Inhibition de la traduction des ARN messagers Hox par les éléments de type TIE

Fatima Alghoul

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

IBMC, UPR 9002 du CNRS, Architecture et Réactivité de l'arN (ARN)

THÈSE

présentée par:

Fatima ALGHOUL

soutenue le: 26 septembre 2019

pour obtenir le grade de: **Docteur de l'Université de Strasbourg**
Discipline: Sciences de la vie et de la santé
Spécialité: Aspects moléculaires et cellulaires de la biologie

The mechanism of inhibition of cap-dependent translation by the Translation Inhibitory Elements (TIE) a3 and a11 in Hox mRNAs

Directeur de thèse:
Dr. MARTIN Franck — Directeur de Recherche, UPR 9002 du CNRS, IBMC

Rapporteurs:
Dr. SCHMITT Emmanuelle — Directrice de Recherche, UMR 7654 du CNRS, Ecole Polytechnique
Dr. MORALES Julia — Directrice de Recherche, UMR 8227 du CNRS, LBI2M

Examinateur:
Dr. MOINE Hervé — Directeur de Recherche, UMR7104 du CNRS, Inserm U964, IGBMC

Acknowledgments

First, I would like to thank my thesis jury members: Dr. Hervé Moine who accepted to review my work, Dr. Julia Morales and Dr. Emmanuelle Schmitt who accepted to evaluate my work. It is an honor to have you on my thesis jury and looking forward to discuss my work with you.

I thank Dr. Pascale Romby who was very supportive during my thesis and welcome me in the unit.

I would like to thank Dr. Gilbert Eriani who provided me with the opportunity to join his team first as an intern then as a PhD student. Without his precious support, it would have not been possible to conduct my research. I also thank him for the fruitful discussions and advices. Chef, you are the best!

I would like to express my greatest gratitude to my supervisor Dr. Franck Martin. It was one of my best moments when he accepted me as a master student and prepared me for the Ecole Doctoral competition. I would have not done it without him. Knowing how great he is, I was so excited and passionate to do a PhD. I thank him for the continuous support of my PhD study, for his patience, motivation and immense knowledge. Franck was not just my supervisor; he was my mentor who guided me through this journey. I owe him what I have learnt in science and life in these four years and I thank him for this unforgettable beautiful experience. I also thank his beautiful family Véronique, Amélie and Pauline who were very supportive and welcoming.

I would like to thank our lab 333 members:

Dr. Christine Allmang for the fruitful discussions and advices on my project. It was great to discuss with her my experiments and get advices on future post doc positions.

Laure Schaeffer who has been my tutor and a sister during my thesis. It would've never been the same without her. She didn't just teach me lab management and experimental tricks, but

she also taught me how to act like myself all the time without worrying about anything. Thank you Laure.

Hassan Hayek who has been a brother and extremely supportive in the last few months. I will miss working next to him in the lab! It was so much fun to have him around.

Aurélie Janvier, whom I consider a little sister. Thank you for the cookie competitions that were always exciting and competitive but I always won them!

Aurélie Durand who has been extremely fun to have her around especially the unforgettable experience of the escape game!

Antonin Tidu who joined the lab last year but felt like I knew him since ever! He is always motivating and supportive.

Lab 333 is unforgettable. I had the best memories in my life with these people.

I would like to thank past lab members:
Lauriane Gross for the nice memories in two years, it was great to have her in the lab.
Justine Mailliot who was very motivating and supportive. I had great discussions with her and she has always been so genuine and optimistic.
Onaïs, Lea, Manon, Zina, Anaïs, Bastien. It was great to have you on the team. You were supportive and we miss you all.

I would like to thank the proteomic platform members: Philippe Hammann, Johana Chicher and Lauriane Kuhn. Thanks for arranging my mass spectrometry results. They were very helpful.

I would like to thank my collaborators :Dr. Jean Luc Imler, Dr. Carine Meignin and Loïc Talide.

I would like to thank all members of the unit for making it a pleasant environment to work in.

I would like to thank my best friend in Strasbourg Elie, who was extremely supportive and caring. I'm lucky to have had him during these times. Eliane, an unforgettable person whom I loved to have around. Ahmad who is very supportive and helpful.

Great thanks to my family members:
My mom, Zahra, Zeinab, Hussein, Hassan, Iman, Nader and Sarah who gave me hope in the most difficult times of my life. We have been through a lot in the last few years and I would have not done this without their unlimited support and love. I look up to them in life and I carry all of them in my heart. They encouraged me to travel and fulfill this dream. They are the reason of my success. I'm very lucky to have such a beautiful family.
To my cute nieces Elena and Eva, I hope they chase their dreams one day and become successful women.

To my fiancée, Mohamad. He has been amazing in the last three years. A man with a beautiful soul who is extremely patient and supportive of me. He shared this experience with me and listened to me talking passionately about science all the time. I owe him my strength and love in life. He is one of the reasons of my success. It wouldn't have been the same without him...To my family in law members who have been extremely nice and supportive during these years.

To my dad..
The man who is no longer with us today but lives through me. This work is his dream as much as it is mine. I wouldn't have done this without his support. I dedicate this work to my guardian angel and wish he lived to see this work coming out. This dream is coming true and it's just the beginning like you always imagined. I hope you are proud of me..
"Dad, your guiding hand on my shoulder will remain with me forever..."

To life... for this unforgettable beautiful experience...

TABLE OF CONTENTS

ABBREVIATIONS

3'CITE	3' Cap-Independent Translation Enhancer
4E-BP	4E-Binding Protein
4E-SE	4E-Sensitive Element
AMV-RT	Avian Myeloblastosis Virus Reverse-Transcriptase
Antp	Antennapedia
CAT-1	Cationic Amino acid Transporter 1
CBP	Cap Binding Pocket
cDNA	complementary deoxyribonucleic acid
circRNA	Circular RNA
CL	Cloverleaf
CMCT	1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate
CrPV	Cricket Paralysis Virus
Cryo-EM	Cryo-Electron Microscopy
DAP5	Death-associated protein 5
DCV	Drosophila C Virus
DDR	DNA-Damage Response
DENR	Density regulated protein
DMS	DiMethyl Sulfate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
eEFs	eukaryotic Elongation Factors
eIFs	eukaryotic Initiation Factors
EMCV	Encephalo Myocarditis Virus
eRF	eukaryotic Release Factor
FC	Fold Change
FGF	Fibroblast Growth Factor
FMDV	Foot-and-Mouth Disease Virus
FMRP	Fragile X Mental Retardation Protein
GDP	Guanosine DiPhosphate
GEF	Guanine nucleotide Exchange Factor
GTP	Guanosine TriPhosphate
HCV	Hepatitis C Virus
HD	Homeodomain
HEAT-1	Huntingtin, eEF3, protein phosphatase 2A, kinase TOR1
Hox	Homeobox
HRI	Heme-Regulated eIF2α kinase
HX	Hexapeptide
IFRD1	Interferon-Related Developmental regulator 1
IGR	Intergenic Region IRES
INFG	Ìnterferon gamma
IRE	Iron-Response Element
IRES	Internal Ribosome Entry Site
ITAF	IRES *Trans*-Acting Factor
LGTN	Ligatin

m6A	N6-methyladenosine
MCTS1	Multiple Copies T-cell lymphoma Subunit 1
MDM2	Murine Double Minute 2
MEIS	Myeloid Ecotropic viral Integration Site
mTOR	mammalian Target Of Rapamycin
mTORC1	mammalian Target Of Rapamycin Complex 1
mRNA	messenger RNA
NHEJ	Non-Homologous End Joining
NMD	Non-sense Mediated Decay
nt	nucleotides
ORF	Open Reading Frame
PABP	Poly(A)-Binding Protein
PBC	Pre-B Cell
PCBP2	Poly (rC) Binding Protein 2
PDGF	Platelet-Derived Growth Factor
PERK	(PKR)-like Endoplasmic Reticulum Kinase
PIC	Pre-Initiation Complex
PK	Pseudo-knot
PKR	Protein Kinase R
PTB	Polypyrimidine Tract-Binding Protein
PV	Poliovirus
RACK1	Receptor for Activated C Kinase 1
RBP	RNA Binding Protein
RLU	Relative Light Units
RNA	Ribonucleic acid
RNPs	Ribonucleoprotein complexes
RpL	Ribosomal Protein from large subunit
RpS	Ribosomal Protein from small subunit
RRL	Rabbit Reticulocyte Lysate
RRM	RNA-Recognition Motif
rRNA	ribosomal RNA
SL	Stem-Loop
TALE	Three Amino acid Loop Exension
TC	Ternary Complex
TIE	Translation Inhibitory Element
TWJ	Three-way Junction
tRNA	transfer RNA
$tRNAi^{Met}$	Initiator Methionine tRNA
Ubx	Ultrabithorax
uORF	Upstream Open Reading Frame
UTR	Untranslated Region
VEGF	Vascular Endothelial Growth Factor
WGE	Wheat Germ Extract
WH	Winged Helix
YB1	Y-box protein 1

INTRODUCTION

INTRODUCTION

1. Gene expression: a highly regulated process

Gene expression constitutes an indispensible cellular process for which the genetic information is converted into proteins. The process of gene expression involves two main stages called transcription and translation. Transcription is the process by which the messenger RNA is synthesized by RNA polymerase II from the genomic DNA template. Transcription is a highly regulated process. The control of eukaryotic gene expression is primarily at the level of transcriptional initiation (Lee and Young, 2013). In addition, the packaging of DNA into chromatin and the epigenetic modifications (such as methylation), control the accessibility of the transcriptional machinery to the DNA sequences. Transcription initiation is regulated by transcriptional activators, which are *trans*-acting factors that bind regulatory DNA sequences called enhancers and promote transcription initiation. These factors contain two independent domains, the DNA binding domain and the activation domain that interacts with other components of the transcriptional machinery (Thomas and Chiang, 2006). Gene expression can also be down regulated by repressors that compete with activators for binding to regulatory DNA sequences. When they bind to a promoter or enhancer, these repressors block the binding of activators thereby inhibiting transcription (Reynolds et al., 2013). During transcription, numerous so-called co-transcriptional modifications of messenger RNAs in eukaryotes take place through events including capping at the 5' end of nascent mRNA and polyadenylation a their 3' end to produce the mature mRNA.

2. Mechanism of cap-dependent translation in eukaryotes

2.1. Initiation step

Translation of the majority of cellular mRNAs initiates by a cap-dependent mechanism. It involves a large number of auxiliary proteins termed eukaryotic Initiation Factors (eIFs). They are required for the attachment of the ribosomes on mRNA (Hinnebusch, 2014). Translation initiation results in the positioning of the AUG start codon of mRNA in the Peptidyl site (P-site) of the ribosome, base-paired with the anticodon of methionyl-tRNA$^{Met}_{i}$. First, the 5' m^{7}G cap is specifically bound by a tripartite factor called eIF4F complex, which comprises the eIF4E cap binding protein, eIF4G a platform protein and the RNA helicase eIF4A with its auxiliary factor eIF4B. Simultaneously, the Poly (A) tail at the 3' end is bound by the Poly (A)-Binding Protein (PABP). Then, a so-called 43S Pre-Initiation Complex (PIC)

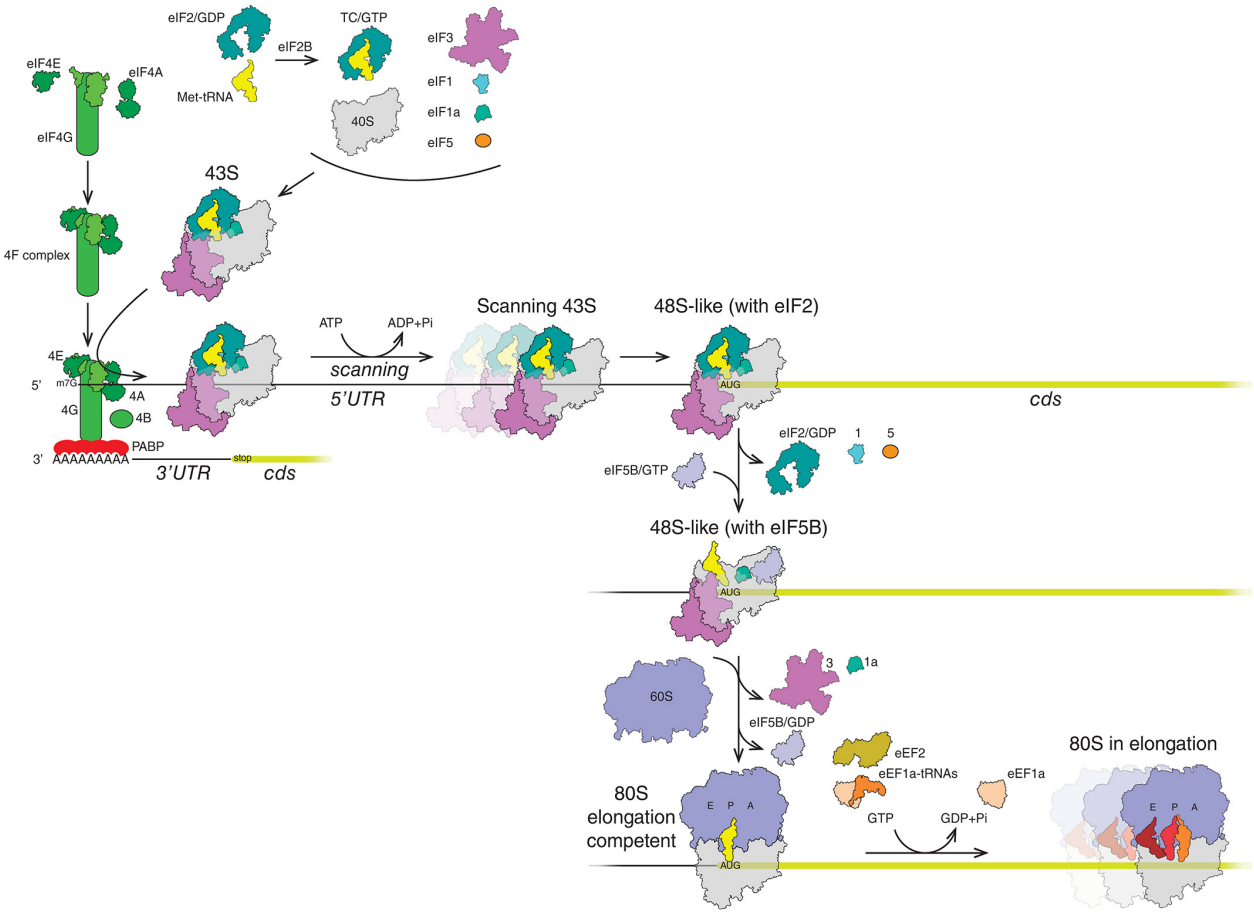


Figure 1: **Cap-dependent translation in eukaryotes.** Figure adapted from (Mailliot and Martin, 2018). The mechanism of canonical cap-dependent translation requires several steps. First, the Ternary Complex in its GTP form (TC-GTP) associates with the small 40S ribosomal subunit together with eIF3, eIF5, eIF1 and eIF1A to form the so-called 43S Pre Initiation Complex (PIC). In parallel, the eIF4F complex comprising eIF4E, eIF4G, eIF4A and eIF4B (and eIF4H in mammalians) is recruited on the 5' m^7G cap. Upon binding, the PIC is recruited to the 5'UTR of the mRNA due to specific interactions between eIF4G and eIF3. Then the 43S PIC scans along the 5'UTR until the AUG start codon is encountered. Upon recognition of the AUG codon, the GTP molecule bound to eIF2 is hydrolysed to form the 48S PIC. Then eIFs are released by the hydrolysis of GTP-eIF5B to GDP which finally allows the recruitment of the 60S large ribosomal subunit to form the 80S complex. This marks the -end of initiation step and start of elongation step.

assembles. The 43S PIC comprises the small 40S ribosomal subunit, the ternary complex containing eIF2–GTP–Met-tRNA$^{Met}_i$ (eIF2-TC), and the following factors eIF3, eIF1, eIF1A and eIF5 (**Fig. 1**). The 43S PIC is recruited to the 5' cap by an interaction between the platform eIF4G and eIF3 from the 43S PIC. Since eIF4G also interacts with PABP, this interaction promotes mRNA circularisation (Paek et al., 2015). However, the circularisation is not observed for all messenger mRNAs (Adivarahan et al., 2018). Then, the 43S particle

undergoes a so-called scanning step through the 5'untranslated region (5'UTR) from the 5' to 3' until it reaches the AUG start codon (Hinnebusch, 2014). The secondary structures that might block scanning step are unwound by the RNA helicase eIF4A and its auxiliary factor eIF4B. The ribosome initiates at AUG start codons that are embedded in a optimal context GCC(A/G)CCAUGG called the Kozak sequence with a purine at -3 and G at +4 (by convention, the A from the AUG codon is numbered +1) (Kozak, 1986). After AUG codon recognition, the complex is now termed 48S PIC. In this 48S complex, eIF5 and eIF5B promote the hydrolysis of eIF2-bound GTP thereby inducing the release of eIFs, which further allows the joining of the 60S large ribosomal subunit. The final 80S complex is now fully assembled marking the end of translation initiation step (Jackson et al., 2010) (**Fig. 1**). The second codon is positioned in the Aminoacyl site (A-site) of the ribosome and is ready to accept the cognate aminoacyl-tRNA. The translation process enters into elongation phase to synthesize the first peptide bond in the Peptidyl-Transferase Center (PTC) of the ribosome. The translation initiation step is the rate-limiting step and therefore it determines the translational landscape in the cell. To ensure fine-tuning of translation, this step is highly regulated by numerous *trans*-acting factors (Sonenberg and Hinnebusch, 2009).

2.2. Elongation step

Following initiation, an 80S ribosome is positioned on the mRNA with the anticodon of Met-tRNA$^{Met}_{i}$ base-paired with the AUG start codon in the P-site. The second codon of the open reading frame (ORF) is then placed in the A-site of the ribosome waiting for the binding of the cognate aminoacyl-tRNA. Then, the first peptide bond is formed in the PTC of the ribosome. Next, the so-called translocation step enables the 80S ribosome to move from one codon. Translocation requires three Elongation Factors (eEFs): eEF1A, eEF1B, and eEF2. The eEF1A protein (the homologue of EF-Tu in prokaryotes) is a G-protein with a low GTP hydrolysis rate that favours high fidelity decoding (Schmeing et al., 2011). In fact, with this low rate, it prevents the non-cognate aa-tRNAs from binding to the ribosome and mediates the recruitment of the cognate tRNA into the A-site of 80S ribosome (Becker et al., 2013). The second factor, eEF1B, is a tripartite complex which acts as a Guanine nucleotide exchange factor. Its function is to recycle eEF1A-GDP into its active eEF1A-GTP form (Cao et al., 2014). The third factor, eEF2 is a G-protein that enables the translocation of codon-anticodon duplex through the ribosome and triggers the movement from A- to the P-site to free the A-site for the next aminoacyl-tRNA (Susorov et al., 2018).

2.3. Termination step

Translation termination occurs when the ribosome reaches the stop codon that is positioned in the ribosomal A-site. There are three types of stop codons, UAA, UAG or UGA. This step is catalysed by two Release Factors (eRFs): class I or eRF1 and class II or eRF3 (Alkalaeva et al., 2006; Blanchet et al., 2015). The first factor eRF1 recognizes all the three stop codons with high fidelity and mediates peptidyl-tRNA hydrolysis. The second factor eRF3 is a translational GTPase which accelerates the efficiency of termination and the release of peptide in a GTP-dependent manner (Atkinson et al., 2008). Interestingly, eRF1 and eRF3 interact with very high affinity and probably enter the ribosome as a complex (Pisareva et al., 2006). The structure of the termination complex has been solved by cryo-EM (Brown et al., 2015) (**Fig. 2**). It seems that the +4 nucleotide position of the stop codon is required for interaction with the release factor eRF1. In fact, in eukaryotes, there is a specific preference for purines (A/G) at +4 position of stop codons (Brown et al., 1990).

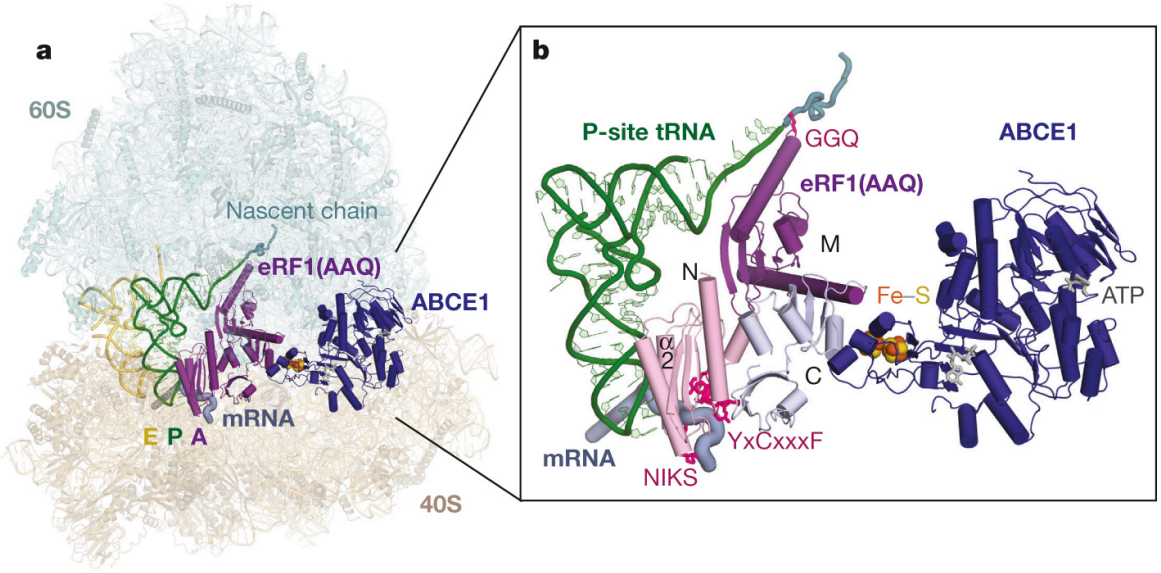


Figure 2: **Overall structure of a eukaryotic translation termination complex.** Figure adapted from (Brown et al., 2015)**.** Overview of the eRF1(AAQ)-stalled mammalian ribosome-nascent chain complex containing the UAG stop codon. The 40S (yellow) and 60S (blue) ribosomal subunits are displayed as well as E- (orange) and P-site (green) tRNAs, eRF1(AAQ) occupying the A-site, and ABCE1 occupying the GTPase center. **b**, Close-up view of eRF1(AAQ) coloured by domain (N, M, C) with the GGQ, NIKS and YxCxxxF motifs highlighted (pink). Also shown are the P-site tRNA, nascent polypeptide, the mRNA containing the UAG stop codon and ABCE1 with its iron–sulfur clusters and ATP-binding sites indicated.

3. Regulation of translation initiation steps in eukaryotes

3.1. Regulation of 43S pre-initiation complex assembly and binding to mRNA

The process of 43S PIC assembly depends on either the integration of a small 40S ribosomal subunit from the pool of free 40S or from 40S subunits generated by ribosome recycling after the termination of mRNA translation (Jackson et al., 2012). The hierarchical or random binding of the factors to the 40S subunit is not fully understood so far. However, studies have shown that the factors eIF1, eIF2, eIF3 and eIF5 can constitute a Multifactor Complex (MFC) without the presence of 40S subunit suggesting that this complex can be formed independently prior to its binding on the 40S subunit (Sokabe et al., 2012). In addition, other studies suggest that some surface-exposed residues in eIF1 are involved in its ability to promote Ternary Complex (TC) recruitment to the 40S subunit, a critical step in 43S PIC assembly (Cheung et al., 2007).

3.2. Regulation of scanning and start codon selection

Recruitment of the 43S PIC might be problematic if structures are in close proximity to the 5' m^7G cap because they prevent proposer eIF4F binding (Babendure et al., 2006; Livingstone et al., 2010). Upon recruitment, the 43S PIC will scan along the 5'UTR for AUG codon selection. The 43S PIC scanning might be also be disturbed by the presence of highly stable secondary structures located within the 5'UTR of mRNAs in between the 5' cap and the AUG start codon. For mRNAs that contain such secondary structures in the 5'UTR, they require the ATP-dependent activity of eIF4A RNA helicase to unwind the structures in order to enable efficient 43S PIC assembly and smooth progression of the 43S PIC during scanning (Pestova and Kolupaeva, 2002). It has been previously established that eIF4A is the only helicase involved in initiation pathway, however, recent studies have revealed the implication of other helicases in specific mRNAs translation initiation (Parsyan et al., 2011). One example is the DEAD-box helicase DDX3 (Ded1 in yeast), a helicase required for the unwinding of secondary structures near the m^7G cap structure in some mRNAs (Soto-Rifo et al., 2012). In yeast, Ded1 has also been found to be required to unwind secondary structures located downstream of start codons from upstream ORF (uORF) in 5'UTR (Guenther et al., 2018). Another example is DHX29 which is thought to predominantly promote translation of mRNAs containing highly stable structures (Pisareva et al., 2008). Interestingly, the silencing of DHX29 has an inhibitory effect on global translation and induces the reduction of polysome formation (Parsyan et al., 2009). The role of RNA helicases will be further

discussed in section (5.1.4). The 43S scanning will pause when a genuine AUG codon is placed in the P-site of ribosome. The proper selection of an AUG codon during scanning is another key step in translation initiation that requires the hydrolysis of GTP from the TC in its eIF2-GTP-Met-tRNA$^{Met}_{i}$ form and the subsequent release of Pi from eIF2-GDP-Pi. Interestingly, eIF1 plays an important role in this step whereby it promotes an open conformation of the mRNA binding site in the 40S subunit when non-AUG codons occupy the P-site of the ribosome (Passmore et al., 2007). Indeed, these non-AUG codons are not recognized because the release of a Pi from the partially hydrolysed eIF2-GDP-Pi in the PIC is blocked. This blocking is abrogated when a genuine AUG codon enters the P-site. Then, eIF1 dissociates from its location near the P site (Algire et al., 2005). In fact, mutations in eIF1 have been showed to increase the possibility of initiation at non-AUG (e.g. UUG) codons by facilitating the premature dissociation of eIF1 (Cheung et al., 2007). Non-AUG codons such as CUGs and GUGs can occasionally be used as start codons. These non-AUG initiation occurrences are used to generate or regulate proteins with key cellular functions in fundamental processes such as development or stress responses (Ingolia et al., 2011; Kearse and Wilusz, 2017).

3.3. Importance of initiation codon flanking sequences

The optimal context for AUG codon in mammals is GCC(A/G)CC**AUG**G with a purine at -3 and with G at +4 being considered as the most important flanking residues (Kozak, 1987). However, as previously mentioned, it has been shown that certain eukaryotic mRNAs contain non-AUG codons as initiation codon with CUG being the most efficient and frequent in mammals. These non-AUG codons are also embedded in the same optimal context as canonical AUG codons (Chen et al., 2008; Ingolia et al., 2011). For example, the mammalian Negative Elongation Factor NELF-B mRNA is translated from a non-AUG codon because it is surrounded by a more favourable context of the Kozak sequence than the further downstream AUG codon (Pan et al., 2015). The rationale of optimal Kozak sequences came from cryo-EM images showing specific contacts between -3 and +4 residues with eIF2α and eIF1a respectively (**Fig. 3**). While cross-linking methods have shown contacts with some ribosomal proteins in the 48S complex, specifically between -3/-4 residues with RpS5/uS7 and +4/+5 with RpS15/uS19 (Hussain et al., 2014; Pisarev et al., 2008). Translation initiator of Short 5'UTR (TISU) elements on the other hand do not follow the rules of canonical initiation since the 5'UTR harbours a median length of 12 nucleotides. The TISU elements consist of SAAS**AUG**GCGGC motif in which S is C or G. (Elfakess and Dikstein, 2008;

Elfakess et al., 2011). These elements direct accurate translation initiation in a so far unknown mechanism.

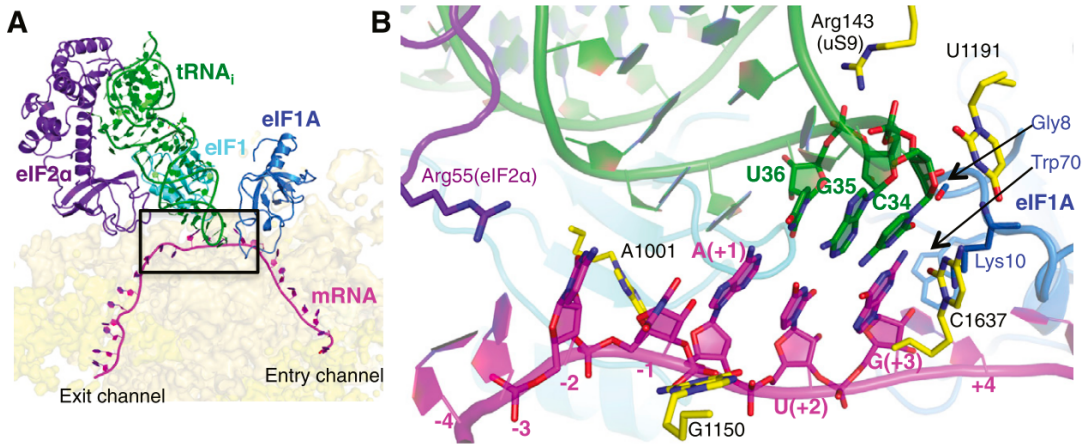


Figure 3: **Codon-anticodon interaction between $tRNA_i$ and mRNA.** Figure adapted from (Hussain et al., 2014). (A) Cross-section of the 40S ribosomal subunit showing the $tRNA_i$ in the P site of the ribosome and the mRNA in its channel. Initiation factors eIF1A, eIF1, and eIF2α are also shown. (B) A close-up view of the codon-anticodon helix and surrounding elements that stabilize this interaction. The α subunit of eIF2 structurally mimics an E-site tRNA and makes contacts with key nucleotides at the positions -2 and -3 in the mRNA. Residue Trp70 from the globular domain of eIF1A makes stacking interactions with the +4 nucleotide of mRNA

3.4. Mechanism of 60S subunit joining

Subsequent to AUG recognition, the mechanism that allows the joining of the 60S ribosomal subunit to the PIC requires GTP-hydrolysis by eIF5B (Cheung et al., 2007). Then, a specific interaction of eIF5B with the C-terminus of eIF1A promotes both ribosomal subunit joining and the GTP hydrolysis. This process then triggers eIF5B release from the 48S PIC (Shin et al., 2002). The interaction of eIF5B with eIF1A helps to recruit the 60S for 80S formation and in turn, eIF5B promotes the release of eIF1A from the formed 80S initiation complex to form the conformation of the complete A-site that is now ready to accept the first aminoacyl-tRNA (Fringer et al., 2007).

3.5. Translational regulation under stress conditions

During cellular stress, canonical cap-dependent translation is inhibited, however, translation of specific pools of mRNAs still occurs using alternative ways of translation initiation. Exposure to certain stresses like nutrient deprivation, temperature shock, DNA damage, and oxidative stress leads to the shut down of canonical cap-dependent protein synthesis and translational reprogramming towards alternative initiation pathways. This leads to the

selective translation of specific mRNAs whose product is involved in adapted stress responses. There are several pathways for stress-induced inhibition of canonical translation, mainly at the level of ternary complex (TC) and eIF4F formation. Such regulation processes are made essentially by phosphorylations of eIFs at specific positions. Four proteins control these modifications: GCN2, PERK, HRI, and PKR kinases. These kinases are activated by different stimuli depending on the nature of the stress (Dever et al., 2016; Golob-Schwarzl et al., 2017) (**Fig. 4**).

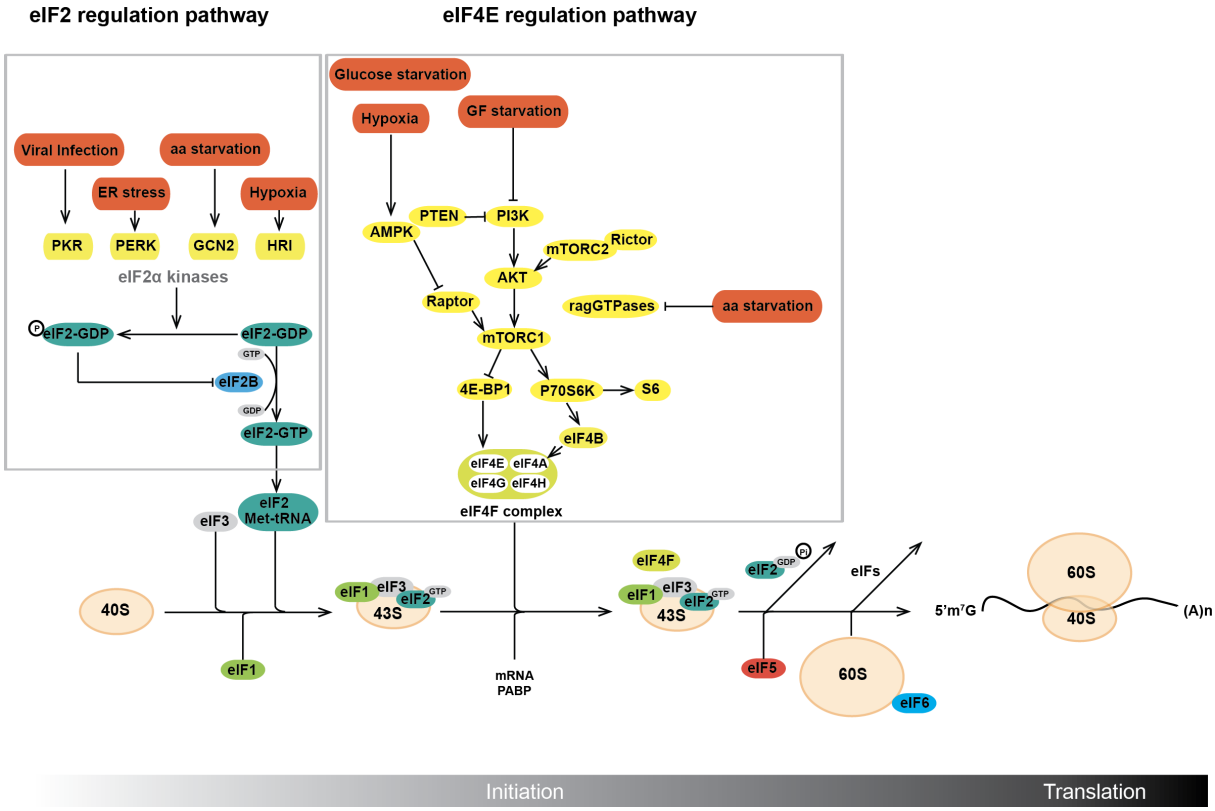


Figure 4: **Regulation of cap-dependent translation in eukaryotes under different stress conditions. Left:** Under stress conditions like viral infection, amino acid starvation, ER stress and hypoxia, different kinases are activated leading to the phosphorylation of eIF2α subunit. This inhibits the formation of eIF4E ternary complex and repress translation initiation. **Right:** Another stress-mediated regulatory pathway affects eIF4F complex. Under normal conditions, 4E-BP is phosphorylated by protein kinase mTORC1 which leaves eIF4E free. This allows the formation of eIF4F complex with eIF4G to initiate cap-dependent translation. However, under certain stress conditions (hypoxia, glucose and amino acid starvation, growth factor starvation), mTORC1 is inactivated and 4E-BP is dephosphorylated. It sequesters eIF4E thereby, inhibits eIF4F complex assembly and represses translation.

Briefly, nutrient deprivation leads to the inhibition of the mTOR (mammalian Target Of Rapamycin) pathway which acts through a kinase termed TOR kinase. The eIF4E binding

proteins (4E-BPs) are proteins that bind to eIF4E and compete with eIF4G thereby blocking eIF4F assembly. This TOR kinase is a component of two large protein complexes mTORC1 and mTORC2 that are able to phosphorylate factors. The mTORC1 complex is able to phosphorylate 4E-BPs. When 4E-BPs are phosphorylated, they cannot bind to eIF4E anymore, this allows proper eIF4F assembly and therefore promotes cap-dependent translation (Sonenberg and Hinnebusch, 2009). The inhibition of mTOR pathway then leads to accumulation of non-phosphorylated 4E-BPs that further induce the dissociation of eIF4F complex thereby leading to a reduction in global protein synthesis (Wang and Proud, 2009).

Regarding cold shock, global translation is mainly inhibited through another phosphorylation event on the alpha subunit of the tripartite factor eIF2 (eIF2α) (Underhill et al., 2007) and reduction of eIF3i expression, a subunit of eIF3 factor which contains 13 subunits (Roobol et al., 2009). These phosphorylation events inactivate these translation initiation factors.

Heat shock, on the other hand, activates the expression of specific proteins called Heat Shock Proteins (HSPs), which act as molecular chaperones to induce protein refolding and increase cell survival following heat shock stress (Westerheide and Morimoto, 2005). Heat shock induces the phosphorylation of eIF4E and eIF2α subunit and reduces global protein synthesis (Coldwell et al., 2001). Heat shock mainly impairs global translation by producing the chaperone HSP27 which specifically binds eIF4G and insolubilizes it by targeting this factor in heat shock granules after its dissociation from eIF4F complex (Cuesta et al., 2000). In fact, heat shock mRNAs are specifically translated under such stress conditions because they do not require functional eIF4F complex, thanks to the presence of specific elements in their 5'UTRs that alternatively regulate their translation initiation without eIF4F (Yueh and Schneider, 2000).

Concerning DNA damage, a DNA-Damage Response (DDR) is activated by the cell and results in activation of apoptosis. In contrast to other stresses, DNA damage does not induce a global response orchestrated by a common pathway but the responses are variable and depend on the type of DNA damage. For instance, UV-induced DNA damages lead to the activation of GCN2 kinase which phosphorylates the eIF2α subunit thereby reducing the level of TC and repression of global translation (Deng et al., 2002). Previous studies suggest that specific mRNAs are selectively translated in order to repair DNA damage induced by UV exposure. These mRNAs are translated by cap-independent mechanisms that use Internal Ribosome Entry Sites (IRES) elements. Such mRNAs like p53 encoding a tumor suppressor protein and their IRES elements will be further discussed in section (4.7). The activation of p53 tumour suppressor protein leads to cell cycle arrest and induces apoptosis (Grover et al., 2008).

During hypoxia, the inhibition of mTOR pathway leads to a set of events including the inhibition of eIF4E through 4E-BP1 binding after its dephosphorylation as previously described, activation of eEF2K which is another kinase that phosphorylates eEF2, and the inhibition of S6 kinase. This leads to the reduction of the activity of the helicase eIF4A and its auxiliary factor eIF4B (Connolly et al., 2006).

Overall, the inhibition of protein synthesis depends on the type of stress. The common feature of all these processes relies on the fact that they either influence the TC formation or the 43S PIC recruitment to the mRNA during the early stages of translation initiation. Interestingly, in each type of stress response, specific mRNAs are still able to overcome such conditions by using alternative non-canonical translation initiation mechanisms to mediate stress response.

4. Alternative translation initiation mechanisms

Some atypical mRNAs do not share the specific features that are found in most of the eukaryotic mRNAs. Indeed, such mRNAs lack the cap structure at the 5' end like viral mRNAs or do not harbour a poly (A) tail at their 3' end like histone mRNAs. However, these mRNAs are still efficiently translated by specific non-canonical mechanisms to initiate translation.

4.1. Translation of H4 mRNA by tethering mechanism

Histone mRNAs have a structural organization that is incompatible with the conventional scanning initiation model. However, they are translated very efficiently to produce very abundant proteins in the eukaryotic cells. Indeed, histone H4 mRNA contains a very short 5'UTR comprising only 9 nucleotides (Meier et al., 1989) and no poly (A) tail at its 3' end (López and Samuelsson, 2008), Instead histone mRNAs harbour a conserved stem loop, which is specifically recognized by a protein called the Stem-Loop-Binding Protein (SLBP) (Jaeger et al., 2006). Our laboratory has shown that the coding sequence of the mouse H4 histone mRNA contains two structural elements critical for efficient translation initiation. One of the structures, termed eIF4E-Sensitive Element (4E-SE), interacts with the N-terminal extremity of eIF4E and allows its recruitment without the need of the cap (**Fig. 5**). Interestingly, a similar 4E-SE was also described in the 3'UTR of Cyclin-D1 mRNA (Culjkovic et al., 2006). The second structure is a Three-Way helix Junction (TWJ) located 19 nucleotides downstream the AUG start codon. The TWJ sequesters the m^7G cap by forming a Cap-Binding Pocket (CBP). These two structures allow the recruitment of the 43S complex

directly on the AUG start codon without any scanning step (Martin et al., 2011). This mechanism has been named ribosome tethering whereby the ribosome is directly deposited on the AUG codon that has to be located at a precise position in the mRNA. The correct positioning of the AUG codon in the P-site of the ribosome is facilitated by an interaction between an AGG triplet located immediately upstream of the TWJ and the loop of helix h16 of the 18S ribosomal RNA (Martin et al., 2016) (**Fig. 6**). Such interaction stabilizes the 48S binding on the AUG codon and explains the lack of scanning. In fact, due to the lack of a time- and energy-consuming scanning step, this tethering mechanism allows a more efficient and faster translation of H4 mRNA compared to canonical translation. This probably explains the massive production of histone proteins during the S phase of the cell cycle.

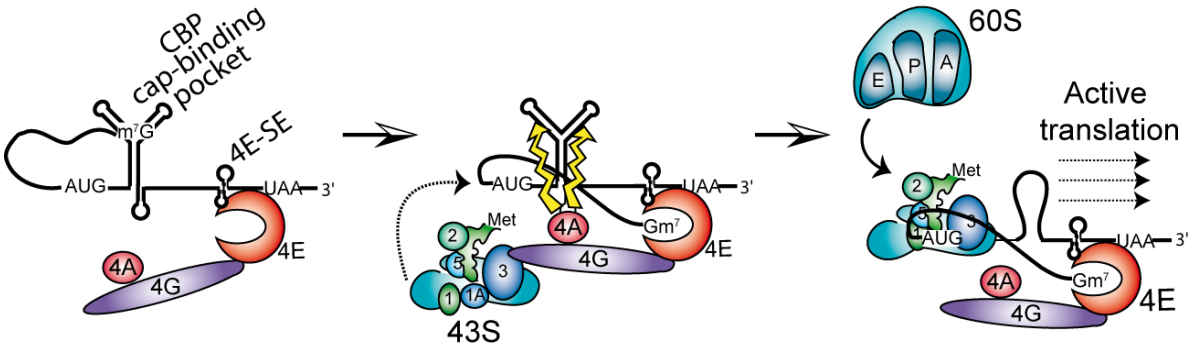


***Figure 5*: Non-canonical cap-dependent translation initiation of H4 mRNA by tethering mechanism.** Figure adapted from (Martin et al., 2011). H4 mRNA contains two structural elements important for its translation. The first element is eIF4E-sensitive element (4E-SE). It interacts with eIF4E without the need of a cap. Then, the 43S PIC is loaded directly on the AUG codon without further scanning step. The second element is a Three-Way helix Junction (TWJ) located 19 nucleotides downstream of the AUG codon which sequesters the m^7G cap in a cap-binding pocket. The two structures facilitate the binding of the ribosome directly on the AUG codon. Then the 60S subunit joins the 40S to allow active translation of H4 mRNA.

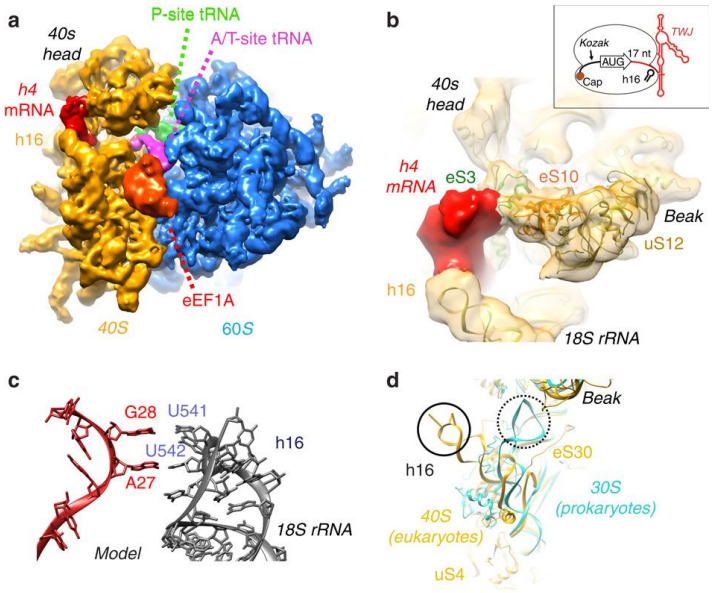


Figure 6: **Structure of 80S ribosome assembled on H4 mRNA during translation initiation.** Figure adapted from (Martin et al., 2016). The correct positioning of the AUG codon in the P-site of the ribosome is caused by an interaction between an AGG triplet located in the Three-Way Junction of H4 mRNA and a $UUUC_{540\text{-}543}$ sequence in the loop helix h16 of the 18S ribosomal RNA**. (a)** Cryo-EM structure of H4-80S complex blocked at translocation step by cycloheximide. **(b)** A closer view of the structure reveals an interaction between H4 mRNA and helix h16 of 18S rRNA and ribosomal proteins eS3 and eS10. **(c)** The interaction model between H4 mRNA and h16 at the nucleotide level. **(d)** Superposition of eukaryotic and prokaryotic ribosomes shows a difference in helix h16 orientation.

4.2. Translation by 3' Cap-Independent Translation Elements (CITEs) in plant RNA viruses

Another example of atypical translation initiation mechanism occurs in the plant viral mRNAs. These viral mRNAs harbour structures in the 3'UTR that replaces the 5' m^7G cap called 3' CITEs (Miller et al., 2007). These elements adopt different pseudo-knot structures from different types such as « I », « T » or « Y » **(Fig. 7)**. These structures are able to promote by distance the recruitment of 43S PIC on the 5'UTR. This occurs by a long range interaction termed « kissing loop » between the 3' CITE and a complementary region located in a stem loop structure in the 5'UTR (Miras et al., 2017; Nicholson et al., 2010). They are classified into at least 8 classes; they do not contain obvious sequence and structural similarities (**Fig. 7**).

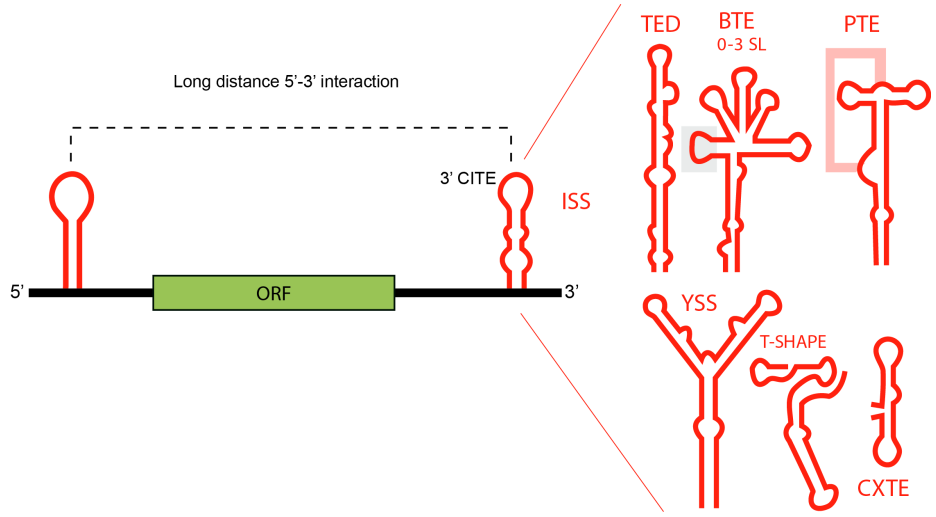


***Figure 7*: Translation initiation mechanism guided by 3' CITEs.** Plant viral mRNAs harbour structures in the 3' UTR that replaces the 5' m^7G cap called 3' CITEs (Miller et al., 2007). These elements adopt different structures of « I », « T » or « Y » pseudo-knot structures and are able to promote the recruitment of 43S PIC on the 5' UTR. This occurs by a long-range interaction termed « kissing loop » between the 3' CITE and the complementary region in a stem loop structure in the 5' UTR.

