a reduced

signal corresponding to ribosomes at the LUC initiation codon were observed in extracts containing high Arg (Figure 2.7C, lane 12 versus lane 11). Neither difference in signal intensity was observed for D12N AAP, which lacks regulatory activity (Figure 2.7C, lane 8 versus lane 7). This is consistent with ribosomes stalled at the AAP_{24} termination codon acting to block leaky-scanning to the downstream LUC initiation codon. As expected, signals corresponding to ribosomes at initiation or termination codons were not observed in toeprint analyses of RNA alone (Figure 2.7C, lane 20) or of extract alone (Figure 2.7C, lane 19). Importantly, when CYH was added at 10 min, stalling occurred at the termination codons for AAP_{23} and AAP_{22} in high Arg but at a reduced level compared to stalling at the termination codon for AAP_{24} (Figure 2.7C, lanes 11-16). High Arg also reduced the signal corresponding to ribosomes at the LUC initiation codon. The AAP_{21} mRNA in parallel reactions showed no effect of Arg on stalling at the uORF termination codon or LUC initiation codon (Figure 2.7C, lane 18 versus lane 17). Thus, with respect to regulating activity for truncated AAPs, toeprinting data was qualitatively similar to the luciferase reporter data (Figure 2.7B). The toeprinting signal showed ribosomes stalled with the stop codon in the A site, indicating that the translocation reaction, which transferred the peptidyl-tRNA from the A site to the P site had occurred in the stalled ribosomes.

The observations that truncated AAPs interfered with PTC activity suggested that the AAP could inhibit PTC activity prior to its complete synthesis. We directly examined this possibility in *N. crassa* extract using $tRNA^{Arg}$-probing. Arg (CGU) codons were strategically placed in mRNAs specifying different truncated AAPs. As

expected, when the last codon was Arg (CGU), L22R AAP_{22}, N23R AAP_{23}, A24R AAP_{24} and *25R AAP_{25} all showed regulation in response to Arg based on both [^{35}S]Met and $tRNA^{Arg}$-probing, while A21R AAP_{21} showed no regulation by either measurement (Figure 2.8A). These data are consistent with the results obtained for wheat germ extract using [^{35}S]Met labeling (Figure 2.7A). In parallel experiments we observed that for truncated mRNA encoding L22R AAP_{24} or N23R AAP_{24}, AAP-$tRNA^{Arg}$ was detected and was resistant to puromycin-release in high Arg but not low Arg. These data indicate that Arg-regulation of PTC activity occurred prior to full synthesis of the AAP. This was not the case for A21R AAP_{24}, which showed regulation by [^{35}S]Met labeling but for which $tRNA^{Arg}$ was not detected (Figure 2.8A). This indicated that translation proceeded efficiently past codon 21 and that a regulatory "window" in which Arg could affect PTC activity spanned from codon-22 to codon-25 of the AAP.

Interestingly, the peptidyl-tRNA of D12N N23R AAP_{24} was detected by $tRNA^{Arg}$ probe (Figure 2.8A), indicating that translation was slowed prior to the complete synthesis of this AAP. This suggested that, independent of the capacity of the AAP to elicit Arg-regulated stalling, slowing occurs near the AAP C-terminus. This could reflect a general inability of the translation system to decode truncated RNA, or it could reflect intrinsic stalling activity of the AAP that was independent of Arg-regulation (*e.g.*, that occurred in the presence of the D12N mutation which eliminates regulation). To examine this, we made a construct specifying a frame-shift AAP (FS AAP) in which only residues 1-3 and 23-24 were identical to those in the AAP (Figure 2.8B). An Arg

(CGU) codon was placed at residues 22, 23 or 24 of WT, D12N and FS AAPs. Intrinsic stalling activity was measured by translation of equal amounts of each RNA in low Arg (Figure 2.8B, lower panel) followed by northern analysis to detect peptidyl-tRNA (Figure 2.8B, upper panel). As expected for ribosome nascent chain complex formation on truncated RNA, each mRNA showed a strong signal when the Arg-codon was the final codon (AAP codon-24) of the truncated RNA (Figure 2.8B, lanes 3, 6 and 9). When the Arg-codon was at AAP codon-22 or codon-23, stronger signals were detected for WT and D12N AAP relative to the FS AAP (Figure 2.8B, lanes 2, 5 versus lane 8 and lanes 1, 4 versus lane 7), with WT AAP showing the strongest signals overall. These data showed that WT and D12N AAP exhibited greater intrinsic stalling activity than FS AAP and indicate that evolutionarily conserved elements in the AAP contribute to intrinsic stalling activity even in the absence of Arg-stimulated stalling activity.

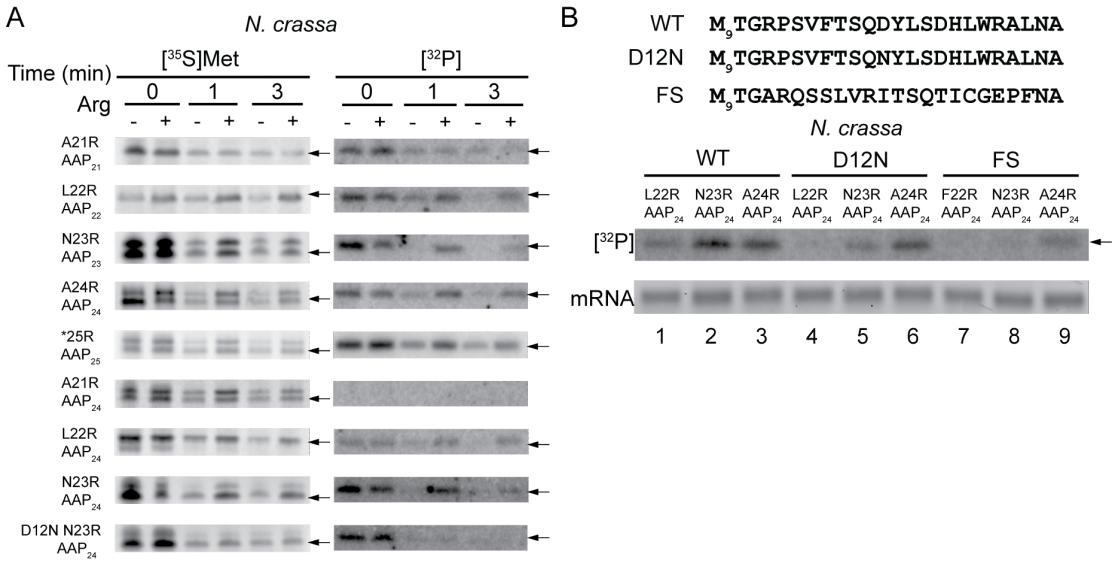


Figure 2.8. Ribosome stalling prior to complete AAP synthesis. (A) A21R, L22R, N23R, A24R and *25R indicate that the original codons at AAP positions 21, 22, 23, 24 or 25 were changed respectively to CGU Arg-codons. AAP_{25}, AAP_{24}, AAP_{23}, AAP_{22} and AAP_{21} indicate that AAP was truncated after position 25, 24, 23, 22 or 21. *N. crassa* extracts were programmed with the indicated RNAs in either low (–) or high (+) Arg. Samples were taken at time 0, and 1 and 3 min after adding 0.1 mM puromycin. Arrows indicate the position of peptidyl-tRNA. Left panel: AAP-tRNA labeled with [^{35}S]Met. Right panel: AAP-tRNA detected with [^{32}P] labeled $tRNA^{Arg}$ (CGU) probe. (B) Truncated mRNAs encoding WT AAP, D12N AAP or frame-shift (FS) AAPs, which contained Arg (CGU) codons at the indicated positions, were translated in *N. crassa* extract for 10 min at 26°C. Upper panel: peptidyl-tRNA (arrow) detected with [^{32}P] labeled $tRNA^{Arg}$ (CGU) probe. Lower panel: Gel analysis to establish similar amounts of input mRNAs were used in translation reactions shown in the upper panel.

A question that arises is whether the addition of Arg codons creates new stalling sites and this accounts for the detection of peptidyl-tRNA corresponding to internal stall sites with the probe. We tested this by making constructs in which the AAP was fused in-frame with the luciferase coding region. In such constructs, for the wild-type AAP, the primary stalling signals observed by toeprinting are in the region downstream of AAP-codon 24. We used toeprinting to analyze ribosome-stalling in constructs that

contained wild-type AAP sequence, or contained mutations D12N, A21R, L22R, N23R, A24R or *25R (the latter representing an additional codon in the construct) (Figure 2.1C and Figure 2.9). In each construct, except for the D12N mutation, increased stalling corresponding to ribosomes downstream of codon-24 was observed. Additional Arg-regulated stall sites in the AAP coding region approximately 30 nt upstream could be attributed to ribosomes stalled behind those primarily stalled by Arg. Importantly, no toeprint signals that would correspond to ribosomes stalled at internal positions in the AAPs substituted with Arg were observed; thus, these Arg-codons did not create new stalling sites.

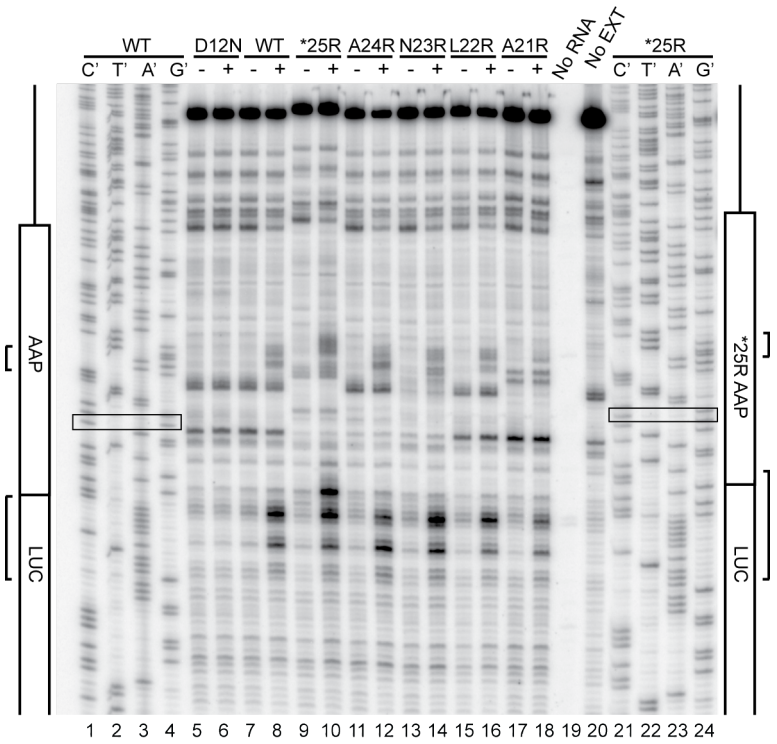


Figure 2.9. Toeprint analysis to indentify ribosome stalling sites for AAP-LUC fusion constructs. *N. crassa* extract was programmed with equal amounts (60 ng) of mRNA specifying WT, D12N, A21R, L22R, N23R, A24R and *25R AAPs fused with in-frame luciferase coding sequence. Cycloheximide was added after 10 mins of incubation. Radiolabeled primer ZW4 was used for primer extension analysis and for sequencing the WT AAP template (lanes 1-4) and *25R AAP template (lanes 21-24). the nucleotide complementary to the dideoxynucleotide added to each sequencing reactions for the template is indicated above the corresponding lane so that the sequence of the template can be directly deduced. Boxes: AAP-codon 24. Brackets (top to bottom): Toeprint products corresponding to ribosomes stalled approximately 30 nt upstream those primarily stalled by Arg and primary toeprint products corresponding to ribosomes stalled in the region downstream of AAP-codon 24. No EXT, RNA without extract. No RNA, extract without RNA.

2.3.4. The effects of other RAPs on peptidyl transferase activity

We compared the function of the *S. cerevisiae* AAP (*Sc* AAP) to the *N. crassa* AAP (*Nc* AAP) using wheat germ extracts (Figures 2.10A and 2.10B). In low Arg, both

Sc WT AAP-tRNA and *Nc* WT AAP-tRNA disappeared rapidly following the addition of puromycin. In high Arg, for both *Sc* WT AAP-tRNA and *Nc* WT AAP-tRNA, the rates of peptidyl-tRNA disappearance were slowed, suggesting that *Sc* AAP functioned similarly to *Nc* AAP to inhibit PTC activity in response to Arg. Synthesis of cytomegalovirus gp48 is inhibited by translation of a 22-codon uORF (gp48 uORF2); the gp48 uORF2 nascent peptide stalls ribosomes at the uORF2 termination codon (*288*). gp48 uORF2 was translated in wheat germ extract and its effects on the PTC were tested using the puromycin-release assay (Figure 2.10C). WT gp48 uORF2-tRNA did not disappear rapidly following the addition of puromycin (Figure 2.10C, lanes 2 and 3 versus lane 1), indicating that WT gp48 uORF2 inhibited PTC activity. However, in striking contrast, for the nonfunctional gp48 uORF2 containing the P22A mutation, uORF2-tRNA was released rapidly by puromycin (Figure 2.10C, lanes 5 and 6 versus lane 4). Thus, in contrast to results with the AAP, where the identity of the C-terminal amino acid had little impact on inhibition of puromycin-release, gp48 uORF2 Pro-22 was crucial for this function. The inhibition of nascent chain release by AAP in high Arg and gp48 uORF2 was directly compared for an extended period of incubation (up to 30 min) in the presence of puromycin. The gp48 uORF2 was more inhibitory than the AAP (Figure 2.11). These analyses of gp48 uORF2 indicate an effect of this RAP on PTC activity beyond its known effect on inhibiting the function of eukaryotic eRF1 (*422*). Preparation of samples for cryo-EM studies by others also indicated that the gp48 uORF2-induced ribosome stall was highly stable (*325*). However, the possibility that the gp48 uORF2-tRNA is released from the P site but remains associated with ribosomes is

another potential explanation for the observed resistance of gp48 uORF2-tRNA to cleavage by puromycin (*433*).

One difference between gp48 uORF2 and AAP is that gp48 uORF2 ends with a Pro residue, and the presence of a Pro residue in the ribosome P site has been associated with reduced reactivity toward puromycin. We tested the A24P AAP and D12N A24P AAP in the puromycin release assay (Figure 2.11). The AAP A24P mutation did not decrease reactivity to puromycin in either the wild-type or D12N AAP context.

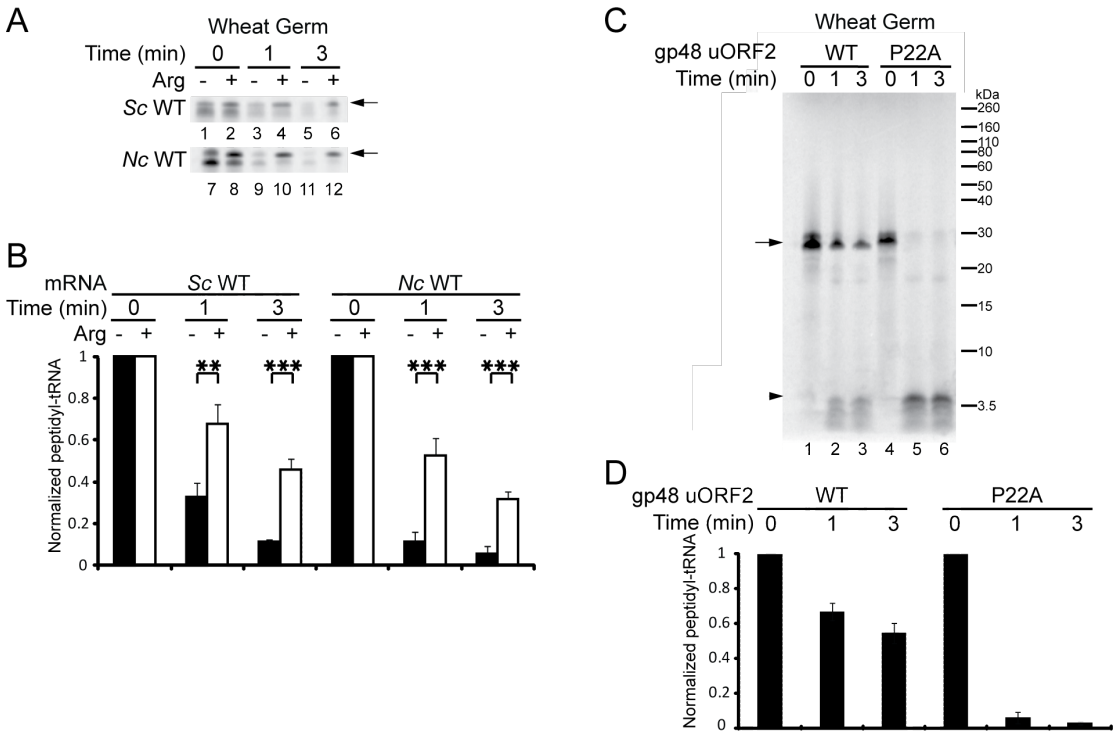


Figure 2.10. The *S. cerevisiae* AAP and CMV gp48 uORF2 peptide interfere with puromycin-release. (A) Truncated mRNAs specifying *S. cerevisiae* (*Sc*) AAP and *N. crassa* (*Nc*) AAP were translated in wheat germ extract. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin. Translation reactions contained low (–) or high (+) Arg. Arrows: [^{35}S]Met-labeled AAP-tRNA. (B) Quantification of puromycin-release assay of *Sc* AAP-tRNA and *Nc* AAP-tRNA in wheat germ extract. The signals representing peptidyl-tRNAs at 1 and 3 min were normalized to the signal at time 0. Mean values and standard deviations from three independent translation reactions are given (**$p < 0.01$, ***$p < 0.001$, Student's t-test). (C) The effect of CMV gp48 uORF2 on peptidyl-tRNA release by puromycin. WT gp48 uORF2 peptide and the nonfunctional mutated P22A gp48 uORF2 were tested in wheat germ extract. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin. Arrow: [^{35}S]Met-labeled gp48 uORF2-tRNA. Arrowhead: [^{35}S]Met-labeled free gp48 uORF2 and gp48 uORF2-puromycin. (D) The signals representing gp48 uORF2-tRNAs at 1 and 3 min were normalized to the signal at time 0. Mean values and standard deviations from three independent translation reactions are given.

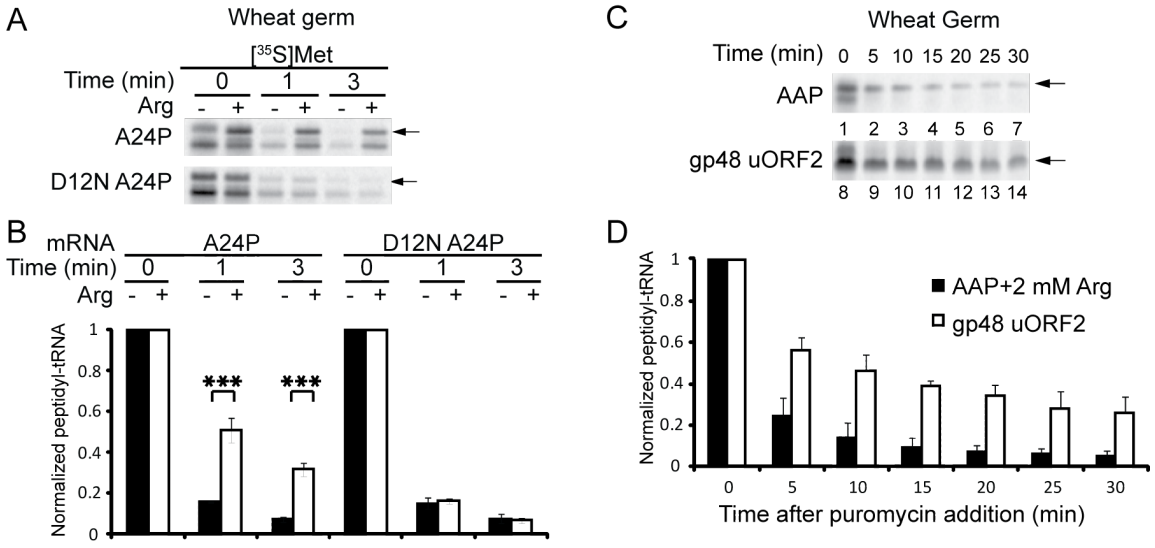


Figure 2.11. Inhibition of puromycin-release of peptidyl-tRNA by the AAP and CMV gp48 uORF2. (A) Truncated mRNAs specifying A24P AAP and D12N A24P AAP were translated in wheat germ extract. Samples were taken at time 0, and 1 and 3 min after adding 1 mM puromycin. Translation reactions contained low (-) or high (+) Arg. Arrows: [35S]Met-labeled AAP-tRNA. (B) Quantification of puromycin-release assay of A24P AAP-tRNA and D12N A24p AAP-tRNA in wheat germ extract. the signals representing peptidyl-tRNAs at 1 and 3 min were normalized ot the signal at time 0. Mean values and standard deviations from three independent translation reactions are given (***$p < 0.001$, Student's t-test). (C) Inhibition of puromycin-release of peptidyl-tRNA by the AAP and CMV gp48 uORF2 assessed over an extended time-frame. Wheat germ extract was programmed with truncated mRNA encoding either the WT AAP or CMV gp48 uORF2 as in Figure 2.10. For the WT AAP, translation reactions were supplemented with 2mM Arg. Samples were taken at time 0, and 5, 10, 15, 20 and 30 min after adding 1 mM puromycin. A representative result of triplicate experiments is shown. (D) Quantification of puromycin-release of WT AAP-tRNA and gp48 uORF2-tRNA. The signals representing peptidyl-tRNAs at 5, 10, 15, 20, 25 and 30 min were normalized to the signal at time 0. Mean values and standard deviations from three independent translation reactions are given.

2.4. Discussion

The nascent AAP interferes with PTC activity in the presence of high Arg. Arg-regulated inhibition of PTC activity was observed for full-length AAP, for AAP extended by one amino acid at its C-terminus (AAP_{25}) and for AAP truncated by one or two amino acids at its C-terminus (AAP_{23} and AAP_{22}) (Figures 2.7A and 2.8A). These truncated AAPs, like the full length AAP, caused regulatory ribosome stalling in response to Arg, albeit with reduced efficiency (Figures 2.7B and 2.7C). Analyses of tRNA-identity also showed the AAP regulated PTC activity prior to its complete synthesis. These results demonstrate that the AAP has the unusual property of interfering with PTC activity across a window spanning at least four consecutive codons (AAP codons 22-25). We also obtained data indicating that both wild-type AAP and D12N AAP (which lacks regulatory function) had higher intrinsic stalling activity than a frame-shifted peptide. Thus the AAP appears to have intrinsic stalling activity detectable even in the absence of Arg-regulated stalling.

Ribosomes containing AAP_{24} synthesized in low Arg, but then subsequently incubated with high Arg, show a change in the relative conformation of the AAP with respect to the ribosome (*324*) and a reduction in PTC activity (Figure 2.2F). These data support the idea that the Arg-induced change in relative conformation of AAP with respect to the ribosome is necessary to provide the capacity for AAP to interfere with PTC activity.

Why does the full-length AAP stall ribosomes more efficiently than AAPs truncated at their C-termini? It is not the absolute length of the nascent AAP, because

while C-terminal truncations AAP_{23} and AAP_{22} reduced the efficiency of ribosome stalling based on reporter synthesis and toeprint assays (Figures 2.7B and 2.7C), N-terminal truncations yielding even shorter AAPs do not (*322*). Possibly, in high Arg, when the ribosome translating AAP_{22} or AAP_{23} is stalled because the ribosome has reached the 3'-end of a truncated mRNA, AAP_{22} or AAP_{23} nascent peptides have sufficient opportunity to obtain proper register with respect to the ribosome so that, like AAP_{24}, they can interfere with PTC activity. Important in this regard, the relative conformation of each of these C-terminally truncated AAPs also changes with respect to the ribosome in response to Arg (*324*). However, if the mRNA extends past codon-22 or codon-23, because AAP_{22} and AAP_{23} might not adapt or maintain register as efficiently as AAP_{24}, the ribosome translates past codon-22 and codon-23. There is some stalling of ribosomes that are engaged in elongation at these codons, however, as determined by the detection of stabilized AAP-$tRNA^{Arg}$ for L22R AAP_{24} and N23R AAP_{24} (Figure 2.8). In contrast, AAP_{21}, which undergoes a Arg-dependent conformational change with respect to the ribosome (*324*), cannot find proper resister to interfere with PTC activity (Figures 2.7 and 2.8). These data suggest a model for AAP-mediated stalling in response to Arg in which the AAP undergoes a change in relative conformation with respect to the ribosome, and this altered conformation must be in proper register with the ribosome to efficiently interfere with PTC activity. In this model, there is a window of AAP chain-lengths for which AAP can find register (with respect to the ribosome and possibly with respect to Arg), with the wild-type AAP length being most efficient at achieving and/or remaining in register. Consistent with this idea, of 120 uORF-encoded AAPs identified

by evolutionary conservation, the C-terminal regions of 119 are at least as long as the *N. crassa* AAP C-terminus, and none are more than two residues longer (*317*).

How is PTC activity affected by AAP and Arg to inhibit the function of puromycin? There are at least several ways that access of puromycin to the ribosome A site could be restricted. There could be failure of translocation (Figure 2.12, step 2 and 3), or the structure of the ribosome could be altered such that access to the empty A site is blocked (Figure 2.12D, step 4). Alternatively, puromycin could enter the ribosome A site (Figure 2.12E), but the peptidyl transferase reaction is inhibited (Figure 2.12, step 5). Toeprint analyses of mRNAs on which the AAP stalls ribosomes at its termination codon show that the termination codon is in the A site, indicating stalling occurs with a configuration of the ribosome depicted in Figure 2.12D. Furthermore, AAP can be fully synthesized in low Arg and still interfere with PTC activity when Arg is subsequently added (Figure 2.2F). These data indicate that Arg's inhibitory effect on PTC activity occurs after translocation (after Figure 2.12, step 3). Thus, the relative conformation of the AAP in the ribosome in high Arg could restrict access to the empty A site (Figure 2.12, step 4) and/or could interfere with the peptidyl transferase reaction (Figure 2.12, step 5). The cryo-EM model of the wheat ribosome containing AAP in the absence of Arg raises the possibility that an altered position of ribosomal protein L10e could be responsible for inhibiting PTC activity (*325*).

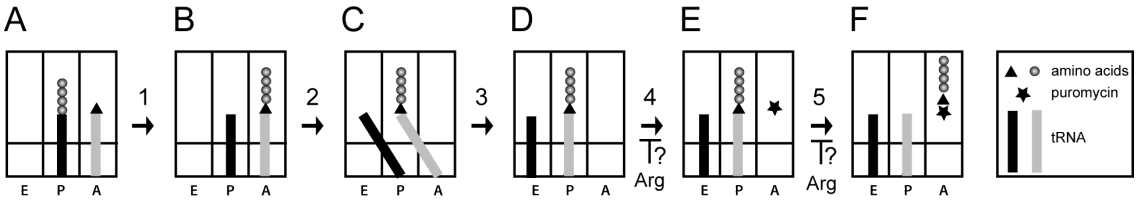


Figure 2.12. Schematic representation of PTC activity and potential steps affected by the AAP and Arg. (A) The peptidyl-tRNA and aminoacyl-tRNA are in the P and the A sites, respectively, of the peptidyl transferase center of the ribosome. (B) Immediately after peptide-bond formation, a peptidyl-tRNA is in the A site and an uncharged tRNA is in the P site. (C) The aminoacyl end of the tRNA moves to the P site, resulting in peptidyl-tRNA in a P/A hybrid site and uncharged tRNA in an E/P hybrid site. (D) The peptidyl-tRNA translocates to the P/P site and deacylated tRNA to the E/E site, together with the movement of the associated mRNA by one codon. (E) Puromycin can enter an empty A site as depicted. (F) The peptidyl transferase reaction transfers the peptide from peptidyl-tRNA to puromycin resulting in release of the nascent peptide from tRNA. In this figure, tRNAs are shown as black and gray bars. The amino acids are depicted as circles and triangles. Puromycin is depicted shown as a star. Steps potentially inhibited by AAP and Arg (see text) are indicated.

Unlike many other RAPs, specific residues of the AAP are not required at or near the PTC for the AAP to cause ribosome-arrest. For example, the C-terminal Pro-22 of CMV uORF2 is crucial for stalling (*289*) and for inhibiting PTC activity based on the puromycin-release assay (Figure 2.10B). SecM, MifM, TnaC and ErmAL1 each require specific amino acids at the PTC for stalling to occur (*346, 355, 371, 434*). ErmCL, ermAL1 and SecM each also require nearby amino acids (-2 relative to the nascent peptide C-terminus) for stalling to occur (*351, 371*). In contrast, AAP residues 9-20 are and sufficient to confer regulatory function and AAP residues 21-24 can all be substituted with Ala and AAP function retained (*317*). Furthermore, the evolutionary conservation of residues of the AAP beyond Arg-20 is relatively low.