A well-characterized example is present in the pea enation mosaic viral mRNA which harbours a pseudo-knot structure at the 3'UTR. This pseudo-knot folds into a peculiar structure in such a way that a guanosine residue is projected out of the structure. This structure allows the recruitment of eIF4E because it binds the protruded G (Wang et al., 2009, 2011). Once bound on the 3'UTR, eIF4E drives the 43S assembly on the AUG start by a looping mechanism. In yellow dwarf viral mRNA, recent studies suggest that the multi-subunit eIF3 is recruited by a structure formed by an interaction between the 3' CITE and 5'UTR to promote cap-independent translation (Bhardwaj et al., 2019).

4.3. Non canonical translation initiation mediated by eIF3

4.3.1. Initiation factor eIF3

Initiation factor eIF3 comprises a versatile scaffold for translation initiation complexes. This 800 kDa complex is the largest translation initiation factor with 13 distinct subunits termed eIF3a to eIF3m (Asano et al., 1997; Browning et al., 2001; Morris-Desbois et al., 2001; Unbehaun et al., 2004). The central core consists of an octamer containing eIF3a, eIF3c, eIF3e, eIF3f, eIF3k, eIF3h, eIF3l and eIF3m. Six of these subunits (a,c,e,k,m and l) comprise a subdomain called PCI (26S Proteosome, COP9 signalosome, Initiation factor 3) while the other two subunits (f and h) form another subdomain called MPN (Mpr1/Pad1 N-terminal). The remaining five subunits eIF3b, eIF3d, eIF3g, eIF3i and eIF3j are considered as peripheral

subunits (Valášek et al., 2017) (**Fig. 8**). Moreover, in yeast, the subunits a, b, g and i are called the 'Yeast-like core' (YLC) since they constitute the minimal functional core of eIF3. In fact, subunits b and g possess an RNA-Recognition Motif (RRM). In addition, subunits b and i possess another motif called the B-helix motif that is formed by the repetition of tryptophan and aspartic acid residues, called WD40. The functions of each subunit in the overall large complex are not very well understood yet. For instance, the implication of subunit eIF3j in the complex has been under investigation for years. In fact, early studies suggested that eIF3j orchestrates the interaction between the 40S ribosomal subunit and the whole eIF3 to form the 43S particle. In the structure of the 43 PIC, eIF3j is located in the decoding centre of the 40S ribosomal subunit at the A-site and the mRNA entry channel. By interacting with eIF1A, eIF3j reduces the ribosome affinity for binding on mRNA (Fraser, 2015). Recently, another report has proposed a new function for eIF3j (also called Hcr1 in yeast), eIF3j might be an auxiliary factor for 60S ribosomal subunit recycling upon its recruitment to the stop codon (Young and Guydosh, 2019). During initiation, the whole eIF3 complex stimulates the binding of the TC to the 40S subunit with the aid of eIF1 and eIF1A (Chaudhuri et al., 1999; Sizova et al., 1998). In addition, eIF3 plays an important role in the attachment of 43S PIC on the mRNA upon its interaction with eIF4G. It forms a bridge between the 40S and the eIF4F-mRNA complex (Siridechadilok et al., 2005). However, the molecular mechanisms that enabled the distinct functions of eIF3 in the translation initiation process remain elusive.

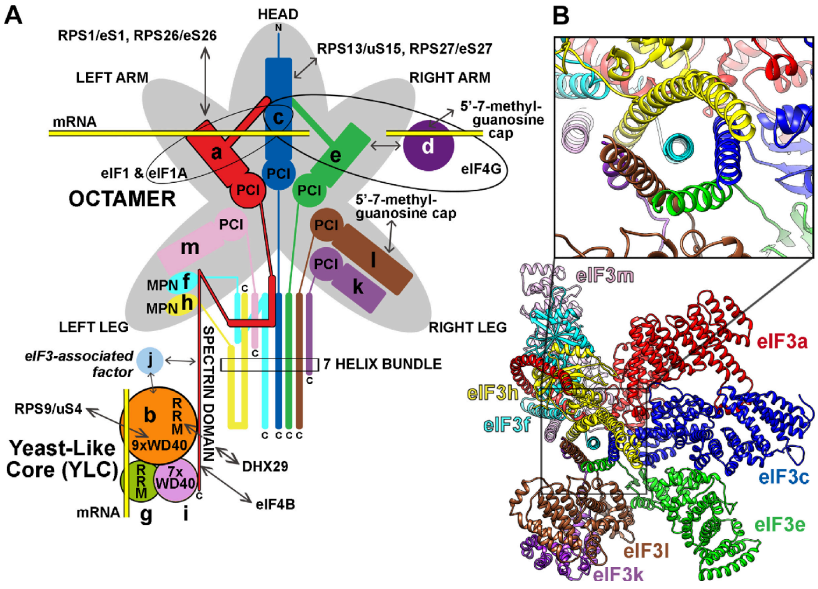


***Figure 8*: Representation of the eIF3 subunits and the interactions among them.** Figure adapted from (Valášek et al., 2017). **(A)** The octamer core of eIF3 constitute the subunits: a, c, e, k, l, m, f and h. The subunits b, d, g, j, and i are located in the peripheral region. The subunits a, b, i, and g are

called the 'Yeast-like core' (YLC) since they constitute the minimal functional core of eIF3 in yeast. **(B)** A close-up view of the 7-helix bundle in the octamer core formed by subunits h, c, e, f, l, and k.

4.3.2. Direct eIF3-mediated recruitment of 43S PIC

Photoactivatable Ribonucleoside-Enhanced Cross Linking Immunoprecipitation (PAR-CLIP) studies from (Lee et al., 2015) have identified a new function of eIF3 as a mediator of cap-independent initiation mechanism. Previous studies on cell proliferation regulators such as c-JUN and B-cell translocation gene BTG1 have shown that these two mRNAs possess in their 5'UTR a conserved long stem loop structure (**Fig. 9**). This long structure is able to recruit the ribosome (Blau et al., 2012). Both structures in each mRNA interact directly with eIF3 by contacting several eIF3 subunits. These interactions enable the recruitment of the 43S PIC in a cap-independent manner. Interestingly the recruitment of eIF3 by both mRNA 5'UTR have opposite effects. In the case of c-JUN, eIF3 activates the translation whereas it represses the translation of BTG1.

***Figure 9*: Secondary structural models of two mRNAs interacting with eIF3.** Figure adapted from (Lee et al., 2015). Sites of eIF3 binding site identified by PAR-CLIP are annotated in the figure. A PAR-CLIP experiment was done on human transcriptome to identify different mRNAs interacting with eIF3. Two examples are shown: c-Jun and BTG1 mRNAs which proved to interact with eIF3 via a long stem-loop structure. This interaction can either induce translational repression or activation.

Moreover, a novel function has been proposed for eIF3d subunit as a novel cap-binding protein, a function which was so far thought to be restricted to eIF4E (Lee et al., 2016). This sets a new role for eIF3 as an mRNA-specific regulator of cap-independent translation activation or inhibition of specific mRNA families.

4.4. Translation regulation by Adenosine methylation (m^6A)

A number of eukaryotic mRNAs possess nucleotide modifications in their 5'UTR and 3'UTR or in the coding sequences. One prevalent example of such modifications is the N6-methyladenosine (m^6A) that is carried out by methyltransferase METTL3, a so-called m^6A writer (Liu et al., 2014). These modifications play a role in mRNA translation and structure (Mauer et al., 2017). It has been shown that a modified nucleotide m^6A located in the 5'UTR promotes cap-independent translation through a direct interaction with eIF3 which further recruits the 43S PIC independent of the eIF4F complex (Meyer et al., 2015) (**Fig. 10A**). However, when present in the coding sequence, these modifications can induce aberrant tRNA selection thereby inducing mis-coding and causing a reduction in elongation dynamics (Choi et al., 2016). Apparently, these m^6A methylation events are preferably implemented in a consensus sequence of GGAC motif (the underlined A being modified to m^6A). In fact, upon heat shock stress, YTH-N^6-methyladenosine RNA-binding protein (YTHDF), a so-called m^6A reader, localises the modified mRNA to the nucleus and maintains the level of m^6A modifications in the 5'UTR of stress-response mRNAs (Wang et al., 2015). It actually protects the m^6A from demethylation by fat and obesity-associated protein (FTO), an m^6A eraser. One of the best examples of modified mRNA is the Heat Shock Protein 70 (HSP70) mRNA which carries a single nucleotide modification at nucleotide A_{103} in its 5'UTR (Zhou et al., 2015). Interestingly, the 3'UTR and the vicinity of stop codon can also harbour m^6A modifications (Dominissini et al., 2012; Meyer et al., 2012). When present in the 3'UTR, m^6A can induce the binding of YTHDF1, an m^6A reader, to promote cap-dependent translation through interaction with eIF3 (Wang et al., 2015) (**Fig. 10B**). In addition to their role in translation, m^6A induce the unwinding of local RNA structures by destabilising RNA-duplexes into less stable structural motifs (Roost et al., 2015). Specific RNA binding sites become more accessible for binding factors when the modification is present and are therefore involved in mRNA decay regulation. For instance, YTHDF2, another m^6A reader, plays a different role in 3'UTR modifications, as it targets the modified mRNA into P bodies for decay through the recruitment of CNOT1, CCR4-NOT complex (Wang et al., 2014). Altogether, this suggests alternative functions of m^6A modifications, depending on their position in the mRNA and the m^6A readers that are implicated.

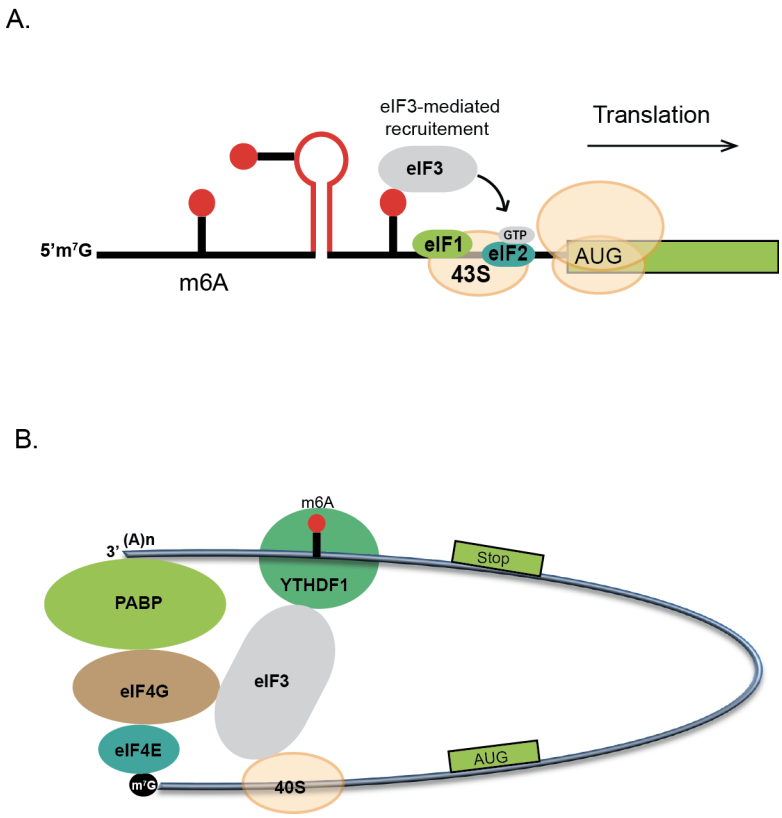


***Figure 10*: Translation initiation mediated by m^{6}A in mRNAs. (A)** m^{6}A modifications (shown by red circle) in the 5'UTR can assemble the translational machinery by interacting with eIF3 **(B)** YTHDF1 binds m^{6}A modification sites in the 3'UTR and interacts with eIF3 to assemble the 40S ribosomal subunit. This is facilitated by loop structure mediated by the interaction between eIF4G and PABP.

4.5. Translation initiation of circular RNA (CircRNA)

CircRNAs are circular RNA molecules comprising intronic sequences that are closed by covalent link. They are produced from transcribed genes undergoing non-canonical splicing through exon shuffling (Barrett and Salzman, 2016; Ragan et al., 2019). Although these RNAs are highly conserved and abundant in the cell however, their functions are still poorly understood. They were initially categorized as long non-coding RNAs, however, recent studies suggest that circRNAs accumulate in the nucleus to regulate transcription (Kramer et al., 2015) or for post-transcriptional functions (Salzman et al., 2012). Interestingly, circRNAs can also act as microRNA (miRNA) sponges. One example is *Sry* circRNA, spliced from developmentally regulated *Sry* gene, which sequesters sixteen molecules of miR-138 (Hansen et al., 2013). The cirRNA can also be exported to the cytoplasm for translation of encoded proteins (Pamudurti et al., 2017). It has been shown that circRNA are translated in a cap-independent manner. Indeed, since they are circular, they lack the 5' cap structure (Legnini et al., 2017). One example is Circ-ZNF609 mRNA which is translated independent of eIF4F

complex recruitment by an internal initiation due to an m^6A methylation in its 5'UTR (Chekulaeva and Rajewsky, 2019; Zhao et al., 2014).

4.6. Initiation by Internal Ribosome Entry Sites (IRESes)

Initially discovered in the 5' UTR of viral mRNAs, IRESes bypass the conventional translation initiation and allow internal ribosomal recruitment in a cap-independent manner. They fold into RNA structural elements in the 5'UTR or intergenic regions of viral mRNAs. IRESes are usually characterized by their high GC content in order to form highly stable secondary structures (Martínez-Salas et al., 2008). The human Poliovirus (PV) and Encephalomyocarditis virus (EMCV) RNA genomes are two viruses from the picornavirus family that harbour the first reported IRES elements (Jang et al., 1988; Pelletier and Sonenberg, 1988). Picornavirus IRESes are by far the most studied, however, more viral IRESes from viruses infecting mammals were discovered in viruses such as Hepatitis C Virus (HCV) (Honda et al., 1996), pestiviruses like swine fever virus (CSFV) (Rijnbrand et al., 1997) and retroviruses such as HIV (Locker et al., 2011). Unlike canonical translation, IRESes do not require eIF4E due to the absence of the 5' m^7G cap in viral mRNAs. Thereby, the initiation occurs using subsets of eIFs and specific RNA binding proteins called IRES *Trans*-Acting Factors (ITAFs). Despite their different structural organization, IRESes allow either the recruitment of the ribosome directly on the AUG start codon or require an additional scanning step to localize the start codon (Lozano and Martínez-Salas, 2015). Interestingly, during viral infection, cap-dependent translation is shut down due to 4E-BP binding to eIF4E and eIF4G proteolysis (Gingras et al., 1996). However, viral translation that is IRES-driven is still active and thereby takes advantage of such event in order to hijack the host translational machinery for viral protein synthesis while cellular translation is impaired.

4.6.1. Four classes of viral IRESes

Despite performing the same function of internal ribosome recruitment in a cap-independent manner, IRESes do not share any sequence similarity, common RNA secondary structures and their mode of action is different. Accordingly, studies have focused on using a combination of structural, biochemical, physical and functional data, mostly of positive single stranded RNA virus for IRES classification. These IRESes can be divided into four distinct classes based on their eIFs requirement, their three-dimensional structure and their mode of ribosome recruitment. Based on this, a classification was established according to several features such as the folding, their requirement of auxiliary factors like eIFs and ITAFs and

their mode of action (Filbin and Kieft, 2009; Johnson et al., 2017; Mailliot and Martin, 2018; Yamamoto et al., 2017). This allowed four classes of IRESes to be distinguished:

4.6.1.1. Class I: extended and largely flexible IRESes

This class of viral IRESes does not contain compact structures but rather possesses quite flexible domains. The majority of these IRESes from this class belong to viral RNA genomes from *Picornaviridae* like Poliovirus (PV) and Human Rhinovirus (HRV). This type of IRESes have different modes of action whereby they do not really recruit the ribosome but rather recruit eIFs in a cap-independent manner. These IRESes do not bind directly to the 40S ribosomal subunit but rather recruit the translational machinery and therefore require almost all the eIFs (except eIF4E) to assemble a PIC upstream of the coding region. The assembled PIC will then undergo a classical 5'-3' scanning to find the AUG start codon. One example is PV which comprises a 5'UTR of 743 nucleotides with the first 100 nucleotides being a cloverleaf (CL) structure that is required for replication (Andino et al., 1990) (**Fig. 11**). The PV IRES covers the region 100-743 and is organized in several structural domains (II to VI). The PV RNA genome encodes for a viral protease 2A which cleaves eIF4G thereby inhibiting the host cap-dependent translation (Byrd et al., 2005; Kempf and Barton, 2008). The 5'UTR contains an AUG at position 586 upstream of the main AUG at position 743. Translation initiates on the second AUG at position 743. However, the upstream AUG at position 586 serves as a cryptic start that participates to the ribosome recruitment with the cooperation of an upstream oligopyrimidine motif (Iizuka et al., 1989; Pilipenko et al., 1992). The PV IRES also requires some ITAFs such as Poly (rC)-Binding Protein 2 (PCBP2) which recognizes a C-rich motif in the cloverleaf and in domain IV (Sweeney et al., 2014; Toyoda et al., 2007). The IRES also interacts with PABP to promote circularisation to favour ribosome recruitment for viral translation (Ogram et al., 2010; Spear et al., 2008). Altogether, IRESes from this class require all eIFs except eIF4E and their structural domains are generally flexible. The recruitment of ribosome occurs upstream of AUG start codon thereby, a scanning step is necessary.

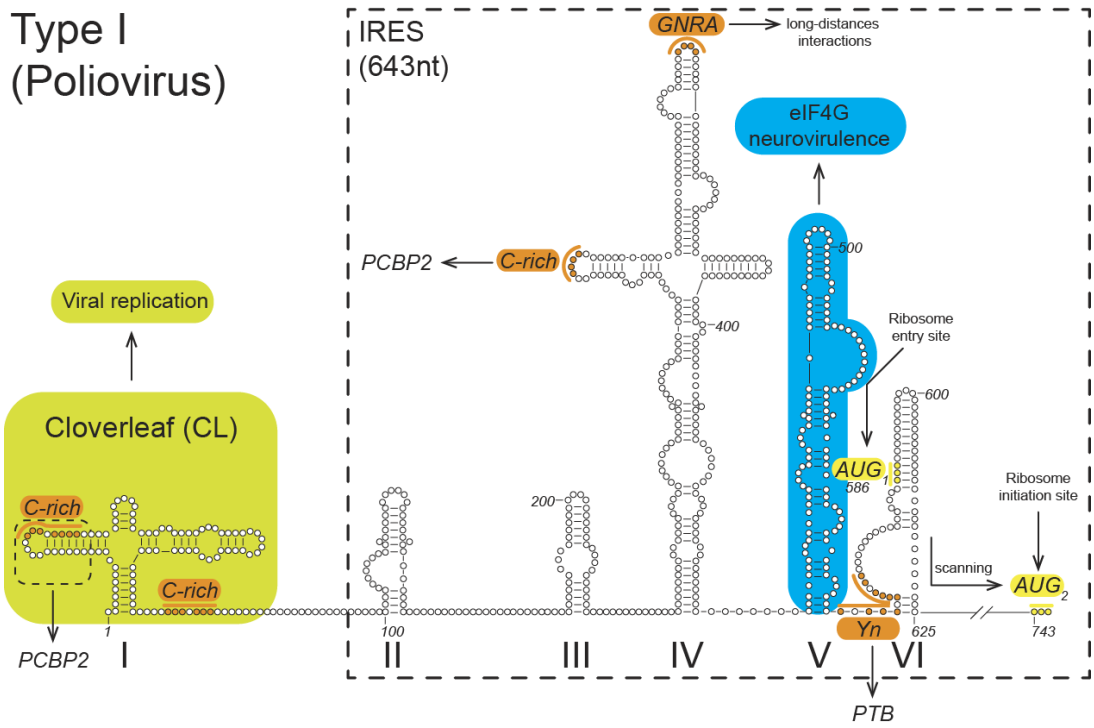


***Figure 11*: Poliovirus IRES is a type I IRES.** Figure adapted from (Mailliot and Martin, 2018). The PV 5'UTR region adopts a secondary structure comprising six distinct domains. Domain I folds into a cloverleaf (CL) structure that is involved in viral replication (light green). Domains II to VI constitute the PV IRES (box in dotted lines). Some nucleotides in these domains are involved in the recruitment of IRES *trans*-acting factors (ITAFs) or are required for the formation of long-distance interactions that contribute to the structural integrity of IRES (these nucleotides are indicated in orange).. The IRES V domain specifically interacts with the 2A protease-cleaved eIF4G translation initiation factor (in blue). The two PV AUG codons are highlighted in yellow.

4.6.1.2. Class II: flexible IRESs not requiring scanning step

Class II IRESes are found mostly in *Picornaviridae* family like for instance Foot and Mouth Disease Virus (FMDV) (Chard et al., 2006) and Encephalo Myocarditis Virus (EMCV) (Lefkowitz et al., 2018). The mechanism for host cap-dependent translation inhibition shut down is different between the two viruses. During FMDV infection, eIF4G is cleaved by viral protease 3C (Mosenkis et al., 1985). In contrast, EMCV infection activates the dephosphorylation of 4E-BP which will subsequently bind to eIF4E and inactive host cap-dependent translation (Gingras et al., 1996). Nevertheless, IRESes of these two viruses do share similar structural organization (domain II-V) despite significant sequence differences. Similar to PV, the 5'UTR of both viruses also comprise a cloverleaf in domain I that is required for viral replication too. EMCV and FMDV IRESes share a common pyrimidine-rich motif in domain II for the binding of Polypyrimidine Tract Binding protein (PTB) which

serves as an ITAF for ribosome recruitment (Andreev et al., 2007). In FMDV IRES, Domain III folds into a long stem loop containing a cruciform structure (Fernández-Miragall and Martínez-Salas, 2003) while domains IV and V recruit eIF4G, eIF4B, eIF3 and PTB due to specific RNA-protein interactions (López de Quinto and Martínez-Salas, 2000). The binding of the ITAFs induce specific structural changes in the IRES that further enhance the binding of eIFs like eIF4G and eIF4A (Song et al., 2005) (**Fig. 12**). In fact, both IRESes encompass two conserved motifs GNRA and RAAA in domain III that are required for proper structural and functional organization of the IRES (Fernández-Miragall and Martínez-Salas, 2003; López de Quinto and Martínez-Salas, 1997; Lozano and Martínez-Salas, 2015). The two IRESes also possess two in frame AUG start codons. However, while EMCV only uses the second one for translation initiation, FMDV uses both (Andreev et al., 2007; López de Quinto and Martínez-Salas, 2000). In contrast to class I IRESes, IRESes from class II do not require a further scanning step. The 43S PIC is directly loaded on the AUG start codon.

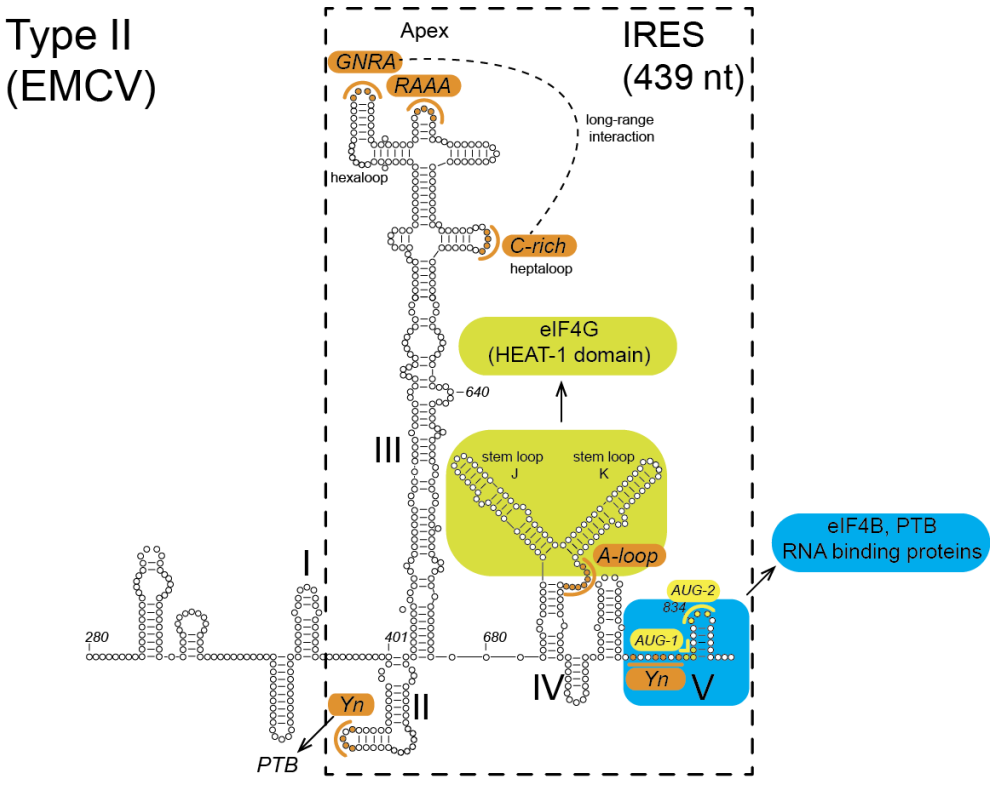


***Figure 12*: EMCV IRES is a type II IRES.** Figure adapted from (Mailliot and Martin, 2018). This figure represents the secondary structure of the 5'UTR region of the EMCV viral RNA. Nucleotides essential for the activity of IRES or for the recruitment of ITAF are indicated in orange. The minimum domain required for the EMCV IRES activity is dashed. Domain IV specifically recruits the eIF4G factor (in dark green) and the V domain interacts with the eIF4B initiation factor and ITAF PTB, as well as with other RNA-binding proteins (in blue). The two EMCV AUG initiator codons are

highlighted in yellow. Structural data obtained by NMR (N) on the isolated domain IV or domain IV complexed with the HEAT-1 domain of eIF4G (by SAXS) are shown on the right.

The ITAFs induce structural changes in the IRES to enhance the binding of eIFs like eIF4G and eIF4A (Song et al., 2005). Altogether, class II IRESes are similar to class I in their requirement of eIFs (except eIF4E) but the scanning step is absent.

4.6.1.3. Class III: IRESes with compact structures

The best-characterized member of this class is the IRES from human Hepatitis C Virus (HCV) viral genome that belong to the *Flaviviviridae* family. IRESes from this class are rather large (~300-400 nt) and bear some highly ordered structures like pseudo-knot structures (Kieft et al., 1999). Among HCV viruses, seven genotypes and seventy subtypes have been described, however the 5'UTR is highly conserved (Echeverría, 2015; Laporte et al., 2000). It spans a region of 341 nucleotides and the structural organization shows four different domains (I to IV) (Dibrov et al., 2007; Lukavsky, 2009; Pérard et al., 2013) (**Fig. 13**). The overall topology is conserved among Flaviviruses whereby domain II consists of a long stem loop structure encompassing two subdomains IIa and IIb. Domain III is more structured with six sub-domains in stem loops IIIa-IIIe that are organized in two three-way junctions adjacent to a four-way junction and a pseudo-knot structure IIIf (Laing and Schlick, 2009). Finally, domain IV is a stem loop structure, which contains the AUG start codon. Domain I, which is not part of the HCV IRES, contains a binding site for micro RNA miR-122 (Kim et al., 2016; Roberts et al., 2011; Schult et al., 2018). The IRES spans domains II to IV and controls the initiation of viral translation. More precisely, Domains II and III orchestrate the 40S ribosomal subunit recruitment through multiple specific RNA-protein interactions with 40S ribosomal proteins. Domain II interacts with the ribosomal protein uS7 and uS11 that are located in the E-site moeity of the 40S ribosomal subunit (Yamamoto et al., 2015), subdomains IIIa and IIIc interact with the small 40S ribosomal proteins eS27 and the 18S ribosomal RNA segment ES7 while subdomain IIIe specifically binds to eS1 (Laing and Schlick, 2009). Subdomain IIIf comprises a pseudo-knot structure which keeps the IRES in a proper conformation to position the AUG in P-site of the ribosome and interacts with eS28 (Berry et al., 2011; Otto and Puglisi, 2004; Spahn et al., 2001). In fact, HCV IRES has a minimal requirement of eIFs for initiation and mainly binds directly to eIF3 (Collier et al., 2002; Khawaja et al., 2015; Sizova et al., 1998). The eIF3 binding domain is located in an internal loop structure (IIIb) as attested by mutational analysis of this subdomain. Indeed,

modifications in this domain drastically reduces the binding of eIF3 and consequently the IRES activity (Collier et al., 2002). Upon the binding of the small 40S ribosomal subunit on the HCV IRES, the 48S is assembled through the recruitment of eIF3 which itself interacts with the ternary complex eIF2–GTP–Met-tRNA$^{Met}_{i}$ in domain IIIb. The interaction between HCV IRES and eIF3 specifically occurs with subunits eIF3a, b, c, and d with the subunit a being the most important one for the binding. This interaction occurs by direct contacts of eIF3a and eIF3c loop residues of their helix-loop-helix (HLH) motif with nucleotides from the HCV IRES and 40S ribosomal subunit. Subunits eIF3b and d are critical for the PIC assembly process (Buratti et al., 1998; Hashem et al., 2013; Sun et al., 2013). The formation of the fully assembled 80S complex is achieved by a conformational change in domain II of HCV IRES (Yamamoto et al., 2015). This step is mediated by eIF5-induced GTP hydrolysis and eIF2-GDP release from the 48S thereby triggering the dissociation of eIFs to constitute an 80S complex (Locker et al., 2007). Altogether, class III IRESes have more sophisticated structures than classes I and II IRESes. These structures mediate direct interactions with the ribosomal 40S subunit. The requirement of eIFs is actually limited to eIF3, eIF2 and the TC, which mediate the positioning of the AUG in the P-site of the ribosome without any scanning step.

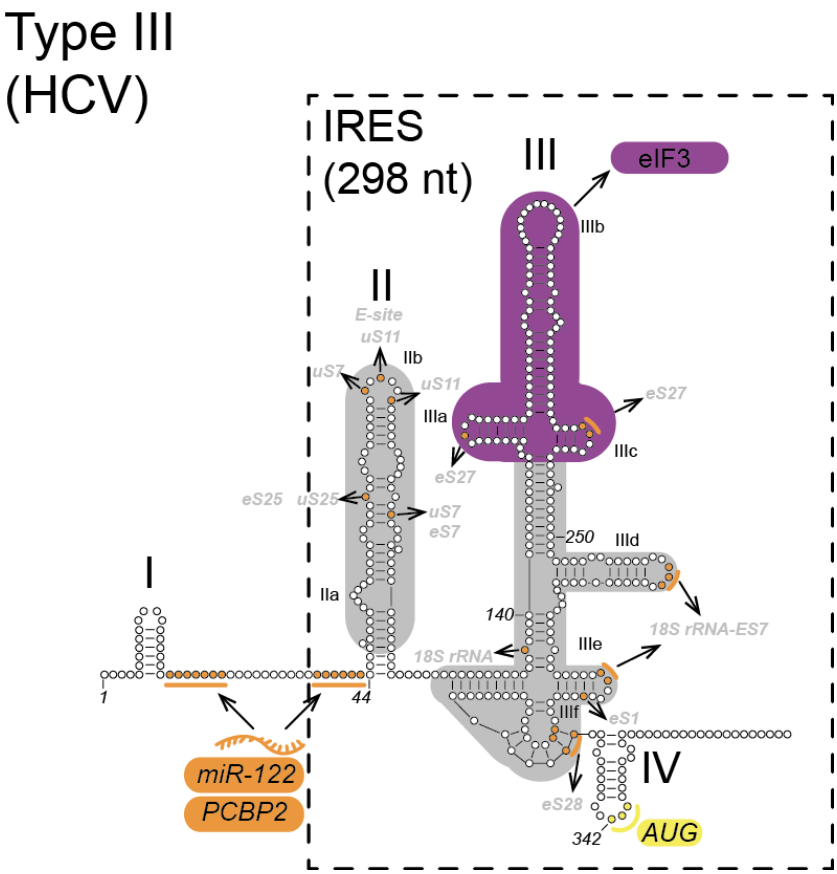


***Figure 13 (previous page)*: HCV IRES is a type III IRES.** Figure adapted from (Mailliot and Martin, 2018). This figure represents the secondary structure of the 5'UTR region of HCV genomic RNA. Nucleotides essential for the activity of IRES or for the recruitment of ITAFs are indicated in orange. The minimum domain of the HCV IRES is boxed. Domains II, IIId, IIIe and IIIf interact directly with the small 40S ribosome subunit (in gray). The apical portion of domain III (highlighted in purple) specifically interacts with eIF3. Contacts between the IRES and the ribosome proteins are indicated by arrows. The AUG initiator codon in the apical loop of domain IV is highlighted in yellow.

4.6.1.4. Class IV: IRESes that contain compact pseudo-knot structures

The most highly structured IRES RNAs yet identified are found in positive single stranded (+ss) RNA viruses from the *Dicistroviridae* family. These viral genomes contain two ORFs. Translation from the second ORF is driven by an IRES that is localised between the two ORFs in the intergenic region (IGR) of the Dicistrovirus genome. Therefore, these IRESes are called IGR. IGR IRESes fold into a small (~200 nucleotides) and very compact three-dimensional structure which contains three pseudoknots (**Fig. 14**). The most characterized member of this class is Cricket Paralysis Virus (CrPV) IGR, the three domains forming pseudo-knots have been named PKI, PKII and PKIII (Jan and Sarnow, 2002; Kanamori and Nakashima, 2001). Domains I and II are required for ribosome recruitment through a direct interaction between stem loops IV and V with ribosomal proteins from the small 40S and large 60S ribosomal subunits (Jan and Sarnow, 2002; Pfingsten and Kieft, 2008; Schüler et al., 2006). Domain III comprises a pseudo-knot (PKI) which mimics a codon-anticodon duplex. When PKI is located in the P-site of the ribosome, the ribosome will allow translation initiation at a non-AUG codon positioned in the A-site of the ribosome.

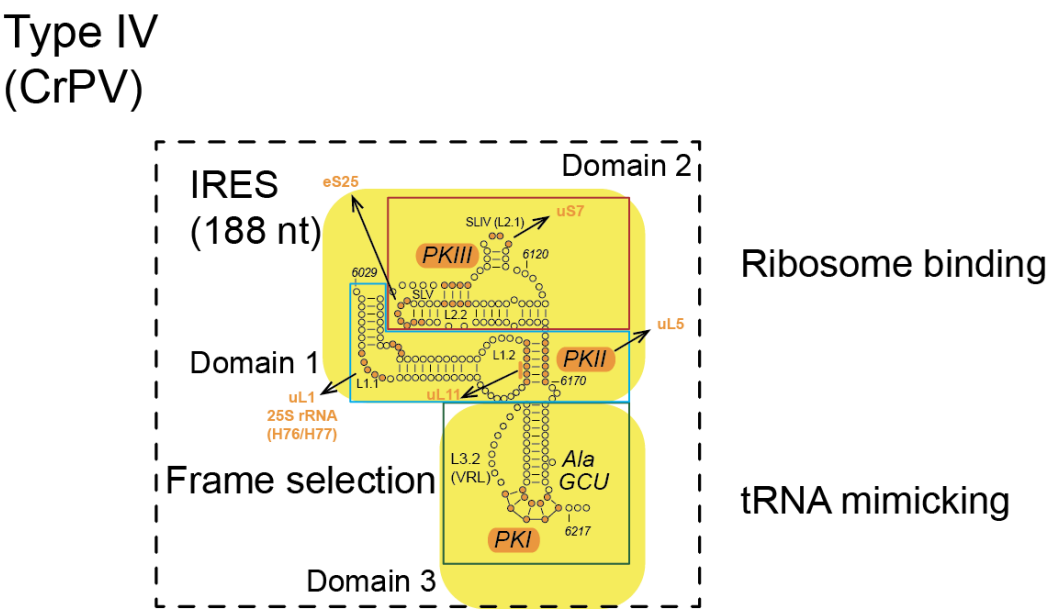


***Figure 14 (previous page)*: CrPV IGR is a type IV IRES.** Figure adapted from (Mailliot and Martin, 2018). This IRES contains three domains forming pseudo-knots named PKI, PKII and PKIII. Domains I and II are required for ribosome recruitment through a direct interaction between stem loops IV and V with ribosomal proteins from the small 40S and large 60S ribosomal subunits (in yellow). Domain III comprises a pseudo-knot (PKI) which mimics a codon-anticodon duplex is used for frame selection.

Therefore, the IGR IRES allows translation initiation without the need of either initiator tRNA, AUG start codon and t any eIFs (Butcher and Jan, 2016; Hussain et al., 2014; Pfingsten and Kieft, 2008). Indeed, the IGR actively manipulates the ribosome. Since no ternary complex is required, the IRES does not require GTP hydrolysis either (Jan and Sarnow, 2002). This molecular mechanism has been elucidated in great details by many functional studies and structural investigations by X-ray crystallography and more recently by cryo-EM studies (Murray et al., 2016; Pisareva et al., 2018). Altogether, IRESes from class IV contains highly sophisticated structures that can manipulate the ribosome to ensure translation initiation without eIFs, without AUG and without GTP. IGR IRESes only require the two ribosomal subunits to promote translation initiation (Pestova and Hellen, 2003; Pestova et al., 2004).

4.6.2. IRES *trans*-Acting Factors (ITAFs)

To complement the lack of some eIFs, IRESes use additional other auxiliary factors termed IRES *trans*-Acting Factors (ITAFs) that usually participate to translation initiation enhancing mainly by causing specific conformational changes in the structures. So far, several ITAFs have been identified to play a role in the translation initiation mechanisms driven by various IRESes. For instance, the factors PTB, ITAF45, PCBP2, SRP20 and Gemin5 have been shown to be ITAFs (Komar and Hatzoglou, 2011; Lee et al., 2017; Martínez-Salas et al., 2015). An interesting feature is that many ITAFs belong to the family of heterogeneous nuclear ribonuclear protein complexes (hnRNPs). ITAFs serve as either RNA-binding proteins or co-mediators with eIFs in order to stabilize IRES-ribosome association (Flather and Semler, 2015). One example of ITAF that activates several IRESes is Polypyrimidine tract protein (PTB) as described in previous sections. PTB is composed of four RNA-Recognition Motifs (RRM) that bind specific sequences. For instance, PTB enhances eIF4G association to ensure optimal IRES activity by binding to domain V of PV IRES (Kafasla et al., 2009, 2010; Pilipenko et al., 2001). ITAFs can also down regulate IRES activity. An example of such a negative mediator ITAF is Gemin5 with the FMDV IRES. In fact, it has a preferential binding site in the IRES which contains hairpin structure flanked by A/U/C-rich

flanking sequences. The binding of Gemin5 leads to the protection of such residues (Piñeiro et al., 2012). Interestingly, Gemin5 competes with PTB by binding to the same site on domain V of FMDV IRES and leads to structural reorganization in the local structure of the IRES thus, promoting IRES structural modulation (Piñeiro et al., 2013).

4.7. Cellular IRESes

IRESes are also found in cellular mRNAs like c-myc (Stoneley et al., 1998), Apaf-1 (Ungureanu et al., 2006), XIAP (Riley et al., 2010), p53 (Ray et al., 2006) and Hox (Xue et al., 2015) and others. Most of these IRESes have been described as highly structured RNA regulatory sequences that govern cap-independent translation initiation during cellular stress (Spriggs et al., 2010). Because of their apparent low translation efficiency, their existence have been controversial for many years and are still not well characterized (Van Eden et al., 2004; Komar and Hatzoglou, 2005). However, hundreds of putative new cellular IRESes have been discovered by a systematic *in vivo* screening approach (Weingarten-Gabbay et al., 2016). In addition, more evidences support a function for these IRESes. First, under normal physiological conditions, these IRESes are not efficiently used because of their structured 5'UTR which is detrimental to scanning. Second, under specific physiological circumstances where cap-dependent translation is compromised, these IRESes boost translation of specific subclasses of mRNAs. As previously described, under stress conditions, these IRESes act as a fail-safe mechanism to assure the translation of subsets of mRNAs (Graber and Holcik, 2007; Komar and Hatzoglou, 2005). Interestingly, most of IRES-containing mRNAs encode for proteins involved in fundamental processes such as apoptosis, stress or cell proliferation (Mokrejš et al., 2010) emphasizing their possible role as alternative mediators of translation under certain physiological conditions. So far, there hasn't been a classification of cellular IRESes since they do not share any obvious sequence motifs, structural similarities or common modes of action. Therefore, only few cellular IRESes have been characterized so far.

4.7.1. IRES of tumour suppressor gene p53

In this section, I will focus on characterization of p53 IRES. The IRES localised in the 5'UTR of the mRNA coding for the suppressor gene p53 has been extensively investigated (Marcel et al., 2018). As previously mentioned, p53 is a tumour suppressor gene which is not expressed under normal physiological conditions. However, upon DNA damage, p53 gene is activated leading to apoptosis (Giaccia and Kastan, 1998). The sequence of p53 mRNA is conserved and comprises two AUG codons: AUG1 allows the production of a full-length p53 protein

whereas alternative initiation from AUG2 produces an N-terminally truncated isoform p53/47 (Błaszczyk and Ciesiołka, 2011; Swiatkowska et al., 2015). It has been discovered that p53 regulation is controlled by two IRESes (Candeias et al., 2006; Halaby et al., 2015; Yang et al., 2006). IRES I is located in the 5'UTR (from nucleotides 1 to 216) and drives the translation from AUG1 while IRES II is embedded in the coding sequence and spans the 120 nucleotides downstream of AUG1. IRES II guides the translation initiation from AUG2 (Grover et al., 2009, 2011; Swiatkowska et al., 2015). In response to different stress conditions, ITAFs consisting of RNA-binding proteins interact with both IRESes to either activate or inhibit translation (**Fig. 15A**). These ITAFs recognize different motifs in p53 mRNA which are required for its translational regulation, they are detailed in the next paragraph.

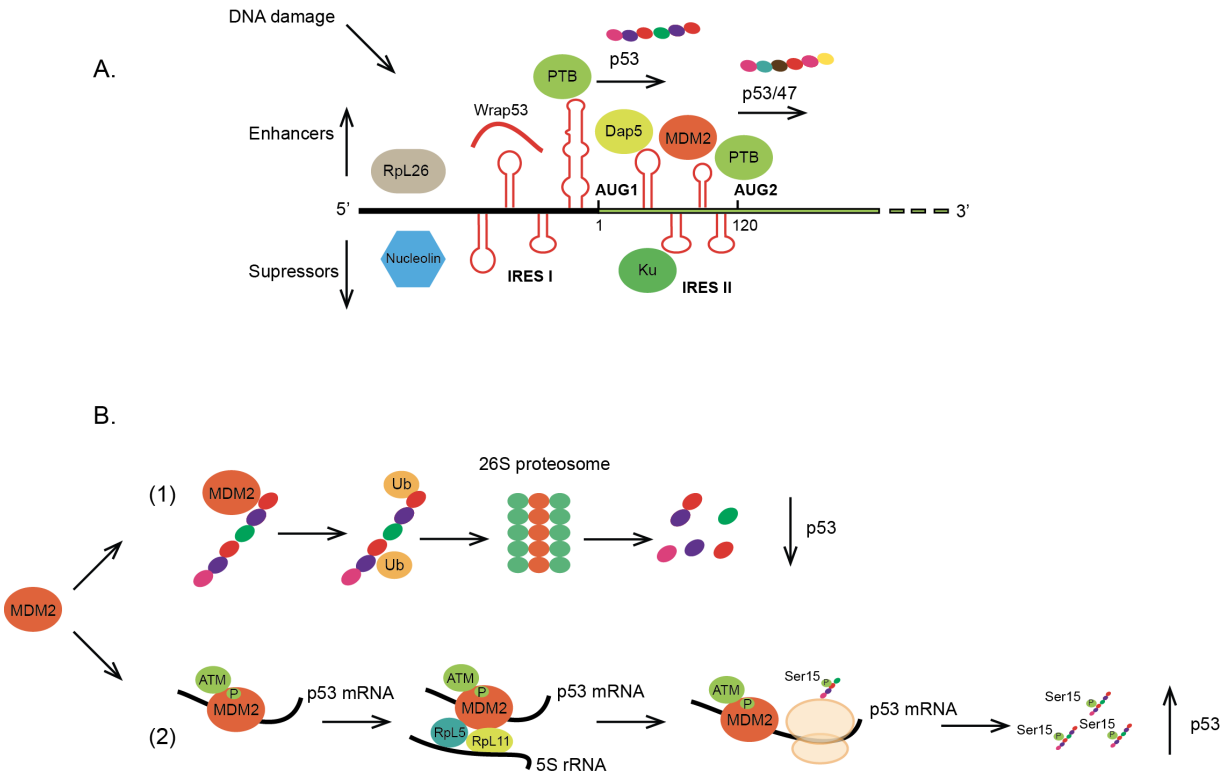


***Figure 15*: the model of p53 mRNA interacting factors. (A)** In response to DNA damage, different RNA-binding proteins and RNA transcripts interact with the p53 mRNA to regulate p53 expression. This occurs by either binding to the 5'UTR or to IRESes. The role of each enhancer and suppressor Is illustrated in **(B)**. MDM2 plays a dual role in p53 regulation. **(1)** Under normal conditions, MDM2 (in orange) binds and catalyses the ubiquitination of p53, which will then be degraded by the 26S proteasomal pathway. **(2)** Following DNA damage, ATM phosphorylates MDM2 at Ser395 (in green). Phosphorylated MDM2 interacts with the p53 mRNA. This promotes the interaction with the 5S RNP complex and an increase in p53 synthesis. The p53 mRNA–MDM2 is required to bring ATM to the

p53 polysome where it phosphorylates the nascent p53 protein to prevent MDM2-mediated degradation of the newly synthesized p53.

4.7.2. Role of ITAFs in p53 translation

Translation initiation of p53 mRNA is positively and negatively regulated by several *trans*-acting factors (**Fig. 15A**):

- The large ribosomal subunit protein RpL26 preferentially binds to the 5'UTR of p53 following DNA damage thereby activating p53 translation (Ofir-Rosenfeld et al., 2008; Takagi et al., 2005).
- Nucleolin is a nucleolar protein which binds to the 5'UTR of p53 and suppresses its translation (Takagi et al., 2005).
- Wrap53 is a long non-coding RNA transcript antisense to TP53 which interacts with IRES I and modulates p53 mRNA levels (Mahmoudi et al., 2009, 2011).
- PTB acts as a chaperone that binds to pyrimidine-rich region in the IRES I or IRES II upon DNA damage and enhances translation of both isoforms (Khan et al., 2013; Pickering et al., 2004). It can also bind to the 3'UTR of p53 mRNA (Grover et al., 2008).
- DAP5 is Death-Associate Protein 5 (also called p97) that activates translation of p53 by interacting with both IRESes with a preferential binding to IRES II. In fact, DAP5 controls the ratio of both p53 isoforms production (Weingarten-Gabbay et al., 2014, 2016).
- Ku protein is a DNA repair protein which binds to a bulge in the stem loop structure located 53 nucleotides upstream from AUG1 in IRES I. It binds upon genotoxic stress and represses p53 translation (Lamaa et al., 2016).
- MDM2 (Murine Double Minute 2) is an important negative regulator of p53. It binds to p53 protein and triggers its ubiquitination to facilitate its degradation by the proteasome (**Fig. 15B**). However, it acts as a positive modulator of p53 mRNA as well. Indeed MDM2 binds to IRES II of p53 and enhances its translation. Another role for MDM2 is to interact with RpL26 and promote its proteosomal degradation thereby attenuating the RpL26-mediated activation of p53 translation (Ofir-Rosenfeld et al., 2008).

Altogether, this suggests that the synthesis of the full-length p53 protein and/or its isoform depends on the interplay between the binding of several *trans*-acting factors on p53 mRNA which presents a platform for these ITAFs according to the stress-conditions.

Interestingly, there exist specific physiological functions for each of the two isoforms. The truncated form p53/47 lacks the N-terminal 40 amino acids that comprise the first transactivation domain (TA1) for MDM2 binding. Following DNA damage, the two isoforms affect the cell cycle progression by targeting different kinase inhibitors (Mlynarczyk and Fåhraeus, 2014). Experimental studies have shown that there is a particular link between p53/47, aging and neurodegenerative diseases like Alzheimer disease (Lee et al., 2001; Pehar et al., 2014).

5. Translational regulatory elements in the 5'UTR of mRNAs

Cellular mRNAs harbour different regulatory elements in their 5'UTR and 3'UTR. These elements can be linear motifs or more sophisticated structural elements which include stable stem loop structures, pseudo-knots or G-quadruplex. Linear motifs are mostly RNA-binding protein motifs and upstream Open Reading Frames (uORFs). These elements can either inhibit or stimulate cap-dependent translation solely by their presence or through the recruitment of specific RNA-binding proteins (**Fig. 16**).

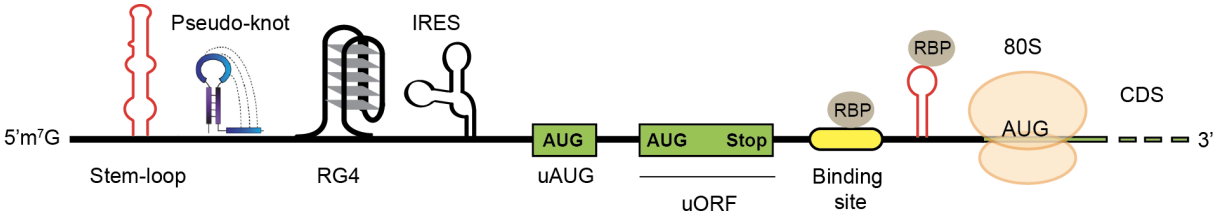


***Figure 16*: regulatory elements and structures in the 5'UTR of mRNA regulate translation.** The mRNA 5'UTR can harbour different structural elements or motifs, that influence mRNA translation. Structural elements like stem-loops, pseudo-knots and RNA G-quadruplexes (RG4) are usually inhibitory. Internal Ribosome Entry Sites (IRESes) initiate translation independently of the 5' cap. Upstream AUGs (uAUGs) and upstream Open Reading Frames (uORFs) are mainly inhibitory of the translation on main ORF. RNA-binding proteins (RBPs) can either bind a sequence motif or a structural motif in the 5'UTR to form RNPs that activate or repress translation.

5.1. RNA Structural elements in the 5'UTR

The average length of 5'UTR of cellular mRNAs ranges from 100 to 200 nucleotides and the GC content plays an important role in determining the stability of folded structures (Gray and Hentze, 1994; Pesole et al., 2001). The 5'UTR can fold into stable secondary structures that play a major role in the translation regulation. In fact, secondary structures seem to be rampant in mRNA subsets encoding for transcription factors, growth factors, oncogenes and more generally for proteins that are poorly translated under normal physiological conditions (Davuluri et al., 2000).

5.1.1. Stable stem-loop structures

Stable stem loop structures have been described as major regulators of translation. More than 90% of mRNA families previously mentioned have secondary structures with a Minimal Free Energy (MFE) of less than -50 Kcal/mol. Among these, 60% of this population comprise stable hairpin structures located in the proximity of the 5' cap (Babendure et al., 2006). In this case, a hairpin with an MFE of -30 Kcal/mol is enough to block the recruitment of 43S PIC onto the mRNA. When the hairpin is present further downstream, the 43S PIC scanning is aborted by a structure with an MFE of -50 Kcal/mol (Araujo et al., 2012). These structures are quite efficient in translation inhibition because they are stable enough to resist to the unwinding activity of eIF4A helicase.

5.1.1.1. The stem-loop in the 5'UTR of TGF-β1 mRNA

One good example of such structural-mediated inhibition has been described in the mRNA coding for Transforming Growth Factor TGF-β1. TGF-β1 is a cytokine, which regulates a number of processes including cell proliferation, differentiation. The TGF-β1 mRNA is poorly translated suggesting the presence of inhibitory element in the 867-nucleotide long 5'UTR (Fraser et al., 2008; Jenkins et al., 2010; Romeo et al., 2014). It was shown that a highly conserved GC-rich stem loop structure (-24 Kcal/mol), spanning from nucleotides 77 to 106 in the 5'UTR, is composed of eleven GC base pairs (**Fig. 17A**). This stem loop is able to dramatically reduce the translation of TGF-β1 mRNA (Jenkins et al., 2010). However, this stem loop structure is not the only mediator of inhibition. Others have shown the involvement of a pool of RNA-binding proteins, among these the Y-Box protein 1 (YB-1) (**Fig. 17B**). YB-1 is a major component of messenger Ribonuclear Protein complexes (mRNPs) that are translationally inactive. This protein also stabilizes the mRNA because it contains a cold shock domain which protects the 5' cap and displaces eIF4G and eIF4E factors (Evdokimova et al., 2001; Nekrasov et al., 2003). This protein has a preferential binding site with a high GC

content rather than specific binding sites on the 5'UTR. The high affinity binding of YB-1 on TGF-β1 mRNA GC motifs promotes their duplex formation leading to low translational efficiency (Fraser et al., 2008).

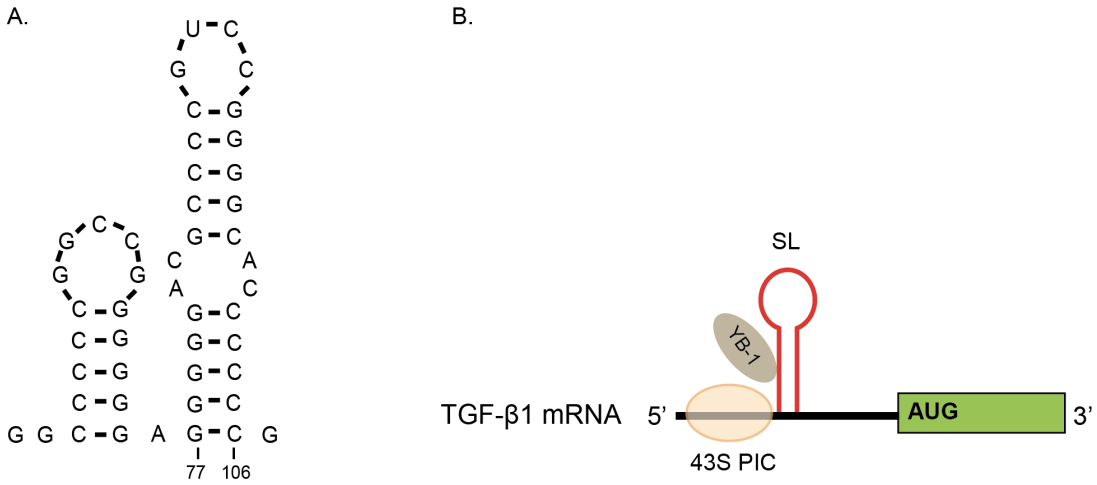


***Figure 17*: the translational repression of TGF-β1 mRNA mediated by a 5'UTR structure. (A)** Secondary structure of a conserved GC-rich stem-loop structure in TGF- β1 mRNA (77-106) that composed of eleven GC base pairs that efficiently inhibits translation. **(B)** YB-1 protein binds a GC-rich motif in the 5'UTR of TGF-β1 mRNA and mediates the formation of duplex leading to translational repression.

5.1.1.2. The Iron-Response Element (IRE)

Regulation of iron homeostasis through storage and metabolism processes is indispensable for cells (Hentze and Kuhn, 2002). One of the mRNAs involved in homeostasis is the ferritin mRNA which encodes the ferritin protein, an iron-storage protein. This mRNA comprises a structural element called Iron-Response Element (IRE) in its 5'UTR. This IRE motif is a 30-nucleotide stem loop structure located 50 nucleotides downstream the 5' m^7G cap. Two Iron-Regulatory Proteins (IRPs), IRP1 and IRP2 bind the loop of IRE with high affinity under iron deprivation conditions (Eisenstein and Munro, 1990) (**Fig. 18**). This prevents the recruitment of the 43S PIC onto the ferritin mRNA thereby inhibiting its translation and preventing iron storage. On the contrary, when iron is sufficient, IRP1 is post-translationally modified whereas IRP2 is degraded by the proteasome leading to their inactivation. In this case, the ferritin protein is produced from the ferritin mRNA which leads iron storage. This mechanism is a representative example of the interplay between structural elements and RNA-binding proteins required to maintain homeostatic physiological conditions.

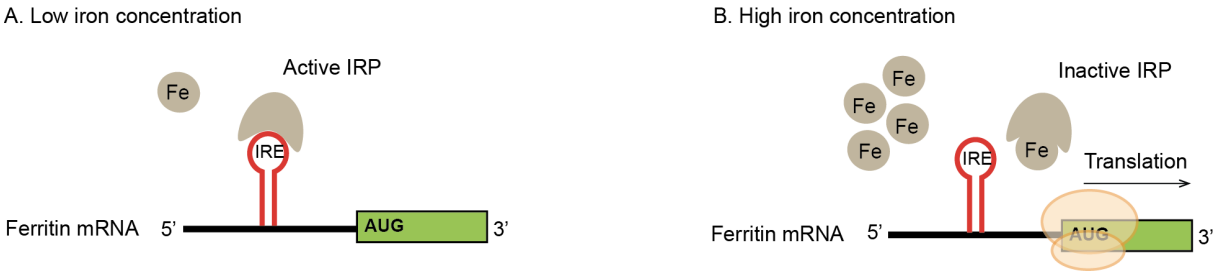


***Figure 18*: regulatory elements in Ferritin mRNA mediate its translation activation or repression. (A)** Under low iron concentration, iron-regulatory proteins called Iron-Response Proteins (IRPs) bind a stem-loop structure, called Iron-Response Element (IRE), in the 5'UTR of Ferritin mRNA with high affinity and inhibit translation of Ferritin mRNA. **(B)** When iron concentration is high, iron binds to the IRPs and inactivates them. This allows translation of Ferritin mRNA.

Altogether, this suggests that stable secondary structures are able to reduce the translational efficiency depending on their stability, size and position in the 5'UTR. The interplay between such structures and specific helicases will be discussed later (section 5.1.4).

5.1.2. Structures containing pseudo-knots

Aside from stable stem loop structures, some mRNAs contain sequence motifs that are able to form long-range interactions thereby creating structures called pseudo-knots (Leppek et al., 2018). Pseudo-knots are tertiary motifs formed when the apical loop from a stem-loop structure is base-paired to another distant sequence. These long-distance interactions create a so-called pseudo-knot structure. Classification of pseudo-knots is quite challenging but a general designation has been attributed where H represents a hairpin and L is a loop or a bulge. According to their arrangement, pseudo-knots are mostly categorized as type "H" because they comprises two base-paired stems and two or three single-stranded loops (van Batenburg, 2001; Pleij, 1994). Double-nested pseudo-knots occur when the loop from a stem-loop pseudo-knot forms another nested pseudo-knot (Peselis and Serganov, 2014). Accordingly, there exist six subclasses of pseudo-knots: H, HH, HL_{out}, HL_{in}, LL, and HHH (**Fig. 19**). Functionally, they represent important RNA regulatory elements because they usually fold in order to form platforms for RNA-binding proteins that are *trans*-acting factors.

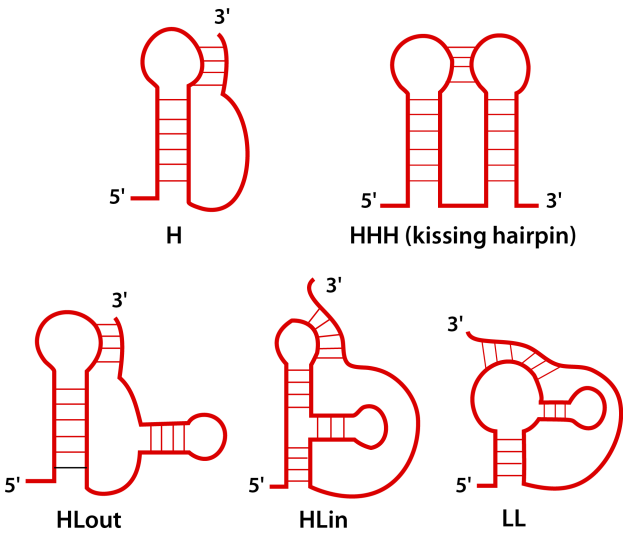


***Figure 19*: types of RNA pseudo-knots structures.**

A representative example of pseudo-knot that regulates translation is found in the mRNA encoding interferon-gamma (INFG) protein. INFG activates the protein kinase PKR. A feedback mechanism takes place that lead to inhibition of INFG mRNA translation. The inhibition mechanism is the following. The 5'UTR and the 5' first 26 codons of the coding sequence of INFG mRNA contain a so-called PKR activating domain. This activator domain folds into a pseudo-knot structure that is composed of short helical segments which are composed of 33 bases pairs (**Fig. 20**). Such a long helical structure is able to trigger PKR activation and thereby assemble a PKR dimer.

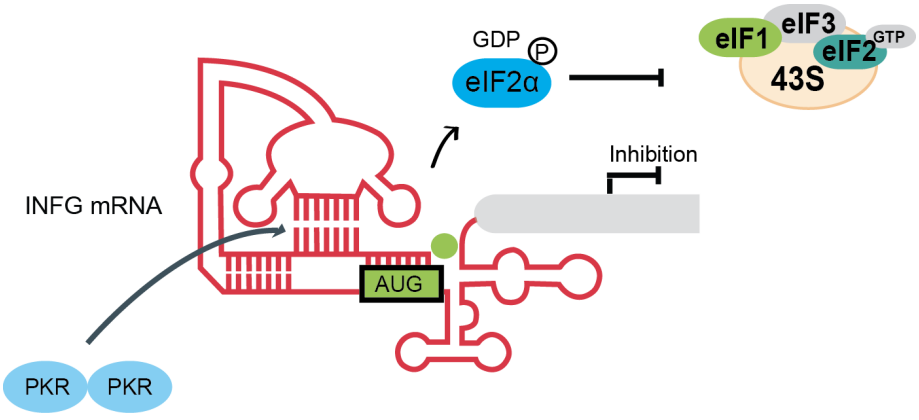


***Figure 20*: interferon-gamma (INFG) mRNA harbours pseudo-knot structures.** INFG activates the protein kinase PKR in which a feedback mechanism takes place and the latter inhibits INFG mRNA translation. The 5'UTR and part of the CDS contain the PKR activating domain. The activator domain constitutes pseudo-knot structures folding into short helical segments responsible for PKR activation through interacting with PKR dimer. PKR then phosphorylates eIF2α resulting in translation repression.

Then, the activated PKR phosphorylates eIF2α resulting in translational repression as previously explained in section 3.5 (Ben-Asouli et al., 2002; Cohen-Chalamish et al., 2009). INFG mRNA translation only takes place when the pseudo-knot is unfolded. Pseudo-knots are also found in the coding sequence. In this case, they can induce ribosome pausing and subsequent frame-shifting to produce different protein isoforms or even promote mRNA decay through destabilization (Advani and Dinman, 2016; Farabaugh, 2002).

5.1.3. RNA G-quartets

Other tertiary interactions can form in mRNAs to produce form so-called G-quartets. These are extremely stable structures formed in G-rich sequences that are found in RNA and also in DNA. G-quartets are formed by four G residues that interact to form a square planar cyclic hydrogen-bond pattern. The G residues interact with each other by making non-Watson-Crick hydrogen bonds at their Hoogsteen face thereby creating Hoogsteen pairing. With these kind of interactions, they form an atypical 3D structure called G-quadruplex or G-quartets. These structures are extremely stable and require a monovalent cation (preferentially K^+ or eventually Na^+) that coordinate the interaction between the G residues in the G-quartets (Balasubramanian and Neidle, 2009; Bugaut and Balasubramanian, 2012). These structures have been shown to play an important role in the translation regulation of specific mRNAs. In Fragile X Mental Retardation mRNA (FMR), a 35-nucleotide G-quadruplex was identified in the coding sequence. FMRP seems to bind on its own mRNA by recognizing the G-quartet structure (Schaeffer et al., 2001). In fact, FMRP is an RNA-binding protein that can repress translation initiation or elongation on several other mRNAs containing G-quadruplexes (Castets et al., 2005; Kumari et al., 2007).

G-quadruplets and RAN translation

Another example of G-quartet regulatory function occurs in the neurodegenerative disorder Amyotrophic Lateral Sclerosis and Fronto-Temporal Dementia (ALS/FTD). In these patients, the C9ORF72 mRNA was aberrantly spliced in such a way that an intron is retained in its 5'UTR. This intronic sequence contains a near-cognate CUG codon located 24 nucleotides upstream of G_4C_2 repeated sequences. In ALS/FTD patients, these repeats occur 30 to several thousands times. They adopt a highly stable G-quadruplex structure that forces the scanning PIC to start translation on this CUG codon by a so-called Repeated Associated Non-AUG (RAN) translation (Tabet et al., 2018). This aberrant initiation event leads to the production of a toxic polyGlyAla protein also called dipeptide repeats protein (DPR) (**Fig. 21**). During

translation, a fraction of the ribosome translating the G quartets shifts to -1 and +1 frame to produce polyGlyArg and polyGlyPro proteins respectively. The production of these three types of DPR are toxic for motor neurons leading to cell death of this particular cell type (Westergard et al., 2019). Altogether, this suggests that not only G-quadruplex can serve as regulators of translation either as stimulators or inhibitors, but can also provoke frame-shifting during translation elongation (Yu et al., 2014).

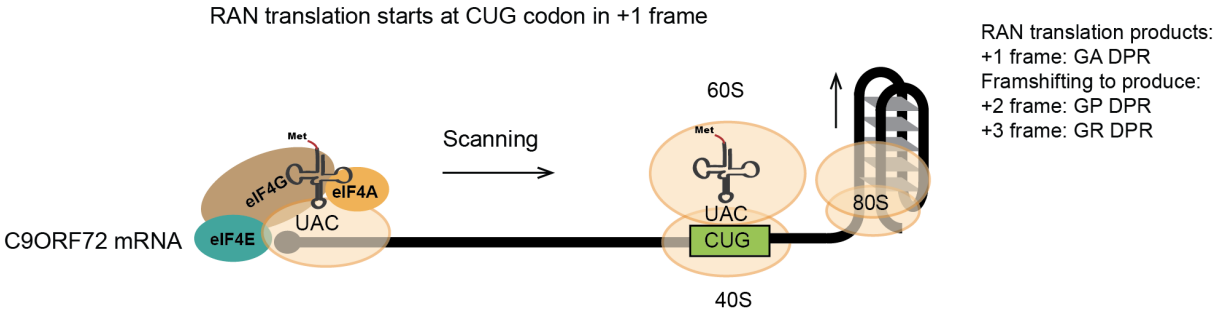


***Figure 21*: RNA G-quartets mediate RAN translation in C9ORF72 mRNA.** In the neurodegenerative diseases Amyotrophic Lateral Sclerosis and Fronto-Temporal Dementia (ALS/FTD), G_4C_2 repeats occur from 30 to several thousands times in the C9ORF72 mRNA. These transcripts accumulate in the cytoplasm of ALS/FTD patient motor neurons and produce toxic dipeptide repeats proteins (DPR). These DPRs are produced by ribosomes by a so-called Repeated Associated Non-AUG (RAN) translation mechanism. The ribosome initiates on a CUG codon located upstream of the G_4C_2 hexamers. In fact, these repeats fold into a highly stable G-quadruplex. The presence of this RG4 forces the ribosome to initiate at an AUG-like codon and mediates aberrant translation initiation. Accordingly, three different DPRs are produced, the poly-GA from the frame +1 and due to frame shifting to the +2 and +3 frames to produce the poly-GP and poly-GR DPRs respectively.

5.1.4. Role of RNA helicases during translation

RNA helicases are enzymes that are able to unwind RNA secondary structures. They are key players in translational control. Indeed, RNA helicases involved in translation initiation regulation are referred to as Superfamily 2 (SF2) that belong to the DEAD-box family (Linder et al., 1989). This family harbours two RecA domains to which the ATP binds. Allosteric conformational changes induced by RNA enable ATP binding. The better characterized helicase is eIF4A, a classical DEAD-box helicase. Together with eIF4E, the cap binding protein and the platform eIF4G, it constitutes the eIF4F complex. After binding the cap eIF4A unwinds encountered structures in the 5'UTR to facilitate ribosome scanning. Eventhough eIF4A is considered as general helicase, recent studies have suggested that it

regulates the translation of specific mRNAs and more particularly those with high GC content, GA-rich and U-rich motifs in their 5'UTR (Modelska et al., 2015). Moreover, repression of eIF4A by specific inhibitors such as silvestrol and rocaglamide have shown to reduce primarily the translational efficiency of mRNAs with long 5'UTRs like those are usually mRNA produced from oncogenes (Iwasaki et al., 2016; Rubio et al., 2014). Similarly, mRNAs containing G-quadruplex appear to be sensitive to eIF4A inhibition too which suggests that they might be also targeted by eIF4A (Wolfe et al., 2014).

In addition to eIF4A, it seems that other helicases are involved in unwinding structures present in 5'UTR structures (Marintchev, 2013). Examples like DHX29, Ded1/DDX3, DHX9/RHA helicases have been shown to be involved in the translation initiation of specific mRNA subsets (Boeras et al., 2016; Cordin et al., 2006; Parsyan et al., 2009). For example, DHX29 is a multi-domain helicase that binds directly to the 40S ribosomal subunit. This interaction is mediated by a double stranded RNA Binding Domain (dsRBD) and a so-called Winged Helix (WH) domain. The binding of DHX29 to the 40S ribosome remodels the mRNA entry channel and enables mRNA positioning. Then, it unwinds hairpin structures in the 5'UTR. However, the molecular mechanism of this process and the requirement of other eIFs remain elusive. Another example is the mammalian DDX3 helicase (Ded1p in yeast) which can bind eIF4G, eIF4E and PABP (Hilliker et al., 2011). DDX3 is required for translation initiation on mRNAs containing 5' proximal secondary structures that are close the 5' cap (15 nucleotides or less). The main function of DDX3 in these cases is to unwind such proximal structures to permit the access of eIF4E to the 5' cap (Soto-Rifo et al., 2012). Altogether, this suggests that helicases are indispensable for either unwinding the secondary structures in the 5' UTR to pave the way for ribosome scanning or to regulate translational activation or repression (de la Cruz et al., 2002). Moreover, it seems that some mRNAs have specific requirements for helicase activity. They recruit the helicases by interacting with motifs located in the mRNA itself. The most important parameters for RNA helicase activity are the 5'UTR structures stability, their length and their 5' cap proximity

5.2. Linear motifs in the 5'UTR

In addition to structural elements, the 5'UTR harbours other unstructured regulatory elements that modulate translation initiation. These can be RNA-binding proteins motifs or upstream Open Reading Frames (uORFs) and upstream AUG start codons (uAUGs).

5.2.1. Binding sites for RNA binding proteins

RNA Binding Proteins (RBPs) are major regulatory elements of translation (Araujo et al., 2012). The human genome encodes for approximately thousand RBPs, most of which are implicated in translation. Accordingly, there exist two categories: RBPs that are essential for translation because they are components of the translational machinery and RBPs which are *trans*-acting regulatory proteins that modulate translation. Generally, RBPs from the latter category recognize specific motifs (linear or structural) in the UTRs and then interact with the translation machinery to govern regulation. I previously explained ITAFs (sections 4.6.1E), as RBPs required for specific IRES activities. In cellular IRESes, a wide range of ITAFs regulates the activity of p53 IRESes which was previously explained (section 4.7B). However, other RBPs regulate the cap-dependent translation initiation. One example is the two iron-regulatory proteins (IRPs), IRP1 and IRP2 that bind the loop of IRE with high affinity under conditions of iron deprivation (section 5.1.1B) (Eisenstein and Munro, 1990). Taken together, this suggests than RBPs can recognize structural or sequence motifs in the 5'UTR to either activate or repress translation.

5.2.2. uAUGs and uORFs

Upstream AUG codons (uAUGs) and upstream Open Reading Frames (uORFs) represent critical regulatory elements located in 5'UTRs (Barbosa et al., 2013). Upstream ORFs are located in the 5'UTRs with a termination codon located either upstream or downstream of the main coding sequence. Upstream AUGs are start codons without an in-frame stop codon. In the human transcriptome, almost 50% of the mRNA contain uORFs and uAUGs in their 5'UTR, the average length is 22 codons and a significant level of sequence conservation is present suggesting a function role of the encoded peptides (Iacono et al., 2005). uORFs are quite frequent in oncogenes, cell cycle and stress response-related mRNAs. So far, their translation have been associated with reduced protein production due to reduced initiation at the main AUG (Calvo et al., 2009; Morris and Geballe, 2000), mRNA decay (Mendell et al., 2004), or translation regulation of mRNAs that are specifically induced by stress conditions (Spriggs et al., 2010).

5.2.2.1. uORFs are translational regulatory elements

As the ribosome scans along the 5'UTR, several factors affect its recognition of the AUG start codon. The context of surrounding nucleotides, known as Kozak consensus sequence defines the efficiency of this recognition. As previously mentioned, the positions -3 and +4

nucleotides define the strength of the context (A from AUG being numbered +1 by convention). Accordingly, A or G at -3 and G at +4 are optimal for translation initiation (Kozak, 1986). If the flanking sequence is not optimal, leaky scanning will occur, meaning the ribosome will scan through the uAUG. However, when it is recognized, translation will be initiated at the uAUG and alternative fates of the ribosome might occur (**Fig. 22**):

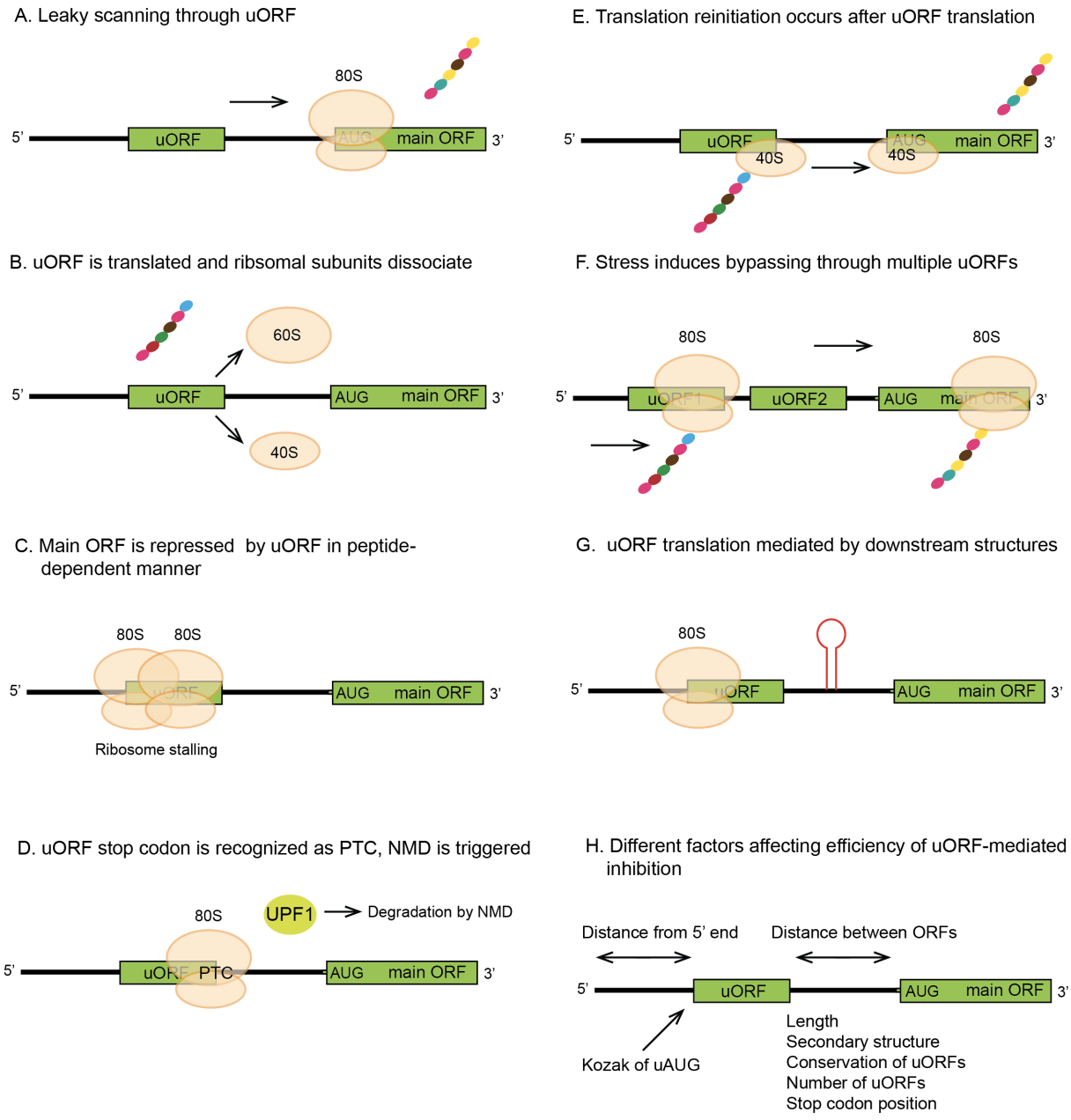


***Figure 22*: uORF-mediated translational regulation. (A)** One mechanism is leaky scanning through uORF and translation of main ORF. **(B)** Alternatively, the uORF is translated, a peptide is produced, released, and then the ribosomes dissociate without reinitiating at the main ORF. **(C)** The uORF can also be translated and the peptide induces a stalling of the ribosomes on the uORF, leading to inhibition of main ORF. **(D)** When the stop codon of uORF is recognized as a premature termination

codon (PTC), it triggers UPF1-mediated Nonsense Mediated Decay (NMD). **(E)** The ribosome scans downstream to reinitiate at the main ORF **(F)** During stress conditions, the ribosome initiates at uORF1 but it requires more time to reinitiate since there is less ternary complex available so it bypasses the next uORF and reinitiates at the main ORF to ensure translation of mRNA. **(G)** Structures present downstream of near-cognate uORFs require specific helicases for unwinding. When these helicases are not present, these uORFs are translated but when structures are unwind by these helicases, it triggers leaky scanning through the uORFs **(H)** In summary, the effect of uORF-mediated translational control depends on several factors: the distance from the 5' m^7G cap, the distance between the ORFs, the Kozak consensus sequence of uAUG, length of the uORF, the secondary structure where the uORF is embedded, conservation of uORFs among species (determinant of functional significance), the number of uORFs in the 5'UTR and the uORF stop codon position with respect to the main AUG codon.

- It translates the uORF in order to produce a peptide which can have a specific function. After translation the ribosome dissociates thereby down-regulating the translation of main ORF further downstream.
- It is stalled during elongation or termination along the uORF. This induces a pause for other translating ribosomes that will create colliding ribosomes and affects mRNA stability by inducing mRNA decay.
- It remains associated with the mRNA after translation of the uORF and reinitiates at the proximal main AUG. Such event, however, is common after translation of short uORFs (MEIJER and THOMAS, 2002). Reinitiation also depends on the length of the uORF, which will determine the translational rate and the initiation and elongation factors which remain associated with the ribosome (Roy et al., 2010).

Several studies have shown that the mode of action of uORFs depends on several factors: (i) the proximity of the uAUG to the 5' cap, (ii) the Kozak context of the uAUG, (iii) the length of the uORF, (iv) the presence of structural element that might disturb the accessibility of the uORF by ribosomes that stall the scanning of PIC, (v) the evolutional conservation of the peptide sequence encoded by the uORF (vi) the number of uAUGs in the 5'UTR and (vii) the distance between the uORF stop codon and the AUG codon of the main coding sequence.

5.2.2.2. Translational repression by uORFs-encoded peptides

One example of uORF-mediated repression by a *cis*-acting uORF peptide is the highly conserved eukaryotic Arginine Attenuator Peptide (AAP). This peptide is implicated in arginine specific translational repression of specific mRNAs that are translated in response to high arginine concentrations. This peptide is translated from a uORF that is present in the

5'UTR of these mRNAs (Wang et al., 1999; Wei et al., 2012). In fact, repression occurs by stalling of the ribosome at the uORF termination codon. This occurs because arginine residues, encoded by the uORF, alters the AAP's conformation in the P-site of the ribosome thereby inducing ribosomal stalling (Gaba et al., 2001; Wu et al., 2012).

5.2.2.3. Regulatory role of multiple uORFs under stress conditions

In *Saccharomyces cerevisiae*, the transcription factor GCN4 activates amino acid biosynthesis (Dever et al., 2016). The 5'UTR of GCN4 mRNA harbours four small uORFs (**Fig. 23**). The rate of translation of uORFs depends on nutritional conditions, except for the first uORF which is constitutively translated. Under normal conditions, the translational machinery favours the translation of the uORFs 2-4 in order to represses the main coding sequence. Two *trans*-acting factors, GCD1 and GCN2, orchestrate the interplay between the uORFs translation. GCD1 inhibits translation initiation at the 5' proximal uAUG (uORF1) under normal conditions. GCN2 phosphorylates eIF2α subunit under amino acid starvation. The latter reduces the loading of the ribosomes on the mRNA due to less availability of ternary complex since eIF2 is activated by phosphorylation.

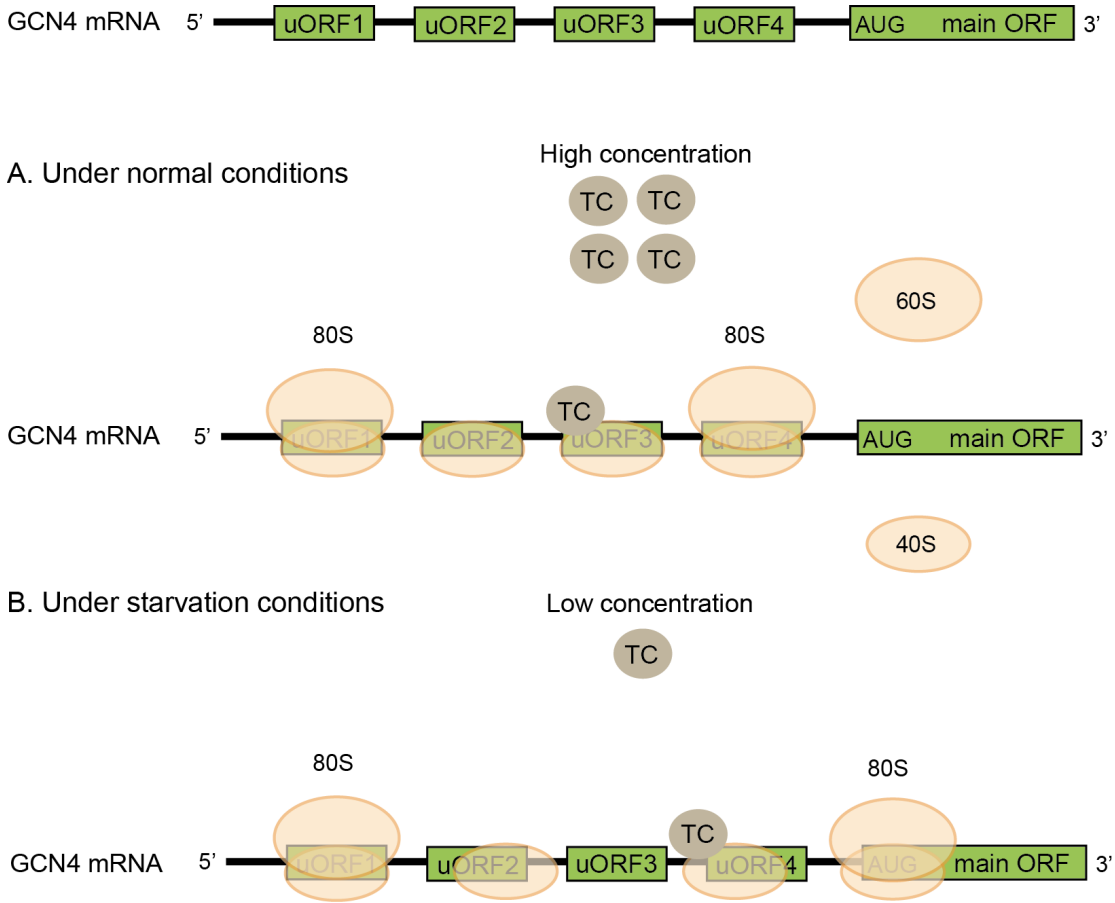


***Figure 23 (previous page)*: translational regulation model by four uORFs in the yeast GCN4 mRNA.** The yeast mRNA GCN4 contains four uORFs in its 5'UTR. These uORFs play an important role in promoting repression or translation of the main ORF under different conditions. **(A)** Under normal conditions, where the level of ternary complex (TC) is high, the ribosome translates from uORF1 in the 5'UTR of GCN4 and reinitiates at uORF4 since TC is available to bind the 40S in short time during scanning. **(B)** Under starvation conditions, TC level is low due to phosphorylation of eIF2 by GCN2 (not shown in the figure). This leads to reinitiation on the main ORF after translation of uORF1 since the 40S scans more distance since it needs more time to assemble the TC.

Accordingly, under amino acid starvation conditions, only the 5' proximal uORF (uORF1) is translated and the ribosome requires more time to assemble the ternary complex in a way that it bypasses the uORFs 2-4 and finally reinitiates only on the main coding sequence (Hinnebusch, 2005; Mueller and Hinnebusch, 1986). In contrast to the general thought about uORF-repressive mechanisms, in the case of GCN4, the uORF1 is indispensable for the PIC assembly and therefore it enhances the translation of the main ORF under conditions where global translation is repressed.

5.2.2.4. Translational regulation by uORF/IRES interplay

Several studies have shown that uORF can modulate the cap-independent translational elements in the 5'UTR, either by favouring or repressing them. Examples include Cationic Amino acid Transporter 1 (CAT1) and Fibroblast Growth Factor 9 (FGF9) mRNAs (Leppek et al., 2018).

CAT1 mRNA regulation by uORF/ IRES

In CAT1 mRNA, an IRES element (192 nucleotides) within the 5'UTR (270 nucleotides) mediates its translation in a cap-independent manner under nutritional stress (**Fig. 24**). This IRES requires another regulatory element, a uORF located within the IRES (from -224 to -78 nucleotides upstream of the main AUG) which encodes a 48-residue peptide (Yaman et al., 2004). In fact, a strong long distance interaction between the 5'UTR and the 3' inactivates the IRES. This occurs under normal nutritional conditions when amino acids are available. During amino acid starvation, the uORF is translated within the IRES, the ribosome induces a structural remodelling to liberate the IRES from the dormant stable conformation (ΔG = -98 kcal/mol) and transform it into an active state (ΔG = -50 kcal/mol). This is not the only requirement since eIF2α phosphorylation is also important. This leads to the selective synthesis of ITAFs specifically required for the CAT1 IRES activation (Bonnal et al., 2003; Fernandez et al., 2002). This cascade of events leads to the translation of CAT1 mRNA. This

model of uORF-mediated regulation is termed 'zipper' because it is a very dynamic activation process. In this case, the uORF plays a dual role; under normal conditions, it is not translated thereby the IRES remains in the dormant stage, and under deprivation, it acts as a positive switcher of the IRES to induce CAT1 expression.

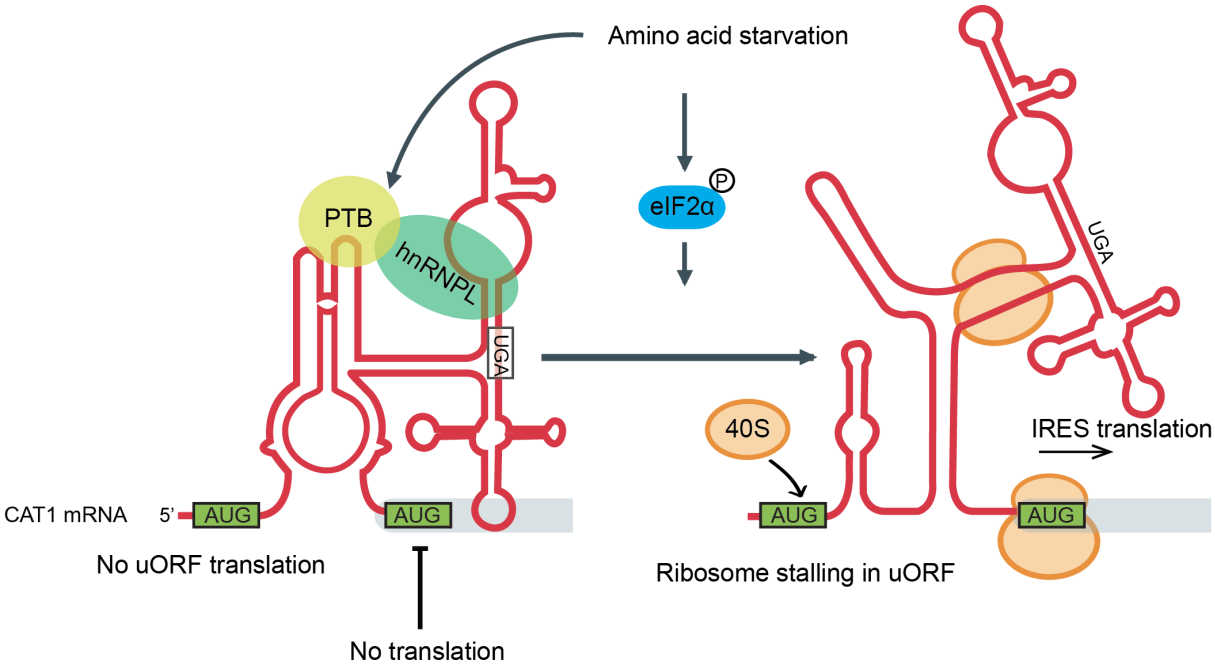


***Figure 24*: regulation of IRES activity by a uORF in CAT1 mRNA.** Under amino acid starvation conditions, a uORF within the IRES is translated which promotes ribosome stalling. This induces a structural remodelling of the IRES to the active state and initiates translation of the main ORF. The association of ITAFs like PTB and heterogeneous nuclear Ribonucleoprotein L (hnRNPL) to the IRES enhances this step and is required for translation initiation during starvation.

VEGF-A regulation by uORFs/ IRESes

The Vascular Endothelial Growth Factor A (VEGF-A) is important for vasculogenesis, angiogenesis and plays an important role in many other physiological processes. Its expression is aberrantly increased in cancer, arthritis and a number of pathological conditions (Ferrara, 2002). The 5'UTR of VEGF-A mRNA comprises two independent IRESes (A and B) and two alternative start codons, an AUG for the main ORF and the AUG-like codon CUG for a uORF (**Fig. 25**). These IRESes mediate translation of VEGF-A under hypoxia-induced stress when cap-dependent translation is repressed.

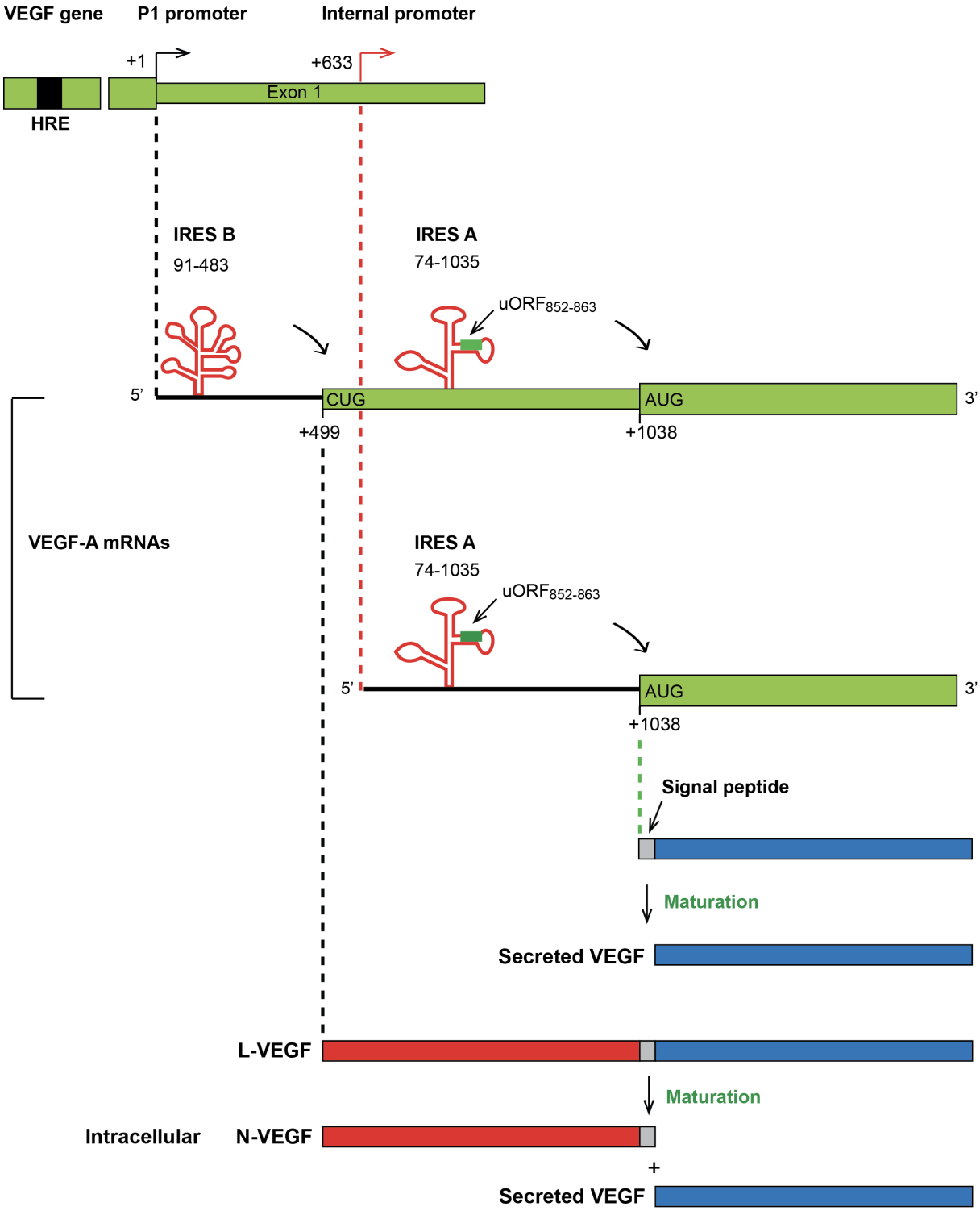


***Figure 25*: regulation of VEGF protein isoforms by two IRESes in VEGF-A mRNA.** The VEGF gene has a Hypoxia Responsive Element (HRE) and two promoter sites: (P1) directs the transcription of full length mRNA and an Internal Promoter (IP) which directs the transcription of a truncated mRNA. The full length mRNA harbours two IRESes that mediate translation on two alternative initiation sites. IRES-A initiates on the main AUG, which starts at position 1038 whereas IRES-B initiates on an in-frame CUG at position 499, which produces a longer isoform called L-VEGF. The cleavage of L-VEGF generates secreted-VEGF and the intracellular N-VEGF. The second isoform

harbours just the IRES-A and produces the secreted VEGF. IRES-A also harbours a uORF starting from position 852, which plays a role in initiation on main ORF.