In summary, we show here that the AAP interferes directly with PTC activity in response to Arg. The AAP is an evolutionarily conserved, uORF-encoded regulatory peptide and there is expanding appreciation of the wider roles for uORFs in controlling gene expression. The AAP represents an unusual example of nascent peptide control of ribosome activity because there is no evident requirement for specific nascent peptide residues to be at or near the PTC for stalling to occur. The AAP stalls ribosomes from fungal, plant, and mammalian sources and thus must exert a regulatory effect through conserved regions of the eukaryotic ribosome. Understanding the function of the AAP provides a basis for gaining insight into fundamental processes by which nascent peptides and metabolites regulate gene expression.

CHAPTER III

THE STRINGENCY OF START CODON SELECTION IN THE FILAMENTOUS FUNGUS *NEUROSPORA CRASSA*[*]

3.1. Introduction

In standard eukaryotic translation initiation, the preinitiation complex that is formed by the small ribosomal subunit, initiator tRNA and multiple initiation factors binds at the mRNA 5'-cap and scans downstream for an initiation codon (*81*). The AUG codon that is closest to the mRNA 5'-cap and that is in a preferred context is typically selected as the initiation site (*69*). Start codon selection is aided through biases in the nucleotides surrounding the start codon and through the actions of initiation factors (*21*). In certain cases, initiation occurs at non-AUG codons, especially near-cognate codons (NCCs) that differ from AUG by a single nucleotide (*372, 377-380*). In most cases, initiation from NCCs uses Met-tRNA$_i^{Met}$ (*372*). There are specific cases where viral internal ribosome entry sites use a different mechanism for non-AUG initiation (*435*), and recently, it was established that leucyl-tRNA can be used for translation initiation at a CUG codon of a mammalian mRNA (*393*). Initiating translation from NCCs may increase coding capacity of transcripts or contribute regulatory functions.

[*] This material has been published in this or similar form in *The Jounral of Biological Chemistry* and is used here with permission of American Society for Biochemistry and Molecular Biology.
Wei, J., Zhang, Y., Ivanov, P.I., and Sachs, M.S. (2013). The stringency of start codon selection in *Neurospora crassa*. J Biol Chem *288*, 9549-9562.

The initiation context surrounding the start codon, which is generally defined by nucleotides from -6 to +4 (the A of AUG is +1), has strong influence on initiation efficiency (*373, 374*). The Kozak consensus [GCC(A/G)CCxxxG] is optimal for initiation in mammalian cells (*374, 436*). A purine at position -3 and a 'G' at position +4 are most important for efficient initiation (*21, 374*). Substitutions at positions -1 and -2 have less impact on translation efficiency, but the effects of changes at these positions are intensified when the -3 nucleotide is also suboptimal.

Codons other than AUG are generally less efficient initiation codons *in vivo*; however, codons differing from AUG in a single position, collectively referred to as near-cognate initiation codons, are known to support translation initiation in eukaryotes (*68, 372*). The presence of a good Kozak context is crucial for efficient use of near-cognate initiation codons in mammals, plants and yeast (*68, 437, 438*). In these studies, the measured efficiencies of initiation from functional NCCs vary, and values between ~1 to 20% of initiation from AUG are reported. AAG and AGG do not serve as initiation codons: a purine at +2 evidently eliminates function as an initiation codon (*372, 377, 378*). A recent study in human cells reveals a hierarchy in initiation efficiencies at NCCs (*379*). The most efficient NCCs are CUG and GUG (19.5% and 9.2% of AUG-initiated translation). ACG (6.6%), AUA (3.3%), AUU (3.2%), UUG (1.9%) and AUC (1.7%) are used as initiation codons at lower efficiencies (*379*). A recent analysis in *Saccharomyces cerevisiae* demonstrated the hierarchy CUG>GUG≈UUG>ACG≈AUA>AUU>AUC (*377*), with efficiencies comparable to those observed in human cells. In plant

protoplasts, CUG is the most active (30%), following by GUG and ACG (each 15%), while AUA, AUU, UUG and AUC are less active (2-5%) (*380*).

Initiation from NCCs increases the capacity to generate protein isoforms with different regulatory functions (*386*). In some cases, initiation from an upstream in-frame NCC produces an alternative N-terminally extended isoform of protein in addition to that produced by initiation at the downstream AUG (*69 and references therein*). Protein isoforms with different N-terminal sequences can serve different functions and/or be targeted to different locations (*439-442*). The synthesis of some proteins is thought to start exclusively at non-AUG codons. These include mammalian eIF4G2, which uses GUG (*443*); *Podospora anserina* IDI-4, which uses CUG (*387*); yeast glycyl-tRNA synthetase, which uses UUG (*56*); *Arabidopsis AGAMOUS*, which uses ACG (*444*), *Arabidopsis FCA*, which uses CUG (*445*); and some RpoT genes in *Nicotiana* spp., which use CUG (*62, 389*). In yeast, genome-wide analysis reveals that NCC-initiated upstream open reading frames (uORFs) can be translated at levels comparable to AUG-initiated uORFs (*385*). Data indicate that, in response to amino acid starvation, levels of translation of these non-AUG yeast uORFs generally increase relative to levels of translation of the downstream coding sequences. Studies in mouse embryonic stem cells identified many unannotated initiation sites started by AUG or NCCs that direct the translation of uORFs with regulatory potential (*384*).

The stringency of start codon selection can affect both the efficiency of initiation at AUG codons in different contexts and the efficiency of initiation at NCCs. Stringency can be modulated by different physiological conditions, such as amino acid starvation

(*385*). eIF1 and eIF5 play important roles in regulating the stringency of start codon selection (*81, 379, 446, 447*). The auto- and cross- regulatory functions of eIF1 and eIF5 that themselves utilize stringency of start codon selection demonstrates that varying the stringency of start codon selection can be used for gene regulation in eukaryotes.

The stringency of start codon selection can also be regulated by other parameters. Two structurally related molecules isoquinoline-1-carboxylic and 7-amino-5-iodo-8-quinolinol were identified in a high throughput screen to decrease the stringency of start codon selection in yeast (*448*). These compounds can also increase initiation at natural uORFs initiated by NCCs. Their mechanism of action is currently not known. The concentration of Mg^{2+} can influence the fidelity of translation initiation: a high concentration of Mg^{2+} increases initiation at non-AUG codons *in vitro* (*68*).

Here, we used a firefly luciferase reporter that was codon-optimized for expression in *N. crassa* to determine the efficiency of translation initiation at NCCs. We stably integrated luciferase reporter genes that initiated with AUG or each of the nine NCCs in preferred context at the *N. crassa his-3* locus and determined luciferase levels *in vivo.* CUG and GUG were the most efficient NCCs, with 11% and 7% of the efficiency of AUG. AUA, AUU, UUG and AUC were less active (1-5%), while AAG and AGG did not function. The hierarchy of initiation efficiency at NCCs was similar to that in human cells. Cell-free translation extracts of *N. crassa* were also used to analyze translation initiation at NCCs. The hierarchy of utilization of NCCs was similar to that observed *in vivo*, but overall efficiency was strongly dependent on the concentration of Mg^{2+}. Our studies, which are one of a handful of analyses that systematically examine

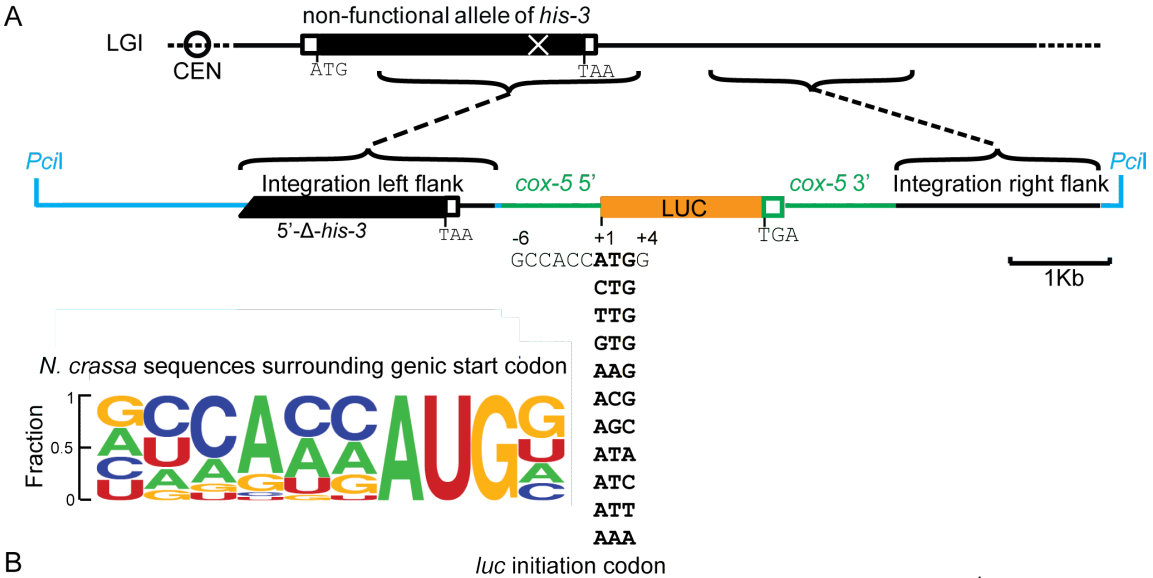
NCC initiation in eukaryotes, and are the first for filamentous fungi, demonstrate that such initiation could substantially increase the coding capacity of mRNAs.

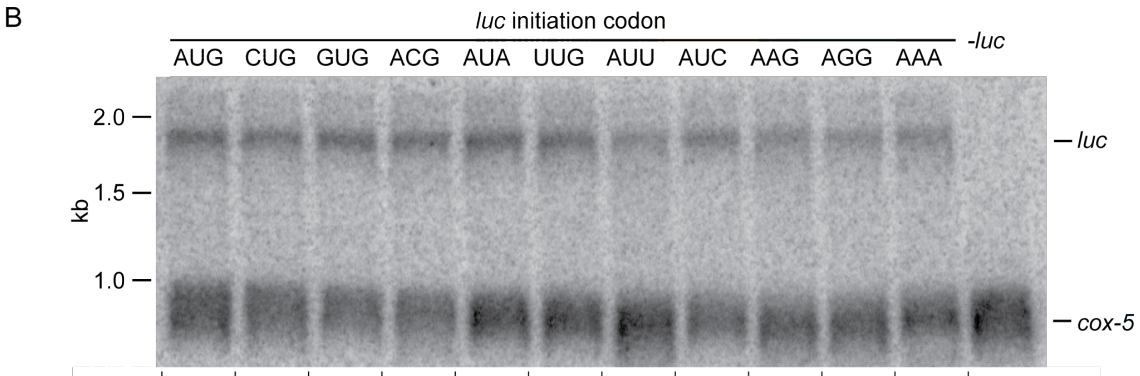
3.2. Materials and methods

3.2.1. Logogram generation

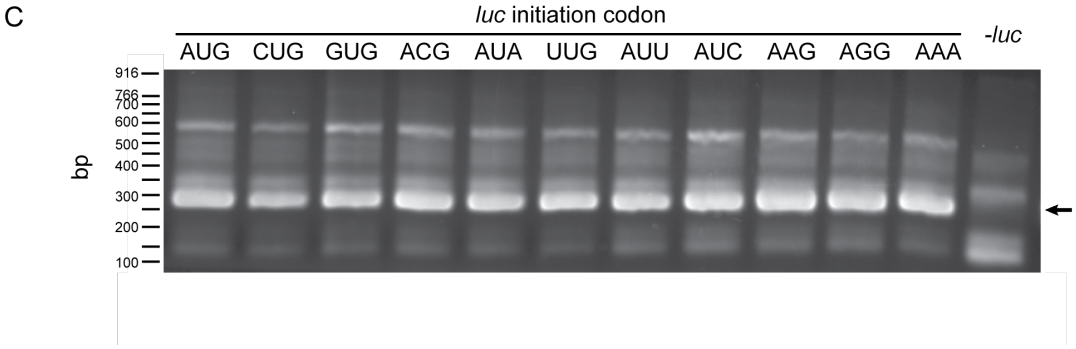
The frequencies for nucleotide occurrence at each position of the *N. crassa* initiation context, which were used to generate the logogram in Figure 3.1, were obtained from the Transterm database (*449*).

Figure 3.1. Site-specific integration and expression of luciferase (LUC) constructs. A. Strategy for placing LUC constructs at the *N. crassa his-3* locus. The *luc* coding sequence is codon-optimized for *N. crassa*. It initiates with an AUG codon, one of the nine NCCs, or an AAA codon. All codons tested were in the same surrounding consensus context as indicated. The LUC coding sequence is preceded by 5' region of the *N. crassa cox-5* gene that has promoter activity, and followed by *cox-5* 3' region, which provides the polyadenylation site to produce reporter transcripts with a *cox-5* 3'UTR. The plasmids used for integration contain a unique *Pci*I site for linearization. The "Integration left flank" on the plasmid contains the distal region of the functional wild-type *his-3* coding sequence and additional downstream genomic sequence; the "Integration right flank" contains additional sequence from the chromosomal region downstream of *his-3*. The positions of the corresponding segments on the *N. crassa* chromosome (Linkage Group I) containing a non-functional *his-3* allele are indicated. LUC coding sequence is orange; *cox-5* sequences are green; Linkage Group I sequences are black; additional plasmid sequences are blue. Left lower panel: frequency logograms of the conservation of the initiation contexts, from -6 to +4, of all predicted ATG-initiated *N. crassa* genes. Letter heights are proportional to the frequency of occurrence of each nucleotide at each position. B. Northern analysis of *luc* and *cox-5* mRNAs [450 ng of poly(A) mRNA/lane] shows *luc* mRNAs containing the indicated *luc* initiation codons are similarly expressed. The primary data for one set of three independent sets of transformants analyzed is shown. Lane 12 was loaded with mRNA from wild type cells lacking *luc*. For quantification, signals representing *luc* mRNA was normalized to signals representing *cox-5* mRNA for all three sets of transformants. Then this ratio for each initiation codon was normalized to the ratio obtained for the AUG initiation codon. For RT-qPCR quantification, cDNA prepared from total RNA from all transformants was analyzed; *luc* mRNA was normalized to *cox-5* or 25S rRNA. Then these ratios for each initiation codon were normalized to the ratios obtained for the AUG initiation codon. C. All *luc*-containing mRNAs are similarly polyadenylated. 3'RACE analysis shows proper polyadenylation at the *cox-5* poly(A) site for the *luc* mRNA. 3'RACE was performed as described in methods with 50 ng poly(A) mRNA template. The major bands of all samples migrate at the position expected from proper polyadenylation (285 bp) (arrow); this was further confirmed by sequencing. The primary data is shown for one set of three independent sets of transformants analyzed.





Lane	1	2	3	4	5	6	7	8	9	10	11	12
Northern *luc/cox-5*	1.0	1.3±.3	1.2±.2	1.3±.4	1.1±.1	1.0±.1	0.7±.2	1.3±.4	0.9±.2	0.9±.4	0.9±.3	
RT-qPCR *luc/cox-5*	1.0	1.1±.3	1.3±.4	0.9±.1	0.8±.4	0.9±.3	1.1±.2	0.7±.3	1.1±.2	1.0±.1	0.7±.6	
RT-qPCR *luc/25S*	1.0	0.7±.3	0.7±.2	0.9±.1	0.8±.4	0.9±.1	0.9±.3	0.7±.2	0.8±.4	0.9±.3	0.8±.1	



3.2.2. Strains and culture conditions

Strains *his-3* (Y234M723) *mat A* strain FGSC 6103 and the *mat A* wild-type (WT) reference sequenced strain FGSC 2489 (74-OR23-1V A) were obtained from the Fungal Genetics Stock Center (FGSC) (*450*).

Firefly luciferase reporters were targeted to the *N. crassa his-3* locus using *his-3* left flank and right flank integration sequences by transformation of FGSC 6103 with *Pci*I-linearized plasmid DNA (pJI301-pJI311). Each plasmid contained the luciferase (*luc*) reporter with an *N. crassa cox-5* promoter and 3'-region (plasmid construction is described below and in Figure 3.2). Transformants were obtained by electroporation (*451*) followed by selection for histidine prototrophy using plates containing Vogel's minimal medium (VM)/0.05% fructose, 0.05% glucose, 2% sorbose (FGS)/2% agar at 30°C. Homokaryons were obtained by microconidiation essentially as described previously (*452*) except that microconidia obtained by filtration were pelleted in an Eppendorf 5415D centrifuge at 12000x*g* for 2 min and, following resuspension, were germinated on VM/FGS/2% agar at 30°C.

Conidia were obtained from cultures in 125 ml flasks containing 25 ml VM/2% sucrose/2% agar (*453*). Cultures were grown at 25°C with 12:12 h Light:Dark cycle for 7 days in a Percival Environmental Chamber (Model I36VL). Conidia were harvested by suspension in VM/2% sucrose and filtration through two layers of cheesecloth. The concentration of conidia was determined using a hemacytometer.

For RNA isolation and preparation of cell extracts to measure LUC activity produced *in vivo*, conidia were inoculated into 25 ml VM/2% sucrose in a 125 ml flask

at a concentration of 10^7 conidia/ml. Conidia were germinated under constant light at 30°C for 6 h with 125 rpm shaking. Germlings were harvested by vacuum filtration onto Whatman 541 filter paper (42.5 mm circle); the pad of cells was washed with 4°C sterile water, cut into approximately 0.1 g pieces with a single-edged razor, transferred to 2-ml screw-cap Eppendorf tubes, quick-frozen in liquid nitrogen and stored at −80°C.

3.2.3. Measurements of LUC cctivity *in vivo*

For measurements of LUC activity *in vivo* using real-time detection of photon emission, conidia were inoculated into 0.15 ml VM/2% sucrose containing 25 nM luciferin (NanoLight Technology, Cat. #306) in a 96-well microtiter plate (Perkin Elmer OptiPlate-96F) and incubated as stationary cultures at 25°C in constant darkness. Light emission was measured with a microplate scintillation and luminescence counter (Topcount NXT, Packard).

For measurements of LUC activity in soluble extracts, *N. crassa* extracts were prepared as described (*309*), quick frozen in aliquots, and stored at -80°C. Protein concentration was determined using the Coomassie Plus (Bradford) Assay Reagent (Thermo Scientific) with BSA as the standard (Albumin Standard, Thermo Scientific) using the microplate procedures (300 µl of the Coomassie Plus Reagent was added to 10 µl of sample). The absorbance at 595 nm was measured by a Victor 3 Multitask plate reader. The typical yield of this method was 10 µg total protein/µl (approximately a yield of 0.1 mg protein/ mg of cells). 10 µl cell extract [diluted to 0.5-1 µg total protein/µl in breaking buffer (*309*)] were further diluted by adding 10 µl 2X passive lysis buffer (Promega) and luciferase activity was measured using a Victor 3 Multitask

plate reader (Perkin Elmer). Firefly luciferase assay reagents were prepared as described (*454*).

3.2.4. Plasmids

The *N. crassa* optimized firefly *luc* coding sequence in plasmid pRMP57 was a gift from Dr. Deborah Bell-Pedersen (R. de Paula and D. Bell-Pedersen, unpublished). Plasmids pJI301-pJI311 (data not shown and Tables 3.1 and 3.3) were used in *in vivo* experiments. Plasmids pJI201-pJI211 and pJI601-pJI606 (data not shown and Tables 3.1, 3.2 and 3.4) were used in *in vitro* assays.

Table 3.1. Vector and oligonucleotides used to introduce different luciferase initiation codons. S and A strands were annealed and cloned into *Bam*HI/*Apa*I sites of pRMP572

Plasmid	Insert		Cloning site	Vector
	Oligo	Sequence (5' – 3')		
pJI001	AGG/S	GATCCTAAGCCACC**AGG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AGG/A	CTTCTTGATGTTCTTGGCGTCCTC **CCT**GGTGGCTTAG		
pJI002	ACG/S	GATCCTAAGCCACC**ACG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	ACG/A	CTTCTTGATGTTCTTGGCGTCCTC **CGT**GGTGGCTTAG		
pJI003	GUG/S	GATCCTAAGCCACC**GTG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	GUG/A	CTTCTTGATGTTCTTGGCGTCCTC **CAC**GGTGGCTTAG		
pJI004	AUG/S	GATCCTAAGCCACC**ATG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AUG/A	CTTCTTGATGTTCTTGGCGTCCTC **CAT**GGTGGCTTAG		
pJI005	AUA/S	GATCCTAAGCCACC**AUA**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AUA/A	CTTCTTGATGTTCTTGGCGTCCTC **TAT**GGTGGCTTAG		
pJI006	AUU/S	GATCCTAAGCCACC**AUU**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AUU/A	CTTCTTGATGTTCTTGGCGTCCTC **AAT**GGTGGCTTAG		
pJI007	AAA/S	GATCCTAAGCCACC**AAA**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AAA/A	CTTCTTGATGTTCTTGGCGTCCTC **TTT**GGTGGCTTAG		
pJI008	UUG/S	GATCCTAAGCCACC**UUG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	UUG/A	CTTCTTGATGTTCTTGGCGTCCTC **CAA**GGTGGCTTAG		
pJI009	CUG/S	GATCCTAAGCCACC**CUG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	CUG/A	CTTCTTGATGTTCTTGGCGTCCTC **CAG**GGTGGCTTAG		

Table 3.1. continued

Plasmid	Insert		Cloning site	Vector
	Oligo	Sequence (5' – 3')		
pJI010	AUC/S	GATCCTAAGCCACC**AUC**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AUC/A	CTTCTTGATGTTCTTGGCGTCCTC **GAT**GGTGGCTTAG		
pJI011	AAG/S	GATCCTAAGCCACC**AAG**GAGGAC GCCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa* I	pRMP 572
	AAG/A	CTTCTTGATGTTCTTGGCGTCCTC **CTT**GGTGGCTTAG		

Table 3.2. Plasmids used in *in vitro* assays.

Plasmid	Insert (initiation codon)	Cloning site	Vector
pJI201	pJI001 (AGG)	*Bam*HI/*Pac*I	pJI101
pJI202	pJI002 (ACG)	*Bam*HI/*Pac*I	pJI101
pJI203	pJI003 (GUG)	*Bam*HI/*Pac*I	pJI101
pJI204	pJI004 (AUG)	*Bam*HI/*Pac*I	pJI101
pJI205	pJI005 (AUA)	*Bam*HI/*Pac*I	pJI101
pJI206	pJI006 (AUU)	*Bam*HI/*Pac*I	pJI101
pJI207	pJI007 (AAA)	*Bam*HI/*Pac*I	pJI101
pJI208	pJI008 (UUG)	*Bam*HI/*Pac*I	pJI101
pJI209	pJI009 (CUG)	*Bam*HI/*Pac*I	pJI101
pJI210	pJI010 (AUC)	*Bam*HI/*Pac*I	pJI101
pJI211	pJI011 (AAG)	*Bam*HI/*Pac*I	pJI101

Table 3.3. Plasmids used in *in vivo* assays.

Plasmid	Insert (initiation codon)	Cloning site	Vector
pJI301	pJI001 (AGG)	*Bam*HI/*Pac*I	pYZ821
pJI302	pJI002 (ACG)	*Bam*HI/*Pac*I	pYZ821
pJI303	pJI003 (GUG)	*Bam*HI/*Pac*I	pYZ821
pJI304	pJI004 (AUG)	*Bam*HI/*Pac*I	pYZ821
pJI305	pJI005 (AUA)	*Bam*HI/*Pac*I	pYZ821
pJI306	pJI006 (AUU)	*Bam*HI/*Pac*I	pYZ821
pJI307	pJI007 (AAA)	*Bam*HI/*Pac*I	pYZ821
pJI308	pJI008 (UUG)	*Bam*HI/*Pac*I	pYZ821
pJI309	pJI009 (CUG)	*Bam*HI/*Pac*I	pYZ821
pJI310	pJI010 (AUC)	*Bam*HI/*Pac*I	pYZ821
pJI311	pJI011 (AAG)	*Bam*HI/*Pac*I	pYZ821

Table 3.4. Vector and oligonucleotides used to introduce different luciferase initiation contexts. S and A strands were annealed and cloned into *Bam*HI/*Apa*I sites of pJI201

Plasmid	Insert		Cloning site	Vector
	Oligo	Sequence (5' – 3')		
pJI601	CCCATG/S	GATCCTAAGCC**C**CCC**ATG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	CCCATG/A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AT**GG**G**GGCTTAG		
pJI602	CCCCTG/S	GATCCTAAGCC**C**CCC**TG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	CCCCTG /A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AG**GG**G**GGCTTAG		
pJI603	CCCTTG/S	GATCCTAAGCC**C**CC**TTG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	CCCTTG/A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AA**GG**G**GGCTTAG		
pJI604	TCCATG/S	GATCCTAAGCC**T**CC**ATG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	TCCATG /A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AT**GG**A**GGCTTAG		
pJI605	TCCCTG/S	GATCCTAAGCC**T**CC**CTG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	TCCCTG/A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AG**GG**A**GGCTTAG		
pJI606	TCCTTG/S	GATCCTAAGCC**T**CC**TTG**GAGGACG CCAAGAACATCAAGAAGGGCC	*Bam*HI/*Apa*I	pJI201
	TCCTTG/A	CTTCTTGATGTTCTTGGCGTCCTC**C** **AA**GG**A**GGCTTAG		

3.2.5. RNA isolation

Cells were kept at -80°C until immediately prior to breakage; cells were not thrawed on ice prior to breaking total RNA from cells was isolated by modification of

the procedure previously described (*309*) using 1 g Zirconia/Silica Beads (Biospec Products, baked at 180°C overnight prior to use) and a Mini-BeadBeater 8 (Biospec Products) with ice-cold 0.5 ml freshly prepared extraction buffer [100 mM Tris-HCl, pH 7.5, 100 mM LiCl, 20 mM DTT in diethylpyrocarbonate (DEPC)-treated water], 0.36 ml phenol, 0.36 ml chloroform, 0.072 ml 10% SDS. Cells (approximately 0.1 g) were extracted by 1 min breakage using the bead-beater at full speed. Tubes were rotated end-over end-for 4 min (New Brunswick TC-6) and then centrifuged (Eppendorf 5415D) at 12,000 x*g* at 4 ºC for 1 min to separate phases. The aqueous phase was removed and was extracted once with 0.69 ml phenol/chloroform, once with 0.69 ml chloroform, precipitated with 0.875 ml ethanol and 0.057 ml 3 M sodium acetate, washed with 80% ethanol twice and dissolved in sterile and filtered DEPC-treated water. RNA concentration was determined using a Nanodrop spectrophotometer. The typical yield of total RNA was 500-1000 µg/ 0.1 g cells. DNA contamination was removed by DNaseI treatment using the Turbo DNA-Free Kit (Ambion). Poly(A) RNAs were purified from 150 µg total RNA using Poly(A) Purist MAG Kit (Ambion) and stored at -80ºC. The yield of poly(A) mRNA was determined using RiboGreen (Invitrogen).

3.2.6. Northens

Northerns were performed as described (*455*) except that dextran sulfate was omitted from the hybridization buffer. [^{32}P]-labeled probes were obtained from the *cox-5* and *luc* templates by random priming and then purified with Mini Quick Spin Columns (Roche).

3.2.7. cDNA synthesis

1.5 µg DNA-free total RNA was used as template to synthesize first-strand cDNA using SuperScript III Reverse Transcriptase (Invitrogen). First, 1.5 µg total RNA, 5 nmol dNTP mix, 125 ng oligo$(dT)_{18}$ and 50 ng random hexamer were mixed together in water to a total volume of 6.3 µl. These components were incubated at 65°C for 5 min and transferred to ice for 2 min. Then the reaction mixture was adjusted to a final volume of 10 µl containing 1X First-Strand Buffer (Invitrogen), 5 mM DTT and 40 units SuperScript III Reverse Transcriptase, and then incubated at 25°C for 5 min and then 50°C for 50 min. The reverse transcriptase was inactivated at 70°C for 15 min, and the cDNA stored at -80°C.

3.2.8. Quantitative PCR (qPCR)

Aliquots of cDNA representing 8 ng and 16 ng of total RNA were used as qPCR templates in triplicate 10 µl reactions containing 1X Platinum Taq PCR Buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 1X ROX Reference Dye (Invitrogen), 1 unit Platinum Taq DNA Polymerase (Invitrogen), 1X SYBR Green I (Invitrogen), 500 nM each primer (chosen using Primer Express software). Thermal cycling performed using an ABI 7300 real-time PCR machine (Applied Biosystems) as follows: 50°C 2 min, 95°C 10 min followed by 40 cycles 95°C 15 s 60°C 1 min.