IRES A mediates translation from the main AUG (at nucleotides 1039-1041) and is actually embedded in the uORF. IRES B controls translation from an in-frame upstream CUG (at nucleotide 499-501) to produce a longer isoform referred to as L-VEGF (Huez et al., 1998, 2014). IRES A, located 300 nucleotides upstream from the main AUG, harbours a 49-nucleotide structural domain (named D4) that is required for the initiation at the AUG without scanning in a similar manner to class II of IRESes. Another uORF is embedded in IRES A and seems to be translated in a cap-independent manner in order to promote translation on the AUG by reinitiation (Bastide et al., 2008; Bornes et al., 2007). In addition, this uORF also acts as a *cis*-regulatory element that suppresses the translation from the CUG codon (L-VEGF isoform). Taken together, these uORF control the levels of VEGF isoforms expressed during hypoxia.

5.2.2.5. uORFs and mRNA decay

In the last decade, several studies have suggested a link between uORFs and mRNA decay. Non-sense Mediated Decay (NMD) is a quality control mechanism that occurs when a premature termination codon (PTC) is present in the coding region (Rehwinkel et al., 2006). NMD aims to eliminate such mRNA that might generate potentially toxic truncated proteins. The consequence of the presence of a PTC in the coding region is in fact that mRNAs have much longer 3'UTRs. The current NMD model suggests that in the case of unusually long 3'UTR, release factor eRF3 that recognizes the PTC, fails to interact with PABP which is bound to the the 3' polyA tail, because of a too big distance. Since this interaction is required for efficient translation termination, the termination complex cannot assemble. Instead, it binds to another factor UPF1 which recruits UPF2 and then UPF3 finally leading to mRNA degradation by the NMD pathway (Mühlemann, 2008; Silva and Romão, 2009). In fact, many of the NMD targets are mRNAs containing uORFs because their stop codons are very far from the 3'UTR end. Indeed, the termination codon of uORF seems to be recognized as a PTC since it is located further away from the 3'UTR thereby creating a long 3'UTR downstream of the uORF, an ideal target for NMD. One example of uORF-triggered NMD is the Interferon-Related Developmental regulator 1 mRNA (IFRD1) (Zhao et al., 2010). IFRD1 mRNA is degraded by the NMD pathway in resting cells because the uORF in the 5'UTR is translated, which also reduces translation of main ORF. Indeed, the instability of IRFD1

mRNA is UPF1-dependent triggered by the length and sequence of the uORF strongly suggests that it involves NMD pathway (Chang et al., 2007; Somers et al., 2013). Under stress conditions, the stability of this mRNA is enhanced together with eIF2α phosphorylation. Consequently, leaky scanning from uORF translation induces translation of the main ORF explaining why IFRD1 is a stress-induced mRNA. In plants, uORFs also induce NMD in a size-dependent manner with a 50 amino acid-long uORF being sufficient for efficient degradation of the mRNA (Nyikó et al., 2009).

5.2.2.6. uORFs translation and reinitiation

Like previously mentioned, after translation initiation at a uORF, a small percentage of the ribosomes will reinitiate translation further downstream at the main ORF (Ingolia et al., 2011). More precisely, at the termination complex on the uORF stop codon, the 60S is released from the complex. The synthesized peptide is also released however, the 40S subunit remains associated on the mRNA in order to reinitiate some nucleotides further downstream. However, the TC at this step is absent, meaning, its recruitement is necessary before the next initiation occurs. Several studies have addressed the mechanism of reinitiation after uORFs translation. In one of the studies, it has been shown that the reinitiation efficiency depends on the length of the uORF and the intercistronic distance between the uORF stop codon and the next AUG codon from the main ORF (Luukkonen et al., 1995). Others have shown that the uORF translation rate is also an important determinant for reinitiation occurance (Kozak, 2001). This suggests two prerequisites for reinitiation: first, several eIFs remain attached to the 40S during elongation and are reused for the following reinitiation event, second, factors such as the ternary complex needs to be recruited *de novo* during the scanning step, highlighting the importance of the scanning time for the efficiency of subsequent reinitiation. Indeed, the time of scanning is directly linked to the eIF recruitment efficiency for reinitiation.

6. Role of DENR/MCT1 complex in reinitiation event

The Density Regulated Protein (DENR) is a non-canonical factor shown to be involved in ribosome recycling and translation initiation in specific viral mRNAs by recruiting Met-tRNA$^{Met}_{i}$ (Dmitriev et al., 2010; Skabkin et al., 2010). DENR is also involved in reinitiation events after translation of short uORFs, by forming a heterodimeric complex with another protein called Multiple Copies T-cell lymphoma protein Subunit 1 (MCTS1) (**Fig. 26**)

(Schleich et al., 2014). In fact, this complex was the first described to be involved in the reinitiation process without being required for the first initiation event. The reinitiation with this complex takes place independently of the distance between the two ORFs precisely becauso it is not implicated in the first round of initiation. However, it is not the case for eIF2. While the recruitment of eIF2 is needed for both initiation events, reinitiation is more efficient when there is a longer distance between the two ORFs. In *Homo sapiens*, there is a preference of the DENR/MCTS1 complex for the regulation mediated by short stuORFs (uORFs with a strong Kozak sequence). Another family of uORFs, which are also preferred by the DENR/MCTS1 complex, contains minimal coding sequences composed of an AUG start codon immediately followed by a stop codon, they are called start-stop sequences (Schleich et al., 2017). Reinitiation requires the delivery of a novel initiator $tRNA^{Met}$ into the P-site of the ribosome and this is performed by the canonical initiation factors. The question is why is there a need for dedicated machinery? Investigations have addressed this question by demonstrating that in some mRNAs, the DENR/MCTS1 complex is required to remove the deacylated tRNAs and the mRNA that are still present in the ribosome in the presence of the dissociating factor ABCE1 after translation of the first ORF. This will efficiently recycle the 40S ribosomal subunit for reinitiation (Skabkin et al., 2010).

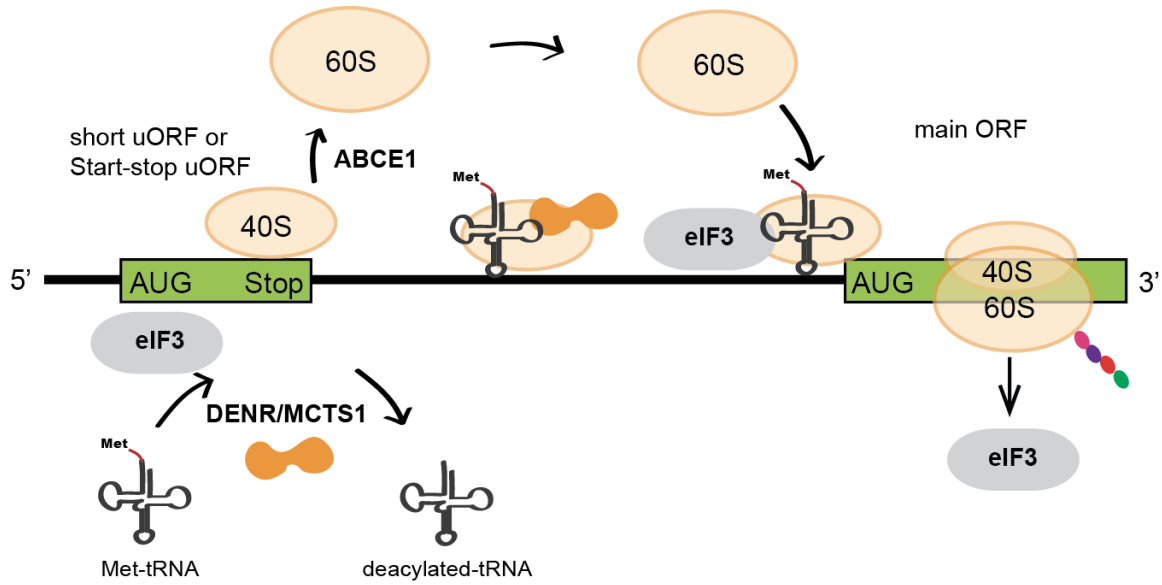


***Figure 26*: model for reinitiation mediated by DENR/MCTS1 complex.** After translation of short uORFs, the 40S ribosome might remain assembled on the mRNA after ABCE1-mediated release of 60S ribosome. The reinitiation complex DENR/MCTS1 recruits the Met-$tRNA^{Met}_i$ into the 40S which will initiate after the assembly of eIF3 on the main ORF. This is preferably performed during translation of short or start-stop uORFs.

7. Non-canonical initiation factor eIF2D

Specific mRNAs are translated using non-canonical translation factors. In this section, I will briefly describe one initiation factor which has been suggested to be involved in the translation of specific mRNAs. The factor eIF2D is a 65 kDa protein, previously identified as Ligatin (LGTN), a membrane receptor of glycoproteins. However, later on, it has been shown that this protein is a eukaryotic initiation factor with two functional domains, an N-terminal domain called PUA and a C-terminal domain termed SUI1 (Dmitriev et al., 2010). Interestingly, PUA is an RNA-binding domain also present in MCTS1. SUI1 is ribosome binding domain homologous to DENR and the initiation factor eIF1 that is important for scanning and start codon selection (Mitchell and Lorsch, 2008; Reinert et al., 2006; Vaidya et al., 2017).

7.1. Structure of eIF2D

Using cryo-EM, the structure of eIF2D bound to the 40S ribosomal subunit in a reinitiation complex has been determined (**Fig. 27**) (Weisser et al., 2017). The structure consists of three globular domains. The first one is in close proximity to mRNA channel near helix 44 of ribosomal RNA suggesting contacts with the ribosome. The second is below the so-called platform of the ribosome. The third domain interacts directly with rRNA helix 44. These modules are connected via flexible domains. The structure also shows that eIF2D interacts with the 40S ribosomal subunit in similar manner like canonical factors eIF2 and eIF1. More precisely, SUI1 domain occupies the same position as eIF1 explaining the fact that eIF2D-mediated translation can be inhibited by eIF1 (Skabkin et al., 2010). In other words, eIF2D and eIF1 can compete for the binding on ribosomal 40S subunit suggesting a similar function for eIF2D and eIF1 in start codon selection.

In this study, important homologies between the DENR/MCTS1 complex and eIF2D are presented (Weisser et al., 2017). For example DENR/MCTS1 and eIF2D interact via the same binding site on the 40S ribosomal subunit although eIF2D has an additional domain, the winged helix (WH) domain, which interacts with helix 44. Another minor difference exists in the SWIB domain of eIF2D, which is slightly shorter than that of DENR. Such differences could explain the different modes of action of both factors. Altogether, this strongly indicates distinct functions, eIF2D being important for ribosome recycling whereas DENR/MCTS1 seems to be involved in reinitiation.

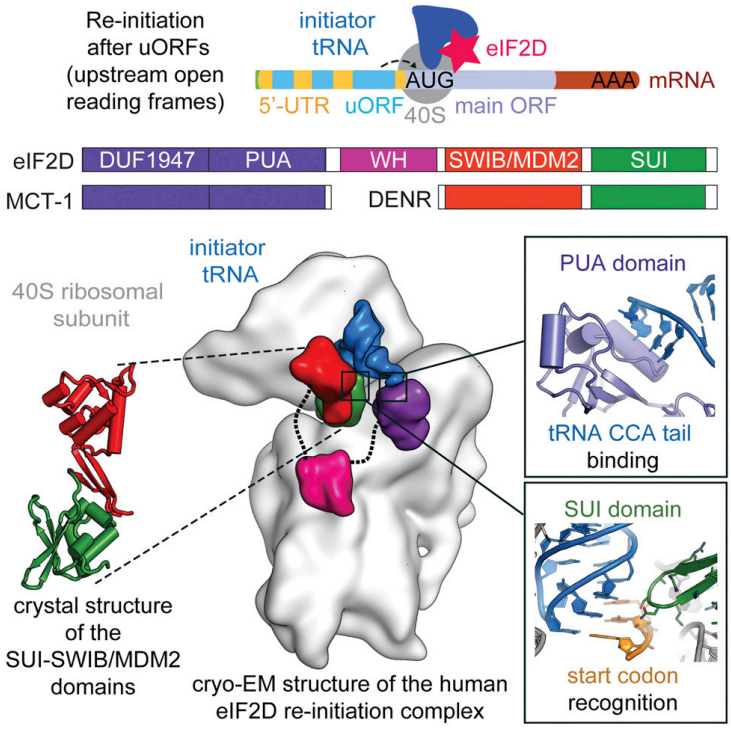


***Figure 27*: cryo-EM structure of eIF2D in a reinitiation complex.** Figure adapted from (Weisser et al., 2017). Topology of eIF2D and its homologous complexes MCT-1 and DENR are shown in the upper pannel. Cryo-EM structure shows the binding site of eIF2D domains (SUI and PUA) on the 40S ribosomal subunit.

7.2. Functions of eIF2D

Using biochemical approaches, it was shown that eIF2D can deliver both initiator and non-initiator tRNA to the P-site of the ribosome without the requirement of GTP hydrolysis. Therefore eIF2D was qualified as a GTP-independent eIF. This unprecedented activity is different from eIF2-mediated delivery of Met-tRNA$^{Met}_{i}$, which strictly requires GTP hydrolysis. It has been proposed that eIF2D has different functions (**Fig. 28**). These include translation initiation on leaderless mRNAs (lacking 5'UTR) for which canonical scanning is seemingly not possible (Akulich et al., 2016). eIF2D might also be required for mRNAs translation of uORFs and formation of initiation complexes on mRNAs with A-rich 5'UTRs (Dmitriev et al., 2010). Due to competition with eIF1 for the same binding site on the ribosome, eIF1 seems to impair the formation of eIF2D mediated complexes. As previously mentioned, eIF2D is able to deliver elongator tRNAs to the P-site of the ribosome, a feature that is distinct from eIF2. In addition, eIF2D plays a role in ribosome recycling after termination. Addition of eIF2D to terminating ribosomes prior to incubation with initiation factors promotes the release of all components (Skabkin et al., 2010). Also, eIF2D is similar

to DENR/MCTS1 complex by promoting the release of deacylated tRNAs and mRNAs from recycled 40S ribosomal subunits in the absence of ABCE-1. Taken together, eIF2D has multiple possible roles in translational control. In some cases, it can promote translation initiation on specific mRNAs, mediate ribosomal subunit recycling or repressing reinitiation on other m

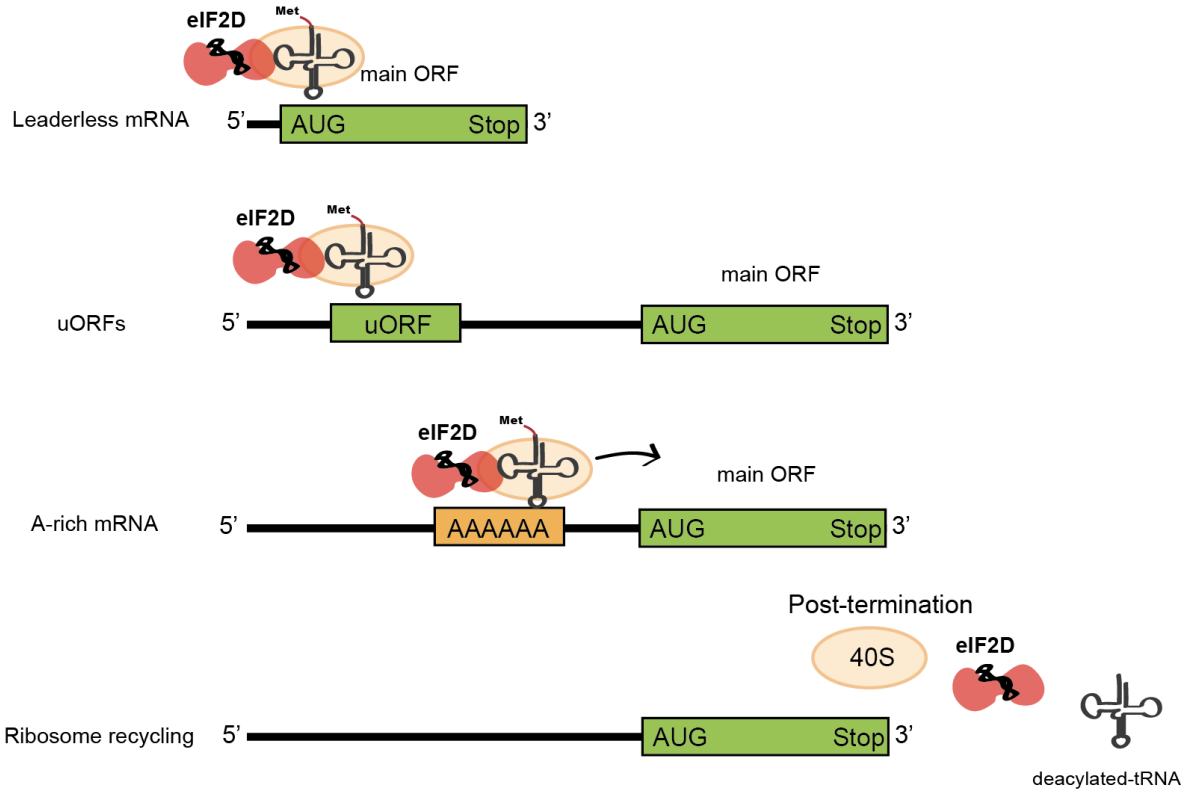


***Figure 28*: functions of eIF2D.** Several functions have been proposed for eIF2D. Initiation on leaderless mRNAs (lacking 5' UTR), initiation on uORFs, formation of initiation complexes on A-rich 5'UTR mRNAs and ribosome recycling after dissociation of 40S subunit and removal of deacylated tRNA.

8. Overview of Hox genes

Homeobox (Hox) genes encode a family of proteins, which are major transcriptional regulators during embryonic development. These genes are evolutionarily conserved in all animals (**Fig. 29**) and are present in every metazoan with no exception. Hox genes encode Hox proteins that possess a 60 amino acids homeodomain and a hexapeptide (HX) motif (Gehring et al., 1994).

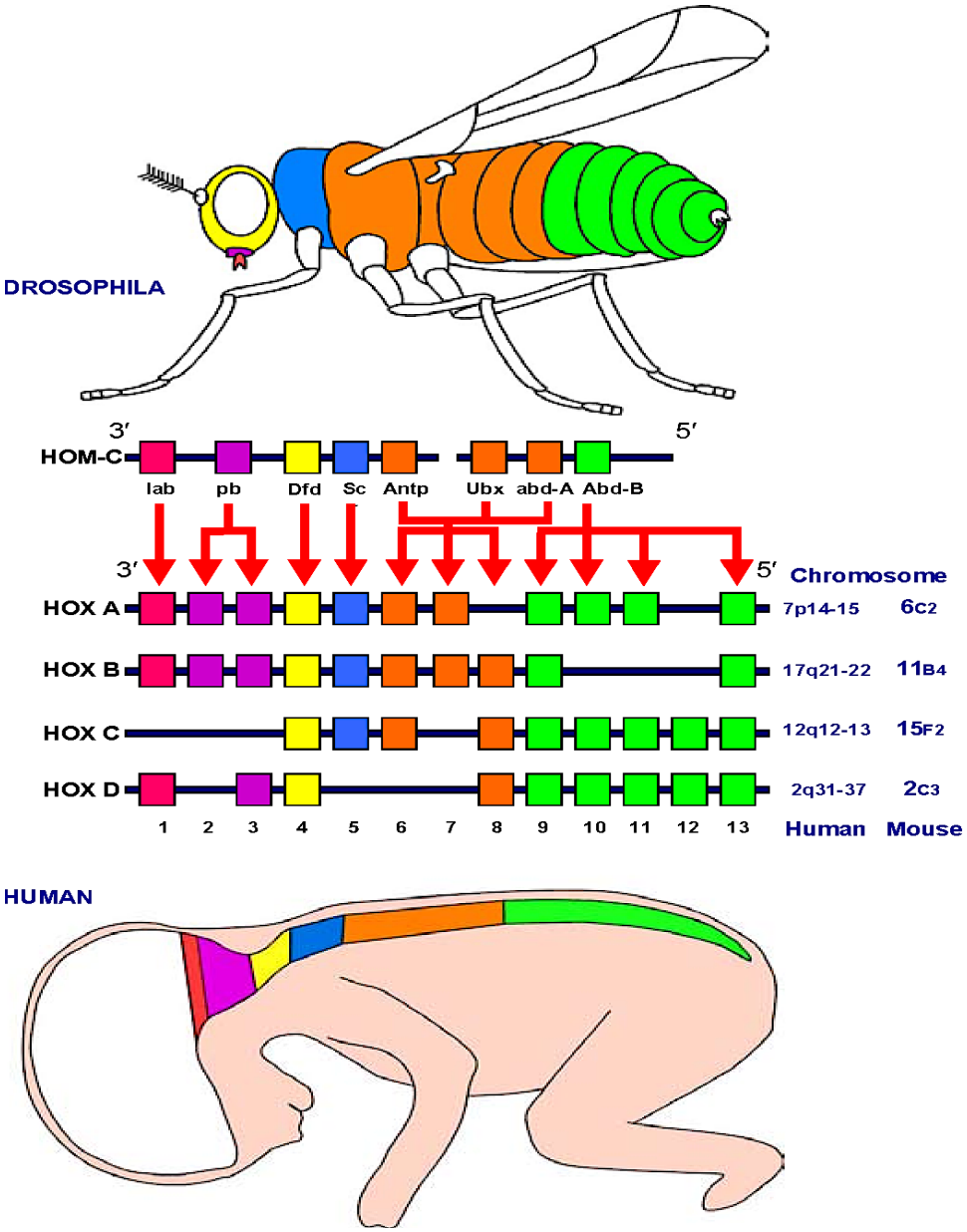


***Figure 29*: conservation of Hox genes clusters in *D. melanogaster* and *H. sapiens*.** Figure adapted from (Lappin et al., 2006). Hox genes in drosophila are arranged in one cluster called the Hom-C cluster composed of eight genes. In mammals, these genes are expanded into 39 genes arranged in four clusters (A-D) on different chromosomes. The expression pattern of these genes correlates with the chromosomal positioning as shown for human and mouse. The 3' genes are expressed more anteriorly and earlier than the 5' genes.

Interestingly, other genes harbour the homedomain motif however they also contain specific modifications thereby they do not belong to the Hox gene family. Hox proteins target genes to activate or repress their transcription. By doing so, they regulate body patterning and organogenesis during embryonic development. The way individual Hox genes are ordered within the Hox cluster reflects their order of expression along the head-to-tail axis of the

embryo (**Fig. 29**) (Deschamps and van Nes, 2005). In other words, genes at the 5' end of cluster are expressed more posteriorly and the genes located at the 3' are expressed more anteriorly in the embryo. This phenomenon is called spatial colinearity (Iimura et al., 2009).
Hox genes provide a paradigm for several areas in modern biology. First, from a developmental perspective, Hox genes constitute a genetic unit governing the fate of segmental identity along the animal body axes during early development. Thereby, they represent a great model to investigate how transcription factors coordinate networks of subordinate genes and how do they guide the behaviour of cell populations during morphogenesis (Mallo and Alonso, 2013; Pearson et al., 2005). Second, given their remarkable evolutionary conservation across distant animal phyla, they represent an abstract system of cardinal information able to operate within a wide spectrum of invertebrates and vertebrates, thus wondering how the same set of developmental genes can be involved in the generation of largely diverse developmental identities. Third, their genomic organization and molecular regulation is highly complex, opening up the possibility to look at the molecular mechanisms by which genomic information is used to enable physiological and morphological processes as development occurs (Lemons and McGinnis, 2006).

8.1. Classification of Hox genes

Hox genes are organised in clusters. They originate from a bilaterian ancestor that carries seven Hox genes (Garcia-Fernàndez, 2005). The first Hox genes that were discovered in *Drosophila melanogaster* possess one cluster termed homeotic complex (Hom-C). It has been proposed that during Cambrian explosion, an expansion of Hox genes occurred and their number increased due to genome duplication events (**Fig. 30**) (Kmita and Duboule, 2003). Interestingly, Hox genes display paralogue groups and more specifically, vertebrates harbour thirteen groups clustered in anterior, central and posterior loci on four different chromosomes. So far, four classes of structural organisation of Hox clusters have been described (**Fig. 31**). In vertebrates, the Hox clusters are highly organized. They can also be totally disorganized like in sea urchins, or splited like the Hom-C in drosophila, or finally atomized like the one in urochordates (Rezsohazy et al., 2015). In humans, the 39 Hox genes are distributed in four clusters (A to D) on different chromosomes at loci 7p15, 17q21.2, 12q13, and 2q31 (Kmita and Duboule, 2003).

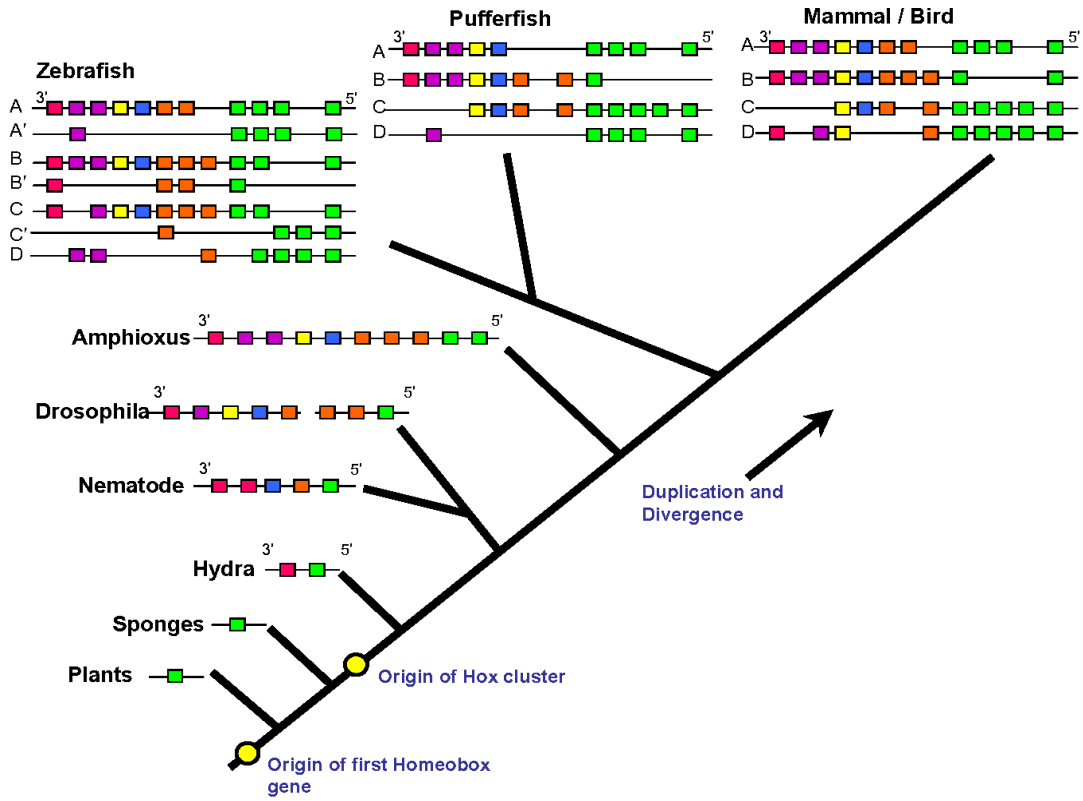


***Figure 30*: a representative dendrogram showing the evolution of Hox genes cluster among different species.** Figure adapted from (Lappin et al., 2006). Hox gene cluster expansion is assumed to occur due to a genome duplication event and divergence from an ancestral gene estimated to have risen 1 billion years ago.

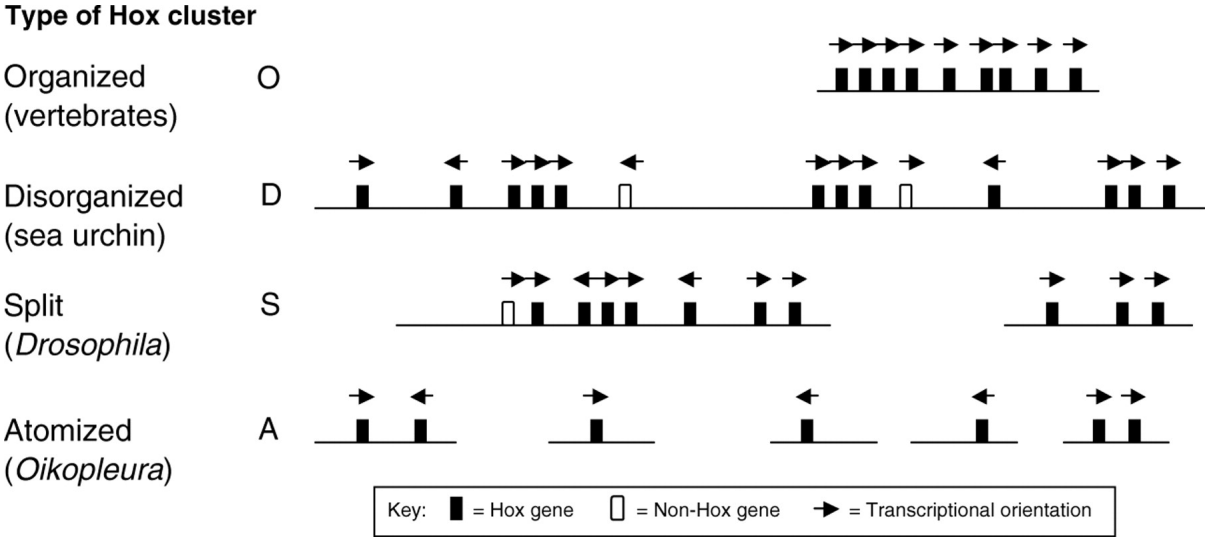


***Figure 31*: structural classification of Hox genes clusters.** Figure adapted from (Kmita and Duboule, 2003). Type O cluster is organized; they are typically present in vertebrates. This type encompasses genes that are tightly arranged on the same DNA strand. Type D cluster is disorganized whereby the genes are much larger and occur in opposite orientations. They are interfered with non-Hox genes. This type is mainly found in sea urchins. Type S is split cluster, which is rather a combination of type O and type D. This is found in drosophila. Type A is Atomized cluster, meaning

that genes are paired at scattered loci in the genome and not clustered as for example in *Oikopleura* (tunicate).

8.2. Functions of Hox genes

The main function of Hox proteins, as previously mentioned, is to mediate transcriptional activation or repression of critical downstream gene networks. By doing so, they define cellular territories and positional information in the developing embryo, in other words, body patterning. Eventhough the general path of Hox genes functions has been characterized, however, the mechanistic pathways that govern the whole developmental program are still poorly understood.

8.2.1. Transcriptional activities of Hox proteins

As previously mentioned, these clustered genes are expressed into proteins which contain a conserved homeodomain (**Fig. 32**). Homeodomain is a DNA-binding domain composed of a triple helicoidal structure. The N-terminal part interacts with the minor groove of DNA and the helix 3-motif is a recognition helix, which interacts with the major groove of DNA. Their main function is to act within a regulatory gene network to activate or repress target genes. During embryonic development, these proteins drive the elaboration of different embryonic structures along the anterior-posterior body axis. They also contribute to organogenesis, cellular differentiation, cell proliferation and apoptosis (Sánchez-Herrero, 2013). It has been proposed that the targets of Hox genes are termed 'realizator' genes (García-Bellido, 1975) because they refer to genes connected to basic cellular functions like cell proliferation, differentiation, and cell adhesion. Thereby, it has been quite important to identify the different individual 'realizator' genes to better understand Hox cellular functions. For instance, Hoxc8-deficient mice have a triggered apoptosis in spinal cord segments, particularly in cervical and thoracic motor neurons (C7 to T1) (Tiret et al., 1998). In contrast, Hoxa13 mediates programmed cell death by regulating retinoic acid pathway (Shou et al., 2013). In vertebrates, Hoxa10, c10 and d10 genes direct vertebra identity to either grow ribs or other bones. In mouse embryos, Hox10 genes (a,c,d) switch off the rib program. They are activated in the lower back where vertebra should not grow ribs, and inactivated in the mid back where ribs should be formed. A selective target for activation by Hox genes is p21. In differentiating myelomonocytic cells, Hoxa10 binds directly to the promoter region of the gene coding for p21 with two partners, the proteins PBX1 and MEIS1 in order to activate transcription from the p21 gene (Bromleigh and Freedman, 2000). Such event results in cell cycle arrest and

differentiation of these cells into monocytes. On the level of organogenesis, a gene localised at the 3' extremity like Hox a3 plays a role in the development of organs such as thymus, thyroid and parathyroid glands while a 5' extremity gene like Hox a11 functions in the uterine development and female fertility (Chojnowski et al., 2014).

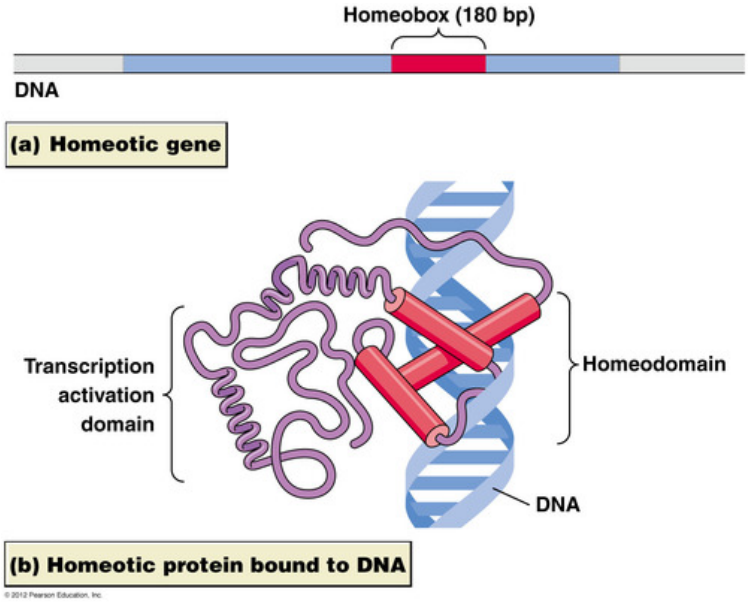


***Figure 32*: the structure of Hox homeodomain.** Figure adapted from (Pearson Education Inc. 2012). Homeodomain is a conserved 60 aa motif that binds the target DNA motif located near the 3' end of the gene. It comprises a stretch composed of a helix-turn-helix structure with three alpha helices. Helix 3 binds the DNA to either activate or repress transcription of the target gene.

8.2.2. Hox are acting with partners

In fact, other DNA-binding proteins drive the specificity of Hox protein for their transcriptional activity. These proteins can also contain homologous homeodomains (HD). Pre-B-Cell (PBC) Leukaemia proteins are general Hox co-factors with an evolutionarily conserved HD. They comprise a Three Amino acid Loop Exension (TALE) in the first helix of homeodomain making them atypical, therefore they belong to the TALE class of HD proteins (Bürglin, 1998). For instance, studies have shown that Hox proteins are either PBC/TALE-dependent or PBC/TALE independent (**Fig. 33**). One example of modulation occurs between Hoxa1 and Pbx1. Pbx1 is a putative Hox co-factor that belongs to the TALE class of proteins and harbours a sixteen-residue C-terminal extension which seems to be required for optimal cooperation with Hox proteins to promote stable interaction with DNA (Bürglin and Ruvkun, 1992; Phelan et al., 1995). In fact, Pbx forms dimers with Hox proteins groups 1-10 (Longobardi et al., 2014). In the Hox-PBC complex, the interaction however, does not rely on HD domain but rather upstream from the Hox side domain (HX) and the

TALE domain from PBC side (Joshi et al., 2007; LaRonde-LeBlanc and Wolberger, 2003). Such an interaction positions the N-terminal extremity of the HD into the DNA minor groove therefore explaining the co-factor effect of PBC. This is referred to as specificity module. Another Hox partner is the Myeloid Ecotropic Viral Integration Site (MEIS) protein which forms dimers with Hox proteins groups 9-13 (Longobardi et al., 2014). This suggests that Hox proteins can have partners that modulate their transcriptional activities responsible for several cellular functions.

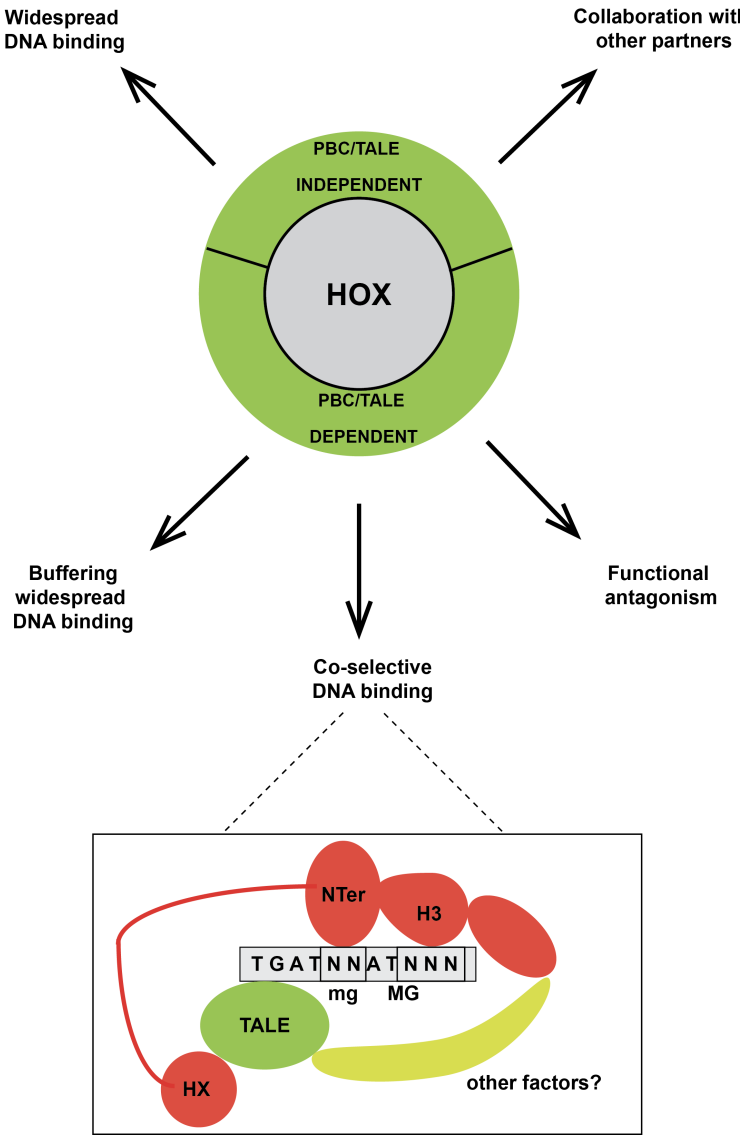


***Figure 33*: multiple pathways for Hox proteins functions.** The modulation of Hox proteins function can be either PBC/TALE independent or PBC/TALE dependent. PBC/TALE proteins act as co-factors with Hox proteins for DNA binding. The Hox HX motif interacts with the hydrophobic pocket formed by the three amino acid loop extension in TALE proteins to direct the positioning of the N-terminal arm (NTer) of the Hox homeodomain (HD) into the DNA minor groove (mg), a contact that defines the identity of the central NN nucleotides. The positioning of the HD recognition helix (helix 3, H3) in

the DNA major groove (MG) through additional factors involves the distal identity (NNN) of the Hox half-sites.

8.2.3. Non-transcriptional activities of Hox proteins

Studies suggest other non-transcriptional functions of Hox including DNA replication and repair, mRNA translation and protein degradation (Lambert et al., 2012; Rezsohazy, 2014).

8.2.3.1. DNA replication

Hoxc13 has been shown to be a member of the DNA replication complexes. This is due to specific interactions between the Hoxc13 HD and lamin B2 replication origin and the replication origins located near TOP1 and MCM4 genes in cells that are in early S phase (Comelli et al., 2009). Another study suggests that Hoxd13, Hoxa13 and Hoxd11 bind human replication origins by interacting with CDC6 loading factor in pre-replication complexes to stimulate DNA synthesis (Salsi et al., 2009).

8.2.3.2. DNA repair

Hox proteins can also be involved in DNA repair. One example is Hoxb7 protein which interacts with DNA-dependent protein kinase to stimulate the DNA repair by Non-Homologous End Joining (NHEJ) in reponse to DNA double-stranded break (Rubin et al., 2007).

8.2.3.3. Translation

Hoxa9 have been shown to interact with eIF4E to enhance cyclin D1 and ornithine decarboxylase (ODC) mRNA export from the nucleus to the cytoplasm and polysome loading to boost its translation. The interaction with eIF4E relies on a Hox-conserved Proline-Rich motif (PRH /Hex) suggesting possible involvement as partners of eIF4E for other Hox proteins (Topisirovic et al., 2005).

8.2.3.4. Protein degradation

Several Hox proteins are also involved in the formation of ubiquitin ligases and their activation. For example, Hoxc10 interacts with a subunit of an E3 ubiquitin ligase complex called CDC27 which is implicated in mitosis exit (Gabellini et al., 2003). These complexes are required for the ubiquitin-dependent proteasome pathway.

8.3. Hox–related disorders and diseases

As previously mentioned, Hox genes are expressed in a temporal and spatial colinear manner during embryogenesis and development (Alexander et al., 2009). Later on, their expression is greatly reduced during adult stage but it is reinforced during early stages of tumour development promoting cancer (Nunes et al., 2007).

Hox genes were originally discovered in *Drosophila melanogaster* due to homeotic transformations observed in the phenotype upon specific loss or gain of function mutations (Maeda and Karch, 2009). Homeotic transformation refers to a condition whereby one organ differentiates to another. One example of such a homeotic transformation is the dominant Antp mutation caused by a chromosomal inversion of hox genes which results in the formation of legs on the head instead of antennas (Kaufman et al., 1990; Lewis, 2007). In humans, ten Hox genes have been linked to genetic disorders so far, these are Hoxa1, Hoxa2, Hoxa11, Hoxa13, Hoxb1, Hoxb13, Hoxc13, Hoxd4, Hoxd10, and Hoxd13 (Boncinelli, 1997; Quinonez and Innis, 2014).

A.

B.

C.

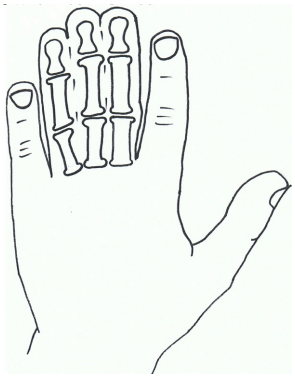


***Figure 34*: Hox-related disorders.** Figures adapted from (A) Texas children hospital (2011) (B) National institute of cancer (2012). **(A)** Microtia is a disorder characterized by a deformation in the external ear canal. Such disorder is due to Hoxa2 mutation. Microtia leads to hearing impairment and

partial cleft palate. There are different levels of severity in microtia-affected individuals. The most severe one is termed anotia, whereby there is a complete absence of the ear and canal. **(B)** Thrombocytopenia, a phenotype that is characterized by the easy bleeding and bruising under skin. It is caused by a point mutation in Hoxa11. **(C)** Synpolydactyly is a phenotype which occurs when extra fingers or toes grow and fuse with other ones. It is caused by a mutation in Hoxd13.

For instance, the mutation of glutamine residue 186, a highly conserved residue of Hoxa2 HD, to a lysine affects its DNA-binding capacity. In fact, such single substitution mutation of Arginine to Lysine in hox proteins leads to a disorder termed microtia, a phenotype characterized by abnormally shaped external ear, hearing impairment and partial cleft palate (**Fig. 34A**) (Alasti et al., 2008). Another disorder is due to defective Hoxa11 termed thrombocytopenia, the phenotype described here is characterized by easy bruising and bleeding (**Fig. 34B**). Molecular genetics of this disorder have shown that this is due to a single point mutation that leads to the deletion of one nucleotide (C872delA) in exon 2 of Hoxa11. This deletion is causing frameshifting and introduces a premature termination codon producing a truncated Hoxa2 protein with an affected homeodomain that is lacking 22 amino acids. (Thompson and Nguyen, 2000). Mutations in Hoxd13 in humans cause a genetic disorder termed synpolydactyly, these patients are growing extra fingers or toes that are fused together (**Fig. 34C**) (Malik et al., 2007).

8.4. Transcriptional regulation

A wealth of experimental data over the last three decades has led to the identification of many *cis*-regulatory elements that control Hox gene transcriptional patterns, thus giving deeper insight into their mechanisms of activity in transcription regulation (Alexander et al., 2009). These studies revealed that transcriptional *cis*-regulation represents a major determinant of Hox gene spatial expression, that *cis*-regulatory elements are functionally autonomous, and that their activity does not absolutely require being inserted within a Hox gene cluster. It was also shown that some Hox transcripts are targeted by microRNAs (miRNAs) for their regulation. One example is the presence of miR-196 target site in the 3'UTR of three Hox paralogues; Hoxb8, Hoxc8 and Hoxd8 (Yekta et al., 2004). In fact, RNA fragments resulting from miR-196 guided endonucleotic cleavages of Hoxb8 were detected in mouse embryos. Also, Hoxa3 is a target for repression by mir-10a, a miRNA expressed from a Hox cluster (Han et al., 2007). However, the effect of this regulation on the developmental patterning has not been clarified so far.

8.5. Translational regulation of Hox mRNAs by two regulons, TIE and IRES

Defects in tissue-specific patterning including homeotic transformations of the axial skeleton, i.e. the replacement of one skeletal muscle by another one, have been linked to mutations in the ribosomal protein from the large 60S subunit RpL38. In mice embryos with mutated RpL38, a subset of Hox mRNAs translation was shown to be disrupted (Kondrashov et al., 2011). Moreover, RpL38 exerts its specialized translational control only when it is associated with the ribosome. From here, it was clear that Hox mRNAs are translated by a atypical mechanism involved in the spatial-temporal expression of Hox proteins.