3.2.9. 3'RACE

First stand cDNA was first synthesized from poly(A) RNA using the Clontech SMART RACE Kit protocol. A 5µl mixture containing 50 ng poly(A) RNA and 1.2 µM

3'–RACE CDS Primer A (oYZ291: 5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$VN-3', N=A, C, G, or T; V=A, G or C) in water was denatured at 65°C for 5 min. The mixture was cooled on ice for 2 min and combined with 2 μl 5X First-Strand Buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 30 mM $MgCl_2$), 0.5 μl 0.1M DTT, 1 μl 10 mM dNTPs and 0.5 μl 200U/μl SuperScript III Reverse Transcriptase (Invitrogen) in a final volume of 10 μl. First strand cDNA was synthesized at 55°C for 1 hr. The reverse transcriptase was inactivated at 70°C for 15 min and the reaction mixture was diluted with 90 μl 10 mM Tris-HCl pH 7.5. Rapid amplification of cDNA Ends (RACE) was performed with 1 μl first strand cDNA as template in a 20 μl reaction containing: 1X TaKaRa PCR Buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 0.2 mM dNTP mix, 0.25 units rTaq DNA polymerase (TaKaRa), 0.5 μM *luc* mRNA specific primer_1 (MSP_1: oYZ365 5'-CGTCTTCGTCGACGAGGTCC-3') and 0.5 μM nested universal primer (NUP: oYZ294 5'-AAGCAGTGGTATCAACGCAGAGT-3'). First-round PCR was accomplished by (i) denaturing at 94 ºC for 30 sec, then (ii) 25 cycles of 94ºC for 30 sec, 55ºC for 30 sec and 72ºC for 60 sec, (iii) a final extension step at 72ºC for 10 min. Then 1 μl of this reaction mixture containing the PCR products was mixed with 99 μl of water and 1 μl of this 100-fold diluted first round PCR product was used as the template for a second round of PCR with different primers using the same reaction components and PCR conditions as the first round. The primers used in the second round were oYZ294 and *luc* mRNA-specific primer_2 (MSP_2: oYZ287 5'-

GGCCAAGAAGGGCGGCAAGATCGCCGTC-3'), which is complementary to the *luc* mRNA 3' distal to the region complementary to oYZ365. After the second round of PCR, 5 µl of the reaction mixture was examined by electrophoresis in a 2% TAE agarose gel. For sequencing, bands of interest were excised from a 2% TAE agarose gel containing 1 mM guanosine to protect DNA from UV damage (*456*). The PCR product representing the equivalent of 30 µl of the reaction mixture was purified from the gel using the QIAquick Gel Extraction Kit (Qiagen), the concentration of recovered DNA determined by A_{260}/A_{280} measurement (Nanodrop) and sequenced with primer 0YZ287 to determine the mRNA 3' sequence including the poly(A) site.

3.2.10. Cell-free translation analysis

Capped and polyadenylated RNAs were transcribed *in vitro* by T7 RNA polymerase from plasmid DNA templates (pJI201-pJI211) that were linearized with *Hin*dIII (*322*). The relative amount of RNA was determined as described (CHAPTER 2). Equal amounts (60 ng) of each *luc* mRNA were used to program *N. crassa* extracts as described (CHAPTER 2) except that Mg^{2+} and K^{+} concentrations were varied as specified in the Results.

For *in vitro* translation using rabbit reticulocyte lysate (Invitrogen), translation reaction mixtures (10 µl) were incubated at 30°C for 30 min, and translation was halted by adding 50 µl of 1.2X passive lysis buffer (Promega). Equal amounts of each *luc* mRNA (6 ng) were used to program extracts.

For cell-free translation in *N. crassa* and in rabbit reticulocyte lysate, 15 µl samples containing 2.5 µl translation reaction and 12.5 µl 1.2X passive lysis buffer were

used to measure luciferase activity with a Victor 3 Multitask plate reader. Firefly luciferase assay reagents were prepared as described (*454*).

Primer extension inhibition (toeprint) assays to map ribosomes at initiation codons were accomplished using ^{32}P-labeled primer oJI105 as described (*320, 431*), except that 0.5 mg/ml cycloheximide or 2 μg/ml harringtonine (Santa Cruz Biotechnology) were added to the reactions as indicated in the results.

3.2.11. Bioinformatica analysis of near-cognate initiated N-terminal extensions in *N. crassa*

The starting point of the analysis was a FASTA file of the *N. crassa* mRNA transcriptome (M. Sachs et. al., unpublished). Genes represented by multiple transcripts were eliminated from the analysis because a pilot indicated they were generating a large number of false positives. As a result only the 6,804 genes represented by unique transcripts were subjected to systematic analysis. In the first step the sequence starting with the annotated AUG start codon of each mRNA and extending 5' to the nearest in-frame stop codon (UAA, UAG or UGA) was extracted. In the next step the coordinates of all in-frame functional near-cognate start codons (i.e. CUG, GUG, UUG, ACG, AUA, AUC and AUU) in these sequences were determined. Finally, the coordinates of those in-frame near-cognate sequences that have A at position -3, or G at position -3 in combination with G at position +4, were extracted. For identifying conserved N-terminal near-cognate initiated extensions to existing *N. crassa* ORFs an approach similar to one previously used to identify such sequences in humans (*457*) was employed. Briefly, all 5' UTRs which in the previous step were shown to have an in-frame near-cognate start

codon in good context at least 25 codons 5' of the annotated AUG start were extracted (representing a total of 1,185 mRNAs). These 5' UTRs were subjected to conceptual translation starting with the annotated AUG start site and extending 5' to the nearest in-frame stop codon. The conceptually translated peptides were then used as query in a BLAST search (*458*) against the genomic sequence of the filamentous fungus *Chaetomium globosum*. It was empirically determined that *C. globosum* was suitably close evolutionarily to allow sufficient sensitivity and at the same time was adequately distant to allow desired selectivity. BLAST hits with expected values of 0.001 or lower were then subjected to manual inspection. In this step all available orthologous sequences from other filamentous fungi (i.e. Pezizomycotina) were obtained and analyzed both for the presence of a homologous N-terminal extension and also for the conservation of the putative near-cognate start codon in good context. For alignment of conserved extensions from Sordariomycetidae species the *N. crassa* sequence (the extension plus 50 to 100 amino acids of the main ORF) were used as queries in a BLAST (tblastn) search in the "whole-genome shotgun contigs" (wgs) databank restricted to Sordariomycetidae. The nucleotide sequences of the positive hits were extracted and then conceptually translated into amino acid sequences. These back-translated sequences were then aligned with ClustalX2, and subjected to minor manual realignment.

3.3. Results

3.3.1. *In vivo luc* reporters containing AUG or near-cognate initiation codons produce similar levels of mRNA

To examine the stringency of start codon selection in *N. crassa in vivo*, plasmids containing *N. crassa* codon-optimized luciferase (*luc*) coding sequences were constructed to enable site-specific integration of *luc* reporter genes at the *N. crassa his-3* locus (Figure 3.1A). In these constructs, either an AUG codon, one of nine NCCs, or an AAA codon was placed at the start of the *luc* coding sequence (Figure 3.1A). The AAA codon serves as a negative control, since it differs by two nucleotides from AUG and is not normally used as an initiation codon. All eleven codons were placed in the most-preferred consensus sequence "GCCAACCxxxG" determined from census of *N. crassa* sequences surrounding the genic start codon (Figure 3.1A); this preferred sequence is nearly identical to the most-preferred consensus sequence for human genes (*436, 459*). All *in vivo* data reported here represent the results of analyses from three independent transformants of each construct.

The *luc* transcripts from *N. crassa* strains containing *luc* genes initiated with different codons were examined by northern blot and RT-qPCR (Figure 3.1B). Northern analysis, using *cox-5* mRNA as internal control, indicated that *luc* mRNA was of the predicted size (1850 nt) and the ratios of *luc* mRNA to *cox-5* mRNA were similar in all eleven strains. The wild-type (WT) strain, which does not contain a *luc* gene, expressed *cox-5* mRNA but not *luc* mRNA as expected (Figure 3.1B lane 12). The similarities in

relative *luc* mRNA levels among strains expressing different genes were confirmed by RT-qPCR, using either *cox-5* mRNA or 25S rRNA for normalizing RNA levels.

To determine that all *luc* transcripts were similarly processed at their 3'ends, poly(A) mRNA was isolated and 3' RACE was performed for all eleven strains containing *luc* mRNAs and for WT. The major 3'RACE product for all *luc*-containing strains, but not the WT strain lacking *luc*, migrated with the size expected (285 bp) for proper polyadenylation in the reporter genes' *cox-5* 3'UTR; this was confirmed directly by sequencing selected 3'RACE products.

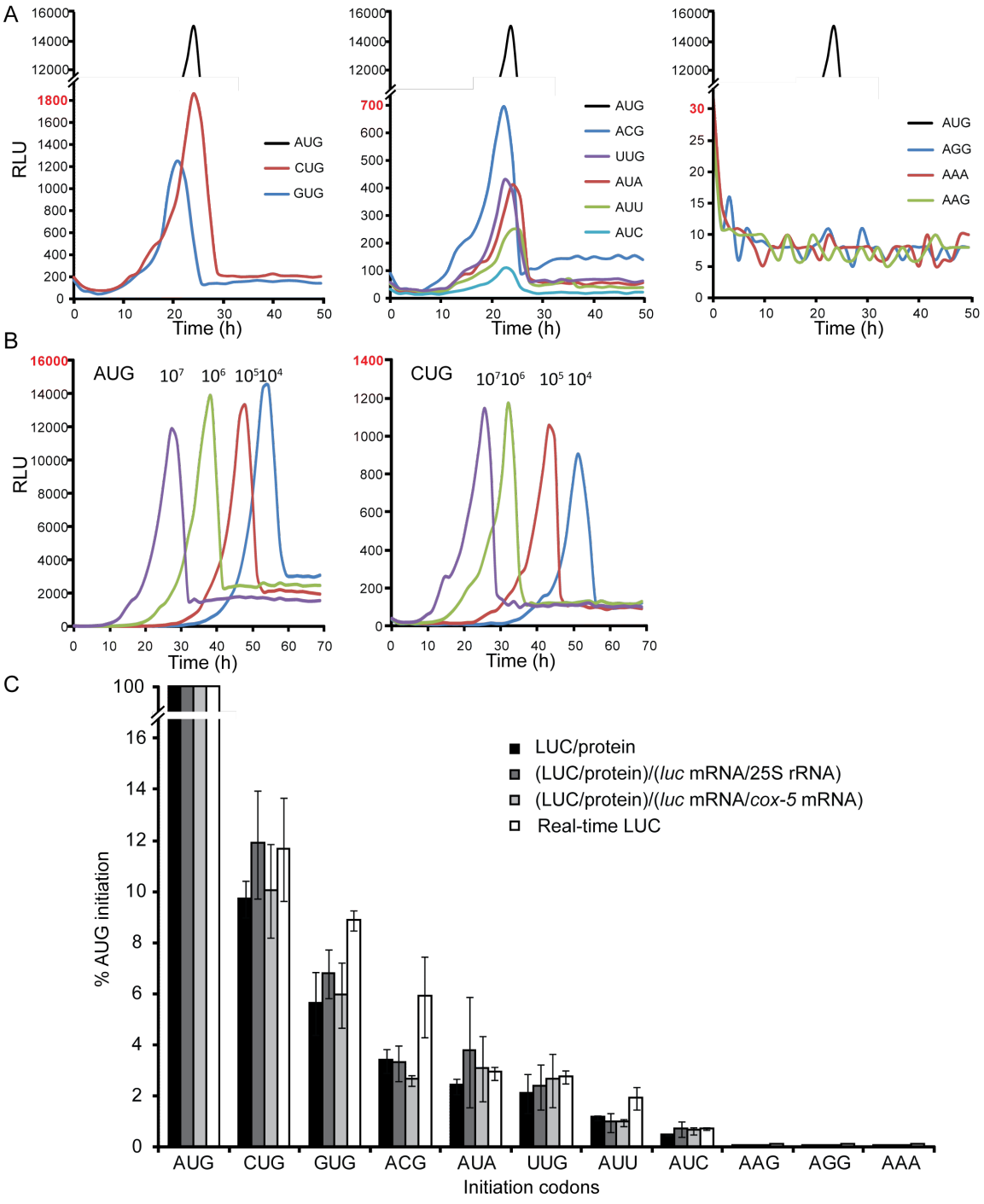
3.3.2. The stringency of selection of NCCs determined by real-time measurements of LUC enzyme activity in *N. crassa* cultures

LUC production *in vivo* was measured using real-time detection of photon emission by growing cells in microtiter plates with luciferin included in the growth medium and imaging with a CCD camera. Different concentrations of conidia (10^7 conidia/ml in Figure 3.2A; 10^4-10^7 conidia/ml in Figure 3.2B) from strains containing AUG-initiated or NCC-initiated *luc* were inoculated into wells containing 0.15 ml medium. During growth at 25°C in constant darkness, LUC activity accumulated, reached a peak value, and then fell-off (Figure 3.2A and B). The magnitude of the peak values depended on the initiation codon of *luc* coding sequence, with an approximately 10-fold higher value for AUG than for CUG. The LUC activity value peaked earlier when more conidia were used for inoculation but the maximum value for each construct did not differ markedly with different amounts of conidia inoculated (Figure 3.2B). We

presume that the falloff in LUC activity arises from depletion of luciferin reagent as a consequence of cell growth but this was not tested.

The real-time assay was used to compare all eleven *luc*-containing strains in parallel inoculated at 10^7 conidia/ml (Figure 3.2A). Eight strains showed the accumulation of LUC enzyme activity and reached peak values between 20 and 24 hours of incubation (Figure 3.2A). The three strains whose *luc* coding sequences initiated with AAG, AGG and AAA did not show detectable LUC production, indicating that these three codons were not used as initiation codon to translate *luc*. Peak values from the eight strains expressing LUC were used to calculate the efficiency of NCC initiation (Figure 3.3C, white bars) by normalizing levels of LUC produced from NCCs to that from the AUG codon. Mean values and standard deviations were derived from three independent experiments, each using one set of independent transformants and performed in triplicate. The real-time detection of LUC activity revealed the following hierarchy: AUG >> CUG (11.6%) > GUG (8.9%) > ACG (5.9%) > AUA (2.9%) ≈ UUG (2.7%) > AUU (1.9%) > AUC (0.7%).

Figure 3.2. Translation initiation efficiency *in vivo* at different initiation codons. A-B. Measurements of LUC levels *in vivo* using real-time detection of photon emission. The lines represent triplicate measurements of one set of three independent sets of transformants. A. Conidia (10^7 conidia/ml) from strains expressing *luc* with initiation codons as indicated were incubated into 0.15 ml of luciferin-containing media in microtiter dish wells and incubated as stationary cultures at 25°C in constant darkness. Light emission was measured with a microplate scintillation and luminescence counter; Relative light units (RLU) are plotted versus hours of incubation time. Left panel: CUG and GUG compared with AUG. Middle panel: ACG, UUG, AUA, AUU and AUC compared with AUG. Right panel: AGG, AAA and AAG compared with AUG. Cultures representing all codons were grown in parallel. B. Different amounts of conidia (10^4-10^7 conidia/ml) from strains expressing LUC were inoculated. Left panel: Cultures expressing AUG-initiated *luc*; Right panel: cultures expressing CUG-initiated *luc*. C. Translation initiation efficiencies of non-AUG codons are calculated relative to the efficiency of the AUG codon using *N. crassa* strains with *luc* reporter genes containing AUG or the indicated codons at the initiation site. White bars: relative real time LUC activities measured *in vivo* using CCD imaging. The mean peak value from triplicate cultures of a given strain inoculated at 10^7 conidia/ml was used for calculation. Black, dark gray, and light gray bars: LUC activities measured from cell extracts normalized in different ways. Black bar, normalized to total extracted protein (determined by Bradford assay); dark gray bar, normalized first to total protein and then calculated relative to the *luc* mRNA/25S rRNA level (determined by qPCR, Figure 3.1B); light gray bar, normalized first to total protein and then calculated relative to *luc* mRNA/*cox-5* mRNA level (determined by qPCR, Figure 3.1B). Mean values and standard deviations for all measurements are derived from three independent experiments, each using one set of independent transformants.


A
RLU
Time (h)
AUG
CUG
GUG
ACG
UUG
AUA
AUU
AUC
AGG
AAA
AAG
B
AUG
CUG
10^7 10^6 10^5 10^4
C
LUC/protein
(LUC/protein)/(*luc* mRNA/25S rRNA)
(LUC/protein)/(*luc* mRNA/*cox-5* mRNA)
Real-time LUC
% AUG initiation
Initiation codons
AUG CUG GUG ACG AUA UUG AUU AUC AAG AGG AAA

3.3.3. The stringency of selection of NCCs *in vivo* determined by measurements of LUC enzyme activity in *N. crassa* extracts

The stringency of start codon selection in *N. crassa* was also analyzed by a measurement of LUC activity in soluble cell-extracts prepared from all eleven *luc*-containing strains (Figure 3.2C). *N. crassa* cells were harvested after 6 hours of germination and a portion of cells was used to make extracts for measuring LUC activity and protein concentration and a portion of cells was used for RNA isolation. LUC enzyme activity was normalized to protein concentration for each extract (LUC/mg protein). The expression of LUC produced from genes containing non-AUG codons was then calculated relative to that from the AUG codon (Figure 3.2C, black bars). To account for (the rather small) differences in *luc* mRNA levels found in these cells (Figure 3.1B), we also calculated differences in the synthesis of luciferase after normalization of the values of LUC/mg protein to relative cellular *luc* mRNA levels (using either 25S rRNA or *cox-5* mRNA as internal RNA controls) (Figure 3.3C, dark-gray and light-gray bars). All of these measurements of relative expression of LUC using soluble extracts (Figure 3.2C, black, dark-gray and light-gray bars) corresponded closely to those obtained by direct measurement of LUC activity in living cells (Figure 3.2C, white bars).

3.3.4. The stringency of selection of NCCs in *N. crassa* cell-free translation extracts

Plasmids were constructed to produce mRNA to test initiation from NCCs in *N. crassa* cell-free translation extracts. The eleven plasmid templates used for *in vitro* studies contain the same *luc* coding and initiation-contexts used for *in vivo* assays.

Synthetic capped and polyadenylated mRNAs produced from linearized templates were used to program translation extracts. Mg^{2+} and K^{+} concentrations can potentially affect the efficiency of translation and/or the stringency of start codon selection (*68, 424*). All combinations of five different Mg^{2+} concentrations and four different K^{+} concentrations were tested in parallel in triplicate in the *N. crassa* cell-free translation system using AUG-initiated *luc* mRNA (Figure 3.3A). The overall production of LUC was highest when 110 mM K^{+} and 2.7 mM Mg^{2+} were added. Next, the same amounts of CUG-initiated *luc* mRNA and AUG-initiated *luc* mRNA were directly compared with the same combinations of Mg^{2+} and K^{+} concentrations to determine the effects of these cations on the stringency of start codon selection (Figure 3.3B). The CUG/AUG ratio was plotted for each condition used. When reaction mixtures contained 110 mM K^{+} and 1.5 mM Mg^{2+}, the efficiency of CUG-initiated translation was 12% of AUG-initiated translation. However, when the concentration of Mg^{2+} increased to 3.1 mM and K^{+} was not changed, the efficiency of CUG-initiated translation increased to 80% of AUG-initiated translation. Three different Mg^{2+} concentrations and a fixed K^{+} concentration (110 mM) were chosen to represent high stringency (1.5 mM Mg^{2+}), intermediate stringency (2.7 mM Mg^{2+}) and low stringency (3.1 mM Mg^{2+}) conditions for subsequent analyses of start codon selection *in vitro*.

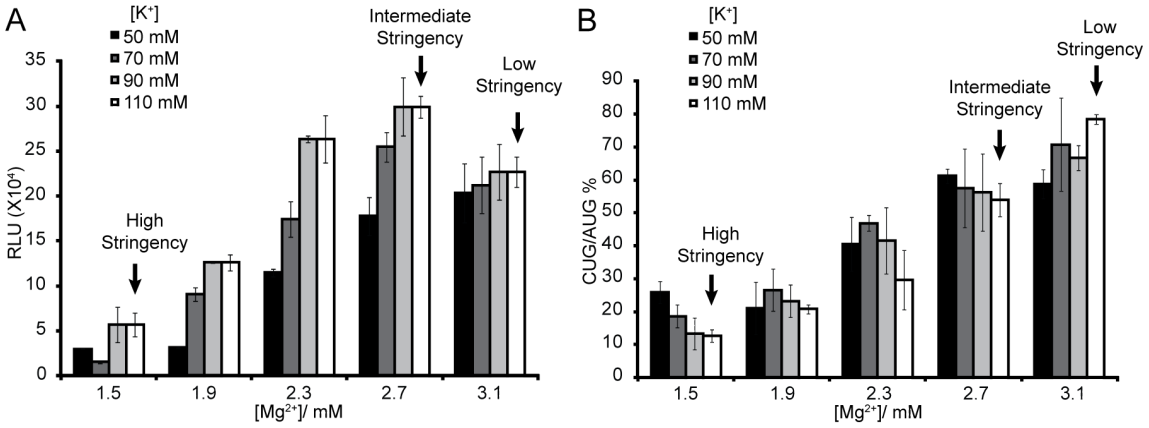


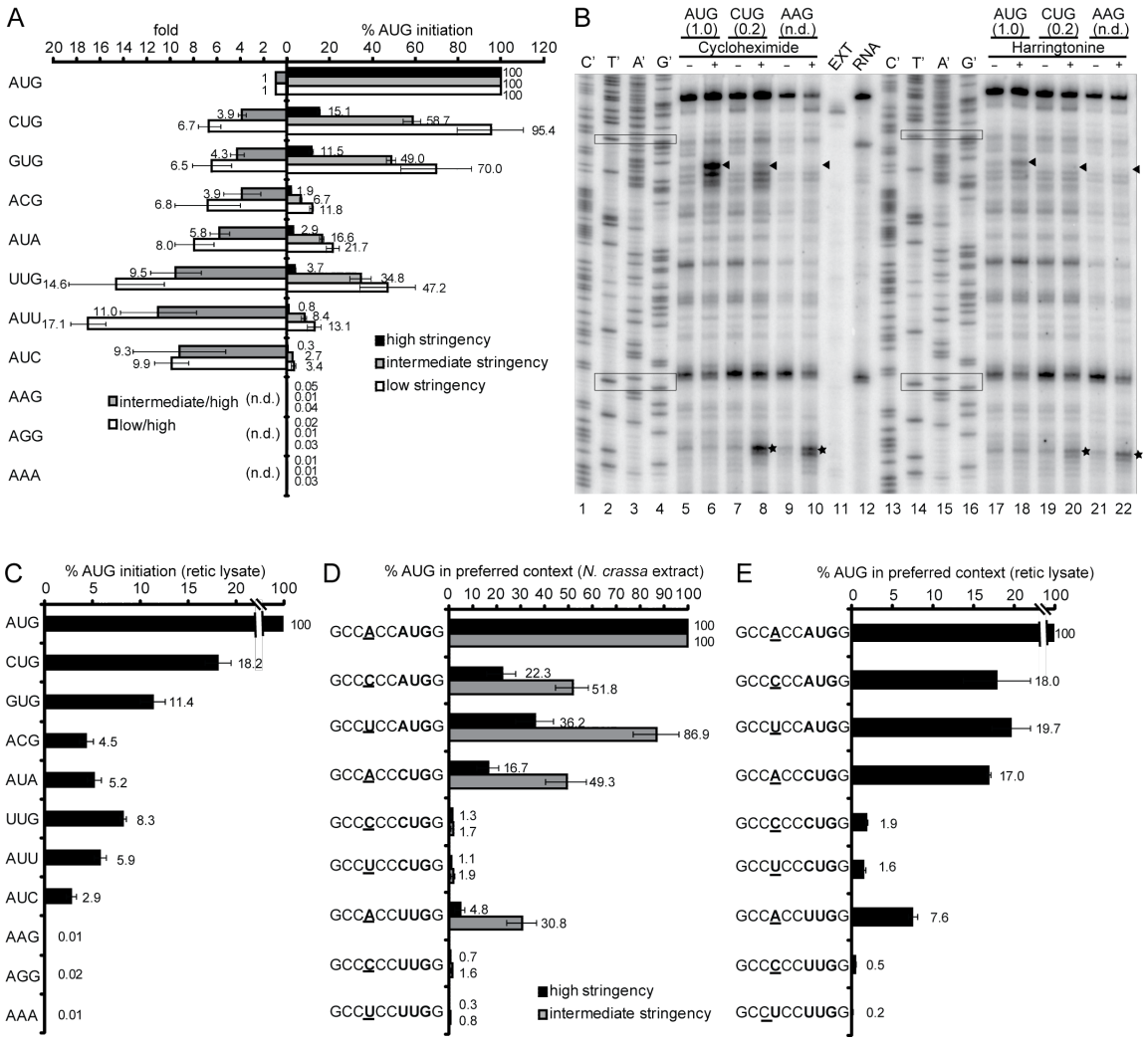
Figure 3.3. Analyses of initiation efficiency using cell-free translation systems. Mean values and standard deviations from three independent experiments, each performed in triplicate, are given. A. Effects of Mg^{2+} and K^{+} concentrations on LUC synthesis from the AUG codon. Capped and polyadenylated *luc* mRNA (60 ng) was used to program *N. crassa* translation reactions (10 µl) with all combinations of five different Mg^{2+} and four different K^{+} concentrations. Mg^{2+} and K^{+} concentrations representing high, intermediate, and low stringency start-codon selection conditions in subsequent experiments are indicated. LUC activity obtained after 30 min incubation at 26°C is given in RLU (relative light units). Low stringency conditions are similar to previously established standard conditions (*315, 322*). B. Effects of Mg^{2+} and K^{+} concentrations on the stringency of start codon selection. Translation reactions were programmed and incubated as described in (A) using either CUG or AUG as the LUC initiation codon. The CUG/AUG ratio was plotted as a function of Mg^{2+} and K^{+} concentrations.

Equal amounts of all eleven mRNAs were tested in parallel in *N. crassa* translation extracts at salt concentrations conferring high, intermediate and low stringency conditions for start codon selection (Figure 3.4A). Similar to the results obtained *in vivo*, the use of AAG, AGG and AAA as start codons was not detected under any condition. While changing [Mg^{2+}] changed the extent of translation from functional NCCs relative to AUG, the relative hierarchy of start codon utilization was not changed. The most efficient NCC was CUG, followed by GUG; codons ACG, AUA, UUG, AUU

and AUC were less efficient as start codons. Under high stringency conditions, the efficiency of utilization of NCCs was generally comparable to those observed *in vivo*.

When conditions changed from high to intermediate stringency, and from high to low stringency, translation from the AUU codon increased 11- and 17.1-fold, respectively; in comparison, ACG initiated *luc* increased 3.8- and 6.8-fold, respectively (Figure 3.4A). This result suggested that different NCCs responded differently to conditions that altered stringency, a phenomenon also observed in human cells (*379*). The biological bases for this differential response are unknown.

Figure 3.4. Analyses of initiation efficiency and stringency of start codon selection using cell-free translation systems. A. Relative initiation efficiency of non-AUG start codons at different Mg^{2+} concentrations (high, intermediate and low stringency conditions). On the right, the black, gray and white bars represent relative initiation at high, intermediate, and low stringency based on LUC activity assays. LUC synthesis from non-AUG codons was calculated relative to synthesis from the AUG codon. On the left, gray and white bars represent the ratios of LUC synthesis between intermediate and high stringency conditions, and low and high stringency conditions, respectively. The ratios for AAG, AGG and AAA are not given because these codons at the *luc* initiation site did not yield detectable LUC. B. Toeprint analysis to assess initiation at AUG, CUG and AAG. *N. crassa* extract was programmed with equal amounts (60 ng) of the indicated *luc* mRNAs. Cycloheximide or harringtonine were omitted (-) or added (+) prior to incubation of translation reactions for 5 min at 26°C. Radiolabeled primer oJI105 was used for primer extension analysis and for sequencing the AUG template (lanes 1-4 and 13-16). The nucleotide complementary to the dideoxynucleotide added to each sequencing reaction is indicated above the corresponding lane. Arrowheads: toeprint products corresponding to ribosomes at the *luc* initiation codon. Signals from CUG were normalized to signals from AUG and the results are shown in parentheses. Values were calculated from two independent experiments (cycloheximide 0.24±0.06 and harringtonine 0.24±0.02). Asterisks: toeprint products corresponding to ribosomes at the first downstream AUG codon within the *luc* coding region. Boxes (top to bottom): *luc* initiation codon, and the first downstream AUG codon. EXT, extract alone; RNA, RNA alone. C. Translation initiation efficiency in rabbit reticulocyte lysate. Equal amounts of each *luc* mRNA (6 ng) were translated in rabbit reticulocyte lysate (10 µl translation reaction incubated for 30 min at 30°C). LUC synthesis from non-AUG codons was calculated relative to synthesis from the AUG codon. D. Relative initiation efficiency of AUG, CUG and UUG in preferred context versus poor contexts in *N. crassa* under high (black bars) and intermediate (gray bars) stringency conditions. E. Relative initiation efficiency of AUG, CUG and UUG in preferred context versus poor contexts in rabbit reticulocyte lysate. LUC synthesis from non-AUG codons and AUG codon in poor contexts was calculated relative to synthesis from the AUG codon in the preferred context. For (A), (C), (D), and (E), mean values and standard deviations from three independent experiments, each performed in triplicate, are given.



We next directly examined the capacity of AUG, CUG and AAG to initiate translation in *N. crassa* extracts under high stringency conditions using the toeprinting assay, which shows the positions of ribosomes engaged in the translation of mRNA (Figure 3.4B). Cycloheximide (CYH), which blocks translation elongation, was added before the translation reaction started to increase the signals from ribosomes at translation initiation sites (*293*). When CYH was added, mRNA encoding AUG-initiated

luc showed a strong toeprinting signal corresponding to that start codon (Figure 3.4B, lane 6). A weaker signal (24% of the signal from AUG) was observed for the CUG codon at the corresponding position (Figure 3.4B, lane 8). For AAG-initiated *luc*, a signal corresponding to ribosomes at the corresponding position was not detected (Figure 3.4B, lane 10), consistent with AAG not serving as an initiation codon. For CUG- and AAG-initiated *luc*, another signal corresponding to ribosomes at the first downstream AUG codon (which is in the 0 frame) within the *luc* coding region was observed (Figure 3.4B, lanes 8 and 10) consistent with ribosomes scanning pass the CUG and AAG codons, and initiating at the first downstream AUG codon. As expected, none of these signals were observed in with extract alone (Figure 3.4B, lane 11) or RNA alone (Figure 3.5B, lane 12). Thus, with respect to the stringency of start codon selection, toeprinting data were similar to the luciferase reporter data both *in vivo* and *in vitro* (Figures 3.2C and 3.4A).

Harringtonine blocks initiation by inhibiting elongation during the first rounds of peptide bond formation following subunit joining, causing ribosomes to accumulate at sites of translation initiation (*460, 461*). We examined AUG, CUG and AAG codons as initiators using harringtonine in the toeprinting assay in parallel with CYH (Figure 3.4B, lanes 17-22). The results were similar to the results obtained for CYH, except that the signals corresponding to ribosomes at initiation codons were weaker than when CYH was used.

3.3.5. The stringency of selection of near-cognate codons in reticulocyte lysate

To examine the stringency of selection of NCCs in a mammalian system, we used rabbit reticulocyte lysate (Invitrogen), which the manufacturer reports to be a high fidelity system with respect to initiation. Similar to the results obtained with *N. crassa*, CUG was the most efficient NCC, followed by GUG. ACG, AUA, UUG, AUU and AUC conferred intermediate translation efficiency and AAG, AGG and AAA yielded no detectable LUC activity (Figure 3.4C).

3.3.6. The effect of altering the initiation context on stringency of start codon selection

To examine the effect of poor initiation context on selection of AUG, CUG and UUG start codons, the nucleotide A at the -3 position, which is the most important upstream position for a preferred context, was mutated to C or U. These poor initiation contexts were compared in parallel to the preferred initiation context in *N. crassa* translation extract under high stringency conditions (Figure 3.4D). As expected, mutating the -3 A to C or U decreased the efficiency of translation initiation from AUG, CUG and UUG codons. The efficiency of LUC synthesis from CUG and UUG decreased more compared to AUG in these poor contexts. Under intermediate stringency conditions, the efficiency of translation initiation increased for AUG in poor contexts and for CUG and UUG in the preferred context. However, translation from CUG and UUG in poor contexts barely improved. In rabbit reticulocyte lysate, similar results were observed (Figure 3.4E). Mutating the -3 A to C or U decreased the efficiency of

translation initiation for CUG and UUG more than for AUG. These results indicated that a preferred context is crucial for efficient translation initiation from NCCs.

3.3.7. Prevalence and relevance of potential initiation at NCCs in *N. crassa*

Experimentally verified or bioinformatically predicted cases of translation initiation at NCCs have been described in animals, plants and fungi, including filamentous fungi (*387, 462, 463*). However, the prevalence of initiation at NCCs in *N. crassa* has not been previously investigated. This is important to consider given the results presented here. When examined systematically, for example by taking advantage of ribosome profiling to identify initiation codons (*384, 464, 465*), initiation at NCCs is seen mostly 5' of the previously annotated AUG start codons. These NCCs either initiate a uORF or an N-terminal extension to the main ORF. Individual examples of conserved non-AUG initiated uORFs are known (*466*), however, a systematic bioinformatics search for such uORFs is complex. Identifying N-terminally extended conserved ORFs initiated with NCCs is a more tractable problem (*457*).

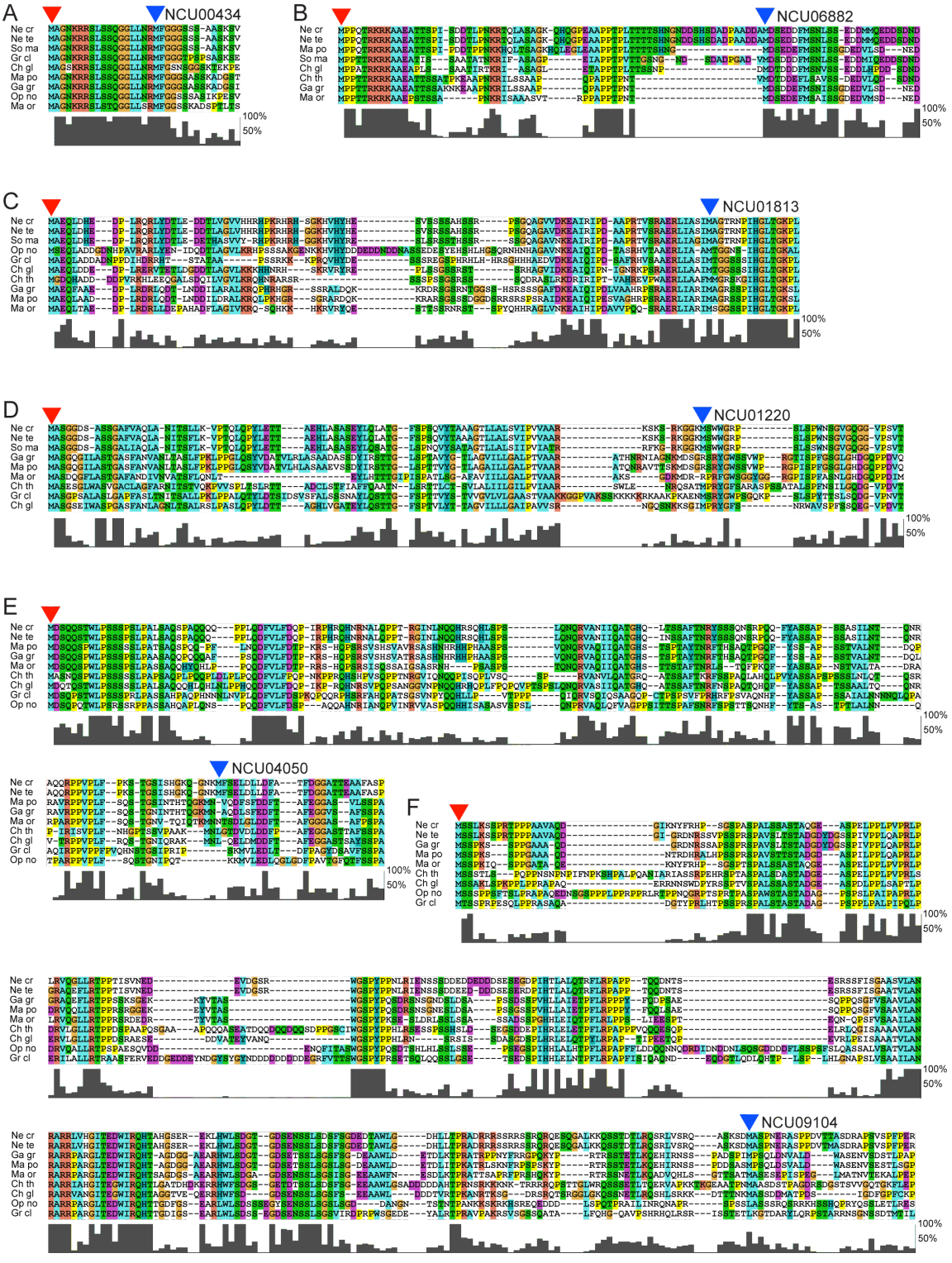
RNA-Seq has been used to determine the *N. crassa* mRNA transcriptome (M. Sachs et. al., unpublished). This data set includes sequences of 10,785 different *N. crassa* mRNA transcripts. Genes represented by more than one transcript were removed from further analysis to avoid redundancy. This yielded a total of 6,804 genes represented by unique transcripts. Using these data, we investigated bioinformatically the prevalence of potential translation initiation at NCCs in *N. crassa*. We defined "optimal" context as having A at position -3, regardless of the identity of position +4, or having G at position -3 in combination with G at position +4. All other contexts were

considered suboptimal. AAG and AGG codons were considered incompatible with initiation. Using these criteria we identified 5,688 NCCs 5' of the annotated starting AUG and in-frame with it, in optimal context and without intervening in-frame stop codons. Because in some instances more than one of these NCCs is present in an mRNA, there are a total of 3,030 (45% of all examined) ORFs with potential N-terminal extensions. Of these, 2,172 in-frame NCCs from 1,185 genes could yield extensions of at least 25 amino acids, with 163 in-frame near codons from 73 genes able to initiate extensions of at least 100 amino acids (data not shown).

We next asked whether cases of physiologically relevant near-cognate initiated N-terminal extensions might be identifiable in this set. We reasoned that many physiologically relevant near-cognate initiated extensions will be conserved, even in more distant relatives of *N. crassa*. We applied a comparative genomics approach similar to one which allowed the identification of conserved near-cognate initiated N-terminal extensions in mammals to search for such extensions in *N. crassa* (Ref. *457, also see Materials and Methods*). We limited the search to extensions that are at least 25 amino acids long. Using a conservative approach, we identified six *N. crassa* genes with conserved near-cognate initiated N-terminal extensions. The genes (and annotations in *N. crassa* genome release NC10 at http://www.broadinstitute.org/annotation/genome/neurospora) are: NCU00434 (protein phosphatase 2C isoform beta); NCU01220 (BAG domain-containing protein); NCU01813 (high affinity glucose transporter); NCU04050 (cross-pathway control protein 1, *cpc-1*); NCU06882 (RING-5); and NCU09104 (hypothetical protein). The

predicted extensions add 18, 88, 91, 153, 71, and 262 N-terminal amino acids, respectively. The identified NCCs are CUG, GUG, AUC, ACG, AUU, and ACG, respectively, and each is highly conserved in the orthologs whose sequences are available. An interesting exception is the N-terminal extension of NCU01813, which is mostly initiated by AUU in homologs, but in a minority of filamentous fungi, for example *Cryphonectria parasitica*, is initiated by a conventional AUG. In all cases the putative N-terminal extensions show phylogenetic conservation in almost all or all Pezizomycotina orthologs for which sequences are available. The conservation of the identified extensions from subclass Sordariomycetidae is shown in Figure 3.5. In all but one case, NCU04050, no out-of-frame AUG codon is located between the conserved in-frame near-cognate and conventional initiation codon of the main ORF. This indicates that both the near-cognate and the conventional initiation codons could potentially be used for initiation, resulting in distinct long and short isoforms of the given protein.

Figure 3.5. Amino acid conservation of the newly predicted non-AUG initiated N-terminal extensions from Pezizomycotina, subclass Sordariomycetidae. The alignments were generated with ClustalX2 using available sequences form Sordariomycetidae. In each case the *N. crassa* sequence is presented on the top row. Amino acids with similar chemical properties are highlighted with the same color. The alignment clusters are as follows: A. NCU00434; B. NCU06882; C. NCU01813; D. NCU01220; E. NCU04050; F. NCU09104. The level of conservation for the amino acids in each column, expressed as a percentage, is indicated on the right side of each alignment. In each case it is assumed that the near cognate start codon is initiated with methionine. This amino acid is indicated by a red arrow. The position of the predominant methionine corresponding to the first in-frame AUG of the main open reading frame is indicated by blue arrow above it. In each case only alignment for the first 15 to 35 amino acids on the main open reading frame is shown. Species name abbreviations (shown on the left of each alignment) are as follows: "Ne cr" = *Neurospora crassa*; "Ne te" = *Neurospora tetrasperma*; "Ma or" = *Magnaporthe oryzae*; "Ma po"= *Magnaporthe poae*; "Ch gl" = *Chaetomium globosum*; "Ch th" = *Chaetomium thermophilum*; "Gr cl" = *Grosmannia clavigera*; "So ma" = *Sordaria macrospora*; "Op no" = *Ophiostoma novo-ulmi*, "Ga gr" = *Gaeumannomyces graminis*.


A
NCU00434
B
NCU06882
C
NCU01813
D
NCU01220
E
NCU04050
F
NCU09104

3.4. Discussion

We examined the stringency of start codon selection in *N. crassa in vivo* and *in vitro* using a codon-optimized firefly luciferase reporter gene. Translation initiation from the nine NCCs in the preferred initiation context was compared in parallel with AUG. CUG and GUG are the most efficient NCCs, followed by ACG, AUA, UUG, AUU and AUC; AAG and AGG are not used for initiation. The efficiency of near-cognate start-codon selection *in vitro* was affected by Mg^{2+} concentration; under the most stringent conditions examined, translation from CUG in a preferred context was approximately 12-15% compared to AUG, similar to the level observed *in vivo*. Additional analyses *in vitro* showed that the preferred -3 nucleotide was important for maintaining translation initiation efficiency, particularly for near-cognate start codons.

There is a good correlation between the efficiency of initiation from NCCs in *N. crassa* and human cells [compare Figure 3.3C with Figure 3B in (Ref. *379*)]. In each case, the efficiency of the NCCs can be grouped in three categories: high – CUG and GUG; intermediate – ACG, AUA, UUG, AUU and AUC; and inactive – AAG and AGG. The difference in initiation efficiency between the most active (CUG) and least active (AUC) NCC in *N. crassa* and human cells is approximately 10-fold. In each case, the context for comparing NCC initiation was GCCACCxxxG. In contrast, one recent analysis in *S. cerevisiae* shows that the efficiencies of the different NCCs do not differ as greatly, with only the least active (also AUC) showing a substantial difference from the others (*377*). The reasons why *N. crassa* more closely resembles mammals than does *S. cerevisiae* in this regard remains to be determined. For at least some initiation factors,

such as eIF3, *N. crassa* more closely resembles mammals than does *S. cerevisiae* (*467*); this might be important for determining initiation efficiency at different start codons. Possibly, additional considerations can impact choice of NCCs as initiation codons. The differences in initiation among NCCs can depend on the specific initiation context (*383*). Earlier studies in *S. cerevisiae* yielded different results with respect to the relative efficiencies of different NCCs (*378, 468*). Thus, some differences may arise because different initiation contexts were used. For example, the respective contexts used in *S. cerevisiae* studies were: CUCUCUxxxC (*377*), GACAAGxxxA (*383*), GAAAAAxxxU (*56, 383*), UGAAUAxxxG (*468*), and CAAAACxxxG (*378*).

Firefly luciferase has been used in *N. crassa* for circadian studies, but has not been used quantitatively. Few reporters have been used quantitatively in this organism. Here we showed that independent transformants containing this reporter integrated as a fusion with *cox-5* 5' promoter and 3' regions showed similar relative levels of expression by both real-time luciferase measurements and assays using cell-free extracts. The *luc* reporter we used can be exchanged between vectors for *in vivo* and *in vitro* expression. The *in vivo* reporter construct is designed to enable exchange of promoter, 5'UTR, and 3'-sequences. This system should thus be adaptable for analyses of other elements that control translation as well as other processes affecting gene expression.

When *N. crassa* extracts are optimized by titrating [Mg^{2+}] for overall translational yield from the AUG-initiated luciferase reporter mRNA (Figure 3.4A), higher concentrations of Mg^{2+} both increase yield and reduce the stringency of start codon selection (Figure 3.4B). It is known that Mg^{2+} levels in cell-free extracts affect

stringency (*68 and references therein*). While the overall hierarchy of NCC utilization did not change, UUG and AUU were used relatively more efficiently under reduced stringency conditions (Figure 3.4A). In cultured human cells, UUG responded similarly to reduced stringency, which was achieved by eIF1 overexpression (*379*). *In vitro* experiments demonstrate that eIF1 is crucial for the stringency of start codon selection, impacting both discrimination between AUG and NCCs and bias between good and poor initiation contexts (*74*). Genetic, biochemical and molecular studies suggest that increased levels of eIF5 can cause eIF1 to be dissociated from the preinitiation complex, decreasing the stringency of start codon selection (*80, 447, 469-471*). We speculate that higher [Mg^{2+}] could increase release of eIF1 from the preinitiation complex *N. crassa* extracts, resulting in increased initiation from NCCs. Increased intracellular Mg^{2+} in *S. cerevisiae* is known to affect the fidelity of translation termination (*472*); effects of Mg^{2+} levels *in vivo* on stringency of initiation have yet to be determined. It is possible that the changes of Mg^{2+} concentration in response to different physiological conditions (*473*) may provide regulatory function by impacting the stringency of start codon selection and therefore gene expression.

In the preferred consensus [GCC(A/G)CCxxxG] that is optimal for initiation (*374, 436*), a purine at position -3 and a G at position +4 are most important (*374*). Mutations that depart from the consensus at position -3 can reduce initiation by more than an order of magnitude in mammalian cells (*374*). The comparison of initiation from AUG, CUG and UUG in the preferred versus poor contexts using *N. crassa* extract indicated that mutation at position -3 reduced translation initiation from each codon. The

reduction of initiation from CUG and UUG was greater than from AUG, indicating that a preferred context is crucial for initiation at non-AUG codons [Figure 3.4D and Refs. (*68, 375, 376*)]. When the stringency of initiation is relaxed by increasing [Mg^{2+}], the efficiency of initiation from these NCCs in poor contexts did not improve although efficiency improved when these codons were in a good context. A possible explanation for this is that NCCs in poor contexts are below a threshold for conversion of the open preinitiation complex (PIC) to a closed PIC, independent of Mg^{2+}. For example, the scanning PIC might fail to recognize NCCs in poor contexts; but when NCCs in a preferred context are recognized, Mg^{2+} levels would increase the likelihood of formation of a closed PIC.

The ramifications of initiation from near-cognate start codons for the biology of eukaryote organisms are beginning to be widely appreciated. Seven of the nine NCCs in a preferred context demonstrably initiated translation in *N. crassa in vivo*, showing that, in this organism, non-AUG codons initiate translation. NCCs can initiate translation of uORFs (*474*) and synthesis of alternative N-terminally extended protein isoforms (*69*). Some proteins are synthesized exclusively from NCCs, including mammalian eIF4G2 (*443*), *P. anserina* IDI-4 (*387*) and *S. cerevisiae* glycyl-tRNA synthetase (*56*). Ribosome profiling has revealed a multitude of eukaryotic near-cognate initiation events (*384, 385, 464*). In *S. cerevisiae*, there is evidence for widespread regulated initiation at non-AUG codons (*385*). Studies using mouse embryonic stem cells provide evidence that NCCs initiate translation of longer and shorter forms of proteins as well as uORFs and that initiation at these codons changes during differentiation (*384*). Our studies, which

demonstrate that NCCs can substitute for AUG to initiate synthesis of a luciferase reporter at substantial levels, provide additional direct experimental support that the traditional view that AUG is the translation initiation codon must be expanded. Evolutionarily speaking, NCC initiation, while generally less efficient than AUG initiation, could serve important roles. Consistent with the idea that high efficiency is not always evolutionarily preferred and low efficiency can be advantageous, recently, it was demonstrated that non-optimal codon preference in the elongation phase is crucial for the synthesis of a functional eukaryotic protein central to establishing a circadian rhythm (*475*). The potential functions of 5' proximal NCCs need to be considered in evaluating the coding capacity of mRNA.

CHAPTER IV

NON-CONONICAL FEATURES IMPACTING THE TRANSLATION OF FILAMENTOUS FUNGAL HOMOLOGS OF *GCN4*

4.1. Introduction

General amino acid control (GAAC) in fungi activates amino acid biosynthetic gene expression in response to amino acid limitation (*290, 476*). This regulatory pathway was originally called cross-pathway control in *Neurospora crassa* and general control in *Saccharomyces cerevisiae* (*303*). *N. crassa cpc-1* and *S. cerevisiae GCN4* specify homologous bZIP transcription factors that were identified using forward genetics based on their function to transcriptionally activate amino acid biosynthetic genes in response to amino acid limitation or imbalance.

Both *N. crassa* CPC1 and yeast GCN4 contain a transcription activation domain, a leucine zipper region involved in dimerization, and a basic DNA binding domain. Genes regulated by CPC1 or GCN4 contain the general control response element (GCRE) sequence TGA(C/G)TCA or a related sequence (*303, 477*). A comparative study of genes regulated by *S. cerevisiae* GCN4, *Candida albicans* GCN4 and *N. crassa* CPC1 revealed that many genes were regulated by these factors in each organism, and that the common core of regulated genes were mostly amino acid biosynthetic genes (*304*). *N. crassa cpc-1*, like *A. nidulans cpcA* and *C. albicans GCN4*, but unlike *S. cerevisiae GCN4*, appears transcriptionally autoregulated in response to amino acid

limitation (*304, 308, 478, 479*), and these fungal *cpc-1* genes contain GCRE sequences in their 5' regions implicated in transcriptional autoregulation.

The translational control of *GCN4* in response to amino acid limitation is the canonical example of how upstream open reading frames (uORFs) mediate regulation of translation via the control of reinitiation (*21, 290, 480*). Four uORFs affect the progression of ribosomes through the 5'-leader of *GCN4* mRNA to regulate *GCN4* expression in response to amino acid limitation. uORF1 acts as a positive regulatory element to facilitate reinitiation, while uORF4 strongly inhibits the translation of *GCN4*. uORF2 and uORF3 play relatively minor roles. *In vivo* experiments (*294*) and cell-free translation assays (*293*) confirm that translation of uORF1 generates reinitiating ribosomes that can start translation either at uORF4 or GCN4 and that translation of uORF4 is incompatible with reinitiation at the GCN4 start codon. Phosphorylation of initiation factor eIF2α by the GCN2 kinase in response to amino acid limitation results in ribosome scanning past uORF2 and increased reinitiation at the GCN4 start codon instead. Homologs of *GCN4* in model fungi described to date contain at least two uORFs and it is generally thought that these perform similar functions to *GCN4* uORF1 and uORF4. *ATF4*, a mammalian homolog of *GCN4*, also contains two uORFs and these are demonstrated to function similarly to *GCN4* uORF1 and uORF4 (*310*).

N. crassa cpc-1 expression is known to be translationally controlled in response to histidine limitation as determined by polysome association analyses (*309*). Also, *N. crassa cpc-3*, the functional homolog of *S. cerevisiae GCN2*, is required for the GAAC response and disruption of *cpc-3* abolishes the increase of CPC1 protein in response to

amino acid starvation (*481*). These studies are consistent with translational regulation of *cpc-1* through its uORFs occurring similarly to *S. cerevisiae GCN4*.

An additional consideration for regulation of *cpc-1* is the recent discovery that the CPC1 reading frame could be extended at its amino terminus if a near cognate non-AUG start codon (NCC) were used to initiate translation (CHAPTER 3). NCCs are known to be used as initiation codons and their significance is actively being explored. In other organisms, the use of NCCs appears to increase in response to conditions that reduces the stringency of start codon selection (*379, 385, 447*).

Here we used *N. crassa* cell-free translation to show that *N. crassa cpc-1* uORF1 and uORF2 act analogously to uORF1 and uORF4 in *S. cerevisiae GCN4* in that ribosomes reinitiate efficiently after translating uORF1 but not uORF2. We also discovered and identified conserved potential N-terminal extensions in the *cpc-1* homologs from a much larger group of fungi including Pezizomycotina and Basidiomycota, but not yeast. Mutiple NCCs, some well conserved and in optimal initiation contexts, which potentially initiate the extension of the *N. crassa cpc-1* homolog were examined both *in vitro* and *in vivo*. The positions of these NCCs indicate that their utilization could bypass the translational inhibitory effect of uORF2. We observed that four of the identified NCCs were used *in vitro* and that, as predicted, their use abrogated the inhibitory effect of uORF2. Evidence for NCC utilization *in vivo* was also obtained. These findings indicate that, in addition to translational control via uORFs, the filamentous fungi possess additional translational mechanisms to produce different CPC1 isoforms.

4.2. Materials and methods

4.2.1. Sequence assembly and analysis

All *cpc-1* sequences were obtained from GenBank using the BLAST algorithm using the *N. crassa cpc-1* sequence as the starting point. In most cases, the sequences were derived from the whole-genome shotgun contigs (WGS) database. WGS sequences were processed manually to predict intron/exon junctions for the mRNA sequence. In a minority of cases, sequences were available from expressed sequence tags (ESTs). EST data were manually assembled into contigs. Additional sequences were obtained from the transcriptome shotgun assembly (TSA) database. All alignments in this study were performed with the ClustalX2 and ClustalW algorithms. Sequences used in this study are available upon request.

4.2.2. Plasmids

The starting point for all constructs was plasmid pPC01 (Z. Wang and M. Sachs, unpublished), which has the 5' UTR of *N. crassa cpc-1* cloned between *Bam*HI and *Xho*I sites (the latter located at the 5' end of the firefly luciferase cassette). First, the sequence, GTCTTC, just upstream of NCC8 ACG codon in the 5' UTR was changed by two-step PCR to a *Sac*I GAGCTC sequence to facilitate making subsequent mutations. This derivative is named pPC100 and is referred to "WT" (i.e. wild-type).

Specifics about plasmids are provided in Table 4.1 and Table 4.2. For *in vitro* experiments, pPC-series plasmids which contained luciferase which was not codon-optimized were used. When two PCR primers are shown in a cell in Table 4.1, one-step

PCR was used to generate inserted regions from corresponding PCR templates. When four PCR primers are shown, two-step PCR was used to generate inserted regions. PCR products and vectors were digested by restriction enzymes, gel-purified and ligated. For pPC176, synthetic complementary oligonucleotides were annealed and ligated to gel-purified vector pPC100 that had been digested with *Age*I and *Xho*I. For *in vivo* assays which contained codon-optimized luciferase, plasmids pJI500, pJI502, pJI501 and pJI576 were made by replacing the small *Bam*HI-*Nsi*I fragment of pJI401 with the small *BamH*I-*Nsi*I fragments of pPC100, pPC102, pPC101 and pPC176, respectively.

Table 4.1. Plasmids used in the study.

Description	Plasmid	PCR Template	PCR primer	Cloning site	Vector
WT	pPC100	pPC01	BGLIIS, SACIA, SACIS, PSTIA	*Bgl*II/*Pst*I	pPC01
UAA	pPC102	pPC100	SACIS; UAA	*Sac*I/*Xho*I	pPC100
ΔuORF1_UAA	pPC187	pPC182	BamHI/S BglII/A	*Bam*HI/*Bgl*II	pPC102
ΔuORF2_UAA	pPC113	pPC100	AUG/ACA; UAA	*Sac*I/*Xho*I	pPC100
ΔuORF1_uORF2_UAA	pPC188	pPC182	BamHI/S BglII/A	*Bam*HI/*Bgl*II	pPC113
ΔuORF2	pPC104	pPC100	AUG/ACA; PSTIA	*Sac*I/*Pst*I	pPC100
ΔmORF	pPC101	pPC100	SACIS; mAUG/ACA	*Sac*I/*Xho*I	pPC100
ΔmORF_UAA	pPC176	AgeI/XhoI_UAA_mORF/S AgeI/XhoI_UAA_mORF/A		*Age*I/*Xho*I	pPC100
Δ8	pPC103	pPC100	ACG/ACA; PSTIA	*Sac*I/*Pst*I	pPC100
Δ5_UAA	pPC175	pPC102	Ppc1/S AUU1UUU/A AUU1UUU/S SACIA	*Bam*HI/*Bgl*II	pPC102
Δ6_UAA	pPC117	pPC100	CUG/CUC; UAA	*Bgl*II/*Xho*I	pPC100
Δ7_UAA	pPC116	pPC100	BGLIIS; AUU/UUU	*Bgl*II/*Sac*I	pPC102
Δ8_UAA	pPC114	pPC100	ACG/ACA; UAA	*Sac*I/*Xho*I	pPC100
8AUG_UAA	pPC115	pPC100	ACG/ATG; UAA	*Sac*I/*Xho*I	pPC100
ΔuORF1_Δ5_UAA	pPC179	pPC175	BamHI/S uORF1/A uORF1/S BglII/A	*Bam*HI/*Bgl*II	pPC102
ΔuORF1_Δ6_UAA	pPC180	pPC102	BamHI/S uORF1/A uORF1/S BglII/A	*Bam*HI/*Bgl*II	pPC117
ΔuORF1_Δ7_UAA	pPC181	pPC102	BamHI/S uORF1/A uORF1/S BglII/A	*Bam*HI/*Bgl*II	pPC116
ΔuORF1_Δ8_UAA	pPC182	pPC102	BamHI/S uORF1/A uORF1/S BglII/A	*Bam*HI/*Bgl*II	pPC114
ΔuORF1_8_UAA	pPC183	pPC102	BamHI/S uORF1/A uORF1/S BglII/A	*Bam*HI/*Bgl*II	pPC115
ΔuORF1_Δ5678_UAA	pPC184	pPC114	CTG/CTC_ATT2/TTT/S XhoI/A	*Bgl*II/*Xho*I	pPC179
ΔuORF1_Δ5678	pPC185	pPC103	CTG/CTC_ATT2/TTT/S XhoI/A	*Bgl*II/*Xho*I	pPC179
Δ5678	pPC186	pPC103	CTG/CTC_ATT2/TTT/S XhoI/A	*Bgl*II/*Xho*I	pPC175
In vivo WT	pJI500	pPC100		*Bam*HI/*Nsi*I	pJI401
In vivo UAA	pJI502	pPC102		*Bam*HI/*Nsi*I	pJI401

Table 4.1. continued

Description	Plasmid	PCR Template	PCR primer	Cloning site	Vector
In vivo ΔmORF	pJI501	pPC101		*Bam*HI/*Nsi*I	pJI401
In vivo ΔmORF_UAA	pJI576	pPC176		*Bam*HI/*Nsi*I	pJI401
Vector	pJI401	pJI304	BamHI_LUC/S BstBI/A BstBI/S NsiI/A	*Bam*HI/*Nsi*I	pJI400
Vector	pJI400	Remove *BstB*I site in pJI304 by *BstB*I digestion, followed by end-repire and self-ligation			

Table 4.2. Primers used in the study.