Until recently, there was no clear evidence for translational regulatory effects on Hox gene expression. The Drosophila genes Antp and Ubx have been suggested to be under translational control during embryonic development (Oh et al., 1992). Precisely, it has been shown that a subgroup of transcripts produced from the Antp and Ubx loci contain functional IRESes that allow their translation using a cap-independent mechanism and that their activity is regulated during development (Ye et al., 1997). However, a recent study also revealed the presence of IRES elements in the 5'UTR of subsets of mice Hox mRNAs including Hoxa3, a4, a5, a9 and a11 mRNAs (Xue and Barna, 2015). Some of these IRESes proved to require the presence of RpL38 in the ribosome to initiate translation which explained the tissue patterning defective phenotype with RpL38 mutations (Kondrashov et al., 2011). Eventhough most cellular mRNAs undergo canonical cap-dependent translation; the activity of Hox IRESes in these mRNAs was found to be critical for their appropriate expression. Following the discovery of the IRES elements in Hox mRNAs, other RNA regulons termed Translation Inhibitory Elements (TIE) were found in the 5'UTR of subsets of Hox mRNAs (**Fig. 35**). TIEs are located upstream of the previously described IRESes (Xue et al., 2015). These elements efficiently inhibit the canonical cap-dependent translation in Hoxa3, a4, a9 and a11 mRNAs, thereby promoting the specialized IRES-mediated translation. In other words, these TIE elements ensure that Hox mRNAs are solely translated by an IRES-mediated translation mechanism. In this context, TIE a3 (170 nts), TIE a4 (484 nts), TIE a9 (342 nts) and TIE a11 (216 nts) were discovered in the 5'UTR of their corresponding Hox mRNAs. These RNA regulons represent the first example of specific elements dedicated to inhibit cap-dependent translation. The mechanism of action of these TIE elements and their structural characterization are still unknown. The objective of my PhD project aimed at better understanding of the molecular basis of TIE-mediated translation inhibition of cap-dependent translation. We focused our investigations on hoxa3 and a11 TIE elements.

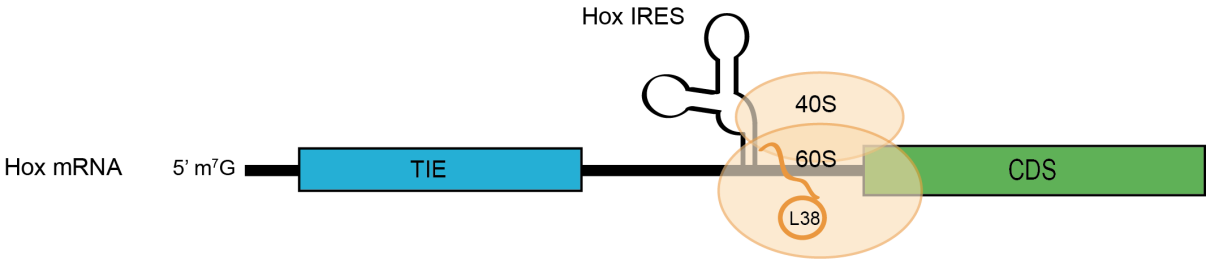


***Figure 35*: subsets of Hoxa mRNAs have two regulatory elements in their 5'UTR.** The first element is the Translational Inhibitory Element (TIE) which inhibits cap-dependent translation. The second element is an IRES which enables the recruitment of the ribosome and is required for the expression of Hox mRNA during embryonic development.

OBJECTIVES OF THE STUDY

Regulation of translation is a critical step to maintain cell integrity. Numerous studies have shown that not all mRNAs that are present in the total mRNA pool are indeed translated by a cap-dependent mechanism. In fact, other alternative initiation elements called cellular Internal Ribosome Entry Sites (IRESes) can mediate translation under specific conditions. As previously mentioned (Introduction section: 4.7), cellular IRESes are not well characterized and their function is generally to promote cap-independent translation in order to produce proteins that enable an adapted response to various cellular stresses (Spriggs et al., 2010). Until recently, the molecular mechanisms that govern translational regulation of Hox mRNAs were not understood. However, studies have shown that Hoxa mRNAs comprise IRESes that are critical for their appropriate expression (Xue et al., 2015). Moreover, the 5' UTR of Hoxa mRNA also contains another element called Translation Inhibitory Element (TIE) that inhibits cap-dependent translation and cooperates with the IRES to regulate Hox protein synthesis.

The evidence of the regulate hox expression in time and space through elements encoded in their messenger RNAs and the fact that Hox genes have been annotated as master regulators of development (Mallo and Alonso, 2013), render the study of Hox RNA regulons of a great interest. The recently discovered TIE RNA regulons in subsets of Hoxa mRNAs pave the way for the IRES element to ensure that Hox mRNA translation is driven solely by IRES-mediated translation (Xue et al., 2015). Interestingly, Hox TIE/IRES interplay allows a tightly controlled translational mechanism for Hox mRNAs during critical developmental stages of embryogenesis. However, the lack of knowledge about TIE elements mechanism of action and their structural characterization were motivating us to initiate this study.

Hence, to gain further insight into the functional and structural characterization of TIE elements, we focused this study on two Hox TIE elements, TIE a3 and TIE a11. The goals of this study were:

- Setting up an efficient *in vitro* cell-free translation assay that faithfully recapitulates TIE-mediated inhibition with a Renilla Luciferase reporter gene and use of this assay to characterize the critical parts of TIEs for their function.
- Establishing the secondary structures of TIE elements by chemical probing.
- Study of the TIE modes of action using various approaches dedicated to translation initiation studies.

Objectives

- Purify pre-initiation complexes blocked by TIE in order to identify *trans*-acting factors.
- Study of the relationships between Hox TIEs and IRESes both at the structural and at functional levels, by addressing this issue, we wondered whether TIE and IRES interact or function as independent modules.

Using an *in vivo* assay using embryonic cells lines that express Hox transcripts, we wanted to validate the results we obtained with our *in vitro* cell-free translation.

MATERIALS AND METHODS

MATERIALS AND METHODS

1. Materials

1.1. Organisms and vectors

1.1.1. *Escherichia coli* strains

TB1

This strain is F^- *ara* Δ *(lac-proAB) [Φ80dlac* Δ *(lacZ) M15] rpsL*(Str^R) *thi hsdR*. These electrocompetent cells were used as transformation hosts. Plasmid DNA was introduced by electroporation.

NEB 5-alpha Competent E. coli

These cells are ultracompetent cells that are highly efficient for assembly of DNA fragments up to 20 Kb. Plasmids were introduced by heat-shock transformation to insert fragments using NEB Builder® *HiFi* DNA assembly kit.

1.1.2. Cell lines

HEK293FT cell line

Human embryonic kidney cell line was cultured according to manufacturer's instructions (ATCC®). This cell line was used for transfection of different vectors and specific siRNAs and luciferase assays.

C3H10T1/2 cell line

Murine mesenchymal cell line, clone 8 (ATCC®) was cultured according to manufacturer's instructions and used for transfection of different vectors for luciferase assays.

1.1.3. Vectors

pUC19

This vector is a commonly used cloning vector. It is a 2.686 Kb size and its origin of replication leads to the production of a high copy number. It conveys tetracycline and ampicilline resistance genes with 54-bp Multiple Cloning Site (MCS) (**Fig. 36**). This vector was used for cloning of TIE a3 (170 nts), TIE a11 (216 nts), full length 5'UTR of Hox a3 (Accession number: NM_010452.3) (333 nts) and full length 5'UTR of Hox a11 (Accesssion number: U20371.1) (496 nts) (Table 1). TIEs and 5'UTRs were inserted upstream of the

5'UTR of human β-globin (Accession number: KU350152) (50 nts) and *Renilla reniformis* Luciferase coding sequence (Accession number: M63501) (936 nts).

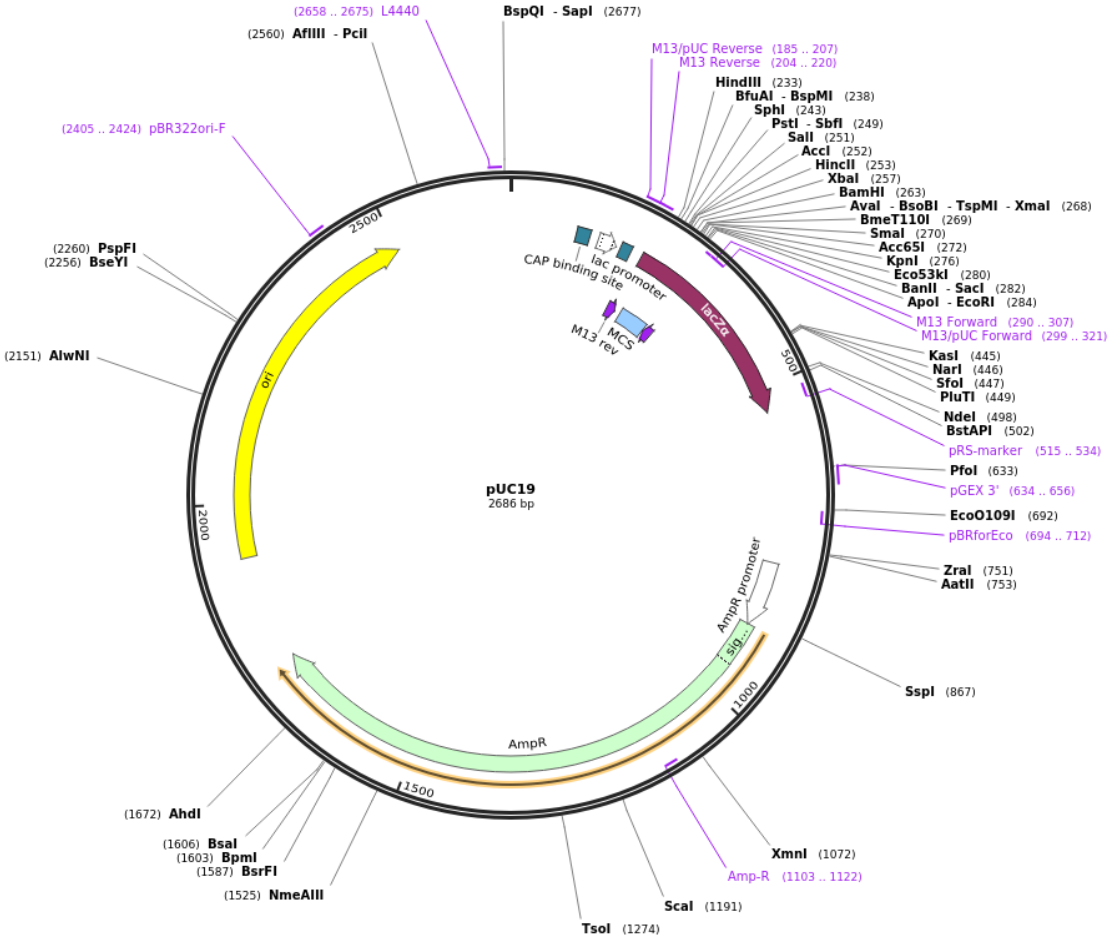


Figure 36: **restriction map of pUC19 vector**

pmirGLO

This 7.35 Kb vector was purchased from Promega, it is based on dual-luciferase technology, with Firefly luciferase (*luc2*) used as the primary reporter to monitor mRNA regulation at 3' UTR and Renilla luciferase (*hRluc-neo*) acting as transfection control reporter for normalization and selection. It has an ampicilline resistance gene, an MCS located in 3'UTR of *luc2* followed by an SV40 late poly (A) signal sequence (**Fig. 37**). Similarily to pUC19, this vector was used for cloning of TIE a3, TIE a11 and their corresponding Hox full length 5'UTRs (Table 1).

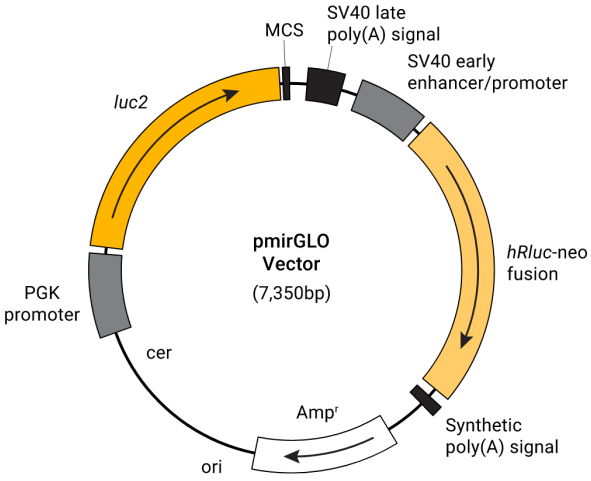


Figure 37: **restriction map of pmirGLO vector**

Name of plasmid	Insert	Reporter gene	Assay
pUC19	TIE a3/a11 FLa3/FLa11	Renilla Luciferase	*In vitro* translation assays, site-directed mutagenesis, transcription
pmirGlo	TIE a3/a11 FLa3/FLa11	Renilla and Firefly Luciferase	*In vivo* transfection and co-transfection (with siRNA) assays, site-directed mutagenesis

Table 1: Plasmids used in different assays. TIE elements and the IRESes were cloned in pUC19 plasmid upstream of Renilla luciferase reporter gene and used for subsequent PCR amplifications and site-directed mutagenesis to obtain mRNA for *in vitro* translation assays and complex assembly. The pmirGlo plasmid has two monocistronic luciferase reporters (Rluc and Fluc), which we have used for transfection assays using different inserts for luciferase assays.

1.2. Oligonucleotides

1.2.1. Gene Blocks

Gene Block	Sequence
XbaI/TIE a3-IRESa3/BstbI	5'CAAGTCTAGACAAAGAAGCTTCTGGGAGCCGCGGTCTGAAGGCTACGTGT GCTGCCTGGTCATTCAAAGTGTCAATTTTAGGTCCAGAAGTGTCCAAACCAC AAGTTCTCAAAACTCTGAAAAATGGCTCCCTCCATGTTAAGAGTGCCTGGAC ACTGGAAAAAAAGATTCCCTGAGCACCTGGAACTGGAGACCATTCCTGGTT CAGATTTAAACATCTTTGGAGGTTTGCGCGCGGCGGCCATTGGCGGCGGAGT GTCACGTGACCGCGGGGGCGTGCCAATGTGCTCCCTTACGGGTGTCAAGCCC

	TTGTCAGAGAGTGTGATCACGATCGTGAAACATCGCGATGACTTCGAAAGTT TATGAT 3’
XbaI/TIE a11-IRESa11/BstbI	5’CAAGTCTAGACAAAAGCTTATCCCTCACCCACCGGGGAGTGGGGGGAGGC GTCAGGAGTTGCGCTCTCTTGCCGCCCTCAGAGCACTCGTGGACCAGGACCA TGTAGCTGAGCAAAGGAGAGCTGCCCGGGGGCGCACCCCAGCCTTTCTGCG CGCGGGGAGGCCCCCCAGCCAACATGAGTTACACCGGCGATTACGTGCTTTC GGTGAGAACACCGAGTGACGATCTGTTGCTTCCCCTGAGGTGGCTACAAAG AAAGGAAGCCGAGGAGGGAGGGGAGGGGAAAAAAGGAAAGGGAGGGGGG TTAAAAAACAAAAAAAAAGCCGGGACTAGCTCGCGGCTTGTCAATTTCAACA TCGGGTCACATGACCAGCACCTCCCTGCTAAGGATGGGGATAGATTTCCACG TCAGCTTACGTCTCCAAATTTCTACTTCACGGATCCGCTTCAAAGAGGCAGC TGCAGTGGAGAATCATGTTAAGCTCGGCTACTGCGGAGAGCCCAAGGTAGC CCAATGACTTCGAAAGTTTATGAT 3’

Table 2: **Gene blocks used for TIE and hox 5’UTR cloning in pUC19.**

1.2.2. Primers for PCR amplification to generate transcription templates by PCR

Name of primer	Sequence
T7-a3	5’**ATATTAATACGACTCACTATA**GGAGAAGCTTCTGGGAGCCGCGGTCTG3’
T7-a3/10	5’**ATATTAATACGACTCACTATA**GGGAGCCGCGGTCTGAAGGCTACGTG3’
T7-a3/28	5’**ATATTAATACGACTCACTATA**GGCTACGTGTGCTGCCTGGTCATTC3’
T7-a3/56	5’**ATATTAATACGACTCACTATA**GGTGTCAATTTTAGGTCCAGAAGTGTCC3’
T7-a3/68	5’**ATATTAATACGACTCACTATA**GGTCCAGAAGTGTCCAAACCACAAG3’
T7-a3/74	5’**ATATTAATACGACTCACTATA**GGAAGTGTCCAAACCACAAGTTCTC3’
T7-a3/77	5’**ATATTAATACGACTCACTATA**GGAAGTGTCCAAACCACAAGTTCTC3’
T7-a3/91	5’**ATATTAATACGACTCACTATA**GGTTCTCAAAACTCTGAAAAATGGCTC3’
T7-a3/106	5’**ATATTAATACGACTCACTATA**GGAAAAATGGCTCCCTCCATG3’
T7-a3/112	5’**ATATTAATACGACTCACTATA**GGCTCCCTCCATGTTAAGAGTGCCTGG3’
T7-a3/138	5’**ATATTAATACGACTCACTATA**GGACACTGGAAAAAAAGATTCCCTG3’
T7-a11	5’**ATATTAATACGACTCACTATA**GGTATCCCTCACCCACCGGGGAGTGGG3’
T7-a11/16	5’**ATATTAATACGACTCACTATA**GGGGAGTGGGGGGAGGCGTCAGGAG3’
T7-a11/23	5’**ATATTAATACGACTCACTATA**GGGGGGAGGCGTCAGGAGTTGCGCTCTC3’
T7-a11/56	5’**ATATTAATACGACTCACTATA**GCCCTCAGAGCACTCGTGGACCAGG3’
T7-a11/78	5’**ATATTAATACGACTCACTATA**GGACCATGTAGCTGAGCAAAGGAGAG3’
T7-a11/99	5’**ATATTAATACGACTCACTATA**GGAGAGCTGCCCGGGGGCGCACCCC3’
T7-a11/110	5’**ATATTAATACGACTCACTATA**GGGGGCGCACCCCAGCCTTTCTGCGCG3’
T7-a11/139	5’**ATATTAATACGACTCACTATA**GGGGAGGCCCCCCAGCCAACATGAG3’
T7-a11/153	5’**ATATTAATACGACTCACTATA**GGCCAACATGAGTTACACCGGCG3’
T7-a11/161	5’**ATATTAATACGACTCACTATA**GGAGTTACACCGGCGATTACGTG3’
T7-a11/171	5’**ATATTAATACGACTCACTATA**GGCGATTACGTGCTTTCGGTG3’

T7-a11/180	5'**ATATTAATACGACTCACTATA**GGTGCTTTCGGTGAGAACACCG3'
T7-a11/188	5'**ATATTAATACGACTCACTATA**GGTGAGAACACCGAGTGACGATCTG3'
T7-β globin 5'UTR	5'**ATATTAATACGACTCACTATA**GGACATTTGCTTCTGACACAACTGTG3'
T7-IRESa3/40	5'**ATATTAATACGACTCACTATA**GGAGGTTTGCGCGCGGCGGCCATTGGC3'
T7-IRESa3/70	5'**ATATTAATACGACTCACTATA**GGAGTGTCACGTGACCGCGGGGGCGTGCC3'
T7-IRESa3/102	5'**ATATTAATACGACTCACTATA**GGTGCTCCCTTACGGGTGTCAAGCCCTTGT3'
T7-IRESa3/141	5'**ATATTAATACGACTCACTATA**GGATCACGATCGTGAAACATCGCGATGAC3'
T7-IRESa11/43	5'**ATATTAATACGACTCACTATA**GGAGGGGAAAAAAGGAAAGGGAGGG3'
T7-IRESa11/89	5'**ATATTAATACGACTCACTATA**GGCCGGGACTAGCTCGCGGCTTGTC3'
T7-IRESa11/134	5'**ATATTAATACGACTCACTATA**GGACCAGCACCTCCCTGCTAAGGATGG3'
T7-IRESa11/184	5'**ATATTAATACGACTCACTATA**GGTCTCCAAATTTCTACTTCACGG3'
T7-IRESa11/221	5'**ATATTAATACGACTCACTATA**GGCAGCTGCAGTGGAGAATCATG3'
T7-IRESa11/252	5'**ATATTAATACGACTCACTATA**GGCTACTGCGGAGAGCCCAAGGTAGCC3'
Rev-Renilla wt	5'GTTCTCAAAAATGAACAATAATTC3'
Rev-Ren/51 (CAA) 10	5'GTGTTGTTGTTGTTGTTGTTGTTGTTGTTGGTTTACATCTGGCCCACCACTGCG G3'
Rev-Ren/151 (CAA) 10	5'GTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCGCCATAAATAAGAAGAGGCCGC GTTACC3'
Rev-glo-(CAA)10	5'GTGTTGTTGTTGTTGTTGTTGTTGTTGTTGGGTGTCTGTTTGAGGTTGCTAGTG AAC3'
Rev-Ren/52	5'GCGGACCAGTTATCATCCGTTTCCTTTGTTCTGG3'
Fwd a3-HindIII	5'CAAAAGCTTCTGGGAGCCGCGGTCTGAAGG3'
Fwd a11-HindIII	5'CAAAAGCTTATCCCTCACCCACCGGGGAG3'
Rev Ren HindIII	5'CAAAAGCTTTTATTGTTCATTTTTGAGAACTCG3'
Rev a11/159	5'AACTCATGTTGGCTGGGG3'
Fwd XbaI/TIE a3	5'CAAGTCTAGACAAAGAAGCTTCTGGGAGCCGCGGTCTGAAGGC3'
Fwd XbaI/TIE a11	5'CAAGTCTAGACAAAAGCTTATCCCTCACCCACCGGGGAGTGG3'
Rev-Renilla 21	5'ATCATAAACTTTCGAAGTCAT3'
Rev a11/159 + GC SL ORF (CAA)	5'AACTCATGTTGGCTGGGGGGCCTCCCCGCGCGCAGAAAGGCTGGGGTGCGCC CCCGGGCAGCTC TTTGGTTTGTTTGTTGTTGTTG3'
Rev a11/79 + ORF (CAA)	5'GCTCTTTGGTTTGTTTGTTGTTGTTGCCTTTGCTCAGCTACATGGTCC3'

Table 3: **Primers used for PCR amplifications.** Different constructs of TIE, IRES and hox 5'UTR were amplified using different primers for cloning, production of 5' truncated constructs, mutations, and CAA tail addition for complex assembly experiments. T7 promoters are shown in bold.

1.2.3. Primers for site-directed mutagenesis

Name of primer	**Sequence**
Fwda3 CUC104/GAG	5'GTGTCCAAACCACAAGTTCTCAAAAGAGTGAAAAATGGCTCCCTCC3'

Reva3 CUC104/GAG	5’GGAGGGAGCCATTTTTCACTCTTTTGAGAACTTGTGGTTTGGACAC3’
Fwda3 GAG132/CUC	5’GAAAAATGGCTCCCTCCATGTTAACTCTGCCTGGACACTGGA3’
Reva3 GAG132/CUC	5’TCCAGTGTCCAGGCAGAGTTAACATGGAGGGAGCCATTTTTC3’
Fwd a3 ATGG111/UACC	5’GTCCAAACCACAAGTTCTCAAAACTCTGAAAATACCCTCCCTCCATG3’
Rev a3 ATGG111/UACC	5’CATGGAGGGAGGGTATTTTCAGAGTTTTGAGAACTTGTGGTTTGGAC3’
Fwd a3 CCAT121/GGUA	5’CTCTGAAAAATGGCTCCCTGGTAGTTAAGAGTGCCTGGACAC3’
Rev a3 CCAT121/GGUA	5’GTGTCCAGGCACTCTTAACTACCAGGGAGCCATTTTTCAGAG3’
Fwd a3 GTCCA79/CAGGU	5’GGTCATTCAAAGTGTCAATTTTAGCAGGTGAAGTGTCCAAACCAC3’
Rev a3 GTCCA79/CAGGU	5’GTGGTTTGGACACTTCACCTGCTAAAATTGACACTTTGAATGACC3’
Fwd a3 AU112/UA and AU124/UA	5’GTTCTCAAAACTCTGAAAATAGGCTCCCTCCTAGTTAAGAGTGCCTGGACAC3’
Rev a3 AU112/UA and AU124/UA	5’GTGTCCAGGCACTCTTAACTAGGAGGGAGCCTATTTTCAGAGTTTTGAGAAC3’
Fwd a3 A111/C and U124/G	5’CAAAACTCTGAAAACTGGCTCCCTCCAGGTTAAGAGTGCCTG3’
Rev a3 A111/C and U124/G	5’CAGGCACTCTTAACCTGGAGGGAGCCAGTTTTCAGAGTTTTG3’
Fwd a3 A111/G and U124/C	5’CAAAACTCTGAAAAGTGGCTCCCTCCACGTTAAGAGTGCCTG3’
Rev a3 A111/G and U124/C	5’CAGGCACTCTTAACGTGGAGGGAGCCACTTTTCAGAGTTTTG3’
Fwd a3 Ins-A- after 220 -Renilla	5’GCAACCTCAAACAGACACCAATGACTTCGAAAGTTTATG3’
Fwd a3 Ins-A- after 220 -Renilla	5’CATAAACTTTCGAAGTCATTGGTGTCTGTTTGAGGTTGC3’
Fwd a3 AAAA107/GGCC	5’CAAACCACAAGTTCTCAAAACTCTGGGCCATGGCTCCCTCCATGTTAAGAGTG3’
Rev a3 AAAA107/GGCC	5’CACTCTTAACATGGAGGGAGCCATGGCCCAGAGTTTTGAGAACTTGTGGTTTG3’
Fwd a3 AAAA107/GGCC	5’CTCCATGTTAAGAGTGCCTGGACACTGGGGCCCCCGATTCCCTGAGCACCTGACATTTGC3’
Rev a3 AAAA107/GGCC	5’GCAAATGTCAGGTGCTCAGGGAATCGGGGGCCCCAGTGTCCAGGCACTCTTAACATGGAG3’
Fwd IRESa3 ΔG333	5’CACGATCGTGAAACATCGCATGACTTCGAAAGTTTATG3’
Rev IRESa3 ΔG333	5’CATAAACTTTCGAAGTCATGCGATGTTTCACGATCGTG3’
Fwd a11 C (4) 120/ G (4)	5’AGCTGCCCGGGGGCGCAGGGGAGCCTTTCTGC3’
Rev a11 C (4) 120/ G (4)	5’GCAGAAAGGCTCCCCTGCGCCCCCGGGCAGCT3’
Fwd a11 G (4) 138/ G (4)	5’CCTTTCTGCGCGCCCCCAGGCCCCCCAGCC3’
Fwd a11 G (4) 138/ C (4)	5’GGCTGGGGGGCCTGGGGGCGCGCAGAAAGG3’
Fwd a11 CGG189/GCC	5’TACACCGGCGATTACGTGCTTTGCCTGAGAACACCGAG3’
Rev a11 CGG189/GCC	5’CTCGGTGTTCTCAGGCAAAGCACGTAATCGCCGGTGTA 3’
Fwd a11 CCG198/GGC	5’ACGTGCTTTCGGTGAGAACAGGCAGTGACGATCTGTTGC3’
Rev a11 CCG198/GGC	5’GCAACAGATCGTCACTGCCTGTTCTCACCGAAAGCACGT3’
Fwd a11	5’GCCAACATGAGTTACACCTCGTATTACGTGCTTTCGGTGAG3’

GGCG171/TCGU	
Rev a11 GGCG171/TCGU	5'CTCACCGAAAGCACGTAATACGAGGTGTAACTCATGTTGGC3'
Fwd a11 TGCT181/GCGG	5'GAGTTACACCGGCGATTACGGCGGTTCGGTGAGAACACCGAGTG3'
Rev a11 TGCT181/GCGG	5'CACTCGGTGTTCTCACCGAACCGCCGTAATCGCCGGTGTAACTC3'
Fwd a11 GTG202/CAC	5'GTGCTTTCGGTGAGAACACCGACACACGATCTGTTGCAC3'
Rev a11 GTG202/CAC	5'GTGCAACAGATCGTGTGTCGGTGTTCTCACCGAAAGCAC3'
Fwd a11 CAC167/GUG	5'CCCAGCCAACATGAGTTAGTGCGGCGATTACGTGC3'
Rev a11 CAC167/GUG	5'GCACGTAATCGCCGCACTAACTCATGTTGGCTGGG3'
Fwd a11 A151-A157 complementary substitution	5'GCGCGGGGAGGCCCCCCTCGGTTCATGAGTTACACCGGCG3'
Rev a11 A151-A157 complementary substitution	5'CGCCGGTGTAACTCATGAACCGAGGGGGGCCTCCCCGCGC3'
Fwd a11 CAUG157/GUCG	5'GGAGGCCCCCCAGCCAAGTCCAGTTACACCGGCG3'
Rev a11 CAUG157/GUCG	5'CGCCGGTGTAACTGGACTTGGCTGGGGGGCCTCC3'
Fwd a11 GUG189/CAC	5'GGCGATTACGTGCTTTCGCACAGAACACCGAGTGACGAT3'
Rev a11 GUG189/CAC	5'ATCGTCACTCGGTGTTCTGTGCGAAAGCACGTAATCGCC3'
Fwd a11 AGAA192/UCUU	5'GCGATTACGTGCTTTCGGTGTCTTCACCGAGTGACGATCTGTTG3'
Rev a11 AGAA192/UCUU	5'CAACAGATCGTCACTCGGTGAAGACACCGAAAGCACGTAATCGC3'
Fwd a11 AUU175/UAA	5'CCAACATGAGTTACACCGGCGTAATCGTGCTTTCGGTGAGAACACC3'
Rev a11 AUU175/UAA	5'GGTGTTCTCACCGAAAGCACGATTACGCCG GTGTAACTCATGTTGG3'
Fwd a11 AUG84/CGC	5'CACTCGTGGACCAGGACCCGCTAGCTGAGCAAAGGAGAG3'
Rev a11 AUG84/CGC	5'CTCTCCTTTGCTCAGCTAGCGGGTCCTGGTCCACGAGTG3'
Fwd a11 A88/G	5'CTCGTGGACCAGGACCATGTGGCTGAGCAAAG3'
Rev a11 A88/G	5'CTTTGCTCAGCCACATGGTCCTGGTCCACGAG3'

Table 4: **Primers used for site-directed mutagenesis.** Different mutations (insertions, deletions, and substitutions) were introduced in TIE and IRES elements using Quick-change mutagenesis II kit (Agilent Technologies®) to study their effect on luciferase expression using *in vitro* translation assays.

1.2.4. Ultramers

Name of primer	Sequence
Rev Hox TIE a3 CDS	5'GTACTCAACGCCATCGGTGCCCAGAGCGGCGGACGGCGCGTATGGCTGCTG ACTGGCATTGTAAGCGAACCCATTGGCTGCTTGGTAGGGGTAGCCACCGTAGAT CGCTGAGCTGTCGTAGTAGGTCGCTTTTTGCATCGCGATGTTTCACGATCGTGAT CACACTCTCTGACAAGGGCTTGACACCCGTAAGGGAGCAC3'
Fwd globin-stem	5'ACATTTGCTTCTGACACAACTGTGGGAGAGCTGCCCGGGGGCGCACCCCAG

loop TIE a11	CCTTTCTGCGCGCGGGGAGGCCCCCCAGCCTTCACTAGCAACCTCAAACAGACA CCATGACTTCGAAAGTTTATGATCCAGAACAAAGGAAACGGATGATAACTGGT CCGCAGTGGTGGGCCAG3’
Fwd a3 U112/C	5’CAAAAGCTTCTGGGAGCCGCGGTCTGAAGGCTACGTGTGCTGCCTGGTCATTC AAAGTGTCAATTTTAGGTCCAGAAGTGTCCAAACCACAAGTTCTCAAAACTCTG AAAAACGGCTCCCTCCATGTTAAGAGTGCCTGGACACTGGAAAAAAAGATTCC CTGAGCACCTG3’
Fwd a3 U124/C	5’CAAAAGCTTCTGGGAGCCGCGGTCTGAAGGCTACGTGTGCTGCCTGGTCATTC AAAGTGTCAATTTTAGGTCCAGAAGTGTCCAAACCACAAGTTCTCAAAACTCTG AAAAATGGCTCCCTCCACGTTAAGAGTGCCTGGACACTGGAAAAAAAGATTCC CTGAGCACCTG 3’
Fwd a11 U160/C	5’CAAAAGCTTATCCCTCACCCACCGGGGAGTGGGGGGAGGCGTCAGGAGTTGC GCTCTCTTGCCGCCCTCAGAGCACTCGTGGACCAGGACCATGTAGCTGAGCAAA GGAGAGCTGCCCGGGGGCGCACCCCAGCCTTTCTGCGCGCGGGGAGGCCCCCC AGCCAACACGAGTTACACCGGCGATTACGTGCTTTCGGTGA 3’

Table 5: Ultramers used for amplification. Fragments of size up to 200 bp were added to TIE elements. The first ultramer was used to introduce the CDS of Hox a3 to show the translation of uORF through the main ORF. The second ultramer was used to introduce the GC-rich stem loop (SL) of TIE a11 in the 5’UTR of β-globin to study its inhibitory effect on luciferase expression. The last three ultramers were used for mutation of uAUG to ACGs in a3 and a11. Mutations of Us to Cs are highlighted in yellow.

2. Methods

2.1. Cell culture and cloning methods

2.1.1. Cell culture

HEK293FT cell line (ATCC®)

HEK293FT is a human embryonic kidney cell line competent for replication of vectors having the SV40 promoter site. This adherent cell line is suitable for efficient transfection; we used it for gene expression and protein production. This cell line is cultured in Dulbecco’s modified Eagle medium (DMEM) with 2 mM of L-Glutamine and 10% Fetal Bovine Serum (FBS) supplemented with 100 units/ml of Penicillin/Streptomycin. Subcultures were performed after treatment with Trypsin-EDTA for dissociation at subconfluent conditions (70%-80%) 1:4 to 1:10 seeding at 2-4.10^4 cells/cm^2 according to manufacturer’s instructions.

C3H10T1/2 cell line (Clone 8, ATCC® CCL-226)

C3H10T1/2 is a murine mesenchymal cell line isolated from C3H mouse embryo cells. We used it in transfection assays with different reporter vectors containing the luciferase gene. This adherent cell line was cultured according to manufacturer’s instructions in basal DMEM medium supplemental with 2 mM Glutamine, 1.5 g/L sodium bicarbonate and 10% FBS supplemented with 40µg/ml Gentamicine. Subcultures were performed after treatment with

Trypsin-EDTA for dissociation at subconfluent conditions (60%-70%). Seeding dilution performed at 2000 cells/cm^2 one time per week.

2.1.2. Cloning and transformation

Cloning

In pUC19: For *in vitro* studies, TIE a3 and TIE a11 and the corresponding full length 5' UTRs were cloned in HindIII site of pUC19 vector.

In pmirGLO: For *in vivo* studies, we first introduced an EcoRI site in the 5'UTR of *hRLuc-neo* fusion sequence in pmirGLO vector (7.35 KB) using Quick-change site-directed mutagenesis kit II XL (Thermo Fischer Scientific®). Then subsequent cloning of TIEs and the corresponding Hox 5'UTRs with their mutants was performed using NEBuilder® *HiFi* DNA Assembly kit with EcoRI digested pmirGLO.

Transformation of E. *coli*

Electroporation To introduce the novel plasmid constructs into *E. coli,* 1 pg of the ligation product was added to 40 µL of electrocompetent TB1 cells in cuvettes. After electroporation (2.5kV, 200 Ω, 25 µF), 90 µL of LB medium was added to each sample and plated on LB-agar supplemented with ampicillin and incubated overnight at 37°C.

Heat-shock Alternatively, we followed another protocol instructed by NEBuilder® *HiFi* DNA Assembly kit. We used 40 µL of NEB 5-alpha *E. coli* with 2 µL of assembled product. The mixture was placed on ice for 30 min followed by heat shock at 42°C for 30 sec then on ice for 2 min. After heat shock, 950 µL of SOC media supplemented provided in the kit was added and plated on LB-agar supplemented with ampicillin and incubated overnight at 37°C. Positive clones were checked by DNA sequencing.

2.2. Polymerase Chain Reaction

The amplification of DNA fragments by PCR was carried out using DNA template (10 ng/µl) with a mixture containing the four deoxyribonucleoside triphosphates (dATP, dCTP, dCTP, dGTP), 10 pmoles of forward and reverse primers, high fidelity Phusion DNA Polymerase, in GC buffer, 3% DMSO (**Table 6**).

Component	Volume/100µl reaction
10 ng/µl DNA Template	2µl
100 µM T7 forward primer	2µl
100 µM Reverse primer	2µl
10 mM dNTPs	1µl
5X GC Buffer	20µl
DMSO (100%)	3µl
Phusion DNA Polymerase (0.6 mg/ml)	1µl

Table 6: **PCR reaction components.**

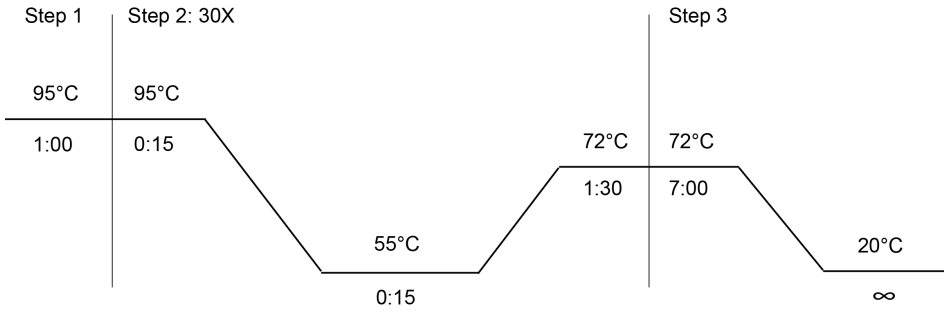


Figure 38: **Standard PCR thermo-cycling.**

Step1: initial Denaturation at 95°C for 1 min

Step 2: 30 cycles of a 15 s denaturing step at 95°C, then an annealing step, the annealing temperature used was generally 55°C for 15 s and an elongation step at 72°C for 1 min 30 s.

Step 3: final primer extension at 72°C-Hold at 20°C.

After PCR, electrophoresis was performed in 1% agarose gels and visualized by addition of ethidium bromide. The amplified DNA is finally precipitated by adding 1/16th volume of NaCl (4M) and 3 volumes of ethanol (100%) at -20°C.

2.3. *In vitro* Transcription

Transcripts were synthesized *in vitro* from DNA templates with recombinant T7 RNA polymerase. For that, the DNA template produced by PCR (20 µL) was added to TMSDT buffer (40 mM Tris/HCl (pH 8.1), 20 mM $MgCl_2$, 5 mM DTT (dithiothreitol), 0.01 % Triton-X100, 1 mM spermidine), 10 µL of each NTP (5 mM), 40 units of RNase inhibitor (RNAsin),

0.2 mg/ml of T7 RNA polymerase, and milliQ water. The reaction mixture was incubated at 37°C for 1 h. Then, 2 µL of pyrophosphatase (Roche®) was added and incubated further for 30 min. Then the DNA template is degraded by addition of 4 units of DNase I RNase-free (New England Biolabs®) and incubated for 1 h at 37°C. To check RNA integrity, an aliquot was mixed with Formamide Dye and loaded on 4% denaturing polyacrylamide gel. The RNA is visualized under UV light after ethidium bromide staining.

To eliminate unincorporated nucleotides, the remaining RNA sample was loaded on a gel filtration Sephadex G25-column (Pharmacia Fine Chemicals). Purified RNA samples were then phenol extracted and precipitated with 250 mM NaCl and ethanol 100%. After centrifugation, RNA pellets were dried and resuspended in autoclaved milli-Q water. The concentration of purified RNA sample was determined by absorbance measurement at 260 nm.

2.3.1. RNA 5' capping methods

Some of the conducted experiments required the incorporation of an m^7G cap (radioactive or non-radioactive) or a non-functional A-cap analog on the 5' end of mRNA transcripts. Capping was achieved by two approaches. The first one consisted of incorporating the cap in a co-transcriptional manner. For that, the transcription protocol was slightly modified. The transcription is initiated in the absence of GTP but in the presence of a cap analog m^7GpppG (0.2 mM) to force transcription start with the cap analog (since G is the first transcribed nucleotide). Then transcription elongation is enabled by gradually addition of GTP (5 mM) every 10 min during the 1h incubation time. The second protocol is the post-transcriptional capping reaction. The cap is added on the 5' end of a transcript by the Vaccinia virus Capping Enzyme (VCE) from the Scriptcap kit (Epicentre®). The VCE catalysis three consecutive reactions, first it hydrolysed the 5' triphosphate moiety thereby releasing PPi from the 5' end of the transcript. Then, GTP is added at the 5' end of mRNA in order to create a 5'-5' link. Finally the 5' end G is methylated at the position 7 by using S-Adenosyl Methionine (SAM) as a methyl-group donor. When radioactive $\alpha^{32}P$-GTP is used, this method allows the synthesis of RNA transcripts containing a radioactive m^7G cap. The capping reactions were performed with pure *in vitro* synthesized RNA transcripts according to the manufacturer's instructions. The radiolabelled RNA transcripts were then purified on a preparative denaturing 4% polyacrylamide gel. Bands corresponding to radioactive RNA were cut and passively eluted by soaking gel slices in cracking buffer (0.3 M NaCl, 0.5 µM EDTA, 10 mM Tris-HCl pH 7.5). Radioactive RNAs were then precipitated, washed and dried as previously described.

2.3.2. RNA purification by electro-elution

This method was used to purify RNA transcripts after separation by electrophoresis on a preparative 4% polyacrylamide gel. RNA bands were visualized by UV shadowing and marked for cutting. Then, by using a Biotrap apparatus (Schleicher&Schull®) filled with autoclaved RNAse-free TBE buffer, RNA samples were electro-eluted from gel slices by electrophoresis. The principle of the method relies on the migration of RNA out of the gel slices and trapping the eluted RNA in a chamber composed of a porous BT2 membrane that allows the passage of RNA and a second rigid BT1 membrane that retains RNA. We repeat twice the elution for each RNA sample for optimal RNA recovery. The pure RNA transcripts are then precipitated as previously described.

2.4. *In vitro* translation assays

In vitro translation was carried out using increasing concentrations of mRNA transcripts with untreated Rabbit Reticulocyte Lysate or other cell-free translation extracts (Wheat Germ Extracts, HeLa or drosophila S2 cell extracts), amino acid mixture containing all the amino-acids except methionine (1 mM of each), RNasin (Promega®), 75 mM KCl, 0.5 mM $MgCl_2$, and ^{35}S-methionine, and autoclaved milli-Q water. Reaction mixture was incubated at 25°C or 30°C depending on the extracts for 1 h. Aliquots of translation mixture were analysed by SDS PAGE (10%) (Laemmli, 1970) and translation products were visualized by phosphor-imaging. The reaction products were also used for luciferase assays.

2.5. Chemical probing

We performed chemical probing to obtain the secondary structural models of TIE and IRES elements. For that, we used two chemical reagents: Dimethyul Sulfate (DMS) and 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) which are both base-specific modifying reagents.

Modification by DMS

DMS is a chemical reagent which modifies N1 of adenosine and N3 of cytosine nucleotides in RNA. Sites of modifications are specific for highly flexible nucleotides that are located in single stranded regions. Modification by DMS was performed on 2 pmoles of each RNA. For that, the RNA is first incubated for 15 min in DMS buffer (50 mM Na Cacodylate, pH 7.5, 5 mM $MgCl_2$ and 100 mM KCl) and 1µg of yeast total tRNA and then modified with 1.25%

DMS reagent (diluted with ethanol 100%) for 10 min at 20°C. Then, the DMS reaction is stopped on ice. Modified RNAs are precipitated with ethanol 100%, 0.250 mM NaCl and 0.1 mg/mL glycogen. Pellets were dried and resuspended in 7 µL autoclaved milli-Q water. Modified nucleotides were detected by primer extension that induces reverse transcription arrests that were quantified and interpreted as reactivity for each nucleotide.

Modifications by CMCT

Similarily, modifications by CMCT were performed on 2 pmoles of each RNA. CMCT is a chemical reagent that modifies N3 of uridines and N1 of guanosines. Each RNA is incubated for 20 min in CMCT buffer (50 mM Na borate 50 mM; 5 mM $MgCl_2$; 100 mM KCl) and 1 µg of yeast total tRNA. Then modifications were performed with 10.5 g/L CMCT reagent for 20 min at 20°C and stopped on ice. Modified RNAs are then precipitated with ethanol 100%, 0.25 mM NaCl and 0.1 mg/mL glycogen. Pellets were dried and resuspended in autoclaved milli-Q water. Sites of modification were detected by primer extension arrests that were quantified as previously described for DMS.

Modified nucleotide detection by primer extension

Reverse transcription was carried out in 20 µl reaction volume with 2 pmoles of RNA and 0.9 pmoles of 5' fluorescently labelled primers. We used Vic and Ned primers (of same sequence for all reverse transcription reactions which are complementary of the 5'UTR β-globin from nts 6 to 37: 5' GGTTGCTAGTGAACACAGTTGTGTCAGAAGC 3'. For probing of IRES elements, we used a different fluorescently labelled primer, which are complementary of the Renilla luciferase CDS from nts 8 to 39: 5'CGTTTCCTTTGTTCTGGATCATAAACTTTCG 3'. First, the modified RNAs are denaturated at 95°C for 2 min. Then, fluoresencent primers are annealed for 2 min at 65°C followed by incubation on ice for 2 min. Primer extension is performed in a buffer containing 83 mM KCl, 56 mM Tris–HCl (pH 8.3), 0.56 mM each of the four deoxynucleotides (dNTP), 5.6 mM DTT and 3 mM $MgCl_2$. Reverse transcription were performed with 1 unit of Avian Myoblastosis virus (AMV) reverse transcriptase (Promega®) at 42°C for 2 min, then 50°C for 30 min and finally 65°C for 5 min. In parallel, sequencing reactions were performed in similar conditions, but containing 0.5 mM dideoxythymidine or dideoxycitidine triphosphate (ddTTP or ddCTP) (protocol adapted from Gross et al, 2017). Then, the synthesized cDNA were phenol–chloroform extracted, precipitated, after centrifugation the pellets were washed, dried and resuspended in 10 µL deionized Hi-Di formamide (highly deionized formamide). Samples were loaded on a 96-well

plate for sequencing on an Applied Biosystems 3130xl genetic analyzer. The resulting electropherograms were analyzed using QuSHAPE software (Karabiber et al., 2013), which aligns signal within and across capillaries, as well as to the dideoxy references of nucleotide at specific position and corrects for signal decay. Normalized reactivities for each nucleotide range from 0 to 2, with 1 to 2 being the range of highly reactive nucleotides. The secondary structure model prediction was initiated using mfold (Mathews et al., 2016) and then edited and refined according to our reactivity values.

2.6. Sucrose gradient analysis

To analyze the pre-initiation complex assembly on RNA of interest, a fractionation by ultracentriguation on 7-47% sucrose gradient was used. Sucrose gradients were prepared in a buffer containing 5 mM $MgCl_2$, 25 mM Tris-HCl (pH 7.5), 1 mM DTT, 50 mM KCl. To address the sequential ordered assembly of pre-initiation complex formation, translation inhibitors specific for distinct steps were used: 4 mM GMP-PNP, 1 mg/mL cycloheximide, 0.7 mM geneticin, 0.5 mg/ml hygromycin, 10 mM edeine. Inhibitors were added to the RRL together with an amino acid mix containing the 20 amino acids at 1.5 mM each, RNasin (Promega®), 35 mM KCl and 0.24-2.4 mM $MgCl_2$, prior to incubation with the previously prepared 5'-capped radioactively labelled mRNA transcript. The RRL were first activated by pre-incubation at 30°C for 5 min and transferred to ice for 20 min. The 5'-radiolabelled RNA was then added to the mixture and the pre-initiation complexes were assembled on the RNA of interest by incubating in RRL at 30°C for 5 min. Then, the complexes were stabilized by addition of 8 mM of $Mg(Ac)_2$ and one volume of 7% sucrose. Samples were layered on the top of a previously prepared 11-mL 7-47% gradient and centrifuged in a SW41 rotor for 2h30 at 37,000 rpm at 4°C. After centrifugation, the whole gradient is collected and fractionated using a peristaltic pump. The radioactive RNA is monitored by Cerenkov counting in a scintillation counter of each fraction of the gradient.

2.7. Toe printing

This assay was used to detect the position of the ribosome during intiation. It is based on the fact that a blocked initiation ribosome will stop reverse transcriptase and lead to the synthesis of a cDNA. The size of the synthesized cDNA allows the precise localisation of the ribosome 3' on the mRNA of interest. A canonical translation initiating ribosome results typically in a reverse transcriptase arrest named a 'toe print' at 16 to18 nts downstream of the AUG initiating codon (**Fig. 39**).

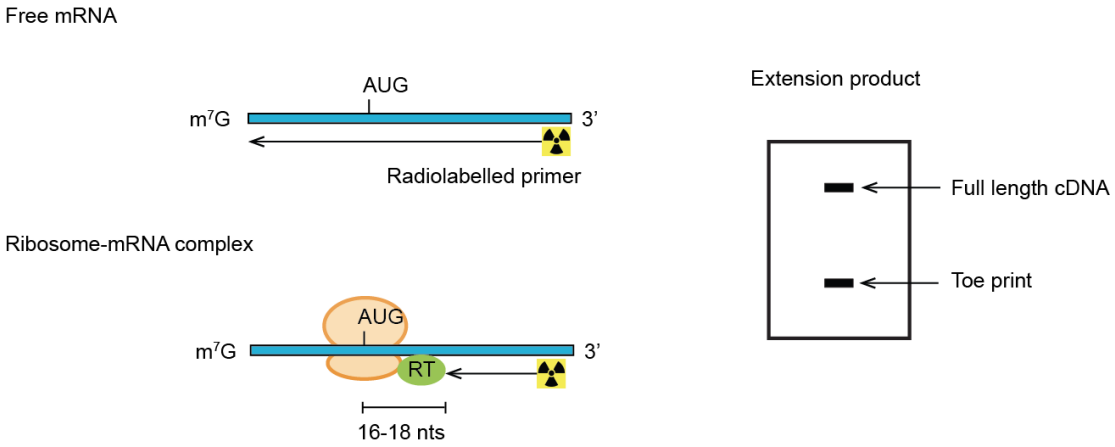


Figure 39: **Toe printing assay.**

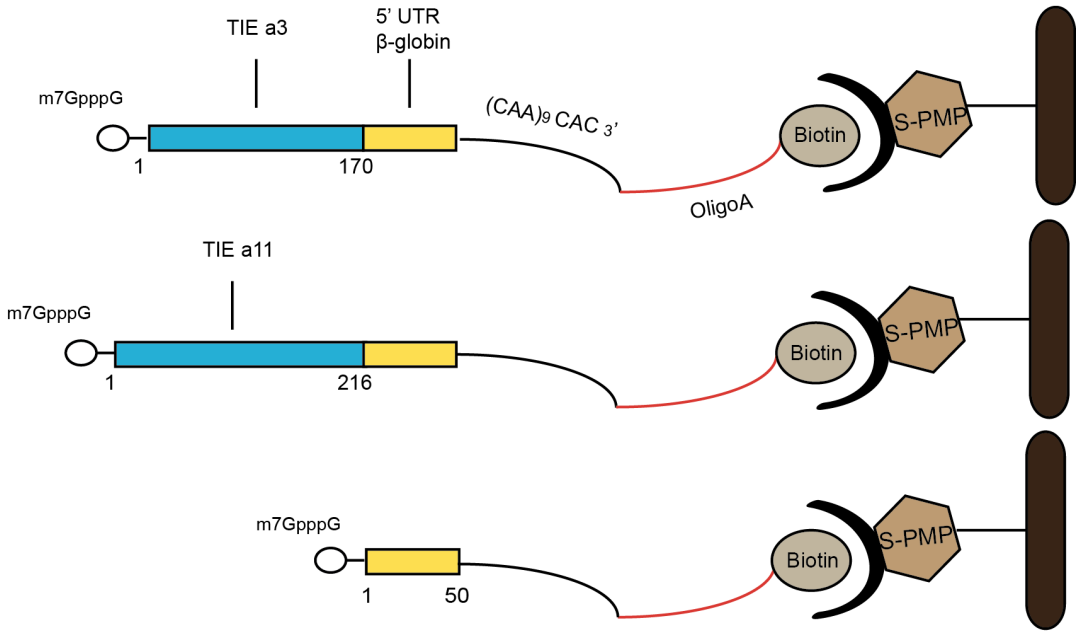
To perform toe printing, RRL were pre-incubated for 5 min at 30°C then 10 min on ice with buffer containing 1 U/µL of RNasin (Promega®), 75 mM KCl , 0.5 mM $MgCl_2$, and 1.3 mM puromycin prior to pre-initiation complex assembly. Then, ribosomal complexes are formed by incubation with 500 nM of the RNA of interest in the presence of specific inhibitors such as 1 mg/mL cycloheximide or 4 mM GMP-PNP for 5 min at 30°C and then 20 min on ice. Then RNA-ribosome complexes were supplemented with one volume of ice-cold buffer A (20 mM Tris-HCl (pH 7.5), 100 mM KAc, 2,5 mM Mg $(Ac)_2$, 2 mM DTT, 1 mM ATP and 0,25 mM spermidine) and placed on ice. In order to separate ribosomal complexes from the non-ribosomal fraction, samples were ultracentrifuged at 88,000 rpm in S100AT3 rotor (Sorvall-Hitachi) at 4°C for 1 hr. Toe printing assay was adapted from previously established protocols in the lab (Martin et al, 2011, Martin et al., 2016). After centrifugation, the supernatant was discarded and the complex in the pellet was resuspended in 30 µL of ice-cold buffer A. Then this complex was incubated with 5' radioactively labelled DNA oligonucleotide complementary to the RNA of interest for 3 min at 30°C. Then, 1 µL of a 320 mM $Mg(Ac)_2$, 4 µL of a dNTP mixture (containing 5 mM of dATP, dGTP, dTTP and dCTP), 10 units of RNAsin (Promega®) and 1 unit of Avian Myoblastosis virus (AMV) reverse transcriptase (Promega®) were added and incubated for 1h at 30°C. The synthesized cDNAs were analysed on 8% PAGE next to sequencing ladders in order to map the reverse transcriptase arrests.

2.8. Preinitiation complex purification for mass spectrometry analysis

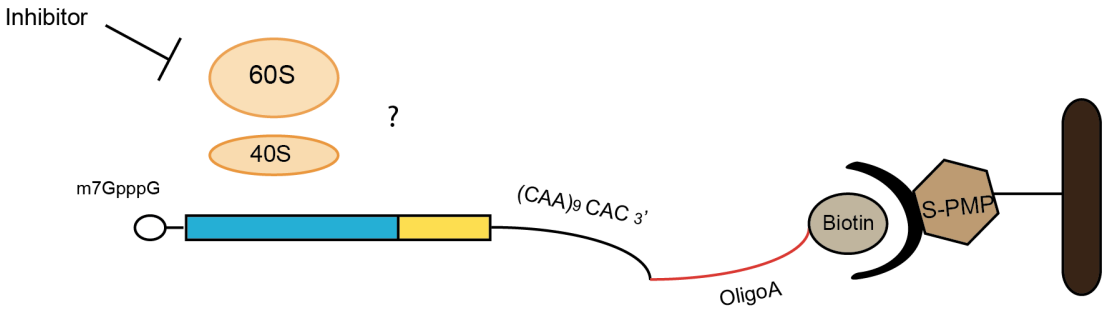
The purification method was adapted from protocols that have been previously established in the lab (Prongidi-Fix et al., 2013; Chicher et al., 2015). First, mRNA purification was

performed as previously described in (2.3.2). Then, a biotinylated DNA oligonucleotide was ligated by splint ligation to the 3' end of each mRNA by T4 DNA ligase. The resulting chimeric mRNAs were immobilized on magnetic streptavidin beads MagSi-DNA 600 (MagnaMedics®) (**Fig. 40**) and used to assemble pre-initiation complexes from ribosomes in RRL. Importantly, the splint DNA oligonucleotide was removed prior incubation in RRL to prevent cleavage by endogenous RNAse H (which cleaves RNA/DNA duplexes). To do so, the immobilized chimeric mRNAs were incubated 2 min at 95°C to unwind mRNA-splint duplexes and then incubated for 10 min at room temperature in the presence of a 10-fold excess of an antisense splint DNA oligonucleotide containing the complementary sequence to trap the splint and avoid re-annealing on the mRNA. The 80S pre-initiation complexes were obtained by adding 1 mg/mL cycloheximide to the RRL prior incubation with mRNA. The complexes were assembled by incubation in the RRL at 30°C for 5 min. Washing and elution of the complexes were performed as previously described (Prongidi-Fix et al., 2013, Chicher et al., 2015). For elution of the RNP complexes, the biotinylated DNA linker moiety of the chimeric mRNA was digested with 10 units of RNase-free RQ1 DNase (Promega®) for 30 min at room temperature. The eluted ribosomal complexes were then anlysed by nano LC-MS/MS as previously described (Chicher et al., 2015).

1. Generating an RNA-DNA biotnylated chimera



2. Assembly of initiation complexes in RRL



3. Elution of complexes by DNase treatment

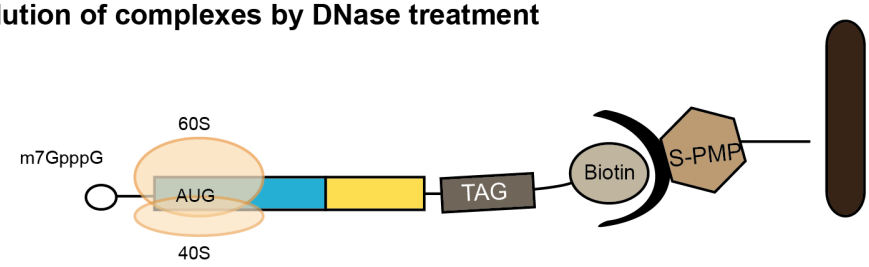


Figure 40: **Experimental strategy for purification of mRNA-translation initiation complexes suitable for Mass Spectrometry analysis.**

Pre-initiation complexes Mass Spectrometry analysis and data processing

Protein extracts were digested with sequencing-grade trypsin (Promega®). Peptide digests were analysed by nano LC-MS/MS. The resulting peptides were identified with the Mascot algorithm by searching the UniProtKB database. Identifications were validated with a protein False Discovery Rate (FDR) of less than 1% using a decoy database strategy. The total number of MS/MS fragmentation spectra was used to quantify each protein from three independent biological replicates. This spectral count was submitted to a negative-binomial test using an edge R GLM regression through the R-package. For each identified protein, an adjusted P-value corrected by Benjamini–Hochberg was calculated, as well as a protein fold-change (FC is the ratio of the average of spectral counts from a specific complex divided by the average of spectral counts from a reference protein complex). The results are presented as volcano plot using protein log2 FC and their corresponding adjusted $\log_{10}$ P-values highlighted proteins up-regulated compared to a reference experiment (TIE a3 versus β-globin mRNA, TIE a11 versus β-globin mRNA and TIE a3 versus TIE a11).

2.9. *In vivo* luciferase assay

HEK293FT cells and C3H10T1/2 cells were transfected in 6-well plate with various constructs of pmirGlo plasmid in triplicates using Turbofect transfection reagent (Invitrogen®). Transfection was performed according to manufacturer's instructions using 4 µg of plasmid with 390 µL of reduced serum medium (Opti-MEM) and 6 µL of Turbofect. Cells were collected 24 hrs post-transfection and lysed using passive lysis buffer (PLB) supplied by Dual-Glo luciferase kit (Promega®). Firefly Luciferase and Renilla Luciferase activity are both measured at the same time. Firefly Luciferase is used to monitor transfection efficiency for normalization and Renilla Luciferase activity is used to measure TIE-mediated translation inhibition.

RESULTS

Chapter 1: Study of Hox TIE elements

Deciphering the mechanism of inhibition of cap-dependent translation by the Translation Inhibitory Elements TIE a3 and a11 in Hox mRNAs

Resulted in

Article 1

-In preparation-

"Translational Inhibitory Elements from Hox a3 and a11 mRNAs use upstream ORFs to repress cap-dependent translation"

Fatima ALGHOUL, Laure SCHAEFFER, Gilbert ERIANI, Franck MARTIN*

1. Summary of article 1

The main objective of my PhD project is to characterize TIE-mediated inhibitory mechanisms in Hox mRNAs. Hox genes, as previously explained (Introduction, section 8), are master regulators of embryonic development. These genes are highly conserved among species and their expression has to be tightly controlled during embryogenesis (Deschamps and van Nes, 2005). Until recently, basic knowledge on the molecular mechanism of translational regulation on Hox genes expression remained elusive. The laboratory of Maria Barna has characterized two RNA regulons in HoxA cluster mRNAs, a TIE element which inhibits cap-dependent translation thereby promoting translation driven by a second regulon, an IRES that is the sole mediator of translation (Xue et al., 2015). However, the mechanism of TIE elements is not yet understood. In our article, we focus on structural and functional characterization of two TIE elements from Hox a3 and a11 mRNAs respectively.

1.1. Structural characterization of TIE a3 and TIE a11 by chemical probing

The structural characterization of TIE elements required the use of chemical probing methods. For that, we used DMS and CMCT, two chemical reagents that modify nucleotides in base-specific manner. DMS modifies A and C residues while CMCT modifies G and U residues that are located in single-stranded regions. Accordingly, we optimized a protocol for probing of TIE elements. Probing results are calculated as reactivity values to the reagent of each nucleotide at a specific position normalized with an untreated sample as a reference. A range of values was determined to assess the level of reactivity of each nucleotide, and thereby, a high reactivity of a nucleotide indicates a high probability of this nucleotide to be located in a single-stranded area of the RNA molecule. Low or lack of reactivity indicates that the nucleotide is most probably involved in a base-pair. The average of reactivities was obtained from three independent experiments for each TIE. Based on this data, structural models were designed with the help of mfold to find putative base pairs (Mathews et al., 2016). Briefly, the two structures are distinct since the two TIEs do not share any sequence similarities or common structural motifs. TIE a3 encompasses a long 5' proximal stem-loop structure and another bigger structure comprising a two-way junction and two additional stem-loops. TIE a11 encompasses four stem-loops and a domain containing a three-way junction. Overall, our models suggest that TIE a11 is more structured than TIE a3 possibly due to a higher GC content.

1.2. Functional characterization by *in vitro* translation assays

1.2.1. Mapping minimal TIE elements

To gain further insight into the inhibitory mechanisms of TIEs, we set an *in vitro* cell-free translation assay using Rabbit Reticulocyte Lysate (RRL). This system allows tracking of translation initiation through the expression of luciferase reporter mRNA. The first step was to design appropriate constructs for our *in vitro* tests. To do so, we inserted each TIE upstream of the 5'UTR of β-globin, a 5'UTR that is known to be short and not structured, followed by the coding sequence of *Renilla reniformis* luciferase. We mapped the minimal TIEs by 5' sequential deletions. For each construct, the impact of the presence of TIE on cap-dependent luciferase activity was measured after *in vitro* translation in RRL. We started with large sequential deletions then with more precise deletions. By using this strategy, we were able to determine minimal domains for a3 between nucleotides 68 and 170 and for a11 between nucleotides 139 and 216.

1.2.2. Study of the upstream AUGs in TIE elements

Minimal TIEs harbour upstream AUGs (uAUGs). Minimal TIE a3 contains two uAUGs: $AUG_{111\text{-}113}$ and $AUG_{123\text{-}125}$ and minimal TIE a11 contains a single $AUG_{159\text{-}161}$. We were interested in determining the functional significance of these uAUGs. For that, we mutated these uAUGs into UACs, to avoid any possibility of AUG-like recognition. We performed translation assay in RRL with these mutants. For a3, $AUG_{111\text{-}113}$ is highly implicated in inhibition, which is not the case for $AUG_{123\text{-}125}$. Interestingly, mutation of $AUG_{111\text{-}113}$ into AUG-like codons CUG and GUG does not recapitulate TIE-mediated inhibition indicating that full inhibition requires an AUG codon. These results were confirmed *in vivo* using embryonic cell lines that efficiently express Hox proteins. Furthermore, we were interested in determining whether a peptide is indeed produced from this $uAUG_{111\text{-}113}$. If translated, the peptide would be difficult to detect due to its small size. Since the uORF and the main ORF are not in-frame, we inserted a single nucleotide insertion upstream the main ORF in order to induce a frame shift that creates a longer uORF that can now be visualized. Indeed, with this single insertion, we could detect the predicted fusion protein between the uORF and the renilla ORF meaning the uORF is efficiently translated from the $uAUG_{111\text{-}113}$. Furthermore, we could detect a strong +16 toe print on $uAUG_{111\text{-}113}$ in the presence of GMP-PNP indicating the pre-initiation complex formation on this AUG codon. For TIE a11, it was not

the case. Mutation of minimal $uAUG_{159\text{-}161}$ had no effect on inhibition. This meant that a different mechanism might be most probably employed by TIE a11.

1.2.3. Study of two regulatory elements in TIE a11

Since $uAUG_{159\text{-}161}$ mutation had no effect on TIE a11 function, we looked into other regulatory elements. TIE a11 harbours a predicted GC-rich stable stem loop (SL) structure (ΔG = -25.00 kcal/mol) at position 104-154. Our chemical probing experiments confirmed that this predicted structure is indeed present. We then transplanted this SL in the 5'UTR of β-globin to assess its inhibitory activity. Interestingly, this SL on its own is sufficient to efficiently inhibit the translation of Renilla luciferase. Nevertheless, another element upstream of the SL is also required for TIE a11 function. An $uAUG_{84\text{-}86}$ is located at a distance of 19 nucleotide upstream of the SL, a distance that is compatible with the assembly of a pre-initiation complex on $uAUG_{84\text{-}86}$ without a steric clash with the SL. Moreover, the $uAUG_{84\text{-}86}$ is immediately followed by a stop codon $UAG_{87\text{-}89}$. We hypothesized that this start-stop blocks the ribosome upstream of the SL. The presence of the highly SL downstream of the uAUG made the use of the classical toe-print technique impossible because the SL strongly provokes premature RT arrests before reaching the toe-printing position. Therefore, to confirm the pre-initiation complex assembly on $uAUG_{84\text{-}86}$, we mutated the stop codon $UAG_{87\text{-}89}$ to UGG, a mutation that creates a longer uORF. Indeed, with this mutation, a small peptide is produced and detected indicating that the ribosome is efficiently assembled on this AUG codon and then stalled at the stop codon. We also confirmed this by sucrose gradient analysis, showing that in the absence of inhibitors, TIE a11 stalls an 80S ribosome, an 80S complex that is greatly reduced upon mutation of upstream $(AUG/UAC)_{84\text{-}86}$. Surprisingly, the minimal TIE a11 determined by 5' deletions does not comprise these elements. We explain this result by the fact when the 5' deletions remove the start-stop and the SL, another $uAUG_{159\text{-}161}$ functions as the next inhibitory element in the truncated mRNA, although less efficiently. However, in full-length context, TIE a11 requires these two elements: a start-stop and the GC SL to stall efficiently an 80S pre-initiation complex.

1.2.4. Mass spectrometry analysis shows different protein requirements for TIE elements

To further characterize the different mechanisms, we were interested in determining the *trans*-acting factors of TIE elements. For that purpose, we performed mass spectrometry analysis. We purified pre-initiation complexes with mRNA transcripts containing each of the TIE after

incubation with RRL in the presence of cycloheximide, using an approach developed in our laboratory (Chicher et al., 2015; Prongidi-Fix et al., 2013). As a negative control, we used the 5'UTR of β-globin mRNA. The results are shown as volcano plots with the significant protein hits for each mRNA. Briefly, we identified several translation-related proteins for TIE a3. Among the strongest hits, is eIF2D, a non-canonical initiation factor which has been shown to be involved in the initiation of uORFs and other specific mRNAs (Akulich et al., 2016; Dmitriev et al., 2010). For TIE a11, we identified different proteins like ASAP1, a GTPase activator protein, Methionine Aminopeptidase MetAP1, and interestingly, eIF3j. By comparing a3 to a11, we can conclude that a3 has more requirement than a11 for scanning factors, like eIF1 and eIF1A-X.

1.2.5. eIF2D, a candidate *trans*-acting factor in TIE a3 mediated inhibition

We identified eIF2D by mass spectrometry analysis of preinitiation complexes programmed with TIE a3. eIF2D was previously described by recent studies (see Introduction section 7). In fact, the functions of eIF2D are still not fully understood however, this non-canonical factor forms preinitiation complexes on A-rich 5'UTRs and has been suggested to be involved in the translation of uORFs and leaderless mRNAs (Akulich et al., 2016; Dmitriev et al., 2010). Interestingly, TIE a3 has at least two A-rich motifs so we wanted to check whether these motifs are involved in the inhibitory mechanism. Therefore, we mutated both motifs upstream and downstream the $uAUG_{111\text{-}113}$. Mutating the motif located upstream to the $uAUG_{111\text{-}113}$ significantly reduced the inhibitory effect of TIE a3 whereas the mutation of the downstream motif had no effect. We concluded that such effect might results from less efficient scanning to the $uAUG_{111\text{-}113}$ and that eIF2D might be involved in such process. However, it remains preliminary and needs to be confirmed by further experiments.

1.2.6. Hox TIE elements inhibit translation through distinct molecular mechanisms

According to our results, we proposed two models for TIE-mediated inhibition by a3 and a11. We showed that TIE a3 inhibits translation due to the presence of a uORF, which assembles a ribosome that further translates through the whole 5'UTR thereby producing a peptide that was efficiently detected. This is sufficient to block cap-dependent translation. Moreover our data suggests a possible involvement of the non-canonical initiation factor eIF2D, which is not fully understand and has yet to be further confirmed. For TIE a11, a start-stop combination blocks an 80S ribosome upstream of a highly stable structure composed of sixteen GC base pairs. The SL is sufficient to block a scanning ribosome however, the

presence of a start codon 19 nucleotides upstream of this structure forces the ribosome to initiate and is blocked at the stop codon leading to the accumulation of one 80S ribosome on TIE a11.

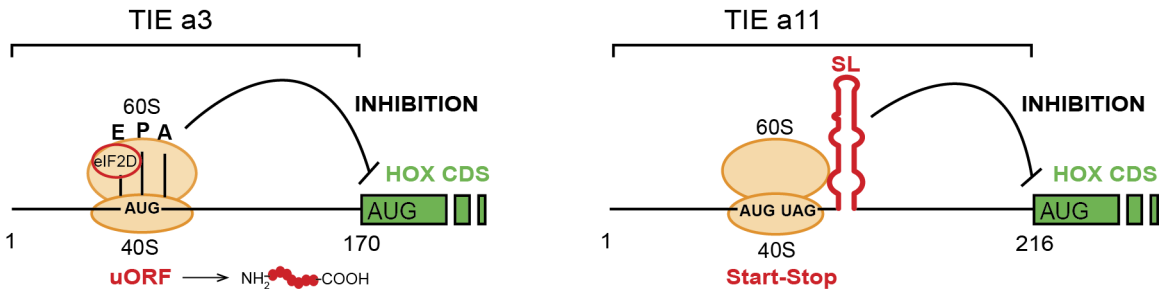


Figure 41: **Two distinct models for translational inhibition by TIE a3 and TIE a11.**
Our model for TIE a3 suggests a ribosomal assembly on the $uAUG_{111\text{-}123}$ with a requirement of eIF2D initiation factor. The model for TIE a11 suggests a stalled 80S ribosome on AUG-stop codons combination upstream of a highly stable structure (SL).

3. Manuscript of article 1

Translation Inhibitory Elements from Hox a3 and a11 mRNAs use upstream ORFs to repress cap-dependent translation

Fatima ALGHOUL, Laure SCHAEFFER, Gilbert ERIANI, Franck MARTIN*

Institut de Biologie Moléculaire et Cellulaire, "Architecture et Réactivité de l'ARN" CNRS UPR9002, Université de Strasbourg, 15, rue René Descartes, F-67084 Strasbourg (France)

*Corresponding authors
Email for correspondence: f.martin@ibmc-cnrs.unistra.fr

Keywords: Hox mRNA, Translation Inhibitory Elements, IRES, uORF

Abstract

During embryogenesis, Hox mRNA translation is tightly regulated by a sophisticated molecular mechanism that combines two RNA regulons located in their 5'UTR. First, an Internal Ribosome Entry Site (IRES) enables cap-independent translation. The second regulon is a so-called Translation Inhibitory Element or TIE, which ensures concomitant cap-dependent translation inhibition. Here, we have dissected the mode of action of mouse Hox a3 and a11 TIE. Both TIEs contain an upstream Open Reading Frame (uORF) that is critical to inhibit cap-dependent translation initiation. However, the molecular mechanisms used by both TIEs are different. In TIE a3, the non-canonical translation initiation factor eIF2D is recruited to assemble a pre-initiation complex on the start of the uORF. Then the uORF is translated into a peptide, which is sufficient to inhibit cap-dependent translation of the Hox main ORF. On the contrary, the TIE a11 contains a minimal uORF formed by AUG codon immediately followed by a stop codon. The a11 'start-stop' sequence is located upstream of a highly stable structure which allows the formation of a stalled 80S ribosome that prevent cap-dependent translation of the Hox main ORF.

Introduction

Gene expression constitutes an indispensible cellular process for which the genetic information is converted into proteins. This process named translation initiates by a cap-dependent mechanism of the majority of cellular mRNAs. It involves a large number of auxiliary proteins termed eukaryotic Initiation Factors (eIFs), they are required for the recruitment of the ribosomes on mRNA (1). To ensure fine-tuning of translation, this step is highly regulated. However, several mRNA subclasses are translated by non-canonical mechanisms. For instance, this is the case for homeobox (Hox) mRNAs. Hox genes encode a family of proteins that are transcription factors. Their main function is to orchestrate specific sequential transcription during embryonic development. A wealth of experimental data over the last three decades has led to the identification of many *cis*-regulatory elements that control Hox gene transcriptional patterns, thus giving deeper insights into the expression of Hox mRNAs (2). However, the translation mechanism leading to the production of Hox proteins has remained unexplored. In fact, the Drosophila genes Antp and Ubx have been suggested to be under translational control during embryonic development (3). More precisely, a subgroup of mRNAs produced from the Antp and Ubx loci contain functional Internal Ribosome Entry Sites (IRES) that allow their translation using a cap-independent mechanism. The IRES activity is modulated during development (4). In addition, the presence of IRES elements in the 5'UTR of subsets of mice Hox mRNAs a3, a4, a5, a9 and a11 have been demonstrated recently (5). Some of these IRESes proved to require RpL38 in the ribosome to initiate translation, thereby explaining the tissue patterning defective phenotype observed with RpL38 knockout mouse (6). The activity of IRES in these hox mRNAs was found to be critical for their appropriate expression. Following the discovery of the IRES elements in Hox mRNAs, other RNA regulons termed Translational Inhibitory Elements (TIE) were also found in the 5'UTR of Hox mRNAs. TIEs are located upstream of the previously described IRES. These elements efficiently inhibit canonical cap-dependent translation in Hox a3, a4, a9 and a11 mRNAs by an unknown mechanism. The action of TIE that ensures efficient blockage of cap-dependent translation promotes IRES-mediated cap-independent translation. TIE and IRES act in synergy to ensure tightly regulated translation during development. Indeed, Hox TIE elements ensure that Hox mRNAs are translated solely by IRES element. TIE elements represent the first example of specific elements dedicated to inhibit cap-dependent translation in such mRNAs. However, the mechanism of action of these elements as well as their structural characterization is still unknown. Here we undergo the functional characterization

of TIE from a3 and a11 Hox mRNAs. We determined their secondary structure in the frame of their complete 5'UTR. Using cell-free translation extracts, we deciphered their mode of action. Interestingly, the translation inhibition mechanism that is mediated by a3 TIE element is radically distinct from the one used by a11. TIE a3 contains a uORF that is translated into a peptide through the IRES and that requires the presence of eIF2D. On the contrary, TIE a11 contains an uAUG followed by a stop codon and a long stable hairpin. These three elements enable a highly efficient inhibition of cap-dependent translation that is achieved through a start-stop stalling mechanism of an 80S ribosome.

Materials and Methods

Plasmids

For *in vitro* studies, TIE a3 (170 nts) and TIE a11 (216 nts) were placed upstream of 5'UTR of human β-globin (Accession number: KU350152) (50 nts) and *Renilla reniformis* Luciferase coding sequence (Accession number: M63501) (936 nts). These constructs were cloned in pUC19 vector in the HindIII site then used as a template for further PCR amplifications and site-directed mutagenesis.

For *in vivo* studies, we introduced an EcoRI site upstream of hRLuc-neo fusion sequence in pmirGLO vector (Promega®) using Quick Change site-directed mutagenesis kit II XL (Thermo Fischer Scientific®). Subsequent cloning of TIE a3, TIE a11 and their mutants were constructed in pmirGLO vector were performed with NEBuilder® *HiFi* DNA Assembly kit. All clones were checked by sequencing.

RNA transcription

We synthesized mRNAs from DNA templates using an *in vitro* transcription reaction with recombinant T7 RNA polymerase. After transcription, we performed G25 column gel-filtration to remove unincorporated nucleotides followed by phenol-chloroform extraction and precipitation with NaCl and 100% ethanol. After precipitation and centrifugation, RNA pellets were dried and resuspended in autoclaved milli-Q water. The concentration of purified RNA sample was determined by absorbance measurement at 260 nm on a nanodrop.

***In vitro* translation assays in cell-free translation extracts**

In vitro translation was carried out using increasing concentrations of mRNA transcripts with untreated Rabbit Reticulocyte Lysate or other cell-free translation extracts (Wheat Germ Extract, HeLa or drosophila S2 cell extract), amino acid mixture containing all the amino-acids except methionine (1 mM of each), RNasin, KCl (75 mM), $MgCl_2$ (0.5 mM), [^{35}S] methionine (Specific Activity= 74 MBq) and autoclaved milli-Q water. Reaction mixture was incubated at 25°C or 30°C depending on the extracts for 1 hr. Aliquots of translation mixture were analyzed by SDS PAGE (10%) (7) and translation products were visualized by phosphore imaging.

Chemical probing

Modification by DMS

Modification by DMS was performed on 2 pmoles of each RNA (TIE a3 and TIE a11). The RNA is first incubated for 15 min in dimethylsufate (DMS) buffer (50 mM Na Cacodylate, pH7.5, 5 mM $MgCl_2$ and 100 mM KCl) and 1 µg of yeast total tRNA and then modified with 1.25% DMS reagent (diluted with ethanol 100%) with 10 min incubation at 20°C and stopped

on ice. Modified RNA is precipitated with ethanol, NaCl and glycogen. Pellets were dried and resuspended in autoclaved milli-Q water. Modified nucleotides were detected by primer extension arrests that were quantified.

Modifications by CMCT

Similarly, modification by 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) was performed on 2 pmoles of each RNA (TIE a3 and TIE a11). Each RNA is incubated for 20 min in CMCT buffer (Na borate 50 mM pH 8,5; $MgCl_2$ 5 mM; KCl 100 mM) and 1 µg of yeast total tRNA. Then modifications were performed with 10,5 g/L CMCT reagent with 20 min incubation at 20°C and stopped on ice. Modified RNA was precipitated with ethanol 100%, NaCl and glycogen. Pellets were dried and resuspended in autoclaved milli-Q water. Sites of modification were detected by primer extension arrests that were quantified.

Primer extension

Reverse transcription was carried out in 20 µl reaction volume with 2 pmoles of RNA and 0.9 pmoles of 5' fluorescently labelled primers. We used Vic and Ned primers of same sequence for all reverse transcription reactions which hybridize in the β-globin 5'UTR: 5'-GGTTGCTAGTGAACACAGTTGTGTCAGAAGC-3'. First, the denaturation takes place at 95°C for 2 min then annealing for 2 min at 65°C then incubation on ice for 2 min. Samples are incubated with 83 mM KCl, 56 mM Tris-HCl (pH 8.3), 0.56 mM each of the four deoxynucleotides (dNTP), 5.6 mM DTT and 3 mM $MgCl_2$. Reverse transcription were performed at 42°C for 2 min, 50°C for 30 min and 65°C for 5 min. In parallel, sequencing reactions were performed in similar conditions, but containing 0.5 mM dideoxythymidine or or dideoxycitidine triphosphate (ddTTP or ddCTP) (protocol adapted from Gross et al, 2017). Samples were phenol–chloroform extracted, precipitated, after centrifugation the pellets were washed, dried and resuspended in 10 µl deionized Hi-Di formamide. Samples were loaded on a 96-well plate for sequencing on an Applied Biosystems 3130xl genetic analyzer. The resulting electropherograms were analyzed using QuSHAPE software, which aligns signal within and across capillaries, as well as to the dideoxy references of nucleotide at specific position and corrects for signal decay. Normalized reactivities range from 0 to 2, with 1.0–2.0 being the range of highly reactive positions. The secondary structure model prediction was initiated using mfold (8) and then edited according to reactivity values.

Sucrose gradient analysis

To analyze the ribosomal preinitiation complexes that were assembled on the RNA of interest, the complexes were loaded on 7-47% sucrose gradients containing 5 mM $MgCl_2$, 25

recombinant RNasin (promega), 75 mM KCl 0.5 mM MgCl2, and 1.3 mM of puromycin prior preinitiation complex assembly. Then, complexes are formed by incubation with 500 nM of the RNA of interest in the presence of specific inhibitors such as cycloheximide (1 mg/ml) or GMP-PNP (4 mM) for 5 min at 30°C and then 20 min on ice. Then RNA-ribosome complexes were complemented with one volume of ice-cold buffer A containing 20 mM Tris-HCl (pH 7.5), 100 mM KAc, 2.5 mM $Mg[Ac]_2$, 2 mM DTT, 1 mM ATP and 0.25 mM spermidine and placed on ice. In order to separate ribosomal complexes from the non-ribosomal fraction, samples were ultracentrifuged at 88,000 rpm in S100AT3 rotor (Sorvall-Hitachi) at 4°C for 1h. Toe printing assay was adapted from previously established protocols in the lab (Martin et al, 2011, Martin et al., 2016). After centrifugation, the supernatant was discarded and the complex in the pellet was resuspended in 30 µl of ice-cold buffer A. Then this complex was incubated with 5' radioactively labelled DNA oligonucleotide complementary to nts 22-51 of renilla luciferase coding sequence for 3 min at 30°C. Then, 1 µl of a 320 mM $Mg(Ac)_2$, 4 µl of a dNTP mixture (containing 5 mM of dATP, dGTP, dTTP and dCTP), 10 units of recombinant RNAsin (Promega) and 1 unit/µl Avian Myoblastosis virus (AMV) reverse transcriptase (Promega) were added and incubated for 1h at 30°C. The synthesized cDNAs were analysed on 8% PAGE next to sequencing ladders.

Results

TIE-mediated inhibition recapitulated in RRL

To address the question of how efficient TIE elements inhibit translation, we first performed *in vitro* experiments using several reporter constructs. To analyse their translation efficiency, capped mRNAs were translated in RRL in the presence of [^{35}S]-methionine to monitor protein neo-synthesis and measured luciferase activity to determine the translation efficiency. As a reference, we used a reporter mRNA containing the 5'UTR of human β-globin mRNA upstream of the Renilla luciferase coding sequence (**Fig. 1A**). We choose the 5'UTR of β-globin because of its small size and its CA-rich sequence that has been shown to be unstructured (11). Then we inserted the TIE a3 and a11 upstream of the β-globin 5'UTR. First, we performed translation experiments with increasing amounts of mRNA and determined that with concentrations higher than 100 nM, the translation yield reached a plateau (**Fig. S1A**). In order to avoid any titration effect due to too high mRNA concentrations, we decided to perform all translation assays at an mRNA concentration of 50 nM, which enables sub-saturating conditions. When the TIE a3 and a11 are placed upstream of the β-globin 5'UTR, translation is inhibited by 79% and 91% respectively (**Fig. 1A**). Therefore our cell-free translation assay using RRL efficiently recapitulates the previously described translation inhibition of TIE elements (12). Likewise, TIE-mediated inhibition was efficiently recapitulated in other cell-free translation systems like Wheat Germ Extracts, drosophila S2 cell extracts, human HeLa cell extracts (**Fig. S1B**). Moreover, the same constructs were introduced in a plasmid that allowed us to monitor translation inhibition *in vivo*. We tested two cell lines expressing Hox genes, namely human HEK293T and mouse mesenchyme cell line C3H10T1/2 (13). In these cells, translation inhibition by TIE a3 and a11 is 98% and 88% respectively (**Fig. S1C**). In order to define the minimal active domains of a3 and a11, sequential 5' deletions were performed. In each experiment, translation for each construct was compared to the control (w/o TIE) (**Fig. 1B**). First, we tested large deletions in order to roughly determine the functional domain (**Fig. S1D**), then more precise deletions to map exactly the 5' end of functional domains for a3 and a11 (**Fig. 1B**). Deletion of 68 nts in TIE a3 does not affect inhibition (99%). A further deletion of only 6 additional nucleotides, to 74, reduces the translation inhibition (78%). Deletion of 113 nts completely abolishes the translation inhibition of TIE a3 (10%). Therefore the minimal a3 element encompasses nucleotides 68 to 170. Concerning TIE a11, the inhibition is partially reduced (70%) when 153 nts are deleted and completely lost when 161 nts are deleted. The minimal

TIE a11 element is located between residues 139 to 216. In conclusion, these experiments allowed the characterization of minimal a3 and a11 RNA regulons that retain their inhibitory function.

TIE elements have distinct secondary structures

To gain further insights into the structural and functional properties of TIE elements, we performed chemical probing using DMS and CMCT reagents for both TIE a3 and TIE a11 (**Fig. 2**). Since modifications were performed in triplicates, the average of reactivity was calculated for each nucleotide (**Fig. S2**). With this reactivity, we build a 2D model for both TIE a3 and a11. The TIE a3 contains a 5' proximal stem-loop and another bigger structure comprising a two-way junction and two hairpins (**Fig. 2A**). For TIE a11, the higher GC content (64%) than TIE a3 (45%) suggests more structured model. Our probing experiments confirmed this statement. It comprises four stem-loops and a three-way junction structure (**Fig. 2B**). The secondary structures of both TIEs were also probed in the frame of the full-length Hox 5'UTR containing the IRES. The reactivities obtained with isolated TIEs and with TIEs embedded in full-length 5'UTR were highly similar suggesting that the TIEs fold independently from the IRES (**Fig. S2**). Using these two models, we wanted to characterize further the structure-function relationships of both TIE elements.

TIE a3 contains a uORF that inhibits translation

The 5' truncations experiments allowed us to map the minimal inhibiting element. Interestingly, in the minimal a3, a putative uORF is present starting from two putative uAUGs at positions 111-113 and 123-125 (**Fig. 3A**). In order to test their implication in the translation inhibition, both uAUGs were mutated into UAC thereby eliminating any possibility of AUG-like codon recognition. Interestingly, the mutation of $AUG_{111\text{-}113}$ completely abolishes the inhibition confirming its implication. This is not the case for $AUG_{123\text{-}125}$ since its mutation had no effect (**Fig. 3A**). This is compatible with the fact that $AUG_{111\text{-}113}$ has an optimal kozak sequence (A at -3 and G at +4), thereby the ribosome initially recognizes it during scanning along the 5'UTR (14). To further confirm the recognition of $uAUG_{111\text{-}113}$, we mutated it into AUG-like codons. Since the $AUG_{111\text{-}113}$ is involved in a stem, we mutated the AUG into CUG and GUG and inserted at the same the compensatory mutations that enable the formation of the stem with the AUG-like codons. In both mutants, the inhibition was fully destroyed indicating that efficient inhibition requires a genuine AUG codon (**Fig. S3A**). Next, we confirmed the use of $uAUG_{111\text{-}113}$ by toe print assay. As expected, a canonical +16 reverse transcription arrest from $uAUG_{111\text{-}113}$ was detected with GMP-PNP and less efficiently with cycloheximide. When the uAUG is mutated to UAC, the +16 toe print disappears (**Fig. 3B**).

Altogether, this data confirm that a pre-initiation complex efficiently assembles on the $uAUG_{111\text{-}113}$. We then wondered whether the ribosome assembled on this AUG codon actually proceeds to translation elongation in order to produce a peptide. By performing sucrose gradient analysis, we detected polysome formation suggesting that efficient translation is actually occurring from $AUG_{111\text{-}113}$ (**Fig. S3B**). Accordingly, mutation of $uAUG_{111\text{-}113}$ drastically reduces the amount of polysomes and the formation of 48S complexes in the presence of GMP-PNP thereby confirming that the translation is starting on $uAUG_{111\text{-}113}$ (**Fig. S3B**). Interestingly, the alignement of Hox a3 TIE element shows a conservation of the $uAUG_{111\text{-}113}$ among the species (**Fig. S3C**). In the Hox a3 5'UTR, the uORF starting from $uAUG_{111\text{-}113}$ is extending through the whole IRES. To check further whether the uORF is indeed translated through the full length Hox 5'UTR we first deleted a single nucleotide (G_{333}) thereby producing a fusion protein formed by the peptide produced from the uORF and Renilla Luciferase. Indeed, with this single frameshifting mutation, we could detect a longer protein demonstrating that the pre-initiation complex assembled on the $uAUG_{333}$ indeed proceeds to translation elongation and is efficiently translating through the full length 5'UTR of Hox a3 mRNA (**Fig. 4A**). We also verified the translation from $uAUG_{111\text{-}113}$ with our reporter constructs containing only the TIE a3 element. Likewise, the insertion of a single frameshifting nucleotide (A_{220}) allows the detection of a longer fusion protein (**Fig. S3C**). Remarkably, the small peptide encoded by the uORF is also detected with the wild-type TIE a3 reporter construct (**Fig. S4A**) and in Hox a3 mRNA (**Fig. S4B**). Altogether, our cell-free translation assays demonstrate that Hox TIE a3 translation is achieved by translation of a uORF that starts at $AUG_{111\text{-}113}$ codon that extends through the whole Hox a3 5'UTR. To confirm these results *in vivo*, we generated reporter constructs in the plasmid pmirGlo. As expected, Wt TIE a3 blocks very efficiently translation both in HEK293T and C3H10T1/2 cells. When the $AUG_{111\text{-}113}$ is mutated to UCA, the inhibition is significantly affected both in HEK293T and C3H10T1/2 at respectively 57% and 45% compare to Wt TIE a3 (**Fig. 4B**). Although the inhibition is not fully abolished, these experiments confirmed that the codon $AUG_{111\text{-}113}$ is critical for efficient translation inhibition in Hox a3.

TIE a11-mediated inhibition is mediated by a stalled 80S ribosome

We next asked whether TIE a11 has a similar inhibitory mechanism. The TIE a11 contains two putative uAUGs at positions 84-86 and 159-161. Mutations of $AUG_{84\text{-}86}$ had no impact on translation inhibition as expected since this codon is not present in the minimal TIE a11 that was mapped in figure 1. Likewise, mutation of $AUG_{159\text{-}161}$, which is located in the minimal TIE a11, has no effect either on translation inhibition (**Fig. 5A**). According to our 2D model,

a long GC-rich stable stem loop (SL) structure (ΔG = -25.00 kcal/mol) spans nucleotides 104-154 (**Fig. 2B**). This long hairpin comprises sixteen GC base pairs that putatively can interfere with a pre-initiation scanning complex. To test the inhibitory efficiency of this stem-loop, we transplanted it in a strictly cap-dependent reporter mRNA containing the 5'UTR of β-globin upstream of the renilla coding sequence. The a11 SL was inserted in the middle of the 5'UTR thereby keeping 25 nucleotides from the 5' proximity to ensure proper access the 5' cap. Interestingly, the translation of this mRNA was significantly abolished showing that this stem loop is sufficient to inhibit cap-dependent translation when transplanted in any mRNA (**Fig. S5A**). Strikingly, $uAUG_{84-86}$ is located 19 nts upstream of the bottom of the inhibiting SL. This distance is compatible with a ribosome initiating on $uAUG_{84-86}$ without clashing with SL. Moreover, the distance between the uAUG and the SL is optimal for favouring AUG recognition by scanning arrest forced by the SL. Interestingly AUG_{84-86} is followed by a stop codon UAG_{87-89}. This unique combination of start-stop codon upstream of stem loop structure raised the question of whether the ribosome is forced to recognize this AUG despite a suboptimal Kozak context (A_{-3} and U_{+4}). To address this hypothesis, we mutated the stop codon UAG_{87-89} to UGG_{87-89}. When the stop is mutated, a small peptide presumably resulting from the translation from $uAUG_{84-86}$ through 5'UTR, which we called a11 uORF (**Fig. 5B**). The presence of the highly stable SL did not allow us to confirm the assembly of the 80S on $uAUG_{84-86}$ by toe-print assay. Indeed, premature RT arrests due to the presence of the highly stable SL render the toe-print assay on a11 impossible. In order to demonstrate that the ribosome is assembling on $uAUG_{84-86}$, we performed sucrose gradient analysis with radiolabeled mRNAs. With Wt TIE a11, an 80S complex accumulates with and without inhibitor confirming the assembly of a stalled ribosome (**Fig. 5C**). The mutation of $uAUG_{84-86}$ to UAC drastically reduces the amount of 80S; in contrast, mutation of $uAUG_{159-161}$ to UAC does not affect 80S accumulation. Interestingly, alignment of Hox a11 TIE element shows a conservation of the $uAUG_{84-86}$ among most of the species (highlighted in red box) with exceptions of mutations to AUG-like codons (such as AGG or GUG) with the lack of conservation of the UAG_{87-89} in other species (mutation to CAG) (**Fig. S5B**). In humans, the start-stop codon is not present however, TIE a11 harbours a downstream insertion of another start-stop AUG-UAG upstream of the predicted SL (highlighted in red box). Moreover, it remains true that a11 SL (104-154) is conserved among species suggesting a functional significance (**Fig. S5B**). Altogether, this data show that a stalled ribosome is indeed assembled on $uAUG_{84-87}$ and the stalling is caused by the synergistic effect of a stop codon next to the AUG and a stable SL in a11.

a long GC-rich stable stem loop (SL) structure (ΔG = -25.00 kcal/mol) spans nucleotides 104-154 (**Fig. 2B**). This long hairpin comprises sixteen GC base pairs that putatively can interfere with a pre-initiation scanning complex. To test the inhibitory efficiency of this stem-loop, we transplanted it in a strictly cap-dependent reporter mRNA containing the 5'UTR of β-globin upstream of the renilla coding sequence. The a11 SL was inserted in the middle of the 5'UTR thereby keeping 25 nucleotides from the 5' proximity to ensure proper access the 5' cap. Interestingly, the translation of this mRNA was significantly abolished showing that this stem loop is sufficient to inhibit cap-dependent translation when transplanted in any mRNA (**Fig. S5A**). Strikingly, $uAUG_{84-86}$ is located 19 nts upstream of the bottom of the inhibiting SL. This distance is compatible with a ribosome initiating on $uAUG_{84-86}$ without clashing with SL. Moreover, the distance between the uAUG and the SL is optimal for favouring AUG recognition by scanning arrest forced by the SL. Interestingly AUG_{84-86} is followed by a stop codon UAG_{87-89}. This unique combination of start-stop codon upstream of stem loop structure raised the question of whether the ribosome is forced to recognize this AUG despite a suboptimal Kozak context (A_{-3} and U_{+4}). To address this hypothesis, we mutated the stop codon UAG_{87-89} to UGG_{87-89}. When the stop is mutated, a small peptide presumably resulting from the translation from $uAUG_{84-86}$ through 5'UTR, which we called a11 uORF (**Fig. 5B**). The presence of the highly stable SL did not allow us to confirm the assembly of the 80S on $uAUG_{84-86}$ by toe-print assay. Indeed, premature RT arrests due to the presence of the highly stable SL render the toe-print assay on a11 impossible. In order to demonstrate that the ribosome is assembling on $uAUG_{84-86}$, we performed sucrose gradient analysis with radiolabeled mRNAs. With Wt TIE a11, an 80S complex accumulates with and without inhibitor confirming the assembly of a stalled ribosome (**Fig. 5C**). The mutation of $uAUG_{84-86}$ to UAC drastically reduces the amount of 80S; in contrast, mutation of $uAUG_{159-161}$ to UAC does not affect 80S accumulation. Interestingly, alignment of Hox a11 TIE element shows a conservation of the $uAUG_{84-86}$ among most of the species (highlighted in red box) with exceptions of mutations to AUG-like codons (such as AGG or GUG) with the lack of conservation of the UAG_{87-89} in other species (mutation to CAG) (**Fig. S5B**). In humans, the start-stop codon is not present however, TIE a11 harbours a downstream insertion of another start-stop AUG-UAG upstream of the predicted SL (highlighted in red box). Moreover, it remains true that a11 SL (104-154) is conserved among species suggesting a functional significance (**Fig. S5B**). Altogether, this data show that a stalled ribosome is indeed assembled on $uAUG_{84-87}$ and the stalling is caused by the synergistic effect of a stop codon next to the AUG and a stable SL in a11.