Oligo	Sequence (5' – 3')
BGLIIS	CCCTTCAGCTCTTCCTTCACAGGT
SACIS	GCCAAGACTTTCCAGAGCTCACCACGGATTCCC
SACIA	GGGAATCCGTGGTGAGCTCTGGAAAGTCTTGGC
PSTIA	GCCTGGATAATGTTTGCAACTCGCTG
ACG/ACA	TCCAGAGCTCACCACAGATTCCCAAC
AUG/ACA	TCCAGAGCTCACCACGGATTCCCAACAGTCAACACAGCT TCCCTCC
mAUG/ACA	GTGACTCGAGTGAATGTCTTGTTGCCCT
UAA	GTGACTCGAGTGAACATCTTGTTGCCTTACTTTCCGTGCG
ACG/ATG	TCCAGAGCTCACCATGGATTCCCAAC
CUG/CUC	GGTAGATCTCAACTTCAGCACCCAG
AUU/UUU	GGTGAGCTCTGGAAAGTCTTGGCCGTGAAAGGTGGCGG
BamHI/S	TATAGATCGGATCCTTCCTTTCTCTTCTCTG
BglII/A	GAAGTTCAGATCTACCTGTGAAGGAAGAG
uORF1/S	ATCCATCAAGATGCGTTAAATCGCTCCCA
uORF1/A	TGGGAGCGATTTAACGCATCTTGATGGATGCTTC
CTG/CTC_ATT 2/TTT/S	ACAGGTAGATCTGAACTTCAGCACCCAGCTTCGTAGCTCG CGCTCAAGTTCTCTTACCCCCCACCGCCACCATTCACGGC CA
XhoI/A	TCGAGTGAACATCTTGTTGCCCTGCTTTCCGTGCGAAATA CTA
AgeI/XhoI_UA A_mORF/S	CCGGTAGTATTTCGCACGGAAAGTAAGGCAACAAGACAT TCAC

Table 4.2. continued

Oligo	Sequence (5' – 3')
AgeI/XhoI_UAA_mORF/A	TCGAGTGAATGTCTTGTTGCCTTACTTTCCGTGCGAAATACTA
BamHI_LUC/S	AACTAGTGGATCCTAAGCCACCAT
BstBI/S	CGAGTACTTCGAAATGTCCGTCC
BstBI/A	GGACGGACATTTCGAAGTACTCG
NsiI/A	AATGATAATGCATTGTCAGCTGTACAGTATTTAC
Ppc1/S	GGTGCATGCTAATACGACTCACTATAG
AUU1UUU/S	AATCGTCAACTTTAAACAATTTTACGTTC
AUU1UUU/A	AAATTGTTTAAAGTTGACGATTACCGAAC
SACIA	GGGAATCCGTGGTGAGCTCTGGAAAGTCTTGGC

4.2.3. RNA synthesis and cell-free translation

Capped and polyadenylated RNAs were transcribed *in vitro* by T7 RNA polymerase from plasmid DNA templates that were linearized with *Eco*RI and the relative amounts of RNA determined as described (CHAPTER 3). *In vitro* translation and gel analysis for visualizing [^{35}Met]-labeled proteins using *N. crassa* extracts and wheat germ extract was accomplished as described (CHAPTER 3), except that 10 µl of translation reactions were incubated for 30 min at 25˚C and samples were mixed with 10 µl 2X NuPAGE LDS Sample Buffer (Invitrogen) and put on ice to stop reactions. *In vitro* translation for luciferase activity assays using *N. crassa* extracts was accomplished as described (CHAPTER 3) using 6 ng of each mRNA to program extracts. Primer extension inhibition (toeprint) assays were accomplished using ^{32}P-labeled primers CPC101 and ZW4 as described (CHAPTER 3) except that 0.5 mg/ml cycloheximide were added to the reactions as indicated in the Results.

4.2.4. Strains, culture conditions, and *in vivo* measurements

Strain FGSC 6103 (*his-3* (Y234M723) *mat A*) and the wild type (WT) reference strain FGSC 2489 (74-OR23-1V *mat A*) were obtained from the Fungal Genetics Stock Center (FGSC) (*450*). Targeting of firefly luciferase reporters to the *N. crassa his-3* locus by transformation of FGSC 6103 with *Pci*I-linearized plasmid DNA (pJI301-pJI311), culture conditions, and conditions for luciferase assays, were as described (CHAPTER 3). Total RNA was prepared from cells, cDNA prepared, and RT-qPCR performed as described (CHAPTER 3).

4.3. Results

While studying the regulation of *cpc-1* by its uORFs in *N. crassa* extracts, using a construct in which the wild-type 5' UTR of *cpc-1* was fused with the open reading frame of firefly luciferase, we observed a band of predicted size and a band ~20 kDa larger than predicted (Figure 4.1A). This prompted a more careful examination of the mRNA 5'UTR sequence. We found the CPC1 reading frame extended all the way upstream (Figure 4.1B), without any in-frame stop codons, to the major mapped transcription initiation site, located 703 nt 5' of the predicted AUG for the main open reading frame (designated mAUG and mORF, respectively). Published results reporting the presence of putative N-terminal extentions in *N. crassa cpc-1* postdate these findings (CHAPTER 3). We next compiled partial or complete sequences of *cpc-1* homologs from 108 Pezizomycotina species. 100 of these sequences that include the region spanning from uORF1 to a position downstream of the mORF AUG were analyzed

further. All homologs contain two AUG-initiated uORFs, with uORF1 spanning 3 to 6 codons and uORF2 spanning 35 to 70 codons, including their stop codons. Surprisingly, in all cases, the reading frame for CPC1 could be substantially N-terminally extended without encountering an in-frame stop. We note that some recent automated annotations of CPC1 homologs include this N-terminal extension but do not resolve how initiation occurs. The presence of this feature in both Sordariomycetes and Eurotiomycetes suggests that it was present in their last common ancestor and possibly earlier; the last common ancestor of all Pezizomycotina is estimated to have lived at least 320 million years ago (*482*). The shortest extension of the CPC1 ORF without encountering an in-frame stop codon is 160 codons (*Leptosphaeria maculans*).

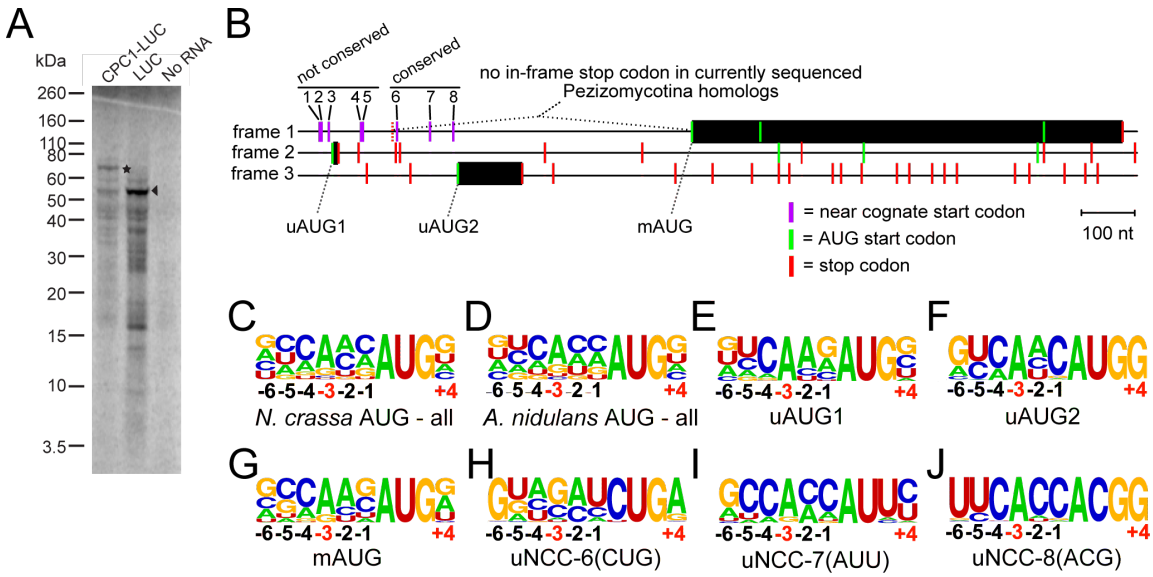


Figure 4.1. N-terminal extensions of Pezizomycotina CPC1. (A) cpc-1-luc mRNA produces a larger product *in vitro* than *luc* mRNA. mRNA templates were used to program *N. crassa in vitro* translation extracts and [^{35}S]Met-labeled products were analyzed on a 12% NuPAGE gel. Translation of mRNA for *N. crassa cpc-1-luc* and *luc* were compared to a no mRNA control. The positions of full-length firefly LUC (arrowhead) and the larger CPC1-LUC product (bracket) are indicated. (B) Schematic diagram of the *N. crassa cpc-1* mRNA. Each reading frame is on a separate line. Frame 1 specifies CPC1 (black rectangle). uORF1and uORF2 (blue rectangles) initiate from uAUG1, and uAUG2, respectively, in other reading frames. AUG codons are indicated by green bars and stop codons by red bars. NCCs in frame 1 upstream of uORF2 are indicated by magenta bars and are numbered 1-8. The approximate position of the 3'-most stop codon upstream of uORF2 and in-frame with the main ORF (present in *Cordyceps bassiana*) is indicated (dashed red bar). The number of sequences used for comparisons is shown in parentheses. Features are drawn to scale. The nucleotide sequence of the *N. crassa cpc-1* 5'UTR is given in Spplemental Figure. 7. (C-J) Frequency logograms of the conservation of the initiation contexts, from -6 to +4, of: (C) all *N. crassa* genes initiated with AUG; (D) all *A. nidulans* genes initiated with AUG; (E) Pezizomycotina *cpc-1* uAUG1; (F) Pezizomycotina *cpc-1* uAUG2; (G) Pezizomycotina *cpc-1* mAUG; (H) Pezizomycotina NCC 6 (CUG); (I) Pezizomycotina NCC 7 (AUU); (J) Pezizomycotina NCC 8 (ACG). Letter heights are proportional to the frequency of occurrence of each nucleotide at each position. The crucial positions -3 and +4 are indicated in red underneath the frequency plots.

Based on the mechanism of translational control of *S. cerevisiae GCN4*, control of *cpc-1* would involve ribosomes initiating at uORF1 and reinitiating at uORF2 under amino acid sufficient conditions. When eIF2α phosphorylation levels increase in response to amino acid limitation, ribosomes would reinitiate at the downstream *cpc-1* mAUG instead of uORF2. Remarkably, without exception in the Pezizomycotina, there is no stop codon in the reading frame of the mORF between the uORF2 AUG and the mAUG. The in-frame stop codon closest to uAUG2 (*Cordyceps bassiana*) is 101 nt upstream of it. Thus, the potential amino terminal extensions of CPC1 end upstream of uORF2 (Figure 4.1B).

Initiation upstream of the uORF2 AUG could produce N-terminally extended isoforms of CPC1, whose synthesis would not be subject to inhibition by uORF2. We searched for potential start codons in this region of *N. crassa cpc-1* mRNA that were in-frame with the predicted mAUG. Eight NCCs fulfilling these criteria were identified - three AUC (NCCs 1, 3 and 4), two ACG (NCCs 2 and 8), two AUU (NCCs 5 and 7) and one CUG codon (NCC 6) (Figure 4.1B). We next searched for potential NCCs in similar regions of *cpc-1* transcripts from all Pezizomycotina (Figure 4.2) and compared their conservation. Three NCCs in *N. crassa* showed particularly deep conservation – the closest to the CPC1 AUG, an ACG (NCC 8), is perfectly conserved in 98 of 100 species; the next closest, an AUU (NCC 7), is conserved in 74 of 100 (as AUU in 64 and AUC or AUA in 10); the next closest, a CUG (NCC 6), is conserved in 77 of 100 (as CUG in 73 and UUG in 4) (Figure 4.2 and data not shown). Notably, in no case was there an in-frame stop codon between these three conserved NCCs and the mAUG. The three

conserved NCCs also show a clear pattern of fungal branch-specific distribution: the minority of homologs lacking both AUU and CUG NCCs clustered separately from the other homologs (Figure 4.2). The other five NCCs showed only sporadic conservation and were found only in species that were more closely related to *N. crassa*. None of the *N. crassa* NCCs appeared conserved in the two most distant Pezizomycotina, *A. oligospora* and *T. melanosporum*, (Figure 4.2). However, even these two species' CPC1 homologs contain multiple NCCs upstream of uORF2 and in-frame with the mORF – 6 NCCs in *A. oligospora* and 7 NCCs in *T. melanosporum* (data not shown).

Figure 4.2. Phylogeny of Pezizomycotina based on CPC1 amino acid sequences and conservation of NCCs 6, 7 and 8. The unrooted phylogenetic tree is based on the amino acid sequences (from the 5' end to the mORF stop codon when no in-frame stop codon is present, or from the stop codon in the 5' region to the mORF stop codon) of the *cpc-1* homologs from the Pezizomycotina species indicated on the right of each branch. The tree was generated using ClustalX. Bootstrap values are given for key nodes. The scale represents the divergence rate at each residue. Species names are coded with respect to which NCCs are identical to *N. crassa*: black, NCCs 6, 7 and 8; blue: 7 and 8; green: 6 and 8; red: 8; orange: no identical NCCs.


0.05
1000
994
940
994
974
911
975
913
1000
992
994
999
990
978
998
1000
941
477
1000
Microsporum_canis
Arthroderma_gypseum
Arthroderma_benhamiae
Trichophyton_tonsurans
Trichophyton_verrucosum
Trichophyton_rubrum
Thermomyces_lanuginosus
Talaromyces_stipitatus
Penicillium_marneffei
Penicillium_digitatum
Penicillium_chrysogenum
Aspergillus_nidulans
Aspergillus_kawachii
Aspergillus_niger
Aspergillus_clavatus
Neosartorya_fischeri
Aspergillus_fumigatus
Aspergillus_terreus
Aspergillus_sojae
Aspergillus_oryzae
Coccidioides_immitis
Coccidioides_posadasii
Paracoccidioides_brasiliensis
Ajellomyces_dermatitidis
Histoplasma_capsulatum
Ajellomyces_capsulatus
Verticillium_dahliae
Colletotrichum_gloeosporioides
Colletotrichum_higginsianum
Glomerella_graminicola
Claviceps_fusiformis
Claviceps_paspali
Aciculosporium_take
Metarhizium_acridum
Metarhizium_anisopliae
Periglandula_ipomoeae
Neotyphodium_gansuense
Epichloe_festucae
Epichloe_typhina
Epichloe_glyceriae
Epichloe_brachyelytri
Epichloe_amarillans
Nectria_haematococca
Fusarium_pseudograminearum
Gibberella_zeae
Gibberella_fujikuroi
Gibberella_moniliformis
Fusarium_oxysporum
Cordyceps_militaris
Beauveria_bassiana
Cordyceps_bassiana
Trichoderma_atroviride
Hypocrea_jecorina
Trichoderma_reesei
Hypocrea_lixii
Hypocrea_virens
Grosmannia_clavigera
Ophiostoma_novo-ulmi
Ophiostoma_piliferum
Daldinia_eschscholzii
Magnaporthe_grisea
Magnaporthe_poae
Gaeumannomyces_graminis
Chaetomium_thermophilum
Podospora_anserina
Chaetomium_globosum
Myceliophthora_thermophila
Thielavia_terrestris
Sordaria_macrospora
Neurospora_tetrasperma
Neurospora_crassa
Marssonina_brunnea
Glarea_lozoyensis
Ascocoryne_sarcoides
Amorphotheca_resinae
Sclerotinia_homoeocarpa
Botryotinia_fuckeliana
Geomyces_destructans
Geomyces_pannorum
Cryphonectria_parasitica
Macrophomina_phaseolina
Leptosphaeria_maculans
Phaeosphaeria_nodorum
Cochliobolus_heterostrophus
Alternaria_arborescens
Pyrenophora_teres
Pyrenophora_tritici-repentis
Cladosporium_sphaerospermum
Mycosphaerella_populorum
Mycosphaerella_fijiensis
Passalora_fulva
Septoria_passerinii
Mycosphaerella_graminicola
Tuber_melanosporum
Arthrobotrys_oligospora

We next examined the conservation of the initiation contexts for the three conserved NCCs and for the uORF1 AUG (uAUG1), uORF2 AUG (uAUG2) and mAUG. The preferred initiation context in *N. crassa* (Figure 4.1C), which is considered optimal, is similar to the preferred context in the relatively distant *A. fumigatus* (*459*) and *A. nidulans* (as we tabulated as shown in Figure 4.1D). uAUG1 and uAUG2 are in conserved optimal contexts (Figures 4.1E and F and data not shown), consistent with their presumed roles in regulating CPC1 translation through controlling reinitiation. The conservation of the context of the mAUG is weaker but the consensus is still near optimal (Figure 4.1G). Of the three conserved NCCs, NCC8, which is closest to the mAUG and is the most conserved, also showed the highest context conservation (Figure 4.1J and data not shown). The consensus initiation context of NCC8 in examined species is in a nearly optimal context for initiation (nucleotides -4, -3, -1 and +4 match the consensus). The most important nucleotides, A at position -3 and the G at position +4, are perfectly conserved in all Pezizomycotina species that have NCC8. Lower context conservation is observed for NCC7 and NCC6 (Figures 4. 1H and I) although their consensus initiation contexts remain near-optimal.

To investigate the effects of uORF1, uORF2 and upstream NCCs on the translation of *N. crassa* CPC1 in cell-free extracts, the 5'UTR of *cpc-1*, including the first two codons of the mORF, was fused in frame to firefly luciferase (*cpc1-luc*, designated as WT, Figure 4.3). The functions of initiation codons identified by bioinformatics approaches were tested by mutational analyses of this construct. A UAA mutation was introduced in-frame 12-nt upstream of mAUG to terminate translation and

therfore truncate translation products that initiated from upstream NCCs (designated as UAA). The start codon of uORF1 was mutated to AAA (ΔuORF1), that of uORF2 to ACA (ΔuORF2), and that of the mORF to CTC (ΔmAUG), to eliminate their initiation.

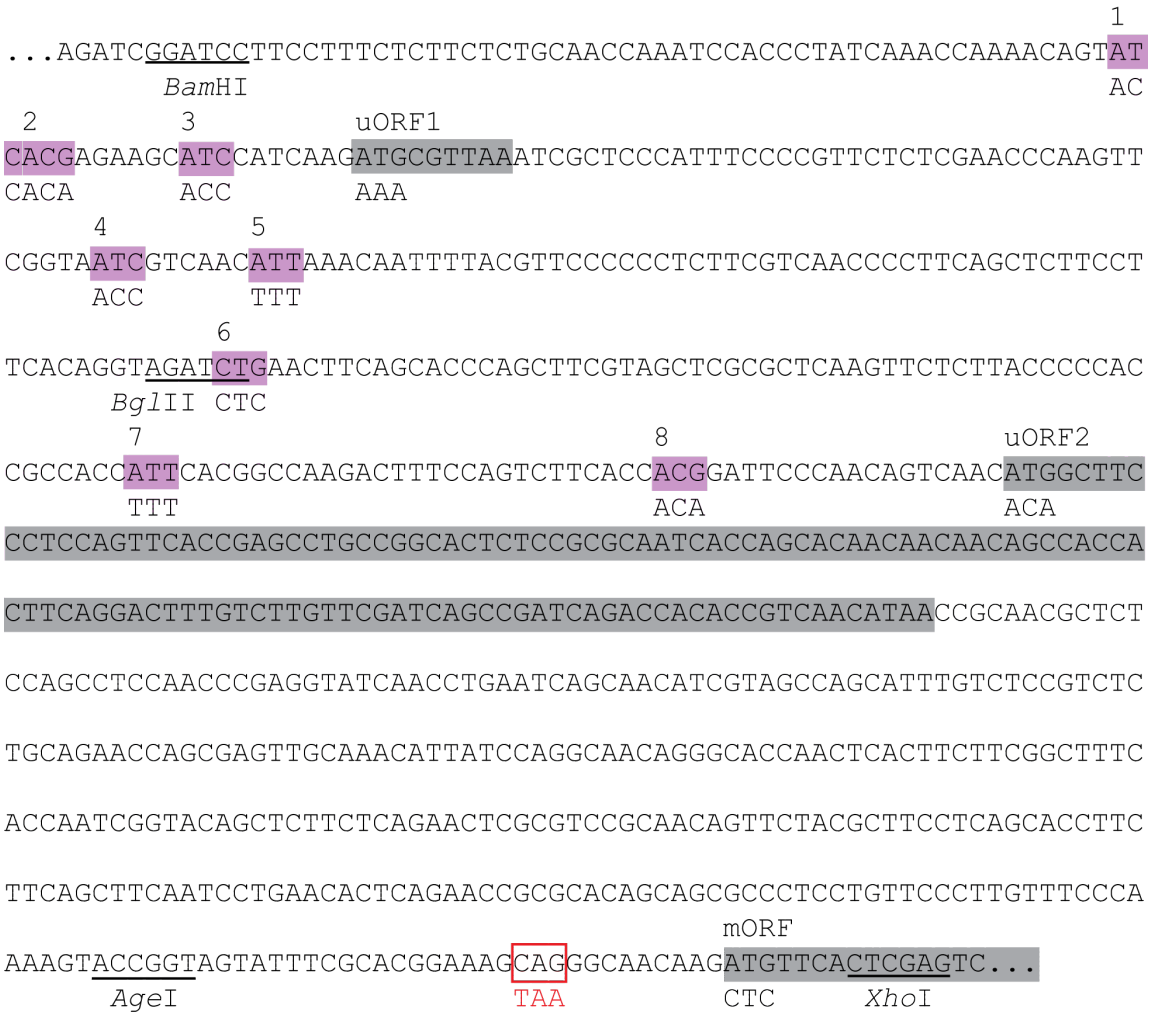


Figure 4.3. The sequences of *cpc1-luc* fusion construct used for *in vitro* experiments. Mutations and selected unique restriction sites are shown below the sequence. uORF1, uORF2 and the beginning of the main ORF (mORF) are shaded gray. NCCs are shaded magenta and are numbered 1-8. The position of the introduced TAA mutation that terminates the translation from NCCs is indicated by a red box. Luciferase was placed in-frame with CPC1 at the *Xho*I site.

The functions of uORF1 and uORF2 were examined by mutating their start codons separately or together. First, we examined these mutations in constructs containing the UAA mutation to look specifically at luciferase synthesis from the mAUG. Luciferase synthesis was measured by enzyme activity assay and by [^{35}S]-Met labeling (Figure 4.4). Compared to a construct containing both uORFs, the ΔuORF1 mutation diminished translation of the mORF as indicated by a reduced level of luciferase activity (15%) and decreased production of [^{35}S]-Met labeled polypeptides (compare lanes 1 and 2, Figure 4.4). The ΔuORF2 mutation increased translation from the mAUG approximately 2.9 fold (compare lanes 1 and 3, Figure 4.4). For the ΔuORF1 ΔuORF2 double mutant, the synthesis of luciferase increased (compare lanes 1 and 4, Figure 4.4), but this increase was less than for ΔuORF2 alone. These data suggest reinitiation occurs after translation of uORF1, that translation of uORF2 is inhibitory, and that a fraction of ribosomes that translate uORF1 reinitiate at uORF2. In the absence of uORF1 and uORF2, synthesis of luciferase is lower than in the absence of uORF2 alone. This could be explained if the NCCs are used more efficiently in the absence of uORF1 (see below).

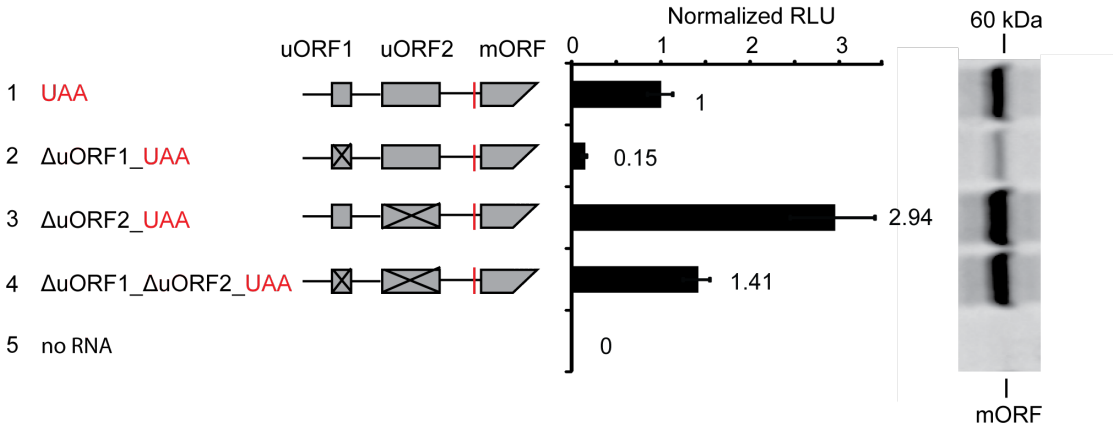


Figure 4.4. The effects of eliminating *cpc-1* uORF1 and uORF2 on the translation from the mAUG in *N. crassa* cell-free extracts. Constructs (numbered 1-4) contained the UAA stop codon (red bar) to eliminate translation from upstream in-frame NCCs, and the indicated mutations to uORF start codons to eliminate initiation from them (uORF1 AUG to AAA and/or uORF2 AUG to ACA). Capped and polyadenylated mRNA (6 ng) was used to program *N. crassa* translation reactions (10 µl). LUC activity produced from mRNAs 2-4 obtained after 30 min incubation at 26°C was calculated relative to the activity produced from mRNA 1. Mean values and standard deviations from three independent experiments, each performed in triplicate, are given as normalized RLU. In addition, [^{35}S]Met-labeled translation products from translation reactions programmed with mRNAs 1-4 or with no mRNA were analyzed on 12% NuPAGE gels and a representative gel is shown. The position of radiolabeled LUC produced from the mAUG is indicated.

In earlier studies on *S. cerevisiae GCN4*, we used toeprint analyses to demonstrate reinitiation following uORF1 but not uORF4 translation in *S. cerevisiae* extracts (*293*). We adapted a similar approach to examine *cpc-1* uORF1 and uORF2 in *N. crassa* extracts (Figures 4.5 and 4.6). Adding cycloheximide (CYH) to reactions at time 0 (T_0) allows toeprint mapping of initiation codons where 80S ribosomes first initiate translation following initial scanning. Adding CYH at ten minutes after transalation (T_{10}) allows toeprint mapping of initiation sites in the steady state, for

example at additional sites where ribosomes have reinitiated. At T_0 and T_{10}, ribosomes are seen at the uORF AUG start codon; mutation to AAA eliminated this signal. This is as expected since the uORF1 AUG is in an optimal initiation context. At T_0, a reduced toeprint signal is seen at the uORF2 AUG relative to the signal at the uORF1 AUG. When the uORF1 AUG is mutated, the uORF2 AUG signal increased substantially; mutation of the uORF2 AUG to ACA eliminated this signal. These data indicate that most ribosomes initiate at uORF1, but when it is absent, they scan to uORF2. At T_0, a relatively low signal was observed at the mORF AUG except when uORF1 and uORF2 AUGs were mutated, as expected from scanning. When CYH was added at T_{10}, the most dramatic change in signal was an increase at the mAUG in the ΔuORF2 construct. This increase of the mAUG was not seen in the ΔuORF1 construct or the ΔuORF1 ΔuORF2 construct. These data are consistent with ribosomes reinitiating at the mAUG following uORF1 translation *in vitro*. They suggest that uORF1 and uORF2 of *N. crassa cpc-1* function similarly to uORF1 and uORF4 of *S. cerevisiae GCN4*.