Mass spectrometry analysis pre-initiation complexes programmed by TIE a3 and TIE a11

To further characterize the two different modes of action employed by a3 and a11, we were interested in identifying the possible factors specifically acting in such mechanisms. For that, we used an approach to purify pre-initiation complexes programmed by TIE a3 and a11 suitable for Mass Spectrometry analysis. Briefly, ribosomes were assembled on chimeric biotinylated mRNA-DNA molecules and immobilized on streptavidin-coated beads after incubation with RRL in the presence of cycloheximide. Complexes were then eluted by DNAse treatment as previously described (9, 10) and analysed by mass spectrometry analysis. Three different mRNA constructs were used, TIE a3, a11 and β-globin mRNA as a negative control. Comparison between the three mRNAs showed specific factors for each RNA (**Fig. 6**). Interestingly, for TIE a3, we identified a set of translation-related proteins. Among the strongest hits, we found eIF2D, a non-canonical initiation factor which has been shown to be involved in the initiation of uORFs and other specific mRNAs (15, 16). Another interesting candidate is methionine aminopeptidase MetAP1 which removes N-terminal methionine from nascent proteins in a co-translationnally manner (17). Then, other factors that are linked to translation have been selected such as the scanning factor eIF1A and its isoform eIF1A-X, eIF3A, Arginyl- and Leucyl-tRNA synthetases QARS and LARS, DEAD-box helicases DHX36 and DDX39B, elongation factor HBS1L and RpL38 ribosomal protein. For TIE a11, we identified different translation-related proteins among which are ASAP1, a GTPase activator protein, methionine aminopeptidase Met AP1, RpL38, Valyl-tRNA synthetase VARS, and interestingly, eIF3j, a subunit of initiation factor eIF3 usually dissociating at early stages of initiation to allow mRNA entry (18, 19) thereby, unlikely to be present in initiation complexes. When comparing TIE a3 with TIE a11, we could detect some initiation factors and translation-related proteins specific for a3 like eIF2D, eIF1A, eIF1A-X, eIF5B, LARS, QARS, DHX21 and HBS1L (**Fig. 6**). These results show a variation in factors involved in the two distinct TIE-mediated inhibitions. It is noteworthy to mention that TIE a3 seems to require more scanning factors than TIE a11.

Translation inhibition by TIE a3 requires eIF2D

After identifying eIF2D specifically by mass spectrometry analysis for TIE a3, we were interested in getting more insights into how this factor might be involved. eIF2D has been shown to be a non-canonical initiation factor which delivers tRNA to the P-site of the ribosome in a GTP-independent manner (15). It is involved in the initiation step on specific mRNAs among which those having uORFs, A-rich sequences, leaderless mRNAs and

reinitiation on main ORF (16). Interestingly, we have shown that TIE a3 has a uORF and A-rich motifs upstream and downstream of $uAUG_{111-113}$. We tested the implication of such motifs in TIE-a3 mediated inhibition. For that, we mutated both A-motif, namely $A_{107-110}$ usptream of the AUG and $A_{147-151}$ downstream of the AUG (**Fig. 7**). In order to avoid any side effects due to inappropriate Kozak sequence, we mutated the As at position 107-110 to GGCC thereby keeping a purine residue at -3 position. The second A-motif downstream of AUG was similarly mutated to GGCCC. Interestingly, the mutation of upstream A motif had a 2-fold reduction effect on translation inhibition of TIE a3, in contrast mutation of the A-motif downstream of the AUG does not affect TIE a3 inhibition (**Fig. 7**). Therefore, the A-motif upstream of the AUG is critical for translation inhibition suggesting that eIF2D is required for efficient inhibition. Indeed, such an A-motif is required for proper eIF2D recruitment (15).

Then, to further characterize the molecular mechanism of TIE a3-mediated inhibition, we performed sucrose gradient analysis with TIE a3 in the presence of different inhibitors of ribosome (**Fig S6A**). Interestingly, with geneticin, an A-site inhibitor that prevents 80S ribosome translocation (20), we saw the accumulation of a homogenous 80S complex. Surprisingly, this is not the case with cycloheximide, another translocation inhibitor that binds the E-site of the ribosome (21), whereby we had an unusual profile of bigger complex than 80S possibly corresponding to multiple stalled ribosomes on TIE a3 uORF. This suggests that cycloheximide blockage is partially inefficient which is not the case for the control (w/o TIE) (**Fig. S6B**). When using GMP-PNP, a non-hydrolysable analogue of GTP that prevents subunit joining, a canonical 48S ribosome accumulates (22). With edeine, an inhibitor that prevents codon-anticodon interaction, we detected a 43S accumulation (23). These findings raised several interesting aspects regarding the efficiency of the translocation inhibitors like geneticin and cycloheximide, which binds to A- and E-sites respectively. The fact that TIE a3 is partially resistant to cycloheximide blockage suggests that its binding site is not fully accessible. In fact, these data are compatible with our previous toe printing results showing less efficient toe print with cycloheximide than with GMP-PNP (**Fig. 3B**). Interestingly, cryo-EM structures have shown that eIF2D might be overlapping, and thereby hindering, at least partially, the binding site of cycloheximide (24). Although preliminary, this data goes along with the fact that eIF2D might be required for TIE a3 mediated inhibition. At this stage, we cannot rule out that this might be due to another specific factor required for TIE a3 inhibition.

Discussion

Our study has shown that two HoxA mRNAs, a3 and a11, are regulated by different mechanisms to ensure the inhibition of cap-dependent translation and allowed us to propose a model for their mode of action (**Fig. 8**). First, we have shown that TIE elements can function *in vitro* using cell-free translation extracts as well as *in vivo*. Our findings suggest that a3 inhibits translation by a uORF which is translated through the whole 5'UTR of Hox a3 mRNA producing a small protein of 11 kDa size (**Fig. S4B**). In contrast to the uAUG that is highly conserved, the coding sequence of the uORF is not conserved among species (**Fig. S3C**). Therefore, we suggest that the produced peptide has no other function than blocking cap-dependent translation of Hox a3 mRNA. Indeed, uORFs have been recognized as regulators of translation for number of cellular mRNAs (25). For instance, four uORFs in the 5' leader of *GCN4* mRNA restrict the flow of scanning ribosomes from the cap site to the GCN4 initiation codon (26, 27) and the uORF in AdoMetDC mRNA generates a nascent hexapeptide that interacts with its translating ribosome to suppress translation of AdoMetDC1 mRNA in a cell-specific manner (28). Interestingly, our data suggests the implication of eIF2D in TIE a3 mode of action. In agreement, it has been shown that eIF2α-phosphorylation during stages of embryonic development promotes translation from uORFs (29). Therefore canonical cap-dependent translation initiation with eIF2 is not possible during embryonic development. The cap-dependent translation initiation of uORF from TIE a3 might use eIF2D as an alternative to replace inactive phosphorylated eIF2 to promote uORF translation. So far, no consensus *cis*-acting sequence has been clearly defined on mRNA for eIF2D recruitment. But it was suggested from *in vitro* studies that eIF2D would form initiation complexes on leaderless, short 5'UTR mRNAs or A-rich 5'UTRs (15, 16). In TIE a3, an A-rich motif adjacent to the uAUG is critical for a3 function (**Fig. 7**), suggesting that this motif is important for eIF2D recruitment to the preinitiation complex as previously described (15). Interestingly, cycloheximide partial efficiency with TIE a3 suggests that its binding site is not fully accessible which was not the case for other inhibitors. This has been recently described as cycloresistance (30). It occurs due to queing of canning PICs in response to slowly elongating ribosomes from non-AUG codons. Regarding TIE a11, the ribosome recognizes a combination of three *cis*-acting elements in the 5'UTR (**Fig. 8**). (i) A start codon located at positions 84-86 (ii) a stop codon located immediately downstream of the AUG start codon and (iii) a long highly stable GC-rich helical structure (SL) located at +20 downstream of the AUG (by convention the A from the AUG being +1). These three elements promote the

stalling of an 80S complex upstream of the SL. This is reminiscent of a similar mechanism that has been described in the *Arabidopsis thaliana* NIP5.1 5'UTR mRNA that contains an AUG-stop that regulates translation of the main ORF through a ribosome stalling mechanism and mRNA degradation (31). In the case of TIE a11, an additional *cis*-element is required for ribosome stalling, namely the stable SL that is present downstream of the AUG-stop. A ribosome that is stalled with a stop codon in the A site is usually a signal for recruitment of the release factors to dissociate the ribosomal subunits from the mRNA. With our cell-free translation assay, we showed that the stalled 80S programmed with TIE a11 is very stable and does not dissociate (**Fig. 5C**). A possible explanation to this phenomenon could be that the SL blocks the access of the release factors to the A site thereby preventing ribosome dissociation (32). Interestingly, the start-stop combination lacks conservation among species (**Fig. S5B**). Some species possess a substitution of the start codon AUG to AUG-likes codon such GUG or a mutation of stop codon that leads to a longer uORF. In contrast, the SL remains highly conserved and we have shown that the sole SL is strong enough to impede ribosome scanning (**Fig. S5A**). Indeed, secondary structures in the 5'UTR have been shown to inhibit translation like the case of a conserved stem loop structure in the 5'UTR of TGF-β1 mRNA (33). Therefore, in various species TIE a11 might use different combinations of the three *cis*-acting elements in order to block cap-dependent translation, the common feature between all species being the presence of the SL. As previously described, the IRES in Hox a11 mRNA requires the presence of the ribosomal protein RpL38 while that in Hox a3 mRNA, the IRES is RpL38-independent (6). Our probing experiments revealed that the folding of both a3 and a11 TIE elements is independent of the presence of IRES suggesting that their mode of action does not dependent on the IRES. TIE may have evolved in such a way to favour the translation from the downstream IRES hence justifying why there is variation in terms of sequence and mechanism but same inhibitory effect. This unique combination of an inhibitor of canonical translation mechanism and the activator of specialized translation sets an interesting point on how the 5'UTR elements confer ribosome specificity to translation (12). Importantly, the acquisition of these TIE elements in subsets of Hox mRNAs enables an additional level of regulatory control between the canonical translation and the IRES-dependent one. It will be interesting to determine how other TIE elements a4 and a9 inhibit translation and whether there are common functional features amongst all Hox mRNAs.

Acknowledgments

This work was funded by 'Agence Nationale pour la Recherche' (Ribofluidix, ANR-17-CE12-0025-01, by University of Strasbourg and by the 'Centre National de la Recherche Scientifique'. We would like to thank Maria Barna for sharing sequences of Hox a3 and 11 transcripts. We are grateful to Christine Allmang, Hassan Hayek, Lauriane Gross, Aurélie Janvier, Aurélie Durand, Philippe Hamman, Johana Chicher, Lauriane Kuhn for technical assistance, Mireille Baltzinger and Pascale Romby for support and fruitful discussions on the project. We would also like to thank Sebastian Pfeiffer for the pmirGLO plasmid.

References

1. Hinnebusch,A.G. (2014) The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.*, **83**, 779–812.
2. Alexander,T., Nolte,C. and Krumlauf,R. (2009) Hox Genes and Segmentation of the Hindbrain and Axial Skeleton. *Annu. Rev. Cell Dev. Biol.*, **25**, 431–456.
3. Oh,S.K., Scott,M.P. and Sarnow,P. (1992) Homeotic gene Antennapedia mRNA contains 5'-noncoding sequences that confer translational initiation by internal ribosome binding. *Genes Dev.*, **6**, 1643–53.
4. Ye,X., Fong,P., Iizuka,N., Choate,D. and Cavener,D.R. (1997) Ultrabithorax and Antennapedia 5' untranslated regions promote developmentally regulated internal translation initiation. *Mol. Cell. Biol.*, **17**, 1714–21.
5. Xue,S. and Barna,M. (2015) Cis-regulatory RNA elements that regulate specialized ribosome activity. *RNA Biol.*, **12**, 1083–7.
6. Kondrashov,N., Pusic,A., Stumpf,C.R., Shimizu,K., Hsieh,A.C., Xue,S., Ishijima,J., Shiroishi,T. and Barna,M. (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, **145**, 383–97.
7. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–5.
8. Mathews,D.H., Turner,D.H. and Watson,R.M. (2016) RNA secondary structure prediction. *Curr. Protoc. Nucleic Acid Chem.*, **2016**, 11.2.1-11.2.19.
9. Chicher,J., Simonetti,A., Kuhn,L., Schaeffer,L., Hammann,P., Eriani,G. and Martin,F. (2015) Purification of mRNA-programmed translation initiation complexes suitable for mass spectrometry analysis. *Proteomics*, **15**, 2417–25.
10. Prongidi-Fix,L., Schaeffer,L., Simonetti,A., Barends,S., Ménétret,J.-F., Klaholz,B.P., Eriani,G. and Martin,F. (2013) Rapid purification of ribosomal particles assembled on histone H4 mRNA: a new method based on mRNA-DNA chimaeras. *Biochem. J.*, **449**, 719–28.
11. Fletcher,L., Corbin,S.D., Browning,K.S. and Ravel,J.M. (1990) The absence of a m7G cap on beta-globin mRNA and alfalfa mosaic virus RNA 4 increases the amounts of initiation factor 4F required for translation. *J. Biol. Chem.*, **265**, 19582–7.
12. Xue,S., Tian,S., Fujii,K., Kladwang,W., Das,R. and Barna,M. (2015) RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature*, **517**, 33–8.
13. Phinney,D.G., Gray,A.J., Hill,K. and Pandey,A. (2005) Murine mesenchymal and

embryonic stem cells express a similar Hox gene profile. *Biochem. Biophys. Res. Commun.*, **338**, 1759–1765.

14. Kozak,M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
15. Dmitriev,S.E., Terenin,I.M., Andreev,D.E., Ivanov,P.A., Dunaevsky,J.E., Merrick,W.C. and Shatsky,I.N. (2010) GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor. *J. Biol. Chem.*, **285**, 26779–26787.
16. Akulich,K.A., Andreev,D.E., Terenin,I.M., Smirnova,V. V., Anisimova,A.S., Makeeva,D.S., Arkhipova,V.I., Stolboushkina,E.A., Garber,M.B., Prokofjeva,M.M., *et al.* (2016) Four translation initiation pathways employed by the leaderless mRNA in eukaryotes. *Sci. Rep.*, **6**, 37905.
17. Dummitt,B., Micka,W.S. and Chang,Y.H. (2003) N-terminal methionine removal and methionine metabolism in Saccharomyces cerevisiae. *J. Cell. Biochem.*, **89**, 964–974.
18. Fraser,C.S., Berry,K.E., Hershey,J.W.B. and Doudna,J.A. (2007) eIF3j Is Located in the Decoding Center of the Human 40S Ribosomal Subunit. *Mol. Cell*, **26**, 811–819.
19. Kolupaeva,V.G., Unbehaun,A., Lomakin,I.B., Hellen,C.U.T. and Pestova,T. V (2005) Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA*, **11**, 470–486.
20. Vicens,Q. and Westhof,E. (2003) Crystal structure of geneticin bound to a bacterial 16 S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.*, **326**, 1175–1188.
21. Schneider-Poetsch,T., Ju,J., Eyler,D.E., Dang,Y., Bhat,S., Merrick,W.C., Green,R., Shen,B. and Liu,J.O. (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat. Chem. Biol.*, **6**, 209–217.
22. Eliseev,B., Yeramala,L., Leitner,A., Karuppasamy,M., Raimondeau,E., Huard,K., Alkalaeva,E., Aebersold,R. and Schaffitzel,C. (2018) Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.*, **46**, 2678–2689.
23. Hinnebusch,A.G. (2011) Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.*, **75**, 434–467.
24. Weisser,M., Schäfer,T., Leibundgut,M., Böhringer,D., Aylett,C.H.S. and Ban,N. (2017) Structural and Functional Insights into Human Re-initiation Complexes. *Mol. Cell*, **67**, 447-456.
25. Barbosa,C., Peixeiro,I. and Romão,L. (2013) Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet.*, **9**, e1003529.
26. Hinnebusch,A.G. (1993) Gene-specific translational control of the yeast GCN4 gene by

phosphorylation of eukaryotic initiation factor 2. *Mol. Microbiol.*, **10**, 215–23.

27. Dever,T.E., Kinzy,T.G. and Pavitt,G.D. (2016) Mechanism and regulation of protein synthesis in Saccharomyces cerevisiae. *Genetics*, **203**, 65–107.
28. Uchiyama-Kadokura,N., Murakami,K., Takemoto,M., Koyanagi,N., Murota,K., Naito,S. and Onouchi,H. (2014) Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in Arabidopsis thaliana. *Plant Cell Physiol.*, **55**, 1556–1567.
29. Friend,K., Brooks,H.A., Propson,N.E., Thomson,J.A. and Kimble,J. (2015) Embryonic stem cell growth factors regulate eIF2α phosphorylation. *PLoS One*, **10**, e0139076.
30. Kearse,M.G., Goldman,D.H., Choi,J., Nwaezeapu,C., Liang,D., Green,K.M., Goldstrohm,A.C., Todd,P.K., Green,R. and Wilusz,J.E. (2019) Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev.*, **33**, 871–885.
31. Tanaka,M., Sotta,N., Yamazumi,Y., Yamashita,Y., Miwa,K., Murota,K., Chiba,Y., Hirai,M.Y., Akiyama,T., Onouchi,H., *et al.* (2016) The Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent Ribosome Stalling and mRNA Degradation. *Plant Cell*, **28**, 2830–2849.
32. Brown,A., Shao,S., Murray,J., Hegde,R.S. and Ramakrishnan,V. (2015) Structural basis for stop codon recognition in eukaryotes. *Nature*, **524**, 493–496.
33. Jenkins,R.H., Bennagi,R., Martin,J., Phillips,A.O., Redman,J.E. and Fraser,D.J. (2010) A conserved stem loop motif in the 5'untranslated region regulates transforming growth factor-β1 translation. *PLoS One*, **5**, e12283.

Figure Legends

- **Figure 1: TIE-mediated inhibition is recapitulated in RRL and does not require full length TIEs.**

A. Three capped mRNAs were used to test TIE-mediated inhibition *in vitro*. TIE a3 and TIE a11 were placed upstream of the 5'UTR of β-globin and the Renilla luciferase coding sequence. Translation assays were performed *in vitro* using RRL at an mRNA concentration of 50 nM, which enables sub-saturating conditions. The relative expression of Luciferase protein reflects the efficiency of translation inhibition by TIE a3 and TIE a11. Values were normalized to that of the control (w/o TIE) which corresponds to normal expression without inhibition and was set to 100%. **B.** Sequential deletions in the 5' extremity of TIE a3 and TIE a11 constructs were performed to assay their effect on translation. Values of translation expression were normalized to that of the control (w/o TIE). Experiments were performed in triplicates. The percentages of inhibition for each TIE are indicated in the histogram.

- **Figure 2: the secondary structural models of TIE a3 and TIE a11 reveal distinct structures.**

The structures of (A) TIE a3 (170 nucleotides) and (B) TIE a11 (216 nucleotides) were obtained by chemical probing using base-specific reagents, DMS and CMCT. After modifications, reverse transcription was performed using fluorescently labelled primers to determine the position of modified nucleotides. These experiments were performed in triplicates, the reactivities are shown as average reactivity from three independent experiments. A representation of reactivities is assigned as colour code depending on a range of values as shown in the figure legend on the right. Reactivity values for each nucleotide with standard deviations are shown in supplementary figures (S2A, S2B, S2C, S2D).

- **Figure 3: upstream $AUG_{111\text{-}113}$ in TIE a3 is essential for inhibition.**

(A) Substitution mutations in $uAUG_{111\text{-}113}$ and $uAUG_{123\text{-}125}$ to UAC were performed. Constructs were translated in RRL to evaluate the effect of mutation on translation efficiency as previously described. **(B)** Toe printing analysis of ribosomal assembly on two mRNAs, TIE a3 Wt and the mutant of $u(AUG/UAC)_{111\text{-}113}$. Initiation complexes were assembled in RRL extracts in the absence or presence of translation inhibitors, cycloheximide and GMP-PNP. Reaction samples were separated on 8% denaturing PAGE together with the appropriate

sequencing ladder. Toe-print positions were counted starting on the A+1 of the AUG codon at +16 position. Full-length cDNAs are indicated by an arrow at the top of the gel.

- **Figure 4: the uAUG$_{111\text{-}113}$ in TIE a3 is translated through 5'UTR of Hox a3.**

(A) Three transcripts were used for this experiment: full length 5'UTR of Hox a3 (right), a deletion mutant at nucleotide G_{333} in IRES a3 (left) and a control transcript without TIE. To test the translation of uORF in TIE a3 starting form uAUG$_{111\text{-}113}$, a deletion of G in IRES a3 at position 333 was performed to create a longer uORF that is in the same frame as the ORF of Renilla luciferase in order to create an N-terminally extended luciferase. Transcripts were translated *in vitro* in RRL and products were loaded on 10% SDS-PAGE in the presence of ^{35}S-Methionine **(B)** *In vivo* luciferase assays in two embryonic cells lines; HEK293FT (left) and C3H10T1/2 (right). Reporter constructs in pmirGlo containing TIE a3, uAUG$_{111\text{-}113}$/UAC, and without TIE, were transfected in the two indicated cell lines. Renilla luciferase expression was normalized to the control (w/o TIE), which was set to 100%. Experiments were performed in triplicates.

- **Figure 5: a start-stop uORF in TIE a11 stalls an 80S upstream of a highly stable structure.**

(A) Mutational analysis of uAUGs in TIE a11. Three transcripts with AUG/UAC mutations were used: M1: (AUG/UAC)$_{84\text{-}87}$ + (AUG/UAC)$_{159\text{-}161}$, M2: (AUG/UAC)$_{84\text{-}87}$ and M3: (AUG/UAC)$_{159\text{-}161}$. Transcripts were translated in RRL at 50 nM concentrations and the luciferase expression was normalized to the control (w/o TIE) as previously described. Experiments were performed in triplicates. **(B) A** single substitution A/G_{88} mutation destroys the stop codon UAG$_{87\text{-}89}$ to UGG in TIE a11, TIE a11 wt and control (w/o TIE) were also translated as references in RRL. Translation products were loaded on 15% SDS-PAGE. **(C)** Ribosomal pre-initiation complexes were assembled and analysed on 7-47% sucrose gradient with [alpha-^{32}P]GTP-radiolabeled TIE a11 as well as the two mutants of uAUG/UAC at the previously indicated positions in the absence or presence of cycloheximide. Heavy fractions correspond to polysomes while lighter correspond to free RNPs.

- **Figure 6: distinct profiles for factors involved in TIE-mediated inhibition.**

Mass spectrometry analysis of cycloheximide-blocked translation initiation complexes on three transcripts: TIE a3, TIE a11 placed upstream of 5'UTR of β-globin and the 5'UTR of β-globin (control). Graphical representation of proteomics data: protein $\log_2$ spectral count fold

changes (on the x-axis) and the corresponding adjusted $\log_{10}$ P-values (on the y-axis) are plotted in a pair-wise volcano plot. The significance thresholds are represented by a horizontal dashed line (P-value = 1.25, negative-binomial test with Benjamini–Hochberg adjustment) and two vertical dashed lines (−1.0-fold on the left and +1.0-fold on the right). Data points in the upper left and upper right quadrants indicate significant negative and positive changes in protein abundance. Protein names are labelled next to the off-centred spots and they are depicted according to the following color code: proteins colored in green are translation-related, by red are non-translation-related and by black are non-significant with <10 spectra. Data points are plotted on the basis of average spectral counts from triplicate analysis. Three profiles were produced by comparing the proteomics of two transcripts.

- **Figure 7: mutational analysis of A-motif sequences in TIE a3 shows a requirement for an upstream A-motif for efficient inhibition.**

Two sets of mutations were performed on distinct A-motif sequences in TIE a3. The first mutation is $(AAAA/GGCC)_{107\text{-}111}$ and the second mutation $(AAAAA/GGCCC)_{147\text{-}151}$. The transcripts were *in vitro* translated in RRL with TIE a3 Wt and control (w/o TIE). Luciferase activities were normalized to control (w/o TIE). Experiments were performed in triplicates.

- **Figure 8: two distinct models for translational inhibition by TIE a3 and TIE a11.**

A model for TIE a3 suggests a ribosomal assembly on the $uAUG_{111\text{-}123}$ with a requirement of eIF2D initiation factor. The model for TIE a11 suggests a stalled 80S ribosome on AUG-stop codons combination upstream of a highly stable structure.

- **Supplemental figure S1: TIE-mediated inhibition recapitulated in different *in vitro* and *in vivo* systems.**

(A) (Upper pannel) The intensity of light emitted by Renilla Luciferase protein as a function of control mRNA (w/o TIE) concentration (nM). Emission of light was measured under sub-saturating conditions (50-100 nM mRNA concentration). **(Middle pannel)** The intensity of light emission by RLuc protein relative to different concentrations (50 nM and 100 nM respectively) of tested mRNA samples TIE a3, TIE a11 and control (w/o TIE). **(Lower pannel)** Analysis of translation products after *in vitro* translation in RRL on 10% SDS PAGE to monitor RLuc expression. Visualization of protein bands was achieved by incorporation of radiolabelled ^{35}S-methionine, which are detected by autoradiography. **(B)** *In vitro* translation of TIE a3 and TIE a11 transcripts in the presence of m^7G_{ppp} cap or a non-functional analog

A_{ppp}. Three *in vitro* systems were used: Wheat Germ Extract (WGE), drosophila embryonic cell extract (S2) and HeLa cell extract. All mRNAs were translated *in vitro* at 50 nM concentrations and RLuc expression was normalized to control (w/o TIE) in each condition. **(C)** Transfection of reporter plasmids with TIE a3 or TIE a11 in two embryonic cell lines, kidney HEK293FT and mesenchymal C3H10T1/2 cell lines. Renilla luciferase expression was normalized to the control (w/o TIE). Experiments were performed in triplicates **(D)** SDS PAGE of Renilla expression with 5' deletions in TIE a3 and TIE a11. The gels at the bottom refer to the minimal functional region compared to longest construct that is inactivated by 5' deletions.

- **Supplemental figure S2: average of reactivities of DMS and CMCT for TIE a3 and TIE a11.**

Probing experiments by DMS and CMCT were performed in triplicates. Figures S1.1 and S1 .2 represent averages of reactivities of DMS and CMCT for TIE a3 respectively. Figures S1.3 and S1.4 represent averages of reactivities of DMS and CMCT for TIE a11 respectively. A scale of 40 nucleotides range was included to determine position of every nucleotide. Standard deviations are also shown.

- **Supplemental figure S3A: AUG-like mutations of $uAUG_{111\text{-}113}$ are not effective for TIE a3-mediated inhibition.**

Representation of $uAUG_{111\text{-}1113}$ context in the secondary structure of TIE a3 is shown. A_{111} and U_{112} base pair with the A_{123} and U_{124} of the second $AUG_{123\text{-}125}$. Mutants of the uAUG into AUG-like (ACG, CUG and GUG) were constructed, with the corresponding compensatory mutations ACG and AGG. To the right, Rluc expression with different transcripts was normalized to control (w/o TIE). Experiments were performed in triplicates.

- **Supplemental figure S3B: sucrose gradient analysis of TIE a3 and mutant of upstream $(AUG/UAC)_{111\text{-}113}$. Alignment of TIE a3 sequences between different species.**

Ribosomal pre-initiation complexes were assembled and analysed on 7-47% sucrose gradient with radioactive m^7G-capped TIE a3 and the mutants of upstream $(AUG/UAC)_{111\text{-}113}$ in the absence (upper panel) or in the presence of GMP-PNP (lower panel). Heavy fractions correspond to polysomes while the lightest correspond to free RNPs.

- **Supplemental figure S3C: sequence alignments of hox a3 5'UTRs**

Nucleotides 1-170 from the mouse TIE a3 sequence were aligned with different species. The position of uAUG in *Mus musculus* is highlighted in a red box. Accession numbers: *F. catus*: XR_002740526.1, *M. musculus*: NM_010452.3, *M. flaviventris*: XM_0279, *H. sapiens*: XM_011515343.3, *M. mulatta*: XM_028845931.1, *P. Anubis*: XM_017956197.2, *E. caballu*s: XM_023639416.1, *U. horribilis*: XM_026508219.1, *C. dingo*: XM_025464436.1

- **Supplemental figure S4: translation of TIE a3 uORF in different constructs.**

(A) A single nucleotide deletion of A_{220} in TIE a3 is performed to frame-shift the uORF frame to the same as the main Rluc ORF. Another mutation was performed combining this deletion with uAUG mutation of $(U/C)_{112}$. Transcripts were *in vitro* translated in RRL and products were loaded on 10% SDS PAGE. **(B)** Translation of TIE a3 uORF in Hox a3 ORF. Hox a3 mRNA (5'UTR+CDS) was translated *in vitro* to check uORF translation in this context. Products were analysed on 15% SDS-PAGE.

- **Supplemental figure S5: transplanting TIE a11 stem loop structure in globin 5' UTR efficiently inhibits translation of RLuc.**

(A) The stem loop structure of TIE a11 (104-154) was transplanted in the 5'UTR of β-globin with 25 nucleotides spanning from each extremity. The three shown transcripts were *in vitro* translated in RRL at two concentrations, 25 nM and 50 nM. Results are represented in relative light unit of Rluc expression. **(B)** Nucleotides 1-216 from the mouse TIE a3 sequence were aligned with different species. The position of uAUG-UAG in *Mus Musculus* and the insertion of AUG-UAG in *Homo sapiens* are highlighted in red boxes. Accession numbers: *H. sapiens*: AF071164.1*, M. Mulata*: XM_015133828.2, *C. dingo*: XR_003143092.1, *C. griseus*: XM_027436289.1, *R. norvegicus*: XM_008762951.2, *E. caballus*: XM_023639423.1, *M auratus*: XM_00508585871.3, *M. musculus*: NM_010450.

- **Supplemental figure S6: sucrose gradient analysis of TIE a3 with different inhibitors shows an atypical profile with cycloheximide.**

(A) Ribosomal pre-initiation complexes were assembled and analysed on 7-47% sucrose gradient with radioactively capped TIE a3 in the presence of different inhibitors: cycloheximide, geneticin, GMP-PNP and Edeine. **(B)** Sucrose gradient analysis of control (w/o TIE) in the presence of cycloheximide.

Figure 1

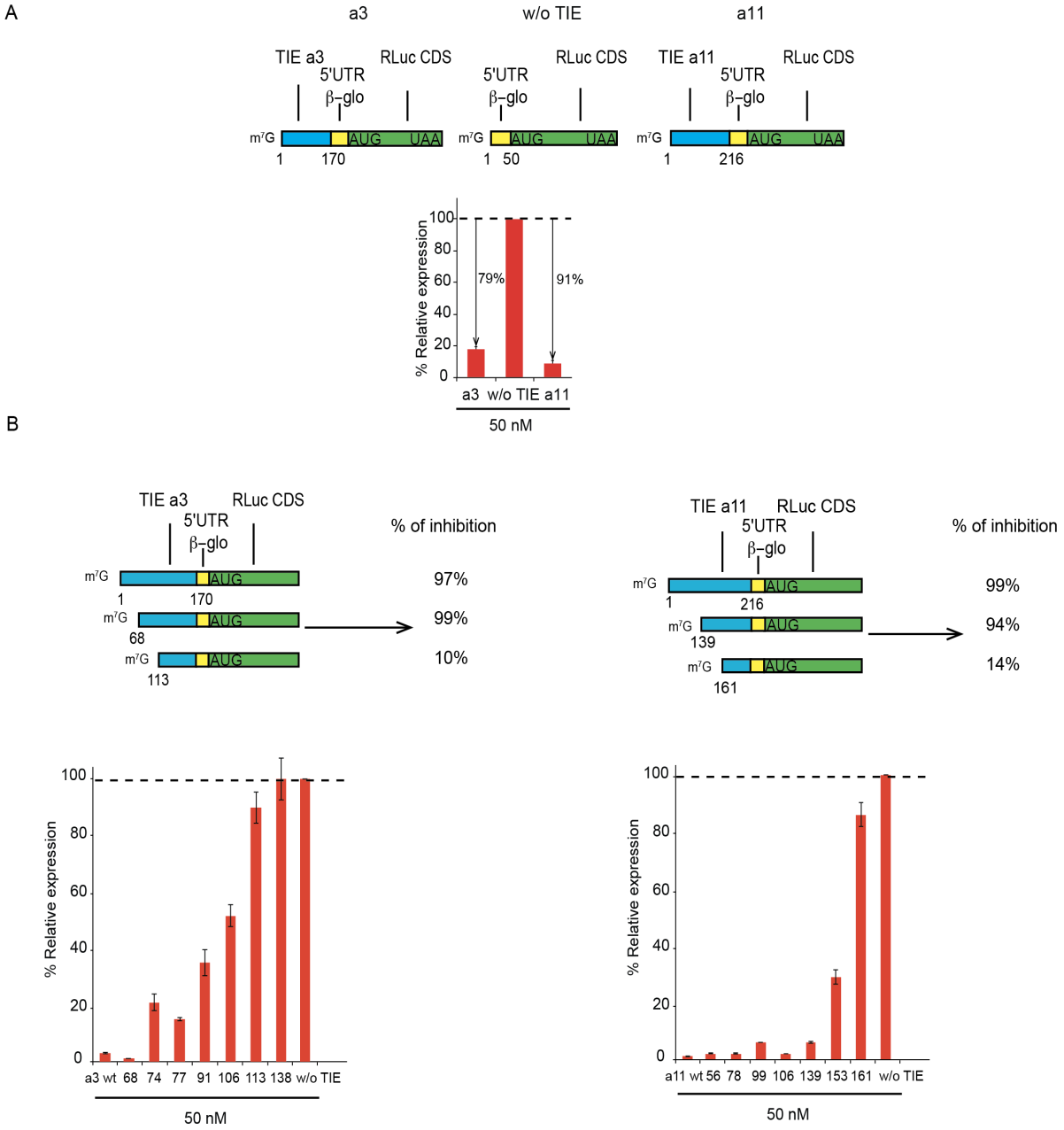

A
a3
w/o TIE
a11
TIE a3
RLuc CDS
5'UTR
β-glo
m7G
AUG
UAA
1
170
1
50
TIE a11
1
216
% Relative expression
100
80
60
40
20
0
79%
91%
a3 w/o TIE a11
50 nM
B
TIE a3
RLuc CDS
5'UTR
β-glo
% of inhibition
97%
99%
10%
68
113
TIE a11
139
161
94%
14%
a3 wt 68 74 77 91 106 113 138 w/o TIE
a11 wt 56 78 99 106 139 153 161 w/o TIE
50 nM

Figure 2

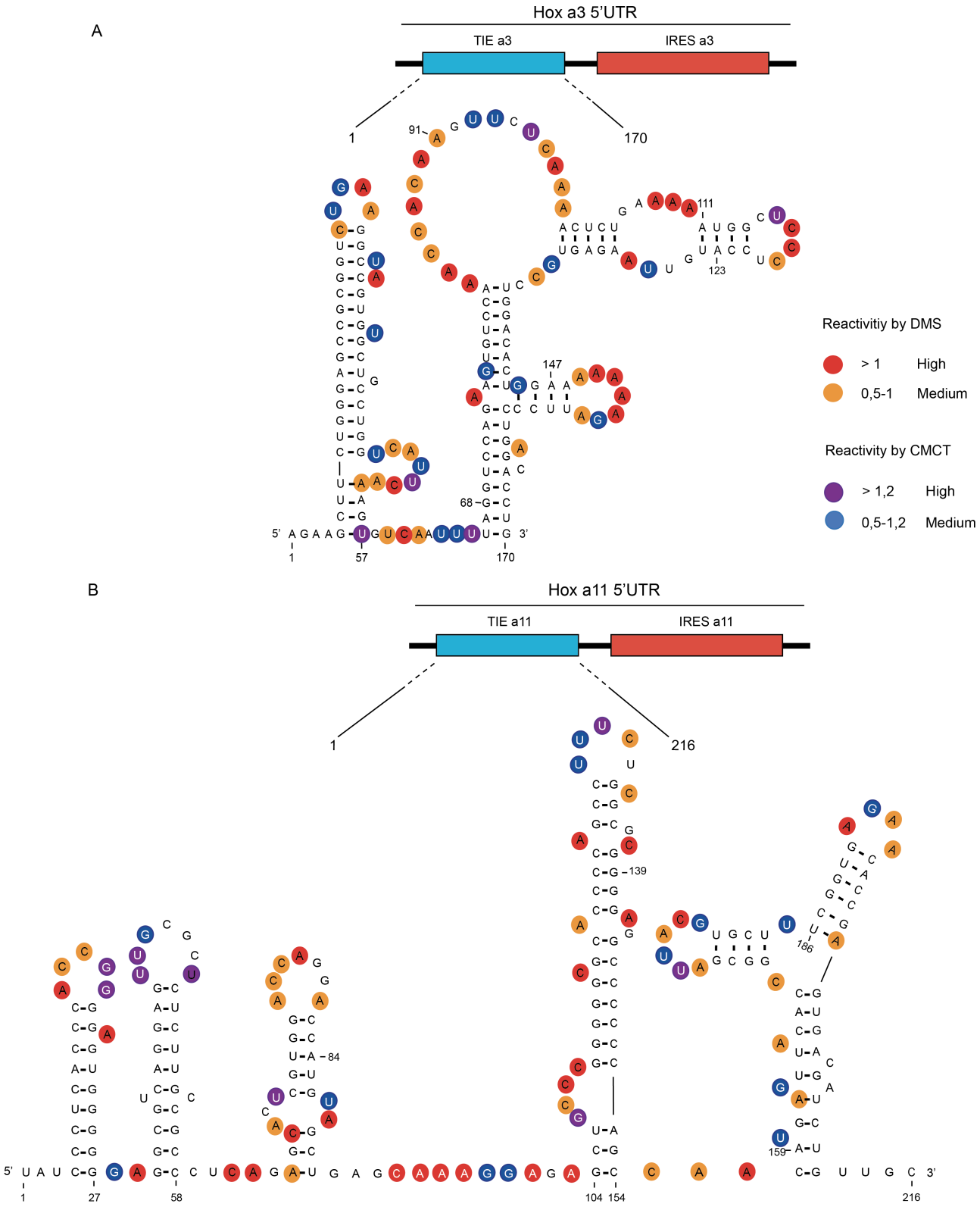

A
Hox a3 5'UTR
TIE a3
IRES a3
1
170
91
111
123
147
68
5' A G A A G U G U C A A U U U G 3'
1
57
170
Reactivitiy by DMS
> 1
High
0,5-1
Medium
Reactivity by CMCT
> 1,2
High
0,5-1,2
Medium
B
Hox a11 5'UTR
TIE a11
IRES a11
1
216
139
84
186
159
5' U A U C-G G A G-C C U C A G A-U G A G C A A A G G A G A G-C C A U U G C-G U U G C 3'
1
27
58
104 154
216

Figure 3

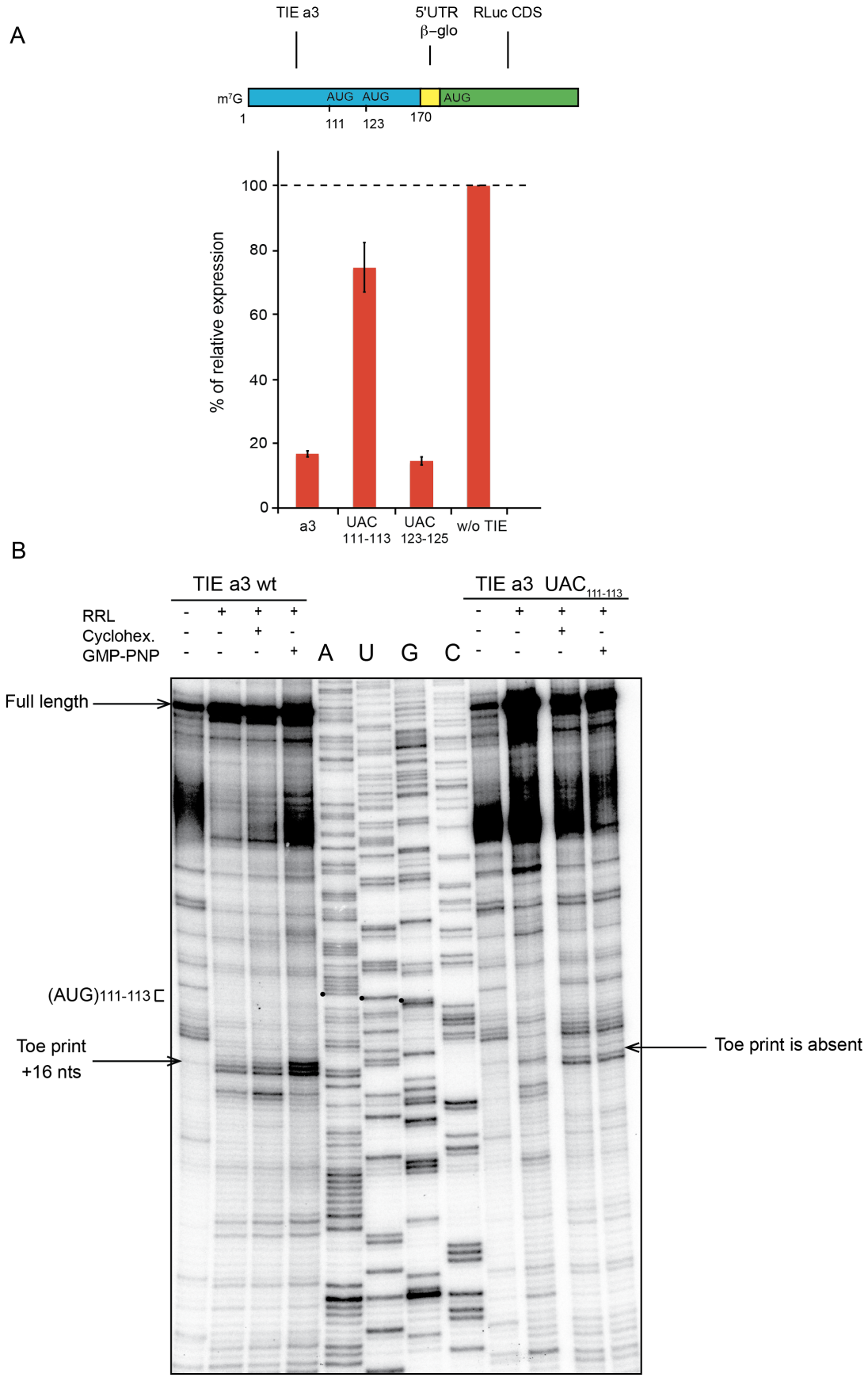
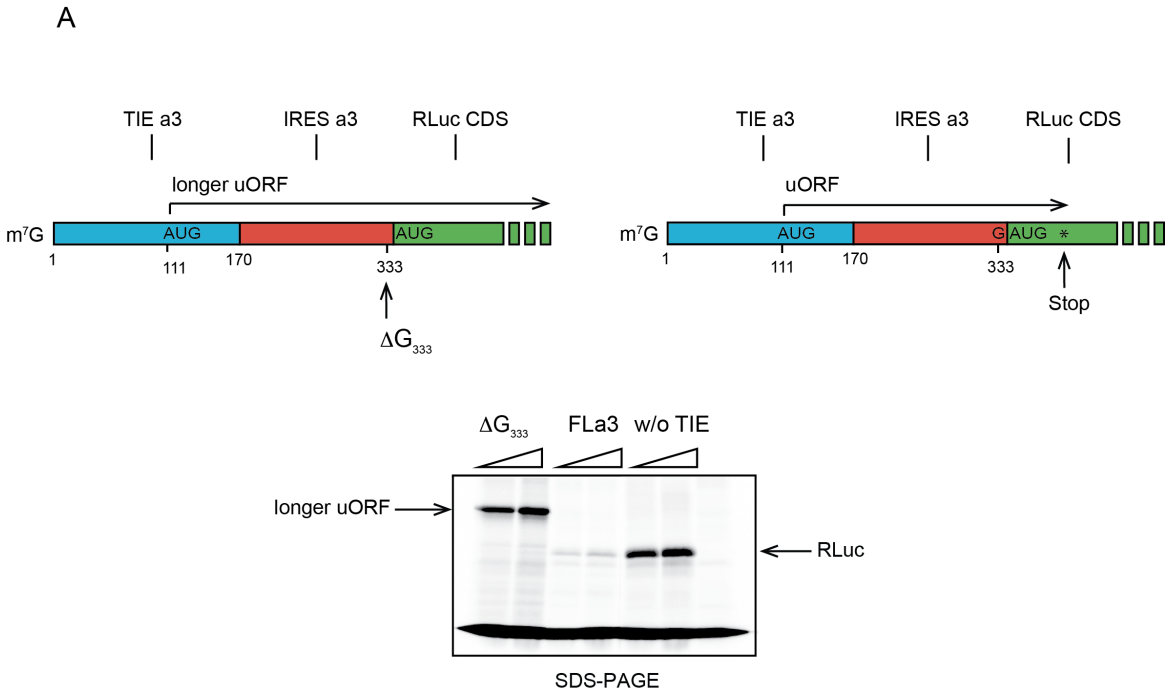