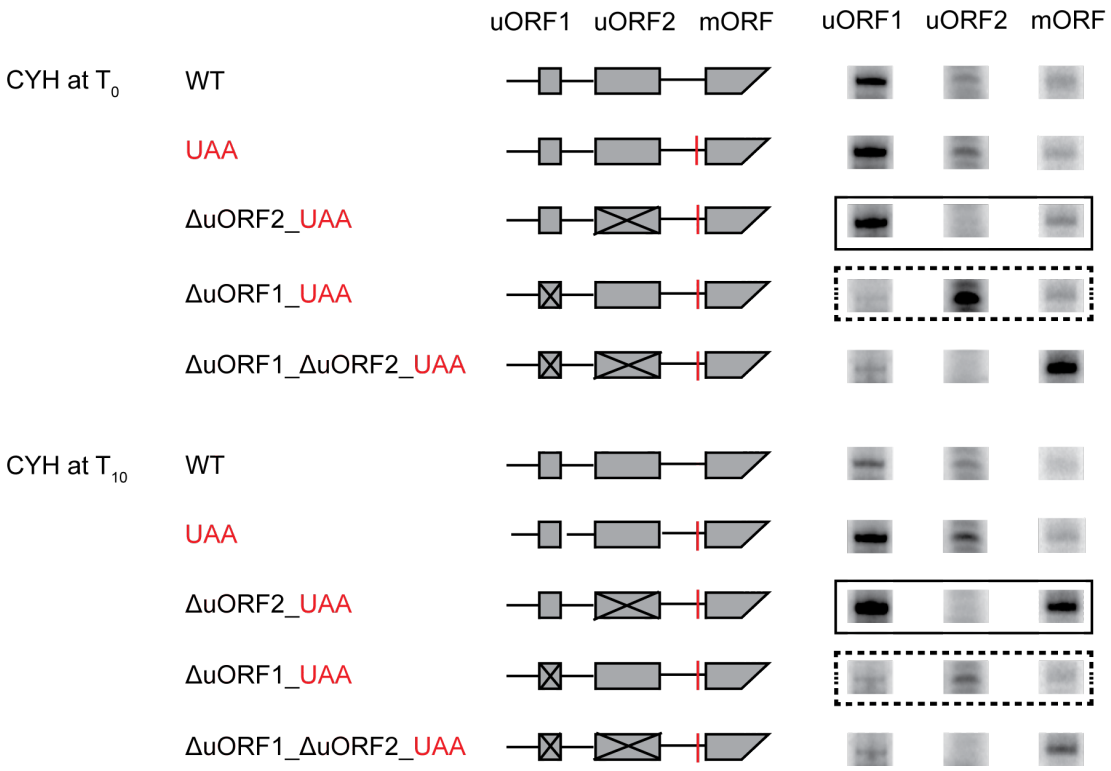
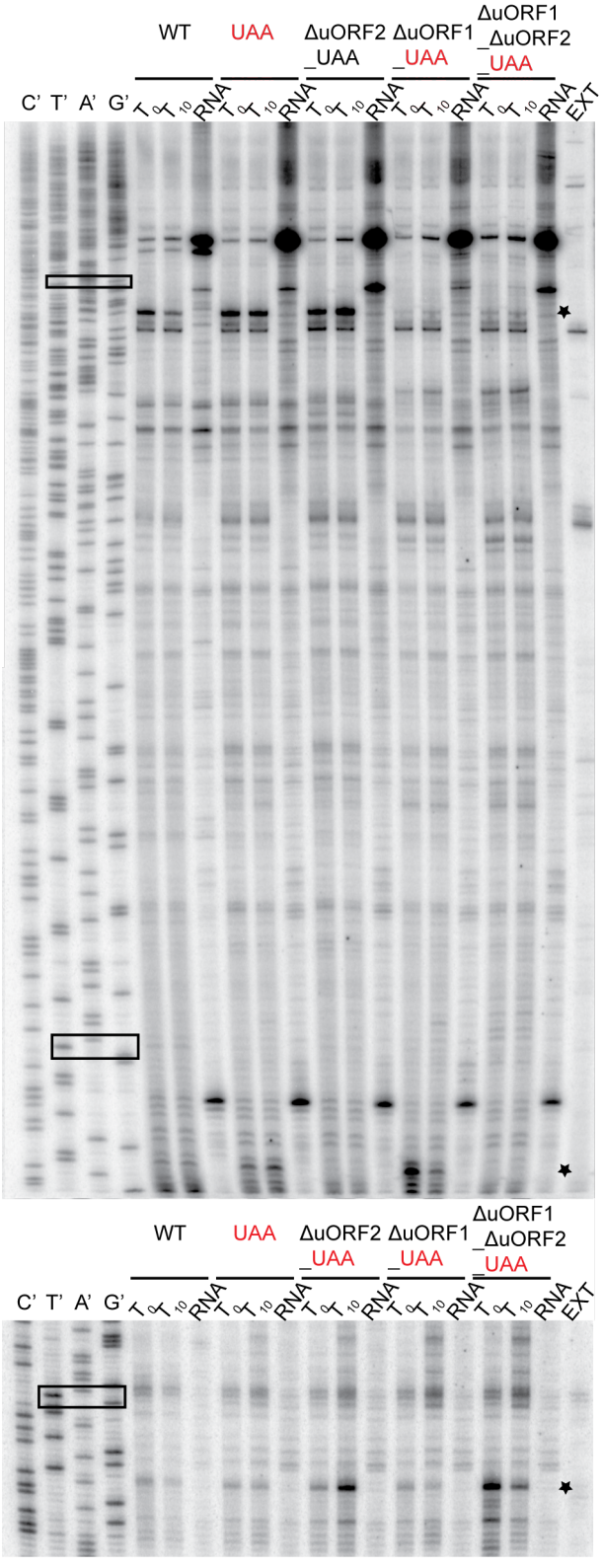


Figure 4.5. Toeprint analysis indicates reinitiation following translation of *cpc-1* uORF1 but not uORF2. *cpc1-luc* mRNA (60 ng) were used to program 20 µl *N. crassa* cell-free translation reactions. WT mRNA containing the wild-type *cpc-1* 5'UTR, and the mRNAs used in Fig. 3, were analysed in parallel along with controls. Reactions were incubated at 26°C min with cycloheximide (CYH) added either prior to incubation (T_0) or after 10 min of incubation (T_{10}) as indicated. Radiolabeled primer CPC101 was used to examine ribosomes at uORF1 and uORF 2; primer ZW4 was used to examine ribosomes at the mORF. The original data from which the toeprint signals were excised are shown in Figure 4.6.

Figure 4.6. Toeprint analysis indicates reinitiation following translation of *cpc-1* uORF1 but not uORF2 (primary data for Figure 4.5). The description of the experiment is given in the text. For markers, either primer CPC101 (upper panel) or ZW4 (lower panel) was used to sequence the WT template (the four left-most lanes). The nucleotide complementary to the dideoxynucleotide added to each sequencing reaction for the wild-type *cpc1-luc* template is indicated above the corresponding lane so that the sequence of the template can be directly deduced; the 5'-to-3' sequence reads from top to bottom. Asterisks (top to bottom): toeprint products corresponding to ribosomes bound at uAUG1, uAUG2 and mAUG on the *cpc1-luc* RNA. Boxes (top to bottom): uAUG1, uAUG2 and mAUG in the sequence markers. The ribosome protects approximately 16-nt 3' of a start codon in the ribosome P-site.


WT
UAA
ΔuORF2 _UAA
ΔuORF1 _UAA
ΔuORF1 _ΔuORF2 _UAA
C' T' A' G'
T
T 0
T 10
RNA
EXT

We next compared luciferase activity obtained from constructs with and without the introduced in-frame UAA stop codon to examine translation from NCCs upstream of the mAUG (Figure 4.7). The production of luciferase decreased when the UAA was present, indicating that polypeptides with luciferase activity were produced using NCCs upstream of the mORF (compare 1 and 3, 2 and 4, and 3 and 6, Figure 4.7). The UAA mutation decreased luciferase synthesis in the presence or absence of uORF2 (compare 1 and 2, and 3 and 4, Figure 4.7). Elimination of uORF2 resulted in overall increased luciferase synthesis as expected from its proposed inhibitory role on initiation at mAUG. As expected, NCCs and the mAUG have separate roles in initiation, combining ΔmAUG and UAA mutations yielded no detectable luciferase (Figure 4.7).

Interestingly, ΔmAUG showed a relatively small decrease in luciferase activity compared to WT (ΔmAUG 71% of WT; compare 1 and 5, Figure 4.7). This observation is consistent with the differences between the WT and the UAA constructs (UAA 22% of WT; compare 1 and 3, Figure 4.7). These data indicate that upstream NCCs are used to initiate translation efficiently *in vitro*.

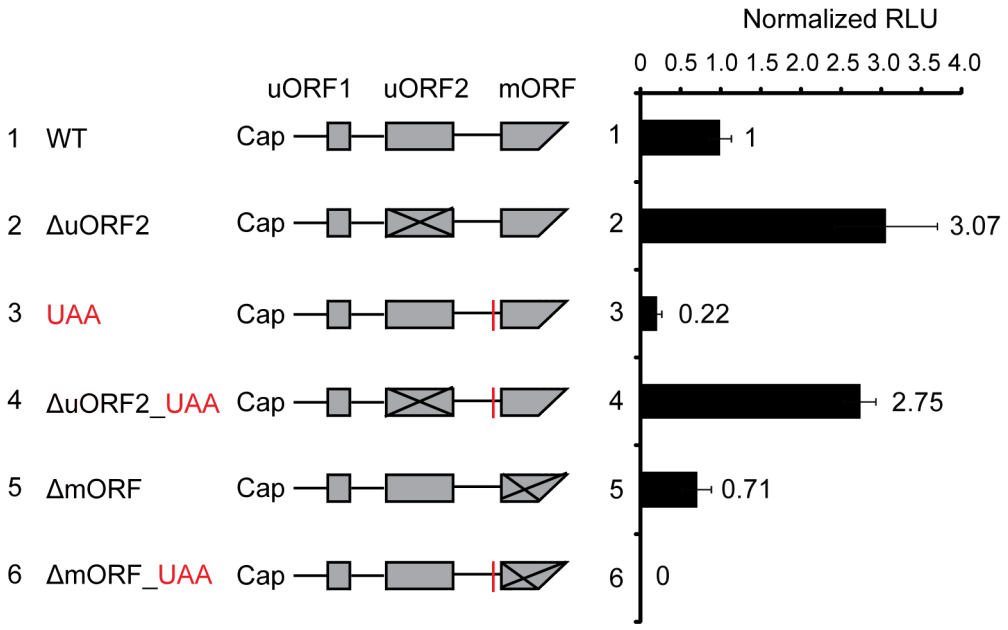


Figure 4.7. Discriminating translation from *N. crassa cpc-1* NCCs and mAUG *in vitro*. Capped and polyadenylated mRNA (6 ng) was used to program *N. crassa* translation reactions (10 µl) with the indicated constructs. Firefly luciferase activity from each mRNA obtained after 30 min incubation at 26°C was calculated relative to synthesis from the WT construct. Mean values and standard deviations from three independent experiments, each performed in triplicate, are plotted.

The eight NCCs identified bioinformatically (Figure 4.3) were individually eliminated and the consequences examined by [^{35}S]-Met labeling in *N.crassa* and wheat germ extracts (Figure 4.8 and 4.9). These mutations were also combined with the UAA mutation so the resulting polypeptides produced from upstream initiation could be better resolved by SDS-PAGE. Elimination of NCC 1, NCC 2, NCC 3 or NCC 4 did not yield any detectable differences compared to UAA (Figure 4.9, lanes 3-7). In contrast, elimination of NCC 5, NCC 6, NCC 7 or NCC 8 resulted in disappearance of specific truncated polypeptides (Figure 4.9, lanes 8-11 and 3, and Figure 4.8, lanes 8-11 and 3),

indicating that NCCs 5-8 initiated translation in *N. crassa* and wheat germ systems. When NCC 8 (ACG) was changed to AUG, the signal in the corresponding band increased as expected (lanes 12 and 3, Figure 4.9, and lanes 8 and 3, Figure 4.8).

Cell-free translation extracts programmed with CPC1-LUC (WT) produce polypeptides migrating more slowly than luciferase synthesized from an mRNA specifiying LUC alone, or a CPC1-LUC mRNA with the UAA mutation (Figure 4.1A, Figure 4.8 and Figure 4.9). When NCCs 5, 6, 7 and 8 were eliminated together in the absence of the UAA mutation, polypeptides larger than luciferase were still observed, although the amount was reduced compared to WT (Figure 4.8, compare lanes 10, 11 and 3). This suggests that other upstream codons in the *cpc-1* upstream region can be used to initiate polypeptide synthesis. This was observed in the presence or absence of uORF1 (lanes 10 and 11, Figure 4.8).

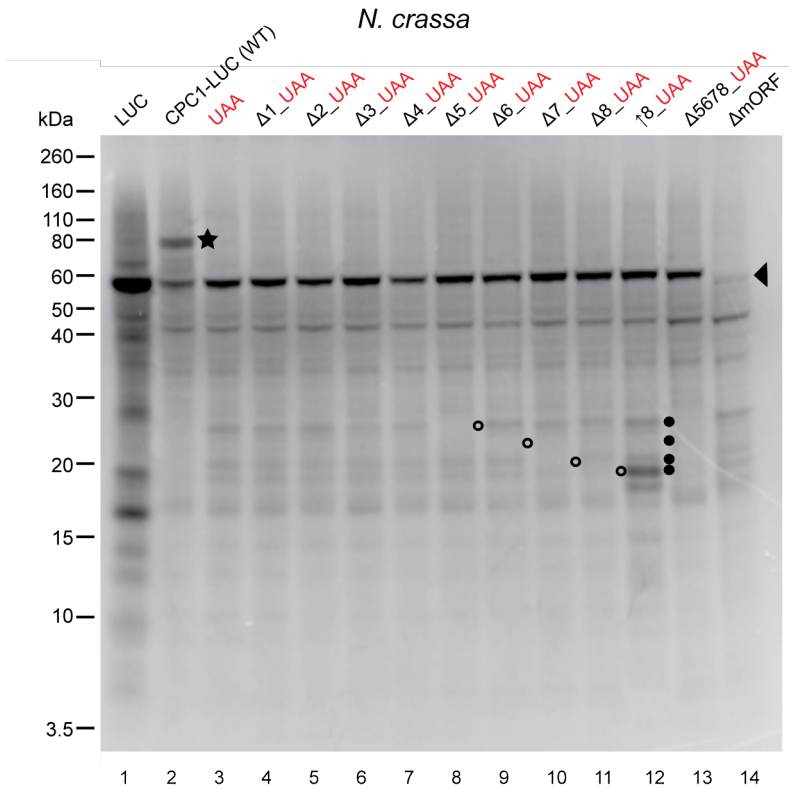


Figure 4.8. Evidence that *N. crassa* NCCs 5-8 initiate translation *in vitro*. Synthetic RNAs (60 ng) for the indicated constructs were used to program 10 µl of cell-free translation systems from (A) *N. crassa* or (B) wheat germ. Reactions were incubated for 30 min at 26°C. Radiolabeled products were analyzed on 12% NuPAGE gels. Open circles: translation products eliminated upon mutation of NCCs 5-8; the product predicted to be initiated from NCC 8 also increased when NCC 8 was changed to AUG (Lane 8). Arrowhead: position of mAUG-initiated translation product (mORF). Brackets: translation products larger than the mORF produced in the absence of an in-frame UAA stop codon.

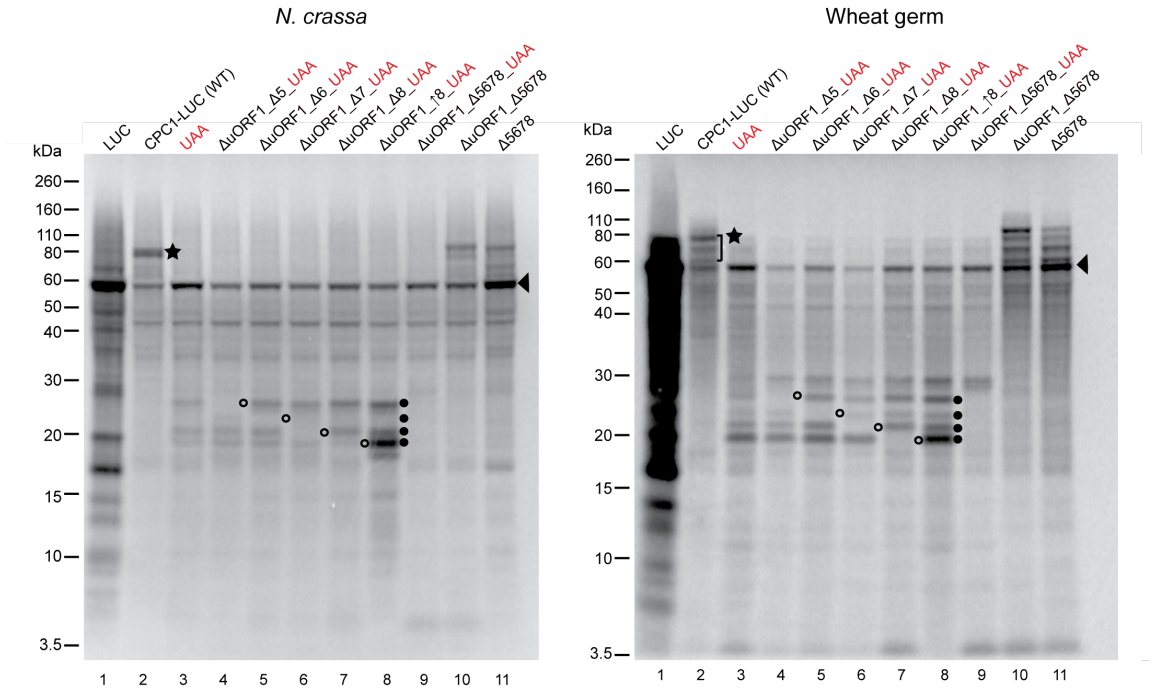


Figure 4.9. Evidence from [^{35}S]Met labeling showing that NCCs 5-8, but not 1-4, initiate translation in an *N. crassa* cell-free system. Synthetic RNAs (60 ng) for the indicated constructs were used to program 10 µl of cell-free translation reactions from (A) *N. crassa* or (B) wheat germ. Reactions were incubated for 30 min at 26°C. Radiolabeled products were analyzed on 12% NuPAGE gels. Open circles: translation products eliminated upon mutation of NCCs 5-8; the product predicted to be initiated from NCC 8 also increased when NCC 8 was changed to AUG (Lane 8). Arrowhead: position of mAUG-initiated translation product (mORF). Brackets: translation products larger than the mORF produced in the absence of an in-frame UAA stop codon.

To examine the roles of upstream NCCs in translation *in vivo* in *N. crassa*, strains containing *N. crassa* codon-optimized luciferase fused in-frame with wild type or mutated *cpc-1* 5' sequences were constructed. Three independent transformants containing each construct were used to measure LUC activity and LUC mRNA levels. We examined WT, UAA, ΔmAUG, and the ΔmAUG UAA double mutant. Luciferase activity was measured and normalized to reporter mRNA levels to account for the small

differences in luciferase mRNA levels observed. As shown in Figure 4.10, expression levels of WT and UAA reporters were similar. Luciferase activity from ΔmAUG was much lower but this activity was higher than for the ΔmAUG UAA construct. For the ΔmAUG construct, higher luciferase activity was observed than for the ΔmAUG UAA double mutant. Thus, although the amount of luciferase activity derived from upstream NCCs was less than 1% of activity from the mAUG *in vivo*, detectable luciferase was nevertheless observed (compare 3, 1 and 2 in Figure 4.10 and compare 5, 1 and 3 in Figure 4.7). Possibly, NCCs were not used as efficiently *in vivo* as *in vitro*. Alternatively, N-terminally extended luciferases are less stable *in vivo*.

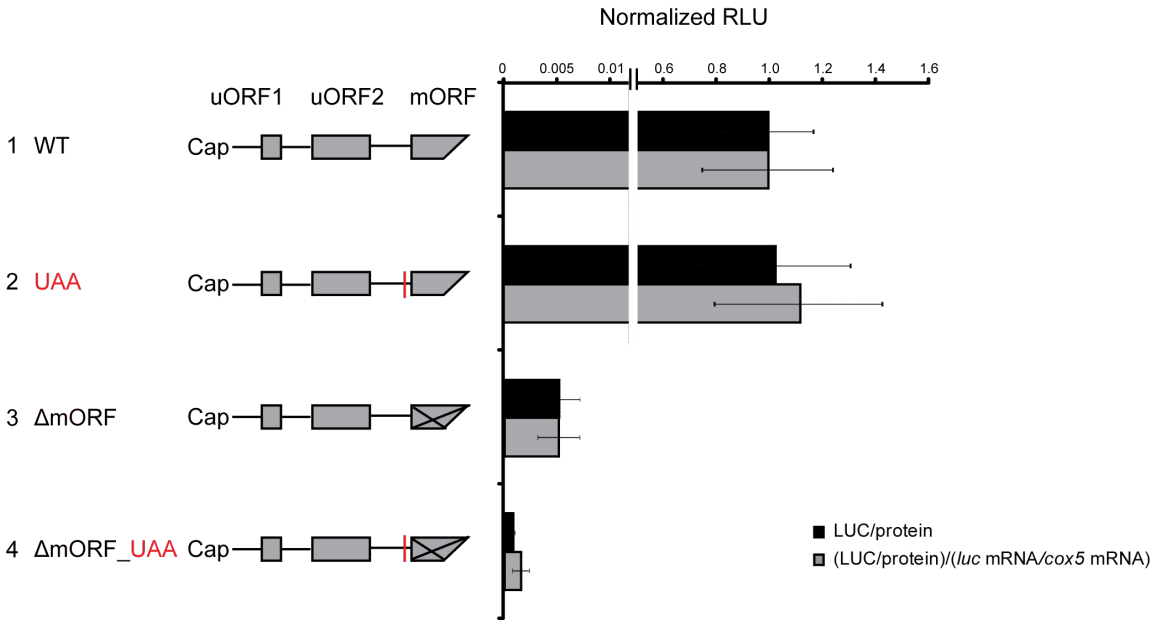


Figure 4.10. Discriminating translation from *N. crassa cpc-1* NCCs and mAUG *in vivo*. Constructs 1-4 were placed at the *N. crassa his-3* locus (three independent transformants of each). LUC activities were measured and values plotted relative to WT. Black bars: LUC activities normalized to total extracted protein; gray bars, LUC activities normalized first to total protein and then normalized to *luc* mRNA/cox-5 mRNA levels. Mean values and standard deviations for all measurements are derived from three independent experiments, each using all independent transformants.

4.4. Discussion

We examined the structures of *N. crassa cpc-1* homologs in fungi for which sequence was available. In the Pezizomycotina, all *cpc-1* genes could specify two uORFs, uORF1 and uORF2, within long mRNA 5' UTRs. The data obtained here with *N. crassa* are consistent with uORF1 and uORF2 functioning analogously to *S. cerevisiae GCN4* uORF1 and uORF4, respectively, to control initiation at the predicted mAUG start codon. Surprisingly, a long, conserved coding region upstream of this AUG start codon that was in-frame with CPC1 was present in all homologs from

Pezizomycotina, and in some cases this open reading frame extended to the predicted mRNA 5' ends. While no AUG codons were observed that could produce N-terminally extended isoforms of CPC1 (excepting the possibility of frameshifting from a uORF AUG), near-cognate start codons (NCCs), some well conserved, were present in the CPC1 reading frame that potentially could initiate translation of such isoforms. Translation initiating from four conserved NCCs in the *N. crassa cpc-1* 5'UTR was observed *in vitro* in *N. crassa* and wheat germ translation extracts. Such utilization of NCCs *in vivo* would result in synthesis of alternative isoforms of CPC1; these isoforms might have similar or different functions than CPC1 produced from the main AUG. N-terminal extensions could also influence protein stability. The synthesis of these alternative isoforms from NCCs upstream of uORF2 would also bypass the inhibitory effect of uORF2, which reduces synthesis of CPC1 from the downstream main AUG. These findings suggest a model for additional translational regulation of Pezizomycotina *cpc-1* through the use of NCCs, which could be independent of the uORF-control model elucidated for *S. cerevisiae GCN4*.

We found no fungal homologs of *cpc-1/GCN4* outside of Pezizomycotina and Basidiomycota that have NCC- initiated N-terminal extensions with the potential to preempt the effect of translating a long and inhibitory uORF. Since the other two sub-phyla of Ascomycota, Saccharomycotina and Taphrinomycotina, do not have potential for NCC-initiated extensions, it not entirely clear if the conserved extensions present in Pezizomycotina and in Basidiomycota (a sister phylum of Ascomycota in the subkingdom Dikarya) are homologous and were present in the last common ancestor of

Dikarya, which lived around 500 million years ago (Lucking et al., 2009), or whether they are examples of convergent evolution.

In the studies reported here, there is a discrepancy between luciferase activity produced *in vitro* and *in vivo* from the *cpc-1* NCCs. At face value, this would mean that *in vitro* there is more initiation from NCCs than from the mAUG, while *in vivo* the situation is reversed. For the *in vitro* experiments, we used intermediate [Mg^{2+}] which favors AUG over NCC initiation, but the *in vitro* conditions used here are not expected to be stringent as *in vivo* (CHAPTER 3). Thus, we expect that the *in vitro* conditions would yield more NCC-initiated products than occurs *in vivo*, but the discrepancy in levels of CPC1-LUC activity ovserved in vitro compared to *in vivo* still seems too large to be simply accounted for by known initiation efficiencies *in vivo* and *in vivo*. It is possible that the N-terminally extended forms of the luciferase reporter are unstable *in vivo*. That is, even if NCC translation were high, the level of protein would be low. The luciferase reporter data may thus be providing accurate information on the relative level of N-terminally extended CPC1 isoforms *in vivo*. This level, while low, could nevertheless be all the cell requires for these forms to fulfill their physiological functions.

Taken together, the data from *in vitro* analyses show that conserved CPC1 NCCs can be used to initiate translation and data from ribosome profiling and luciferase reporters indicate that ribosomes are translating extended N-terminal isoforms of CPC1 in *N. crassa* under standard culture conditions. These data raise important new questions regarding the functions of these isoforms and their regulation.

It is worth considering that the translation of another fungal bZIP transcription factor, *Podospora anserina* IDI-4, requires initiation from a CUG and not an AUG (*387*). This is additional evidence supporting the possibility that fungal bZIP transcriptional factors may use NCCs to initiate their translation more generally.

A mammalian bZIP transcription factor family, the CCAAT/enhancer binding proteins (C/EBPs) C/EBPα, C/EBPβ and C/EBPγ may also be instructive to consider with respect to CPC1 in terms of providing a potential paradigm for the function of different CPC1 isoforms. An in-frame AUG in a poor initiation context (C/EBPα) or an NCC in a good context (C/EBPβ) initiates the longest protein isoform (*483*). Leaky scanning past this initiation codon leads to initiation at an out-of-frame AUG codon in good context, producing a short (5-11 codons long) ORF. Two additional C/EBP isoforms (LAP and LIP) are generated by initiation from in-frame AUG codons downstream of this short ORF. The expression of these two isoforms is governed by reinitiation following translation of the short ORF and the relative levels can be altered by changes in eIF2α phosphorylation. LAP functions as a transcriptional activator and LIP as a transcriptional repressor and thus modulate different transcriptional outcomes under "normal" and stress conditions.

The physiological conditions that govern initiation at NCCs are an emerging area of investigation. In *S. cerevisiae*, amino acid limitation increases initiation at NCCs (*385*), as does the shift to the meiotic developmental program (*415*). A chemical screen identified several compounds that increase the efficiency of initiation at NCCs (*394*). The concentration of free polyamines affects initiation from a conserved AUU start

codon of a uORF within the mRNA encoding AZIN1 in mammalian cells (*466*). A number of cellular factors are known to be involved in discrimination between favourable and unfavourable initiation codons and contexts. Changes in the activity or cellular levels of eIF1 or eIF5 have profound effects on translation initiation at NCCs or AUG codons in poor context (*21*). Understanding the physiological conditions that control initiation at NCCs has broad implication for gene regulation and protein synthesis as well as for specific understanding of these aspects of CPC1.

The pivotal role played by CPC1 in transcriptional regulation in *N. crassa*, combined with its known translational regulation and the suspected modularity of stringency of start codon selection strongly raise the exciting possibility that NCC-mediated initiation might provide additional routes for translation regulation and the diversity of mechanism by which CPC1 control gene expression.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

5.1. Summary of research

This research has focused on understanding translational control in the model filamentous fungus *N. crassa*, particularly on processes impacting translation termination and initiation.

The mechanism of uORF-mediated regulation was examined for the *N. crassa* uORF-encoded AAP. Studies in cell-free extracts from *N. crassa* and wheat germ using puromycin-release assays indicated that inhibition of PTC activity by the AAP in response to Arg is the basis for the AAP's function to stall ribosomes at the uORF termination codon. The mode of PTC inhibition appears unusual because neither a specific amino acid nor a specific nascent peptide chain length was required for AAP to inhibit PTC activity.

The use of non-canonical start codon to initiate translation was studied systematically *in vivo* and *in vitro*. A codon-optimized firefly luciferase reporter initiated with AUG or each of the nine NCCs in preferred context was used to examine the stringency of start codon selection in *N. crassa*. *In vivo* and *in vitro* results indicated that the hierarchy of initiation at start codons in *N. crassa* is similar to that in human cells. Analyses of the 5'-leader regions in the *N. crassa* transcriptome revealed examples of highly conserved NCCs in preferred contexts that could extend the N-termini of the

predicted polypeptides. Near-cognate start codons, some well conserved, that potentially initiate conserved N-terminal extensions of the *N. crassa cpc-1* homolog, were identified and their potential use investigated.

The regulatory function of two uORFs in the *N. crassa cpc-1* transcript was also investigated. *In vitro* studies show that *cpc-1* uORF1 and uORF2 are functionally analogous to uORF1 and uORF4 in *S. cerevisiae GCN4*. uORF1 acts as a positive regulatory element to facilitate reinitiation, while uORF2 strongly inhibits the translation of *cpc-1*. The used of NCCs might bypass the translational inhibitory effect of uORF2 of *cpc-1* to regulate translation.

These studies provide evidence for translational control mechanisms in which uORF and NCCs regulate gene expression. The successful use of *N. crassa* to study these cases *in vivo* and *in vitro* establishes *N. crassa* as a powerful system that enables genetic and biochemical approaches to be used to understand the factors and mechanisms contributing to translational processes.

5.1.1. The nascent *N. crassa arg-2* AAP interferes with PTC activity in response to Arg

The nascent AAP functions with Arg to interfere with the PTC activity of *N. crassa* and wheat ribosomes. AAP containing the D12N mutation, which eliminates Arg-induced ribosome stalling, also eliminated Arg's effect on PTC activity. Arg-regulated inhibition of PTC activity was observed for full-length AAP, for AAP extended by one amino acid at its C-terminus and for AAP truncated by one or two amino acids at its C-terminus. Toeprinting and luciferase activity analyses indicated the truncated AAPs

function to stall ribosomes, albeit with reduced efficiency. Analyses of tRNA-identity also showed the AAP regulated PTC activity prior to its complete synthesis. Evolutionarily, AAPs in other species identified to date are all at least the size of the *N. crassa* full-length AAP and our data indicate that truncated AAPs do not function as efficiently as the full-length AAP. These results demonstrate the inhibition of PTC activity is the basis for the AAP's capacity to stall the ribosome. Unlike many other ribosome stalling peptides, specific features of the AAP near the ribosomal catalytic center appear relatively unimportant for stalling.

We also obtained data indicating that both wild-type AAP and D12N AAP (which lacks regulatory function) had higher intrinsic stalling activity than a frame-shifted peptide. Thus the AAP appears to have intrinsic stalling activity detectable even in the absence of Arg-regulated stalling.

While AAP truncated by one or two amino acids at its C-terminus still functions to stall ribosomes, full-length AAP synthesis appeared important for most efficient stalling. Combined with structural analyses of AAP in the ribosome exit tunnel (*324*), these data suggest a model for AAP-mediated stalling in response to Arg in which the AAP undergoes a change in relative conformation with respect to the ribosome, and this altered conformation must be in proper register with the ribosome to efficiently interfere with PTC activity. In this model, there is a window of AAP chain-lengths for which AAP can find register (with respect to the ribosome and possibly with respect to Arg), with the wild-type AAP length being most efficient at achieving and/or remaining in register.

5.1.2. Use of near-cognate initiation codons in *N. crassa*

Codon-optimized firefly luciferase reporters initiated with AUG or each of the nine NCCs in preferred context were used to study the stringency of start codon selection in *N. crassa*. The luciferase reporter genes were incorporated at the *N. crassa his-3* locus. *In vivo* studies to measure initiation activity by measuring LUC enzyme activity indicated that the hierarchy of initiation at start codons is similar to that in human cells (*379*). CUG and GUG were the most efficient NCCs, with 11% and 7% of the efficiency of AUG. AUA, AUU, UUG and AUC were less active (1-5%), while AAG and AGG did not function. Similar results were obtained by translating mRNAs in a homologous *N. crassa in vitro* translation system or in rabbit reticulocyte lysate. The efficiency of initiation at AUG, CUG and UUG codons in a context other than the preferred context was also examined *in vitro*. The results show that the preferred context was more important for efficient initiation from NCCs than from AUG.

Bioinformatic analyses of the 5'-leader regions in the *N. crassa* transcriptome revealed examples of highly conserved NCCs in preferred contexts that could extend the N-termini of the predicted polypeptides. Additional NCCs in all three reading frames are also evident, but they were not analyzed further.

In summary, these studies demonstrated that NCCs can be used for initiation in *N. crassa* and identified potential examples for which initiation at NCCs may have physiological significance.

5.1.3. Near-cognate codons are identified to initiate translation of *N. crassa* CPC1

In vitro studies using *N.crassa* and wheat germ extracts suggest that *N. crassa cpc-1* uORF1 and uORF2 behave analogously to uORF1 and uORF4 in *S. cerevisiae GCN4*. That is *cpc-1* uORF1 and uORF2 would control translation by facilitating reinitiation and by inhibiting the translation of the CPC1 mORF, respectively. Conserved N-terminal extensions in the *cpc-1* reading frame from Pezizomycotina species were identified using bioinformatic approaches. NCCs, some well conserved and in optimal initiation contexts, which potentially initiate translation of extended CPC1 polypeptides, were examined both *in vitro* and *in vivo*. Four identified near-cognate start codons were demonstrated to initiate translation *in vitro* to produce N-terminally extended CPC1 polypeptides. The positions of these near-cognate start codons enable their utilization bypass the translational inhibitory effect of uORF2. These findings suggested that in addition to control of translation through uORFs, as exemplified in *S. cerevisiae GCN4*, the 5' UTRs of *N. crassa* and other fungal homologs of CPC1 may use a separate mechanism, which responds to physiological conditions that change the stringency of start codon selection.

5.2. Future directions

5.2.1. The function of C-terminally shortened and lengthened AAPs *in vivo*

The level of a eukaryotic mRNA can be controlled by linking its stability to its translation. The nonsense mediated mRNA decay (NMD) pathway targets mRNA for degradation through the recognition of premature termination codons (*484*). Our lab

used metabolic labeling with 4-thiouracil (4TU) and a pulse-chase procedure to assess the roles of NMD factors in controlling mRNA stability in *N. crassa* (Ying and Sachs, unpublished data). The *arg-2* mRNA, which contains AAP that causes ribosomes to stall during translation termination, was stabilized in mutants lacking NMD factors UPF1 or UPF2. Analyses of luciferase reporter genes containing wild type AAP demonstrated that the *arg-2* uORF conferred NMD-sensitivity to reporters. The D12N AAP that does not stall ribosomes did not confer NMD-sensitivity, suggesting functional AAP was responsible for triggering NMD of the *arg-2* mRNA.

The present *in vitro* data indicate that AAPs shortened at the C-terminus by one or two residues retain their regulatory function to stall ribosomes in response to Arg. However, it is not clear whether these truncated AAPs can function *in vivo.* To test this, luciferase reporter genes containing C-terminally shortened AAPs can be used to determine their functions in triggering NMD *in vivo*. If NMD mutations (*Δupf1* and/or *Δupf2*) increased the stability of these transcripts, this would indicate that shortened AAPs cause substantial stalling *in vivo*. Similarly, C-terminal lengthened AAPs can also be tested for their functions *in vivo* by examining whether they are regulated by the NMD pathway. Stalling during elongation in the C-terminal regin of extended AAPs shoud not trigger NMD, but stalling at the termination codon should trigger NMD.

5.2.2. The conditions that change the stringency of start codon selection

Ribosome profiling studies in yeast revealed that NCC initiated uORFs can be translated at levels comparable to AUG-initiated uORFs and translation of these non-AUG uORFs generally increases in response to amino acid starvation (*385*). The

compound 3-aminotriazole (3-AT) induces histidine starvation in *N. crassa* and other fungi. However, testing for changes in stringency in response to 3-AT using luciferase reporter genes initiated with AUG or NCCs did not yield conclusive results in preliminary *N. crassa* studies (data not shown). The effects of amino acid starvation could be tested more systematically by ribosome profiling studies in *N. crassa* grown in the presence and absence of 3-AT. If amino acid starvation decreases the stringency of start codon selection in *N. crassa*, we expect that one or more of the four NCCs we identified upstream of the *cpc-1* mAUG could be used at higher levels. This would further support the model that the use of NCCs could response to conditions that change the stringency of start codon selection and bypass the inhibitory effect of uORF2 to create another layer of translational control that regulates *cpc-1* gene expression.

In mammalian cells, overexpression of wild-type eIF1 reduces the use of AUG codons in poor context and the use of non-AUG start codons (*379*). In contrast to eIF1, over-expression of eIF5 decreases the stringency of start codon selection (*400*). In *N. crassa*, expressions of eIF1 and eIF5 could be regulated by placing their expression under the control of a conditional promoter into the genome. *cis*-elements in eIF1 and eIF5 for their auto-regulation need to be removed to increase/decrease the expression of eIF1 and eIF5 with more dramatic effects. Testing the use of NCCs under these conditions might lead to a better understanding of the effects of eIF1 and eIF5 on the stringency of start codon selection in *N. crassa*. Similarly to the 3-AT treatment, if stringency is altered in response to changes in eIF1 or eIF5 and this effects *cpc-1*

expression, then eIF1/5 control is considered to be another way of translational control in *cpc-1* gene in addition to the eIF2α-mediated regulation.

REFERENCES

1. M. B. Matthews *et al.*, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 1-31.
2. S. E. Calvo *et al.*, Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA* **106**, 7507 (2009).
3. R. Green, H. F. Noller, Ribosomes and translation. *Annu Rev Biochem* **66**, 679 (1997).
4. F. Mignone *et al.*, Untranslated regions of mRNAs. *Genome Biol* **3**, REVIEWS0004 (2002).
5. F. Gebauer, M. W. Hentze, Molecular mechanisms of translational control. *Nat Rev Mol Cell Bio* **5**, 827 (2004).
6. J. W. B. Hershey, W. C. Merrick, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 33-88.
7. M. Kozak, Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187 (1999).
8. C. J. Decker, R. Parker, Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem Sci* **19**, 336 (1994).

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

10. N. Cougot *et al.*, Cytoplasmic foci are sites of mRNA decay in human cells. *J Cell Biol* **165**, 31 (2004).

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

12. A. Jacobson, in *Translational control,* J. W. B. Hershey *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996), pp. 451-480.

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

14. N. Iizuka *et al.*, Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from *Saccharomyces cerevisiae*. *Mol Cell Biol* **14**, 7322 (1994).

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

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

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

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

19. N. Sonenberg, A. G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731 (2009).

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

21. A. G. Hinnebusch, Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev* **75**, 434 (2011).

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

23. T. V. Pestova *et al.*, in *Translational control in biology and medicine,* M. B. Mathews *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007), pp. 87-128.

24. K. Asano *et al.*, Conserved bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J* **18**, 1673 (1999).

25. S. Das, U. Maitra, Mutational analysis of mammalian translation initiation factor 5 (eIF5): role of interaction between the beta subunit of eIF2 and eIF5 in eIF5 function in vitro and in vivo. *Mol Cell Biol* **20**, 3942 (2000).

26. J. P. Laurino *et al.*, The beta subunit of eukaryotic translation initiation factor 2 binds mRNA through the lysine repeats and a region comprising the C2-C2 motif. *Mol Cell Biol* **19**, 173 (1999).

27. D. R. Dorris *et al.*, Mutations in GCD11, the structural gene for eIF-2 gamma in yeast, alter translational regulation of GCN4 and the selection of the start site for protein synthesis. *EMBO J* **14**, 2239 (1995).

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

29. N. J. Gaspar *et al.*, Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. *J Biol Chem* **269**, 3415 (1994).

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

31. H. K. Huang *et al.*, GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev* **11**, 2396 (1997).

32. N. C. Kyrpides, C. R. Woese, Universally conserved translation initiation factors. *Proc Natl Acad Sci USA* **95**, 224 (1998).

33. E. Schmitt *et al.*, The large subunit of initiation factor aIF2 is a close structural homologue of elongation factors. *EMBO J* **21**, 1821 (2002).

34. B. S. Shin *et al.*, Initiation factor eIF2gamma promotes eIF2-GTP-Met-tRNAi(Met) ternary complex binding to the 40S ribosome. *Nat Struct Mol Biol* **18**, 1227 (2011).

35. A. G. Hinnebusch, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 185-243.

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

37. N. Sonenberg *et al.*, *Translational control of gene expression*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000).

38. T. E. Dever *et al.*, Phosphorylation of initiation factor 2alpha by protein kinase GCN2 mediates gene-specific translational control of *GCN4* in yeast. *Cell* **68**, 585 (1992).

39. J. W. B. Hershey, Translational control in mammalian cells. *Ann Rev Biochem* **60**, 717 (1991).

40. K. Asano *et al.*, A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. *Genes Dev* **14**, 2534 (2000).

41. J. Chaudhuri *et al.*, 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 (1999).

42. R. Majumdar *et al.*, 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 (2003).

43. Y. Yamamoto *et al.*, The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. *Proc Natl Acad Sci USA* **102**, 16164 (2005).

44. I. B. Lomakin *et al.*, Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev* **17**, 2786 (2003).

45. Y. Yu *et al.*, Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res* **37**, 5167 (2009).

46. D. S. Olsen *et al.*, Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. *EMBO J* **22**, 193 (2003).

47. B. Siridechadilok *et al.*, Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**, 1513 (2005).

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

49. A. V. Jivotovskaya *et al.*, Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. *Mol Cell Biol* **26**, 1355 (2006).

50. A. V. Pisarev *et al.*, Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* **27**, 1609 (2008).

51. J. A. Grifo *et al.*, New initiation factor activity required for globin mRNA translation. *J Biol Chem* **258**, 5804 (1983).

52. T. M. Hinton *et al.*, Functional analysis of individual binding activities of the scaffold protein eIF4G. *J Biol Chem* **282**, 1695 (2007).

53. D. Prevot *et al.*, Conducting the initiation of protein synthesis: the role of eIF4G. *Biol Cell* **95**, 141 (2003).

54. N. Sonenberg *et al.*, A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci USA* **75**, 4843 (1978).

55. B. J. Lamphear *et al.*, 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 (1995).

56. S. Mader *et al.*, The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol Cell Biol* **15**, 4990 (1995).

57. J. D. Gross *et al.*, Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* **115**, 739 (2003).

58. F. Rozen *et al.*, Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol Cell Biol* **10**, 1134 (1990).

59. T. Preiss, H. M. W., Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays* **25**, 1201 (2003).

60. A. Marintchev *et al.*, Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* **136**, 447 (2009).

61. N. L. Korneeva *et al.*, Characterization of the two eIF4A-binding sites on human eIF4G-1. *J Biol Chem* **276**, 2872 (2001).

62. J. Marcotrigiano *et al.*, A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell* **7**, 193 (2001).

63. P. Schutz *et al.*, Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci USA* **105**, 9564 (2008).

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

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

66. A. G. Hinnebusch, J. R. Lorsch, The mechanism of eukaryotic translation initiation: new insights and challenges. *CSH Perspect Biol* **4**, (2012).

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

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

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

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

71. T. F. Donahue, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 595-614.

72. C. A. Fekete *et al.*, The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection in vivo. *EMBO J* **24**, 3588 (2005).

73. T. V. Pestova *et al.*, Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**, 854 (1998).

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

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

76. W. C. Merrick, J. W. B. Hershey, in *Translational control,* J. W. B. Hershey *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996), pp. 31-69.

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

78. M. A. Algire *et al.*, Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol Cell* **20**, 251 (2005).

79. C. A. Fekete *et al.*, N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. *EMBO J* **26**, 1602 (2007).

80. J. S. Nanda *et al.*, eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *J Mol Biol* **394**, 268 (2009).

81. C. E. Aitken, J. R. Lorsch, A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol* **19**, 568 (2012).

82. T. V. Pestova *et al.*, The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**, 332 (2000).

83. B. S. Shin *et al.*, Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. *Cell* **111**, 1015 (2002).

84. J. H. Lee *et al.*, Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proc Natl Acad Sci USA* **99**, 16689 (2002).

85. M. G. Acker *et al.*, Interaction between eukaryotic initiation factors 1A and 5B is required for efficient ribosomal subunit joining. *J Biol Chem* **281**, 8469 (2006).

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

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

88. S. K. Jang *et al.*, A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* **62**, 2636 (1988).

89. T. V. Pestova *et al.*, Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* **16**, 6859 (1996).

90. T. V. Pestova *et al.*, Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol Cell Biol* **16**, 6870 (1996).

91. T. V. Pestova *et al.*, eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J* **27**, 1060 (2008).

92. I. M. Terenin *et al.*, Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat Struct Mol Biol* **15**, 836 (2008).

93. J. E. Wilson *et al.*, Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511 (2000).

94. J. E. Wilson *et al.*, Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol Cell Biol* **20**, 4990 (2000).

95. A. J. Chase, B. L. Semler, Viral subversion of host functions for picornavirus translation and RNA replication. *Future Virol* **7**, 179 (2012).

96. A. C. Gingras *et al.*, Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. *Proc Natl Acad Sci USA* **93**, 5578 (1996).

97. S. Cornelis *et al.*, Identification and characterization of a novel cell cycle-regulated internal ribosome entry site. *Mol Cell* **5**, 597 (2000).

98. M. Holcik *et al.*, Internal ribosome initiation of translation and the control of cell death. *Trends Genet* **16**, 469 (2000).

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

100. C. M. Spahn *et al.*, Structure of the 80S ribosome from *Saccharomyces cerevisiae*--tRNA- ribosome and subunit-subunit interactions. *Cell* **107**, 373 (2001).

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

102. M. V. Rodnina, W. Wintermeyer, Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol* **21**, 435 (2009).

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

104. H. Stark *et al.*, Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy. *Cell* **88**, 19 (1997).

105. R. C. Thompson *et al.*, The reaction of ribosomes with elongation factor Tu.GTP complexes. Aminoacyl-tRNA-independent reactions in the elongation cycle determine the accuracy of protein synthesis. *J Biol Chem* **261**, 4868 (1986).

106. J. M. Ogle *et al.*, Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897 (2001).

107. T. Pape *et al.*, 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 (1998).

108. T. Pape *et al.*, Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J* **18**, 3800 (1999).

109. K. B. Gromadski *et al.*, A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol Cell* **21**, 369 (2006).

110. K. B. Gromadski, M. V. Rodnina, Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* **13**, 191 (2004).

111. L. Cochella, R. Green, An active role for tRNA in decoding beyond codon:anticodon pairing. *Science* **308**, 1178 (2005).

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

113. I. Wohlgemuth *et al.*, Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J Biol Chem* **283**, 32229 (2008).

114. T. M. Schmeing *et al.*, An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520 (2005).

115. D. A. Hiller *et al.*, A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **476**, 236 (2011).

116. M. V. Rodnina *et al.*, Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**, 37 (1997).

117. W. Wintermeyer *et al.*, Mechanism of elongation factor G function in tRNA translocation on the ribosome. *Cold Spring Harb Symp Quant Biol* **66**, 449 (2001).

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

119. A. L. Konevega *et al.*, Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. *Nat Struct Mol Biol* **14**, 318 (2007).

120. S. Shoji *et al.*, Reverse translocation of tRNA in the ribosome. *Mol Cell* **24**, 931 (2006).

121. H. Stark *et al.*, Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* **100**, 301 (2000).

122. D. J. Taylor *et al.*, Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J* **26**, 2421 (2007).

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

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

125. C. B. Andersen *et al.*, Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* **443**, 663 (2006).

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

127. S. L. Qin *et al.*, Sequence analysis of the translational elongation factor 3 from Saccharomyces cerevisiae. *J Biol Chem* **265**, 1903 (1990).

128. M. Anand *et al.*, Functional interactions between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. *J Biol Chem* **278**, 6985 (2003).

129. O. Kovalchuke *et al.*, Competition and cooperation amongst yeast elongation factors. *Eur J Biochem* **258**, 986 (1998).

130. W. M. Kemper *et al.*, Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2Balpha and M2Bbeta. *J Biol Chem* **251**, 5551 (1976).

131. P. Saini *et al.*, Hypusine-containing protein eIF5A promotes translation elongation. *Nature* **459**, 118 (2009).

132. B. Bonetti *et al.*, The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol* **251**, 334 (1995).

133. K. K. McCaughan *et al.*, Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc Natl Acad Sci USA* **92**, 5431 (1995).

134. L. L. Major *et al.*, Is the in-frame termination signal of the Escherichia coli release factor-2 frameshift site weakened by a particularly poor context? *Nucleic Acids Res* **24**, 2673 (1996).

135. Y. O. Chernoff *et al.*, 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 (1996).

136. A. T. Chao *et al.*, Mutations in eukaryotic release factors 1 and 3 act as general nonsense suppressors in Drosophila. *Genetics* **165**, 601 (2003).

137. A. L. Arkov *et al.*, Mutations in RNAs of both ribosomal subunits cause defects in translation termination. *EMBO J* **17**, 1507 (1998).

138. A. L. Arkov *et al.*, Mutational eidence for a functional connection between two domains of 23S rRNA in translation termination. *J Bacteriol* **184**, 5052 (2002).

139. C. T. Caskey *et al.*, Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc Natl Acad Sci USA* **68**, 3163 (1971).

140. A. Seit-Nebi *et al.*, Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucleic Acids Res* **29**, 3982 (2001).

141. A. V. Zavialov *et al.*, Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol Cell* **10**, 789 (2002).

142. E. Scolnick *et al.*, Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci USA* **61**, 768 (1968).

143. M. Dontsova *et al.*, Translation termination factor aRF1 from the archaeon Methanococcus jannaschii is active with eukaryotic ribosomes. *FEBS Lett* **472**, 213 (2000).

144. L. Frolova *et al.*, A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* **372**, 701 (1994).

145. D. S. Konecki *et al.*, Characterization of reticulocyte release factor. *J Biol Chem* **252**, 4514 (1977).

146. H. Song *et al.*, The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311 (2000).

147. L. L. Kisselev, R. H. Buckingham, Translational termination comes of age. *Trends Biochem Sci* **25**, 561 (2000).

148. H. Gao *et al.*, RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors. *Cell* **129**, 929 (2007).

149. A. V. Pisarev *et al.*, Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev* **20**, 624 (2006).

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

151. E. Z. Alkalaeva *et al.*, In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell* **125**, 1125 (2006).

152. H. Fan-Minogue *et al.*, Distinct eRF3 requirements suggest alternate eRF1 conformations mediate peptide release during eukaryotic translation termination. *Mol Cell* **30**, 599 (2008).

153. L. Frolova *et al.*, Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA* **2**, 334 (1996).

154. A. Korostelev *et al.*, Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci USA* **105**, 19684 (2008).

155. M. Laurberg *et al.*, Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852 (2008).

156. A. Weixlbaumer *et al.*, Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953 (2008).

157. R. J. Jackson *et al.*, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113 (2010).

158. A. V. Kononenko *et al.*, Role of the individual domains of translation termination factor eRF1 in GTP binding to eRF3. *Proteins* **70**, 388 (2008).

159. H. Jin *et al.*, Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc Natl Acad Sci USA* **107**, 8593 (2010).

160. L. Frolova *et al.*, Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA* **8**, 129 (2002).

161. A. Seit-Nebi *et al.*, Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep* **3**, 881 (2002).

162. L. Y. Frolova *et al.*, 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 (1999).

163. R. Karimi *et al.*, Novel roles for classical factors at the interface between translation termination and initiation. *Mol Cell* **3**, 601 (1999).

164. F. Peske *et al.*, Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol Cell* **18**, 403 (2005).

165. D. S. Andersen, S. J. Leevers, The essential Drosophila ATP-binding cassette domain protein, pixie, binds the 40 S ribosome in an ATP-dependent manner and is required for translation initiation. *J Biol Chem* **282**, 14752 (2007).

166. Z. Q. Chen *et al.*, The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors. *J Biol Chem* **281**, 7452 (2006).

167. A. V. Pisarev *et al.*, The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell* **37**, 196 (2010).

168. M. A. Skabkin *et al.*, Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev* **24**, 1787 (2010).

169. A. V. Pisarev *et al.*, Recycling of eukaryotic posttermination ribosomal complexes. *Cell* **131**, 286 (2007).

170. L. A. Passmore *et al.*, The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* **26**, 41 (2007).

171. S. Srivastava *et al.*, Eukaryotic initiation factor 3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface. *J Mol Biol* **226**, 301 (1992).

172. U. A. Bommer *et al.*, Eukaryotic initiation factors eIF-2 and eIF-3: interactions, structure and localization in ribosomal initiation complexes. *Biochimie* **73**, 1007 (1991).

173. C. S. Fraser *et al.*, eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol Cell* **26**, 811 (2007).

174. P. Raychaudhuri *et al.*, Ribosomal subunit antiassociation activity in rabbit reticulocyte lysates. Evidence for a low molecular weight ribosomal subunit antiassociation protein factor (Mr = 25,000). *J Biol Chem* **259**, 11930 (1984).

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

176. M. Ceci *et al.*, Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**, 579 (2003).

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

178. U. Basu *et al.*, The Saccharomyces cerevisiae TIF6 gene encoding translation initiation factor 6 is required for 60S ribosomal subunit biogenesis. *Mol Cell Biol* **21**, 1453 (2001).

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

180. S. Uma *et al.*, The N-terminal region of the heme-regulated eIF2alpha kinase is an autonomous heme binding domain. *FEBS J* **267**, 498 (2000).

181. D. Scheuner *et al.*, The double-stranded RNA-activated protein kinase mediates viral-induced encephalitis. *Virology* **317**, 263 (2003).

182. H. P. Harding *et al.*, Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol Cell* **5**, 897 (2000).

183. P. Zhang *et al.*, The PERK eukaryotic initiation factor 2 alpha kinase is required for the development of the skeletal system, postnatal growth, and the function and viability of the pancreas. *Mol Cell Biol* **22**, 3864 (2002).

184. K. S. Montine, E. C. Henshaw, Serum growth factors cause rapid stimulation of protein synthesis and dephosphorylation of eIF-2 in serum deprived Ehrlich cells. *Biochimica et biophysica acta* **1014**, 282 (1989).

185. A. G. Hinnebusch, in *Translational control,* J. W. B. Hershey *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996), pp. 199-244.

186. X. Wang *et al.*, 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 (2003).

187. W. B. Minich *et al.*, Chromatographic resolution of in vivo phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF-4E: increased cap affinity of the phosphorylated form. *Proc Natl Acad Sci USA* **91**, 7668 (1994).

188. A. C. Gingras *et al.*, Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev* **13**, 1422 (1999).

189. J. C. Lawrence, Jr., R. T. Abraham, PHAS/4E-BPs as regulators of mRNA translation and cell proliferation. *Trends Biochem Sci* **22**, 345 (1997).

190. A. Haghighat, N. Sonenberg, eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J Biol Chem* **272**, 21677 (1997).

191. S. R. Kimball, Regulation of translation initiation by amino acids in eukaryotic cells. *Prog Mol and Subcell Biol* **26**, 155 (2001).

192. N. Sonenberg, A. C. Gingras, The mRNA 5' cap-binding protein eIF4E and control of cell growth. *Curr Opin Cell Biol* **10**, 268 (1998).

193. A. C. Gingras *et al.*, Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes Dev* **15**, 2852 (2001).

194. X. Wang *et al.*, Distinct signaling events downstream of mTOR cooperate to mediate the effects of amino acids and insulin on initiation factor 4E-binding proteins. *Mol Cell Biol* **25**, 2558 (2005).

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

196. J. Ling *et al.*, Inhibition of cap-dependent translation via phosphorylation of eIF4G by protein kinase Pak2. *EMBO J* **24**, 4094 (2005).

197. K. C. Orton *et al.*, Phosphorylation of Mnk1 by caspase-activated Pak2/gamma-PAK inhibits phosphorylation and interaction of eIF4G with Mnk. *J Biol Chem* **279**, 38649 (2004).

198. P. T. Tuazon *et al.*, Comparative analysis of phosphorylation of translational initiation and elongation factors by seven protein kinases. *J Biol Chem* **264**, 2773 (1989).

199. R. Kirchweger *et al.*, Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 gamma. *J Virol* **68**, 5677 (1994).

200. B. J. Lamphear *et al.*, Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of the 2A proteases from human Coxsackievirus and rhinovirus. *J Biol Chem* **268**, 19200 (1993).

201. A. M. Borman *et al.*, elF4G and its proteolytic cleavage products: effect on initiation of protein synthesis from capped, uncapped, and IRES-containing mRNAs. *RNA* **3**, 186 (1997).

202. A. G. Ryazanov *et al.*, Phosphorylation of the elongation factor 2: the fifth Ca2+/calmodulin-dependent system of protein phosphorylation. *Biochimie* **70**, 619 (1988).

203. X. Wang *et al.*, Regulation of elongation factor 2 kinase by p90(RSK1) and p70 S6 kinase. *EMBO J* **20**, 4370 (2001).

204. A. Knebel *et al.*, A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *EMBO J* **20**, 4360 (2001).

205. G. Leprivier *et al.*, The eEF2 Kinase Confers Resistance to Nutrient Deprivation by Blocking Translation Elongation. *Cell* **153**, 1064 (2013).

206. S. Park *et al.*, Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron* **59**, 70 (2008).

207. H. Tang *et al.*, Amino acid-induced translation of TOP mRNAs is fully dependent on phosphatidylinositol 3-kinase-mediated signaling, is partially inhibited by rapamycin, and is independent of S6K1 and rpS6 phosphorylation. *Mol Cell Biol* **21**, 8671 (2001).

208. M. Stolovich *et al.*, Transduction of growth or mitogenic signals into translational activation of TOP mRNAs is fully reliant on the phosphatidylinositol 3-kinase-mediated pathway but requires neither S6K1 nor rpS6 phosphorylation. *Mol Cell Biol* **22**, 8101 (2002).

209. C. K. Damgaard, J. Lykke-Andersen, Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes Dev* **25**, 2057 (2011).

210. O. Meyuhas, Synthesis of the translational apparatus is regulated at the translational level. *FEBS J* **267**, 6321 (2000).

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

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

213. T. V. Pestova *et al.*, Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* **16**, 6859 (1996).

214. S. Vagner *et al.*, Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep* **2**, 893 (2001).

215. C. H. Herbreteau *et al.*, HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat Struct Mol Biol* **12**, 1001 (2005).

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

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

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

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

220. E. Martinez-Salas *et al.*, IRES elements: features of the RNA structure contributing to their activity. *Biochimie* **84**, 755 (2002).

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

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

223. B. Stebbins-Boaz *et al.*, Maskin is a CPEB-associated factor that transiently interacts with elF-4E. *Mol Cell* **4**, 1017 (1999).

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

225. A. Nakamura *et al.*, Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. *Dev Cell* **6**, 69 (2004).

226. J. E. Wilhelm *et al.*, Cup is an eIF4E binding protein required for both the translational repression of oskar and the recruitment of Barentsz. *J Cell Biol* **163**, 1197 (2003).

227. F. Gebauer *et al.*, Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol Cell* **11**, 1397 (2003).

228. M. Grskovic *et al.*, A co-repressor assembly nucleated by Sex-lethal in the 3'UTR mediates translational control of Drosophila msl-2 mRNA. *EMBO J* **22**, 5571 (2003).

229. K. Beckmann *et al.*, A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. *Cell* **122**, 529 (2005).

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

231. A. J. Enright *et al.*, MicroRNA targets in Drosophila. *Genome Biol* **5**, R1 (2003).

232. A. Stark *et al.*, Identification of Drosophila MicroRNA targets. *PLoS Biol* **1**, E60 (2003).

233. V. Ambros, The functions of animal microRNAs. *Nature* **431**, 350 (2004).

234. D. P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281 (2004).

235. A. Krek *et al.*, Combinatorial microRNA target predictions. *Nat Genet* **37**, 495 (2005).

236. B. P. Lewis *et al.*, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15 (2005).

237. R. C. Lee *et al.*, The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843 (1993).

238. B. Wightman *et al.*, Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855 (1993).

239. P. Saetrom *et al.*, Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res* **35**, 2333 (2007).

240. Y. Tay *et al.*, MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124 (2008).

241. B. J. Reinhart *et al.*, The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901 (2000).

242. S. M. Hammond *et al.*, Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146 (2001).

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

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

245. H. S. Soifer *et al.*, MicroRNAs in disease and potential therapeutic applications. *Mol Therapy* **15**, 2070 (2007).

246. M. W. Rhoades *et al.*, Prediction of plant microRNA targets. *Cell* **110**, 513 (2002).

247. G. Tang, siRNA and miRNA: an insight into RISCs. *Trends Biochem Sci* **30**, 106 (2005).

248. H. Guo *et al.*, Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835 (2010).

249. S. Djuranovic *et al.*, miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**, 237 (2012).

250. P. Yao *et al.*, Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. *Cell* **149**, 88 (2012).

251. C. Mayr, D. P. Bartel, Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673 (2009).

252. H. A. Meijer *et al.*, Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science* **340**, 82 (2013).

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

254. M. Kozak, A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr* **1**, 111 (1991).

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

256. M. Iacono *et al.*, uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* **349**, 97 (2005).

257. D. E. Neafsey, J. E. Galagan, Dual modes of natural selection on upstream open reading frames. *Mol Biol Evol* **24**, 1744 (2007).

258. I. B. Rogozin *et al.*, Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics* **17**, 890 (2001).

259. M. Kozak, Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* **29**, 5226 (2001).

260. L. Rajkowitsch *et al.*, Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J Mol Biol* **335**, 71 (2004).

261. D. R. Morris, A. P. Geballe, Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**, 8635 (2000).

262. A. P. Geballe, M. S. Sachs, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 595-614.

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

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

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

266. Y. Chiba *et al.*, 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 (2003).

267. H. Onouchi *et al.*, Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the *CGS1* gene of Arabidopsis. *Genes Dev* **19**, 1799 (2005).

268. T. Nyiko *et al.*, Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* **71**, 367 (2009).

269. A. Gaba *et al.*, Ribosome occupancy of the yeast *CPA1* upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell* **20**, 449 (2005).

270. J. T. Mendell *et al.*, Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36**, 1073 (2004).

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

272. C. W. Tabor, H. Tabor, Polyamines. *Annu Rev Biochem* **53**, 749 (1984).

273. G. L. Law *et al.*, Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *J Biol Chem* **276**, 38036 (2001).

274. G. J. Mize *et al.*, 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 (1998).

275. H. Ruan *et al.*, The upstream open reading frame of the mRNA encoding *S*-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. *J Biol Chem* **271**, 29576 (1996).

276. A. Raney *et al.*, Regulated translation termination at the upstream open reading frame in S-adenosylmethionine decarboxylase mRNA. *J Biol Chem* **277**, 5988 (2002).

277. I. P. Ivanov *et al.*, A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res* **38**, 353 (2010).

278. M. A. Olayioye *et al.*, The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J* **19**, 3159 (2000).

279. Y. Yarden, M. X. Sliwkowski, Untangling the ErbB signalling network. *Nat Rev Mol Cell Bio* **2**, 127 (2001).

280. S. J. Child *et al.*, Cell type-dependent and -independent control of HER-2/neu translation. *Int J Biochem Cell Biol* **31**, 201 (1999).

281. S. J. Child *et al.*, Translational control by an upstream open reading frame in the HER-2/*neu* transcript. *J Biol Chem* **274**, 24335 (1999).

282. C. C. Spevak *et al.*, *her-2* upstream open reading frame effects on the use of downstream initiation codons. *Biochem Biophys Res Commun* **350**, 834 (2006).

283. A. Mehta *et al.*, Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes Dev* **20**, 939 (2006).

284. C. P. Chang *et al.*, Identification and expression of a human cytomegalovirus early glycoprotein. *J Virol* **63**, 3330 (1989).

285. C. R. Degnin *et al.*, Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol* **67**, 5514 (1993).

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

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

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

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

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

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

292. P. P. Mueller, A. G. Hinnebusch, Multiple upstream AUG codons mediate translational control of GCN4. *Cell* **45**, 201 (1986).

293. A. Gaba *et al.*, Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *EMBO J* **20**, 6453 (2001).

294. A. G. Hinnebusch *et al.*, Evidence for regulation of reinitiation in translational control of GCN4 mRNA. *Proc Natl Acad Sci USA* **85**, 7279 (1988).

295. P. F. Miller, A. G. Hinnebusch, 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 (1989).

296. K. H. Nielsen *et al.*, Functions of eIF3 downstream of 48S assembly impact AUG recognition and GCN4 translational control. *EMBO J* **23**, 1166 (2004).

297. K. A. Spriggs *et al.*, Translational regulation of gene expression during conditions of cell stress. *Mol Cell* **40**, 228 (2010).

298. T. E. Dever *et al.*, Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* **68**, 585 (1992).

299. F. Zhang, A. G. Hinnebusch, An upstream ORF with non-AUG start codon is translated in vivo but dispensable for translational control of GCN4 mRNA. *Nucleic Acids Res* **39**, 3128 (2011).

300. Y. Arava *et al.*, Dissecting eukaryotic translation and its control by ribosome density mapping. *Nucleic Acids Res* **33**, 2421 (2005).

301. J.-P. Abastado *et al.*, Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. *Mol Cell Biol* **11**, 486 (1991).

302. O. Grundmann *et al.*, Repression of *GCN4* mRNA translation by nitrogen starvation in *Saccharomyces cerevisiae*. *J Biol Chem* **276**, 25661 (2001).

303. M. S. Sachs, in *The Mycota: Biochemistry and Molecular Biology,* R. Brambl, G. A. Marzluf, Eds. (Springer-Verlag, Heidelberg, 1996), vol. III, pp. 315-345.

304. C. Tian *et al.*, Transcriptional profiling of cross pathway control in Neurospora crassa and comparative analysis of the Gcn4 and CPC1 regulons. *Eukaryot Cell* **6**, 1018 (2007).

305. K. Natarajan *et al.*, Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol* **21**, 4347 (2001).

306. J. L. Paluh *et al.*, 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 (1988).

307. C. Wanke *et al.*, The *Aspergillus niger GCN4* homologue, *cpcA*, is transcriptionally regulated and encodes an unusual leucine zipper. *Mol Microbiol* **23**, 23 (1997).

308. B. Hoffmann *et al.*, Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Mol Biol Cell* **12**, 2846 (2001).

309. Z. Luo *et al.*, Translational regulation in response to changes in amino acid availability in *Neurospora crassa*. *Mol Cell Biol* **15**, 5235 (1995).

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

311. H. P. Harding *et al.*, Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell* **6**, 1099 (2000).

312. N. Donnelly *et al.*, The eIF2alpha kinases: their structures and functions. *Cell Mol Life Sci* (2013).

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

314. M. J. Orbach *et al.*, 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 (1990).

315. Z. Wang *et al.*, The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol Cell Biol* **18**, 7528 (1998).

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

317. C. C. Spevak *et al.*, Sequence requirements for ribosome stalling by the arginine attenuator peptide. *J Biol Chem* **285**, 40933 (2010).

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

319. M. Freitag *et al.*, A UV-induced mutation in *Neurospora* that affects translational regulation in response to arginine. *Genetics* **142**, 117 (1996).

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

321. P. Fang *et al.*, A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci USA* **101**, 4059 (2004).

322. P. Fang *et al.*, Evolutionarily conserved features of the arginine attenuator peptide provide the necessary requirements for its function in translational regulation. *J Biol Chem* **275**, 26710 (2000).

323. Z. Wang *et al.*, 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 (1999).

324. C. Wu *et al.*, Arginine changes the conformation of the arginine attenuator peptide relative to the ribosome tunnel. *J Mol Biol* **416**, 518 (2012).

325. S. Bhushan *et al.*, Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **40**, 138 (2010).

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

327. P. Thuriaux *et al.*, Regulation of carbamoylphosphate synthetase belonging to the arginine biosynthetic pathway of *Saccharomyces cerevisiae*. *J Mol Biol* **67**, 277 (1972).

328. M. Werner *et al.*, 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 (1985).

329. P. Delbecq *et al.*, Functional analysis of the leader peptide of the yeast gene *CPA1* and heterologous regulation by other fungal peptides. *Curr Genet* **38**, 105 (2000).

330. P. Delbecq *et al.*, 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 (1994).

331. M. Crabeel *et al.*, 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 (1990).

332. A. Jacobson, S. W. Peltz, in *Translational control of gene expression,* N. Sonenberg *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 827-847.

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

334. M. J. Johansson *et al.*, Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc Natl Acad Sci USA* **104**, 20872 (2007).

335. T. K. Kim, T. Maniatis, Regulation of interferon-gamma activated STAT1 by the ubiquitin-proteasome pathway. *Science* **273**, 1717 (1996).

336. Y. Chiba *et al.*, Evidence for autoregulation of cystathionine gamma-synthase mRNA stability in *Arabidopsis*. *Science* **286**, 1371 (1999).

337. I. Lambein *et al.*, Decay kinetics of autogenously regulated CGS1 mRNA that codes for cystathionine gamma-synthase in *Arabidopsis thaliana*. *Plant Cell Physiol* **44**, 893 (2003).

338. N. Onoue *et al.*, S-adenosyl-L-methionine induces compaction of nascent peptide chain inside the ribosomal exit tunnel upon translation arrest in the Arabidopsis CGS1 gene. *J Biol Chem* **286**, 14903 (2011).

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

340. H. Nakatogawa *et al.*, Control of SecA and SecM translation by protein secretion. *Curr Opin Microbiol* **7**, 145 (2004).

341. M. G. Schmidt *et al.*, Nucleotide sequence of the secA gene and secA(Ts) mutations preventing protein export in Escherichia coli. *J Bacteriol* **170**, 3404 (1988).

342. K. Ito *et al.*, Divergent stalling sequences sense and control cellular physiology. *Biochem Biophys Res Commun* **393**, 1 (2010).

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

344. H. Nakatogawa *et al.*, SecM facilitates translocase function of SecA by localizing its biosynthesis. *Genes Dev* **19**, 436 (2005).

345. M. N. Yap, H. D. Bernstein, The translational regulatory function of SecM requires the precise timing of membrane targeting. *Mol Microbiol* **81**, 540 (2011).

346. H. Muto *et al.*, Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell* **22**, 545 (2006).

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

348. N. Vazquez-Laslop *et al.*, The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *EMBO J*, (2010).

349. K. Mitra *et al.*, Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol Cell* **22**, 533 (2006).

350. C. A. Woolhead *et al.*, Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell* **22**, 587 (2006).

351. M. N. Yap, H. D. Bernstein, The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell* **34**, 201 (2009).

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

353. L. R. Cruz-Vera *et al.*, Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. *Proc Natl Acad Sci USA* **103**, 3598 (2006).

354. F. Gong, C. Yanofsky, Analysis of tryptophanase operon expression in *vitro.* Accumulation of TnaC-peptidyl tRNA ina release factor 2-depleted S-30 extract prevents Rho factor action, simulating induction. *J Biol Chem* **277**, 17095 (2002).

355. F. Gong *et al.*, The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNAPro. *Proc Natl Acad Sci USA* **98**, 8997 (2001).

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

357. A. V. Kamath, C. Yanofsky, Characterization of the tryptophanase operon of *Proteus vulgaris*. Cloning, nucleotide sequence, amino acid homology, and in

vitro synthesis of the leader peptide and regulatory analysis. *J Biol Chem* **267**, 19978 (1992).

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

359. L. R. Cruz-Vera *et al.*, Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of *tna* operon expression. *Mol Cell* **19**, 333 (2005).

360. L. R. Cruz-Vera *et al.*, Ribosomal features essential for *tna* operon induction: tryptophan binding at the peptidyl transferase center. *J Bacteriol* **189**, 3140 (2007).

361. R. Yang *et al.*, 23S rRNA nucleotides in the peptidyl transferase center are essential for tryptophanase operon induction. *J Bacteriol* **191**, 3445 (2009).

362. B. Seidelt *et al.*, Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412 (2009).

363. A. K. Martinez *et al.*, Crucial elements that maintain the interactions between the regulatory TnaC peptide and the ribosome exit tunnel responsible for Trp inhibition of ribosome function. *Nucleic Acids Res* **40**, 2247 (2012).

364. H. Ramu *et al.*, Programmed drug-dependent ribosome stalling. *Mol Microbiol* **71**, 811 (2009).

365. D. Tu *et al.*, Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell* **121**, 257 (2005).

366. F. Schlunzen *et al.*, Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814 (2001).

367. T. J. Gryczan *et al.*, Conformational alteration of mRNA structure and the posttranscriptional regulation of erythromycin-induced drug resistance. *Nucleic Acids Res* **8**, 6081 (1980).

368. S. Horinouchi, B. Weisblum, Posttranscriptional modification of mRNA conformation: mechanism that regulates erythromycin-induced resistance. *Proc Natl Acad Sci USA* **77**, 7079 (1980).

369. N. Vazquez-Laslop *et al.*, Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell* **30**, 190 (2008).

370. N. Vazquez-Laslop *et al.*, Role of antibiotic ligand in nascent peptide-dependent ribosome stalling. *Proc Natl Acad Sci USA* **108**, 10496 (2011).

371. H. Ramu *et al.*, Nascent peptide in the ribosome exit tunnel affects functional properties of the A-site of the peptidyl transferase center. *Mol Cell* **41**, 321 (2011).

372. D. S. Peabody, Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031 (1989).

373. M. Kozak, Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res* **12**, 857 (1984).

374. M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283 (1986).

375. S. J. Chen *et al.*, Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. *J Biol Chem* **283**, 3173 (2008).

376. J. L. Portis *et al.*, Identification of a sequence in the unique 5' open reading frame of the gene encoding glycosylated Gag which influences the incubation period of neurodegenerative disease induced by a murine retrovirus. *J Virol* **68**, 3879 (1994).

377. S. E. Kolitz *et al.*, Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* **15**, 138 (2009).

378. J. M. Clements *et al.*, Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. *Mol Cell Biol* **8**, 4533 (1988).

379. I. P. Ivanov *et al.*, Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc Natl Acad Sci USA* **107**, 18056 (2010).

380. K. Gordon *et al.*, Efficient initiation of translation at non-AUG triplets in plant cells. *Plant J* **2**, 809 (1992).

381. K. J. Chang, C. C. Wang, Translation initiation from a naturally occurring non-AUG codon in Saccharomyces cerevisiae. *J Biol Chem* **279**, 13778 (2004).

382. T. F. Donahue, A. M. Cigan, Genetic selection for mutations that reduce or abolish ribosomal recognition of the HIS4 translational initiator region. *Mol Cell Biol* **8**, 2955 (1988).

383. C. P. Chang *et al.*, A single sequence context cannot satisfy all non-AUG initiator codons in yeast. *BMC Microbiol* **10**, 188 (2010).

384. N. T. Ingolia *et al.*, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789 (2011).

385. N. T. Ingolia *et al.*, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218 (2009).

386. C. Touriol *et al.*, Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell* **95**, 169 (2003).

387. K. Dementhon *et al.*, Characterization of IDI-4, a bZIP transcription factor inducing autophagy and cell death in the fungus Podospora anserina. *Mol Microbiol* **53**, 1625 (2004).

388. J. L. Riechmann *et al.*, Non-AUG initiation of AGAMOUS mRNA translation in Arabidopsis thaliana. *Mol Cell Biol* **19**, 8505 (1999).

389. B. Hedtke *et al.*, Six active phage-type RNA polymerase genes in Nicotiana tabacum. *Plant J* **30**, 625 (2002).

390. Y. Kobayashi *et al.*, Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in Nicotiana sylvestris. *Biochem Biophys Res Commun* **299**, 57 (2002).

391. S. Yamanaka *et al.*, A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes Dev* **11**, 321 (1997).

392. J. Sasaki, N. Nakashima, Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci USA* **97**, 1512 (2000).

393. S. R. Starck *et al.*, Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719 (2012).

394. J. E. Takacs *et al.*, Identification of compounds that decrease the fidelity of start codon recognition by the eukaryotic translational machinery. *RNA* **17**, 439 (2011).

395. H. Yoon, T. F. Donahue, The *sui1* suppressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during $tRNA_i^{Met}$ recognition of the start codon. *Mol Cell Biol* **12**, 248 (1992).

396. B. Castilho-Valavicius *et al.*, Genetic characterization of the *Saccharomyces cerevisiae* translational initiation suppressors *sui1*, *sui2* and *SUI3* and their effects on *HIS4* expression. *Genetics* **124**, 483 (1990).

397. A. K. Saini *et al.*, Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev* **24**, 97 (2010).

398. L. Valásek *et al.*, Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol Cell Biol* **24**, 9437 (2004).

399. H. He *et al.*, 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 (2003).

400. G. Loughran *et al.*, Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res* **40**, 2898 (2011).

401. D. Maag *et al.*, A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell* **17**, 265 (2005).

402. J. Rabl *et al.*, Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730 (2011).

403. A. Unbehaun *et al.*, Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* **18**, 3078 (2004).

404. Y.-N. Cheung *et al.*, Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection *in vivo*. *Genes Dev* **21**, 1217 (2007).

405. S. Das *et al.*, Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J Biol Chem* **276**, 6720 (2001).

406. F. E. Paulin *et al.*, Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr Biol* **11**, 55 (2001).

407. M. D. Jennings, G. D. Pavitt, eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. *Nature* **465**, 378 (2010).

408. M. S. Sachs, A. P. Geballe, Downstream control of upstream open reading frames. *Genes Dev* **20**, 915 (2006).

409. H. M. Hood *et al.*, Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. *Annu Rev Microbiol* **63**, 385 (2009).

410. L. Kurian *et al.*, Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* **477**, 490 (2011).

411. K. Yanagitani *et al.*, Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* **331**, 586 (2011).

412. A. Churbanov *et al.*, Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. *Nucleic Acids Res* **33**, 5512 (2005).

413. J. E. Galagan *et al.*, Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105 (2005).

414. L. R. Cruz-Vera *et al.*, Nascent polypeptide sequences that influence ribosome function. *Curr Opin Microbiol* **14**, 160 (2011).

415. G. A. Brar *et al.*, High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **334**, 522 (2011).

416. S. Chiba *et al.*, Recruitment of a species-specific translational arrest module to monitor different cellular processes. *Proc Natl Acad Sci USA* **108**, 6073 (2011).

417. N. Vazquez-Laslop, A. S. Mankin, Picky nascent peptides do not talk to foreign ribosomes. *Proc Natl Acad Sci USA* **108**, 5931 (2011).

418. D. Dubnau, Induction of ermC requires translation of the leader peptide. *EMBO J* **4**, 533 (1985).

419. M. G. Lawrence *et al.*, Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel. *J Bacteriol* **190**, 5862 (2008).

420. P. S. Lovett, Translation attenuation regulation of chloramphenicol resistance in bacteria--a review. *Gene* **179**, 157 (1996).

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

422. D. M. Janzen *et al.*, Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Mol Cell Biol* **22**, 8562 (2002).

423. Z. Wang, M. S. Sachs, 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 (1997).

424. C. Wu *et al.*, in *Translation initiation: extract systems and molecular genetics,* J. Lorsch, Ed. (Elsevier, San Diego, 2007), vol. 429, pp. 203-225.

425. P. Fang *et al.*, *Neurospora crassa* supersuppressor mutants are amber codon-specific. *Fungal Genet Biol* **36**, 167 (2002).

426. R. L. Davis, in *The Mycota: Biochemistry and Molecular Biology,* R. Brambl, G. A. Marzluf, Eds. (Springer-Verlag, Heidelberg, 1996), vol. III, pp. 347-356.

427. J. L. Hansen *et al.*, Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit. *J Mol Biol* **330**, 1061 (2003).

428. H. Muto, K. Ito, Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. *Biochem Biophys Res Commun* **366**, 1043 (2008).

429. U. Varshney *et al.*, Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutaminyl-tRNA synthetase. *J Biol Chem* **266**, 24712 (1991).

430. A. Pavesi *et al.*, Identification of new eukaryotic tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions. *Nucleic Acids Res* **22**, 1247 (1994).

431. M. S. Sachs *et al.*, Toeprint analysis of the positioning of translational apparatus components at initiation and termination codons of fungal mRNAs. *Methods* **26**, 105 (2002).

432. D. Pitonzo *et al.*, Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum Sec61 translocon. *Mol Biol Cell* **20**, 685 (2009).

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

434. S. Chiba *et al.*, A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J* **28**, 3461 (2009).

435. J. S. Kieft, Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* **33**, 274 (2008).

436. M. Kozak, At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* **196**, 947 (1987).

437. J. N. Vaughn *et al.*, Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA* **18**, 368 (2012).

438. F. Zhang, A. G. Hinnebusch, An upstream ORF with non-AUG start codon is translated in vivo but dispensable for translational control of GCN4 mRNA. *Nucleic Acids Res* **39**, 3128 (2011).

439. E. H. Park *et al.*, Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1*PABP mRNPs in vivo. *EMBO J* **30**, 302 (2011).

440. C. O. Wilke, D. A. Drummond, Signatures of protein biophysics in coding sequence evolution. *Curr Opin Struct Biol* **20**, 385 (2010).

441. H. L. Tang *et al.*, Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. *J Biol Chem* **279**, 49656 (2004).

442. A. C. Christensen *et al.*, Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons. *Plant Cell* **17**, 2805 (2005).

443. K. Takahashi *et al.*, Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). *Genomics* **85**, 360 (2005).

444. J. L. Riechmann *et al.*, Non-AUG initiation of AGAMOUS mRNA translation in *Arabidopsis thaliana*. *Mol Cell Biol* **19**, 8505 (1999).

445. G. G. Simpson *et al.*, Noncanonical translation initiation of the Arabidopsis flowering time and alternative polyadenylation regulator FCA. *Plant Cell* **22**, 3764 (2010).

446. P. Martin-Marcos *et al.*, Functional elements in initiation factors 1, 1A, and 2beta discriminate against poor AUG context and non-AUG start codons. *Mol Cell Biol* **31**, 4814 (2011).

447. G. Loughran *et al.*, Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res* **40**, 2898 (2012).

448. W. Xie *et al.*, The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proc Natl Acad Sci USA*, (2012).

449. G. H. Jacobs *et al.*, Transterm: a database to aid the analysis of regulatory sequences in mRNAs. *Nucleic Acids Res* **37**, D72 (2009).

450. K. McCluskey *et al.*, The Fungal Genetics Stock Center: a repository for 50 years of fungal genetics research. *J Biosci* **35**, 119 (2010).

451. B. S. Margolin *et al.*, Improved plasmids for gene targeting at the *his-3* locus of *Neurospora crassa* by electroporation. *Fungal Genet Newsl* **44**, 34 (1997).

452. D. Ebbole, M. S. Sachs, A rapid and simple method for isolation of *Neurospora crassa* homokaryons using microconidia. *Fungal Genet Newsl* **37**, 17 (1990).

453. H. J. Vogel, A convenient growth medium for *Neurospora* (Medium N). *Microbiol Genet Bull* **13**, 42 (1956).

454. B. W. Dyer *et al.*, A noncommercial dual luciferase enzyme assay system for reporter gene analysis. *Anal Biochem* **282**, 158 (2000).

455. M. S. Sachs, C. Yanofsky, Developmental expression of genes involved in conidiation and amino acid biosynthesis in *Neurospora crassa*. *Dev Biol* **148**, 117 (1991).

456. D. Grundemann, E. Schomig, Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *Biotechniques* **21**, 898 (1996).

457. I. P. Ivanov *et al.*, Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res* **39**, 4220 (2011).

458. S. F. Altschul *et al.*, Basic local alignment search tool. *J Mol Biol* **215**, 403 (1990).

459. S. Nakagawa *et al.*, Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res* **36**, 861 (2008).

460. M. Fresno *et al.*, Inhibition of translation in eukaryotic systems by harringtonine. *FEBS J* **72**, 323 (1977).

461. F. Robert *et al.*, Altering chemosensitivity by modulating translation elongation. *PloS one* **4**, e5428 (2009).

462. H. Conlon *et al.*, The Aspergillus nidulans GATA transcription factor gene areB encodes at least three proteins and features three classes of mutation. *Mol Microbiol* **40**, 361 (2001).

463. S. Gutierrez *et al.*, Characterization of the Cephalosporium acremonium pcbAB gene encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the pcbC gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. *J Bacteriol* **173**, 2354 (1991).

464. S. Lee *et al.*, Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci USA* **109**, E2424 (2012).

465. C. Fritsch *et al.*, Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res* **22**, 2208 (2012).

466. I. P. Ivanov *et al.*, uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc Natl Acad Sci USA* **105**, 10079 (2008).

467. K. A. Borkovich *et al.*, Lessons from the genome sequence of Neurospora crassa: tracing the path from genomic blueprint to multicellular organism. *Microbiol Mol Biol Rev* **68**, 1 (2004).

468. T. F. Donahue, A. M. Cigan, Genetic selection for mutations that reduce or abolish ribosomal recognition of the *HIS4* translational initiator region. *Mol Cell Biol* **8**, 2955 (1988).

469. L. Valásek *et al.*, Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *EMBO J* **20**, 891 (2001).

470. R. E. Luna *et al.*, The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2beta. *Cell Rep* **1**, 689 (2012).

471. C. R. Singh *et al.*, Sequential eukaryotic translation initiation factor 5 (eIF5) binding to the charged disordered segments of eIF4G and eIF2beta stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol Cell Biol* **32**, 3978 (2012).

472. M. J. Johansson, A. Jacobson, Nonsense-mediated mRNA decay maintains translational fidelity by limiting magnesium uptake. *Genes Dev* **24**, 1491 (2010).

473. A. Zhang *et al.*, Extracellular Mg2+ regulates intracellular Mg2+ and its subcellular compartmentation in fission yeast, Schizosaccharomyces pombe. *Cell Mol Life Sci* **53**, 69 (1997).

474. A. V. Kochetov *et al.*, Hidden coding potential of eukaryotic genomes: nonAUG started ORFs. *J Biomol Struct Dyn* **1**, 103 (2013).

475. M. Zhou *et al.*, Non-optimal codon usage affects expression, structure and function of FRQ clock protein. *Nature* **495**, 111 (2013).

476. A. G. Hinnebusch, K. Natarajan, Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. *Eukaryot Cell* **1**, 22 (2002).

477. M. H. Kuo *et al.*, Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Mol Cell* **6**, 1309 (2000).

478. D. J. Ebbole *et al.*, *cpc-1*, the general regulatory gene for genes of amino acid biosynthesis in *Neurospora crassa*, is differentially expressed during the asexual life cycle. *Mol Cell Biol* **11**, 928 (1991).

479. H. Tournu *et al.*, Global role of the protein kinase Gcn2 in the human pathogen Candida albicans. *Eukaryot Cell* **4**, 1687 (2005).

480. A. G. Hinnebusch *et al.*, in *Translational control in biology and medicine,* M. B. Mathews *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007), pp. 225-268.

481. E. Sattlegger *et al.*, *cpc-3*, the *Neurospora crassa* homologue of yeast *GCN2*, encodes a polypeptide with juxtaposed eIF2alpha kinase and histidyl-tRNA synthetase-related domains required for general amino acid control. *J Biol Chem* **273**, 20404 (1998).

482. R. Lucking *et al.*, Fungi evolved right on track. *Mycologia* **101**, 810 (2009).

483. C. F. Calkhoven *et al.*, Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes Dev* **14**, 1920 (2000).

484. S. Kervestin, A. Jacobson, NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13**, 700 (2012).