A
TIE a3
5'UTR
β–glo
RLuc CDS
m7G
AUG AUG
AUG
1
111
123
170
100
80
60
40
20
0
% of relative expression
a3
UAC 111-113
UAC 123-125
w/o TIE
B
TIE a3 wt
TIE a3 UAC111-113
RRL
Cyclohex.
GMP-PNP
A U G C
Full length
(AUG)111-113
Toe print +16 nts
Toe print is absent

Figure 4


A
TIE a3
IRES a3
RLuc CDS
longer uORF
m7G
AUG
AUG
1
111
170
333
ΔG333
TIE a3
IRES a3
RLuc CDS
uORF
m7G
AUG
G
AUG *
1
111
170
333
Stop
ΔG333
FLa3
w/o TIE
longer uORF
RLuc
SDS-PAGE

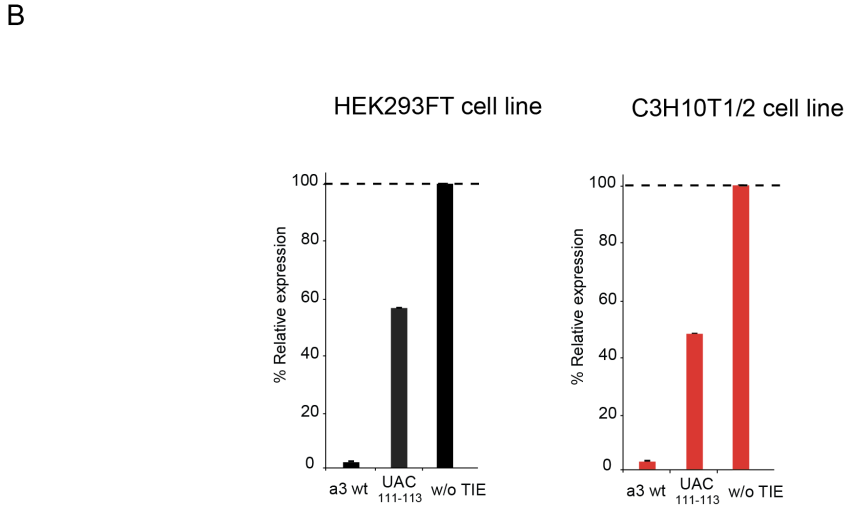

B
HEK293FT cell line
C3H10T1/2 cell line
% Relative expression
100
80
60
40
20
0
a3 wt
UAC
111-113
w/o TIE

Figure 5

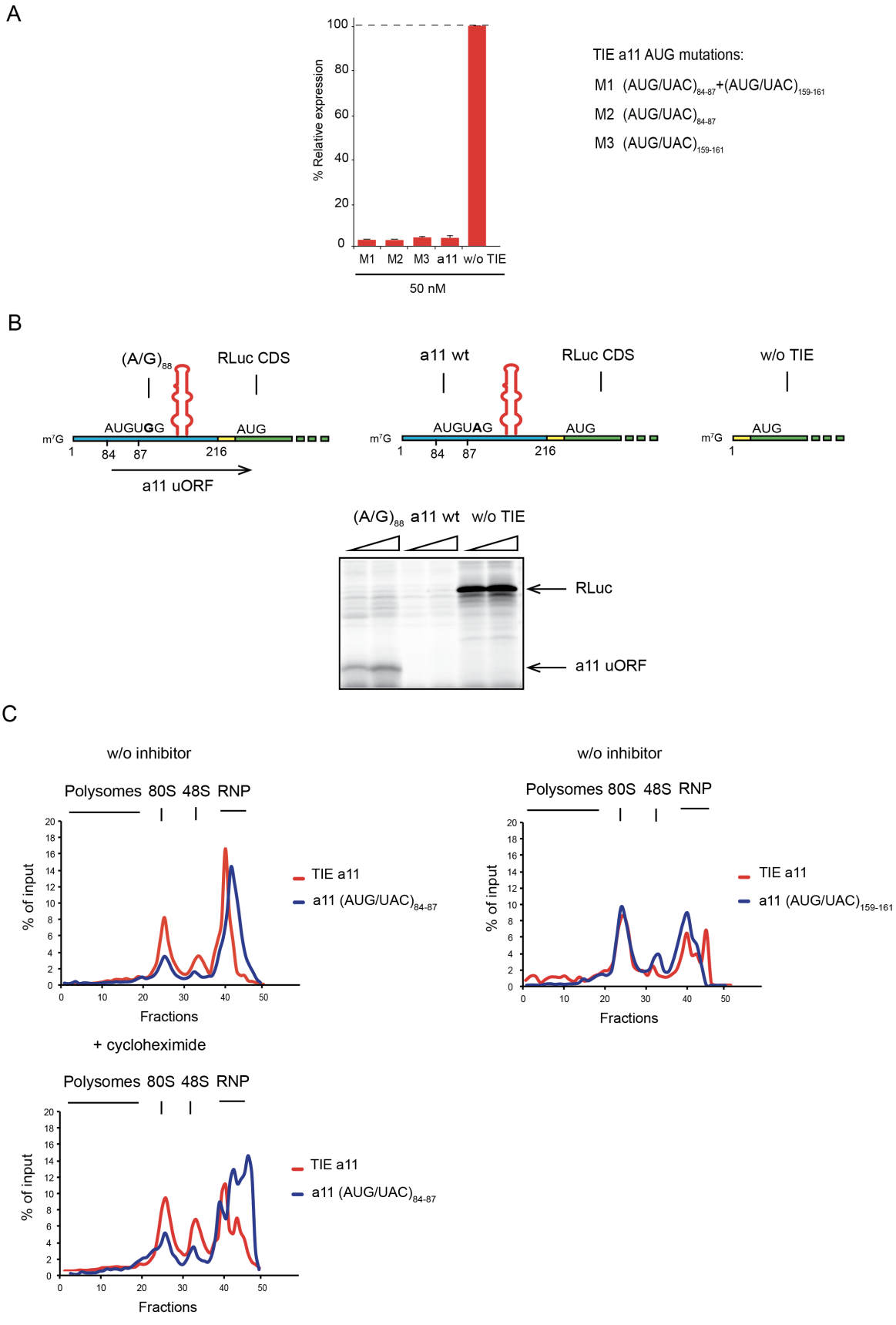
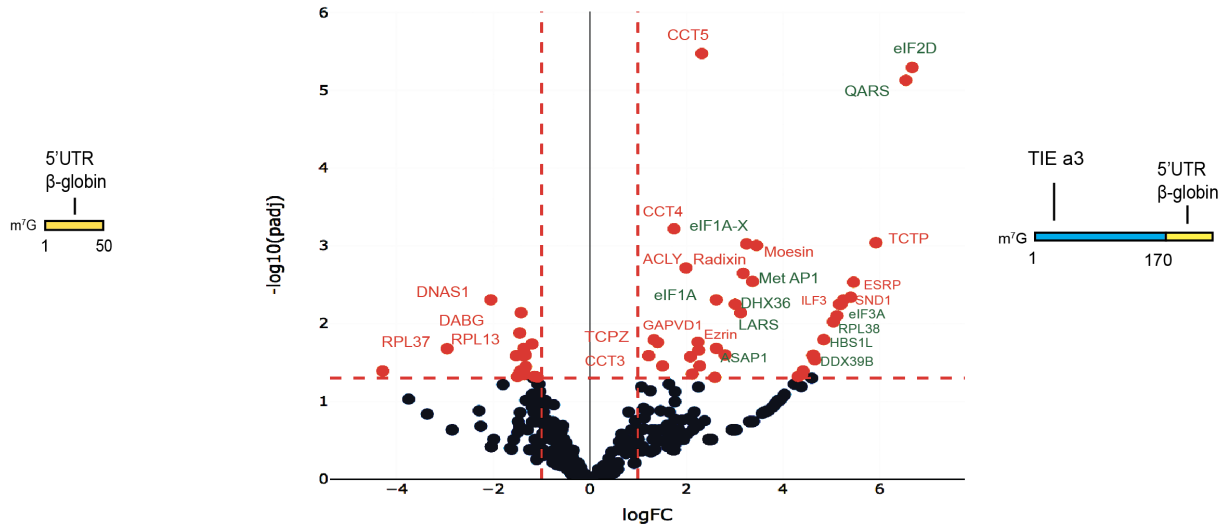
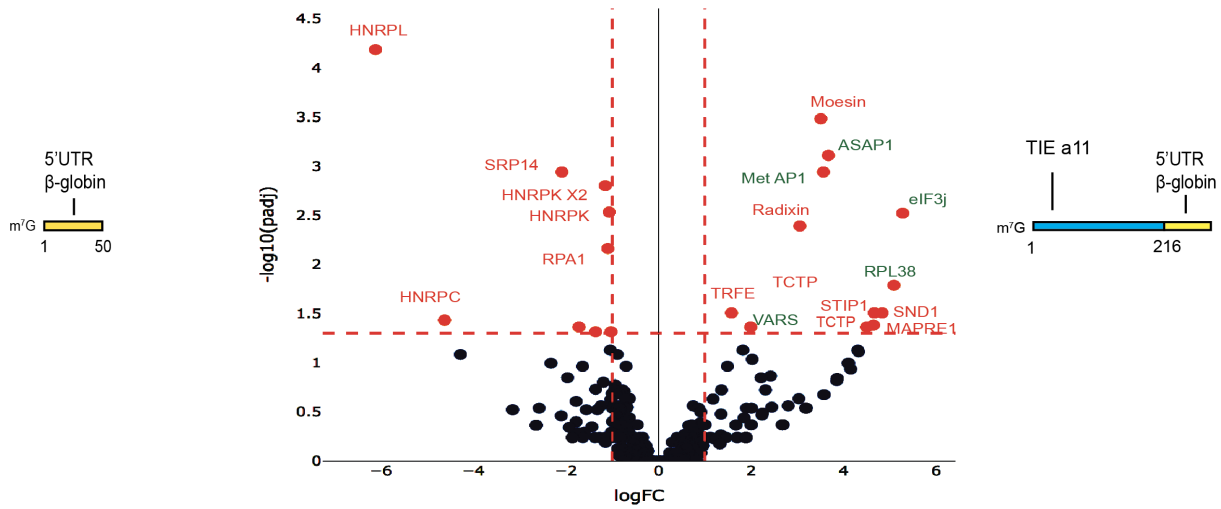

A
% Relative expression
100
80
60
40
20
0
M1 M2 M3 a11 w/o TIE
50 nM
TIE a11 AUG mutations:
M1 (AUG/UAC)84-87+(AUG/UAC)159-161
M2 (AUG/UAC)84-87
M3 (AUG/UAC)159-161
B
(A/G)88
RLuc CDS
m7G
AUGUGG
AUG
1
84
87
216
a11 uORF
a11 wt
RLuc CDS
AUGUAG
AUG
w/o TIE
AUG
(A/G)88 a11 wt w/o TIE
RLuc
a11 uORF
C
w/o inhibitor
Polysomes 80S 48S RNP
% of input
Fractions
TIE a11
a11 (AUG/UAC)84-87
+ cycloheximide
w/o inhibitor
Polysomes 80S 48S RNP
% of input
Fractions
TIE a11
a11 (AUG/UAC)159-161
Polysomes 80S 48S RNP
% of input
Fractions
TIE a11
a11 (AUG/UAC)84-87

Figure 6

Mass spectrometry analysis 5'UTR vs TIE a3



Mass spectrometry analysis 5'UTR vs TIE a11



Mass spectrometry analysis TIE a11 vs TIE a3

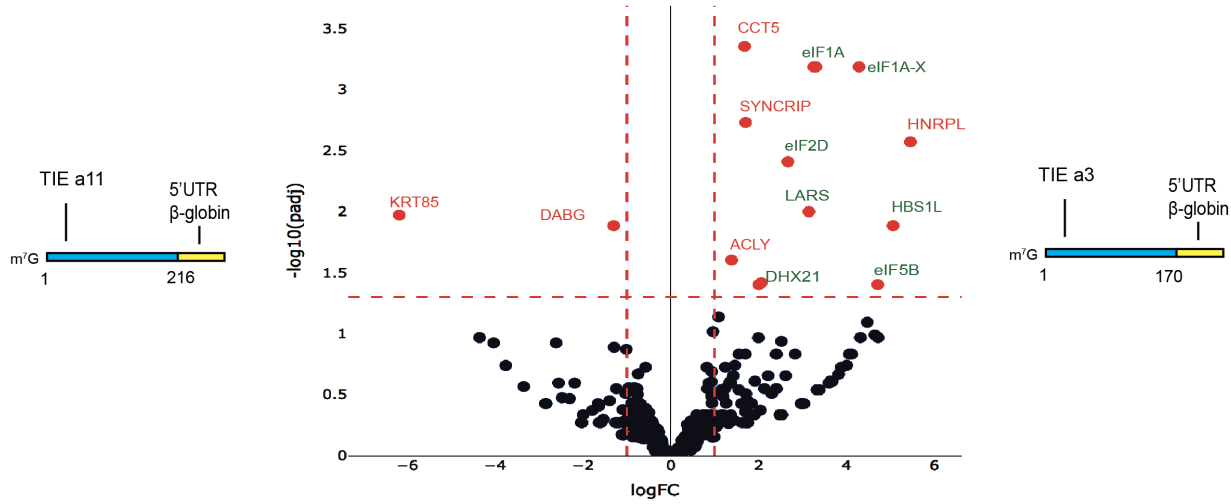


Figure 7

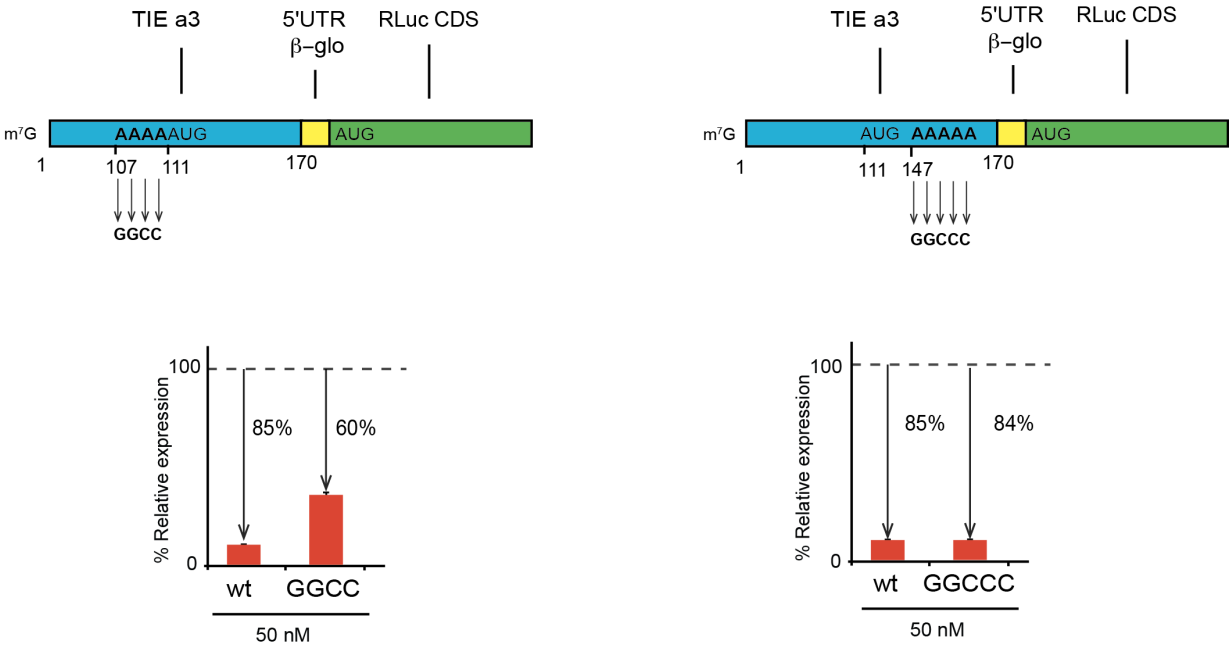
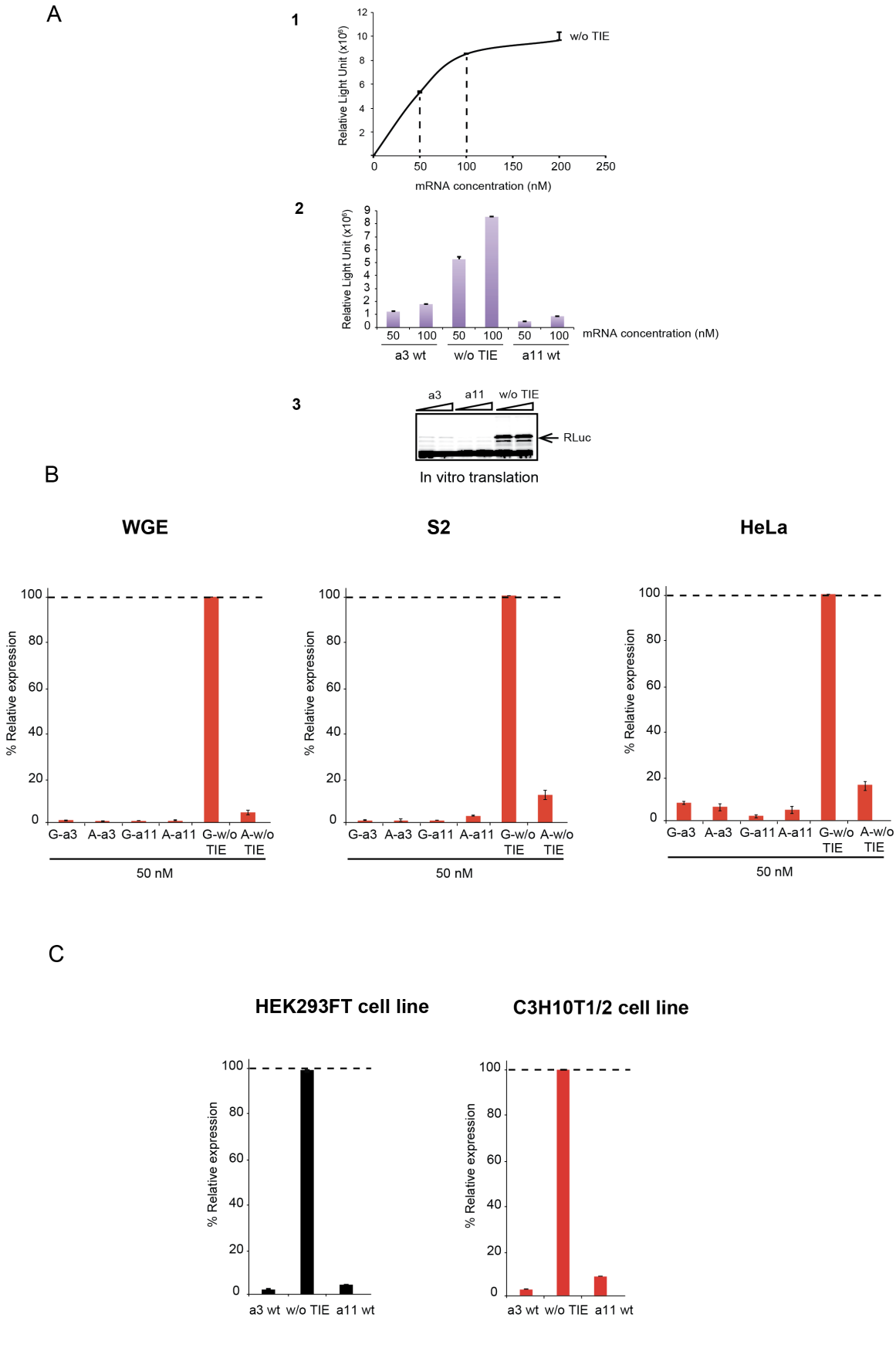

TIE a3
5'UTR β–glo
RLuc CDS
m7G
AAAAAUG
AUG
1
107
111
170
GGCC
% Relative expression
100
0
85%
60%
wt
GGCC
50 nM
TIE a3
5'UTR β–glo
RLuc CDS
m7G
AUG AAAAA
AUG
1
111
147
170
GGCCC
% Relative expression
100
0
85%
84%
wt
GGCCC
50 nM

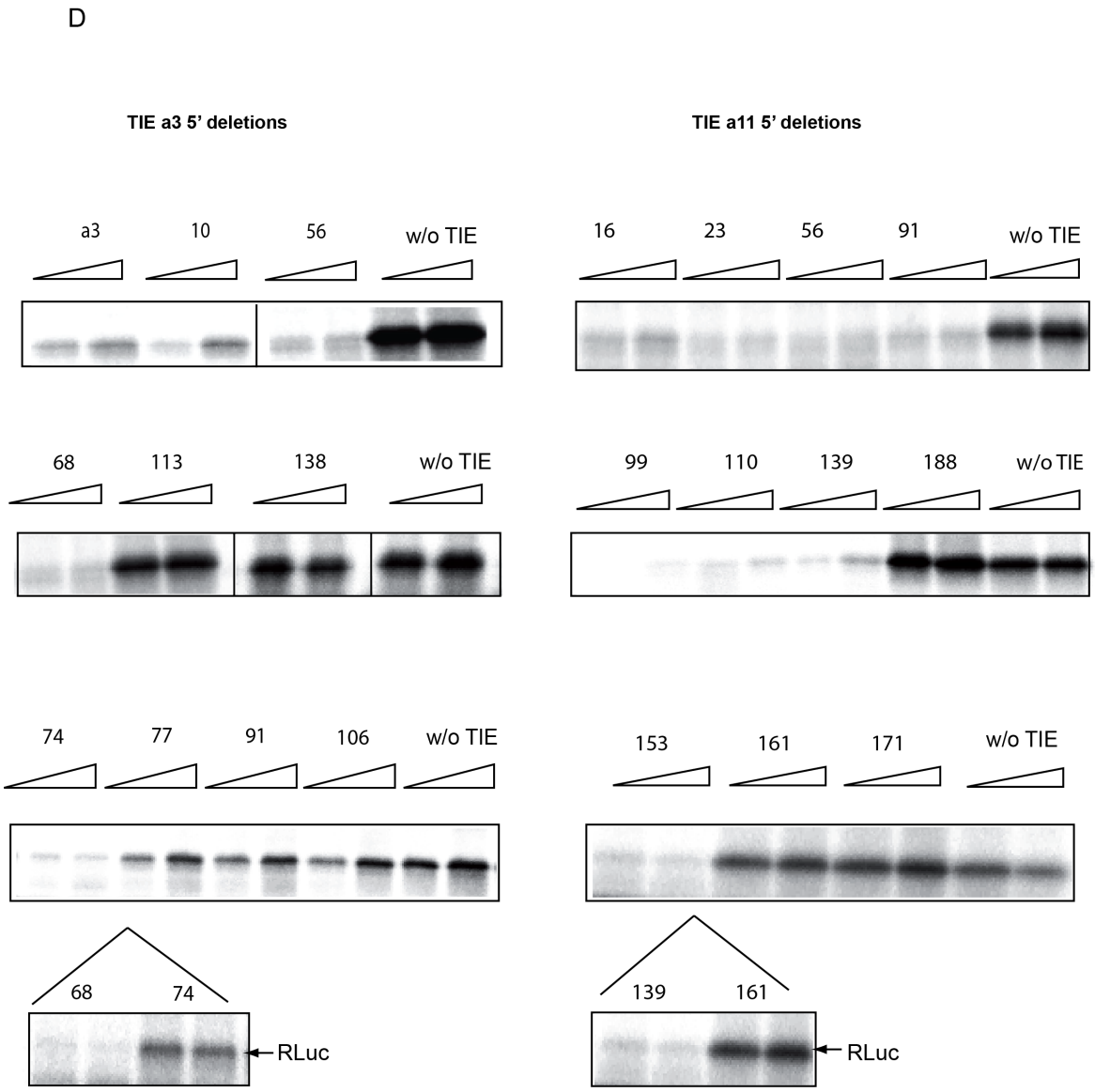
Figure 8


TIE a3
60S
E P A
eIF2D
AUG
1
170
40S
TIE a11
SL
60S
AUG UAG
AUG
1
216
40S
Start-Stop

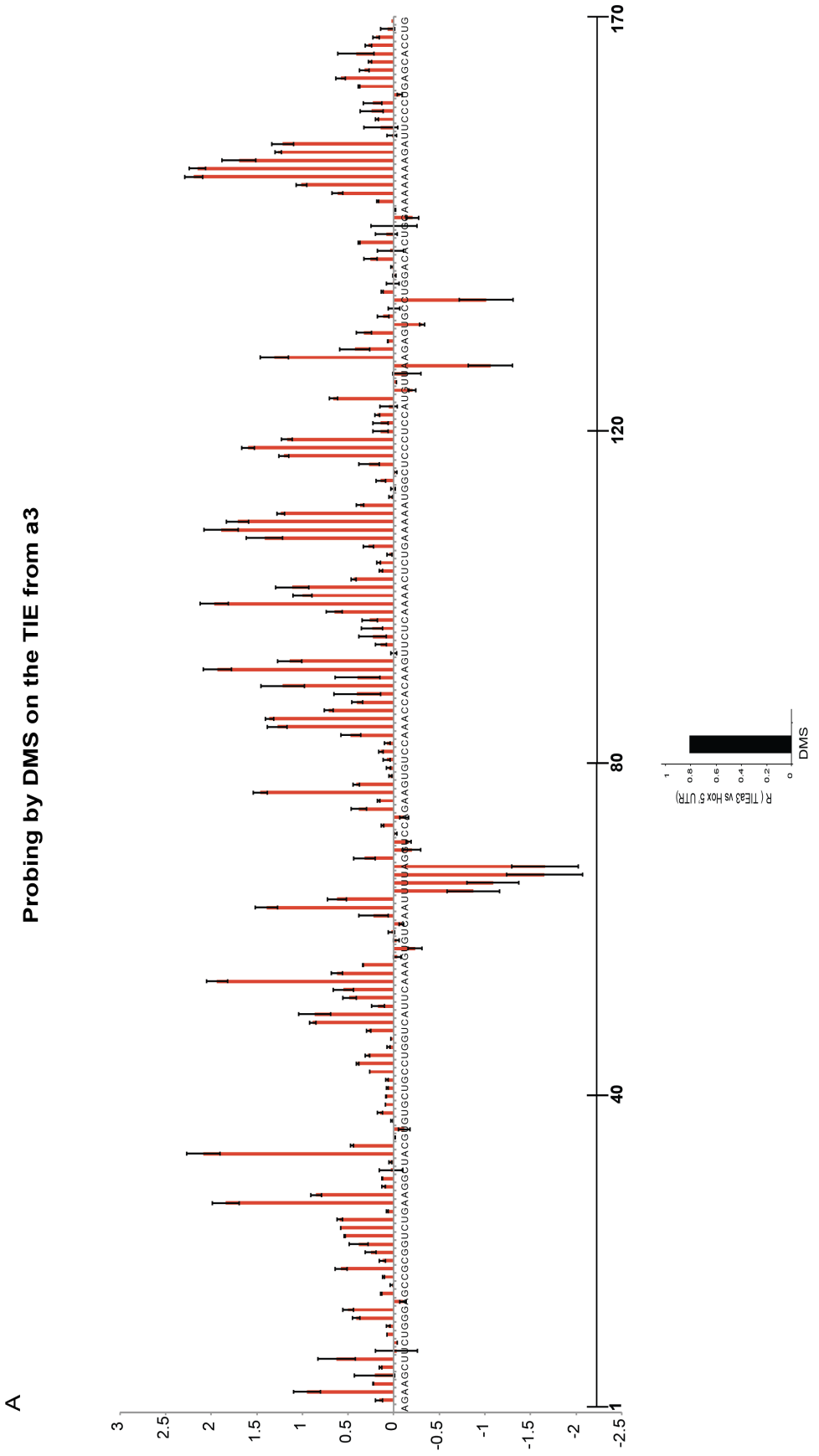
Supplementary figure S1



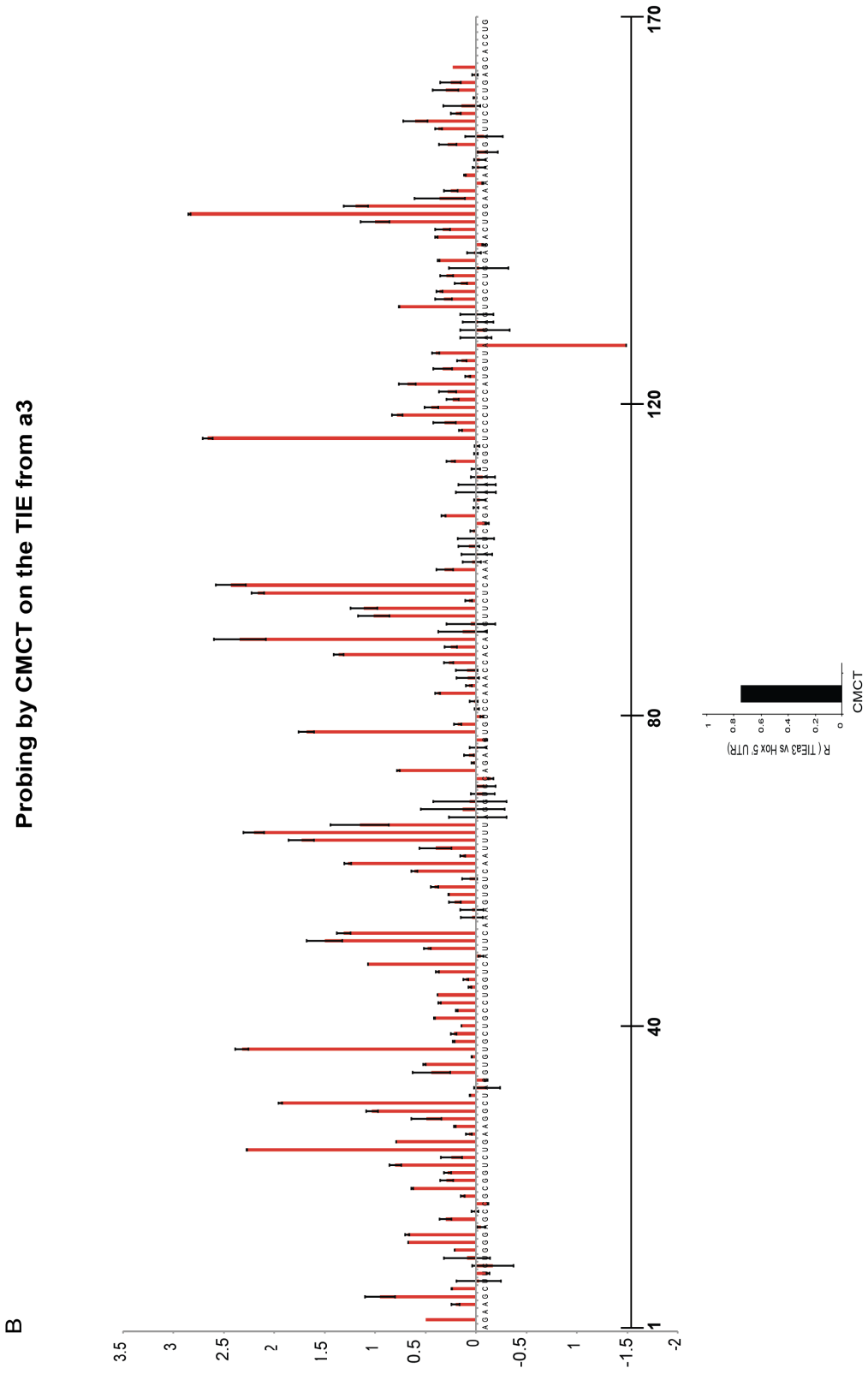
Supplementary figure S1



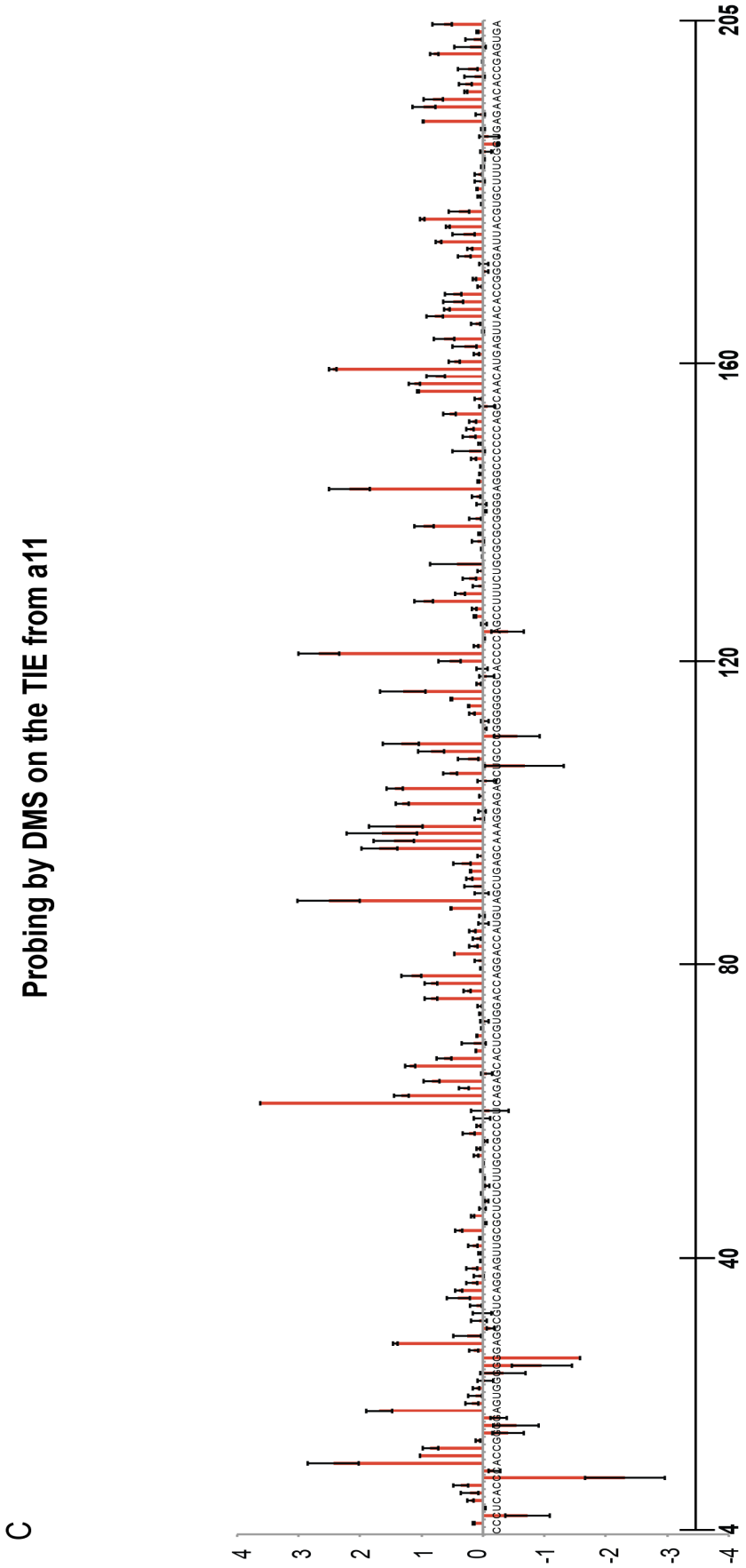
Supplementary figure S2A



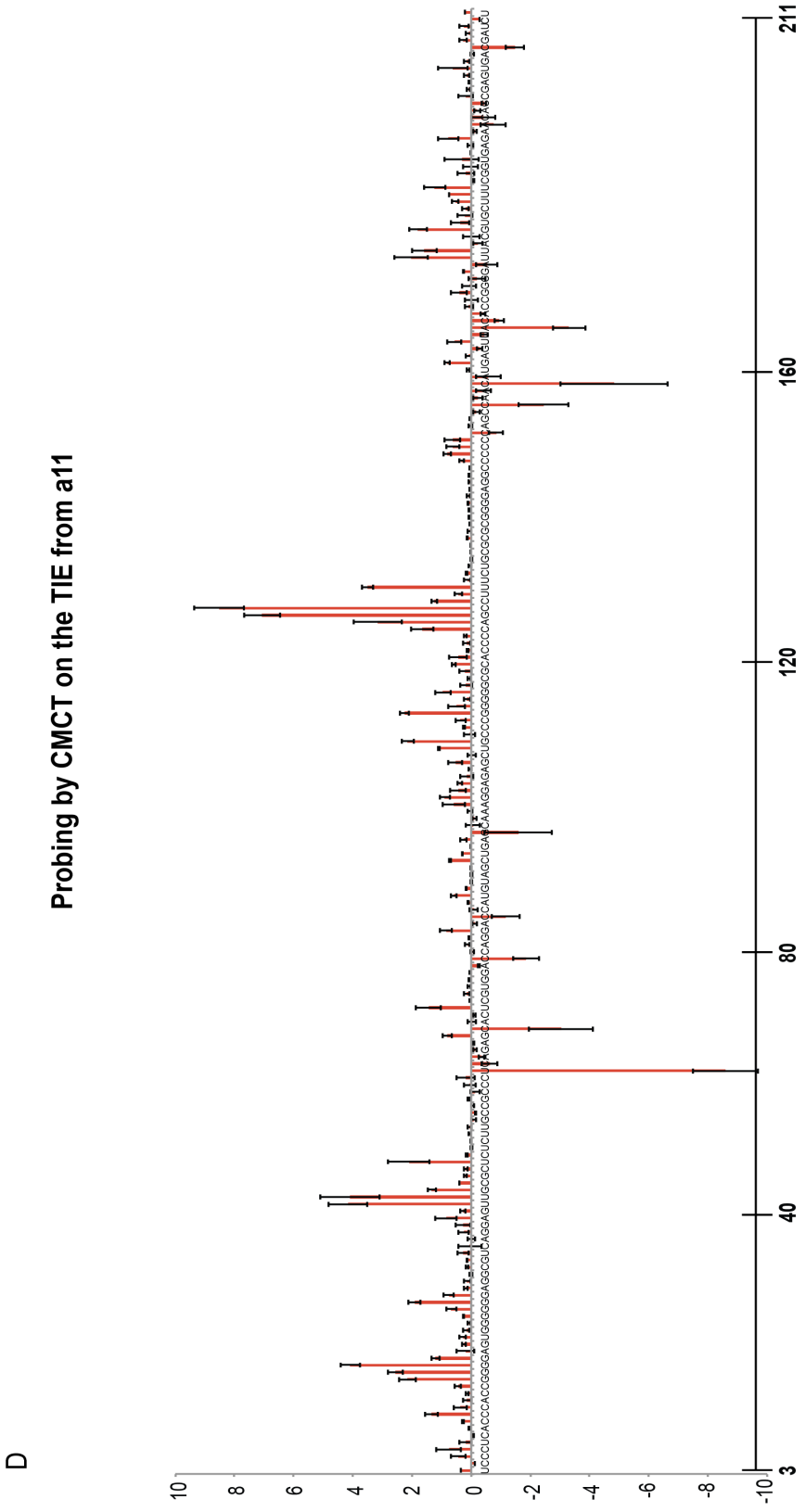
Supplementary figure S2B



Supplementary figure S2C

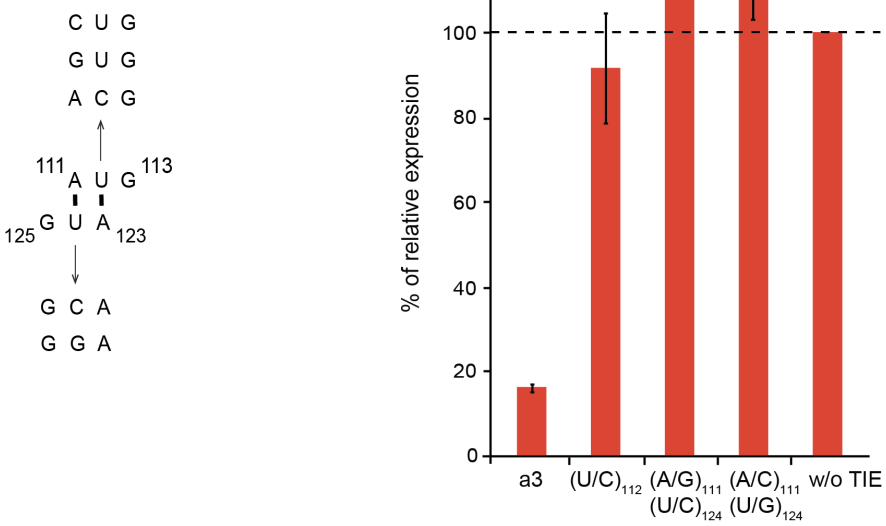


Supplementary figure S2D



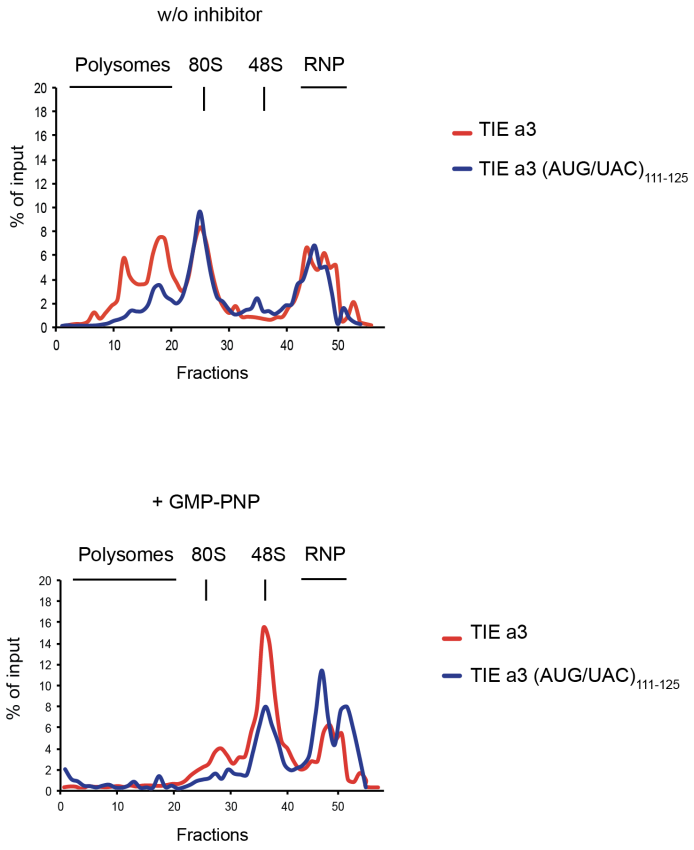
Supplementary figure S3

A

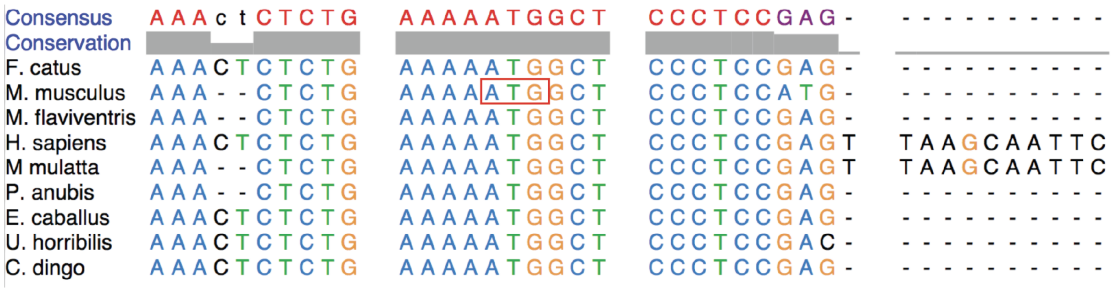


Supplementary figure S3

B

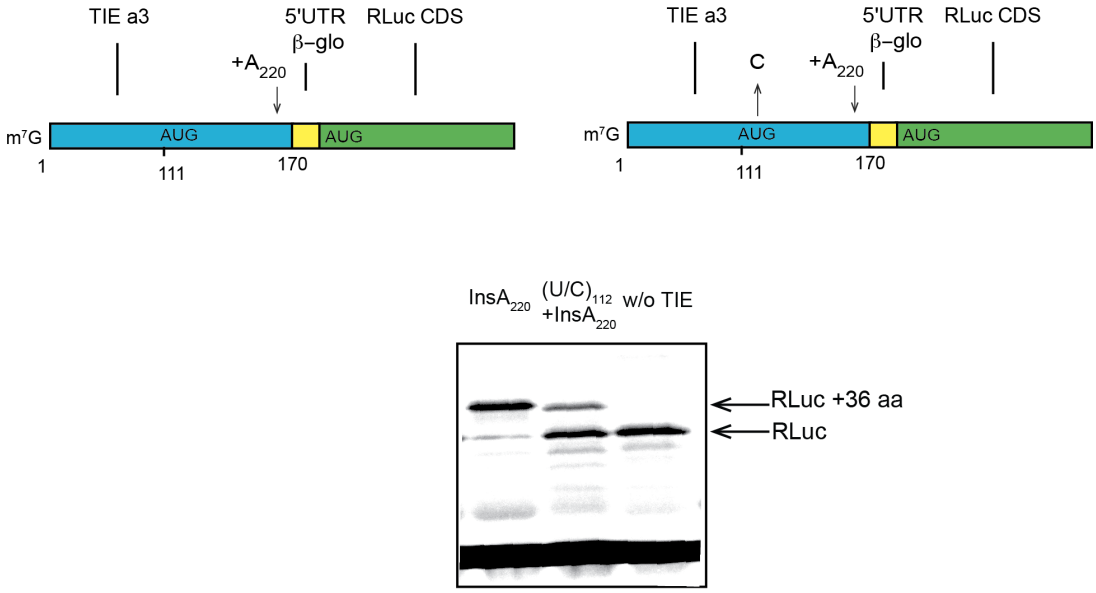


C

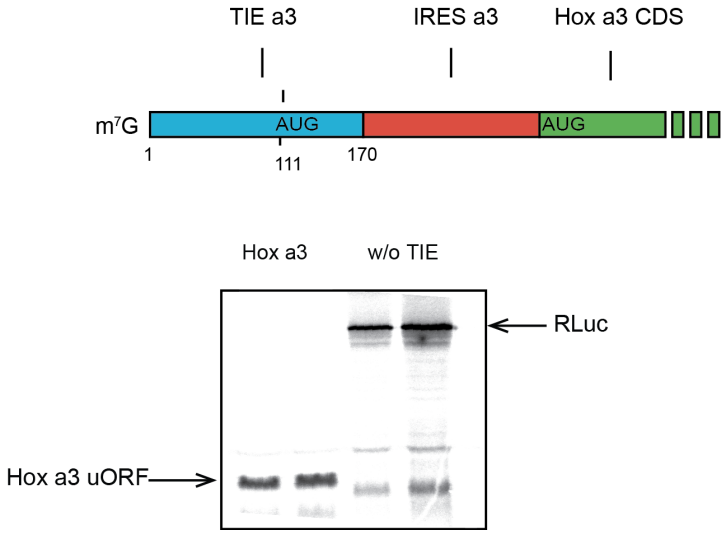


Supplementary figure S4

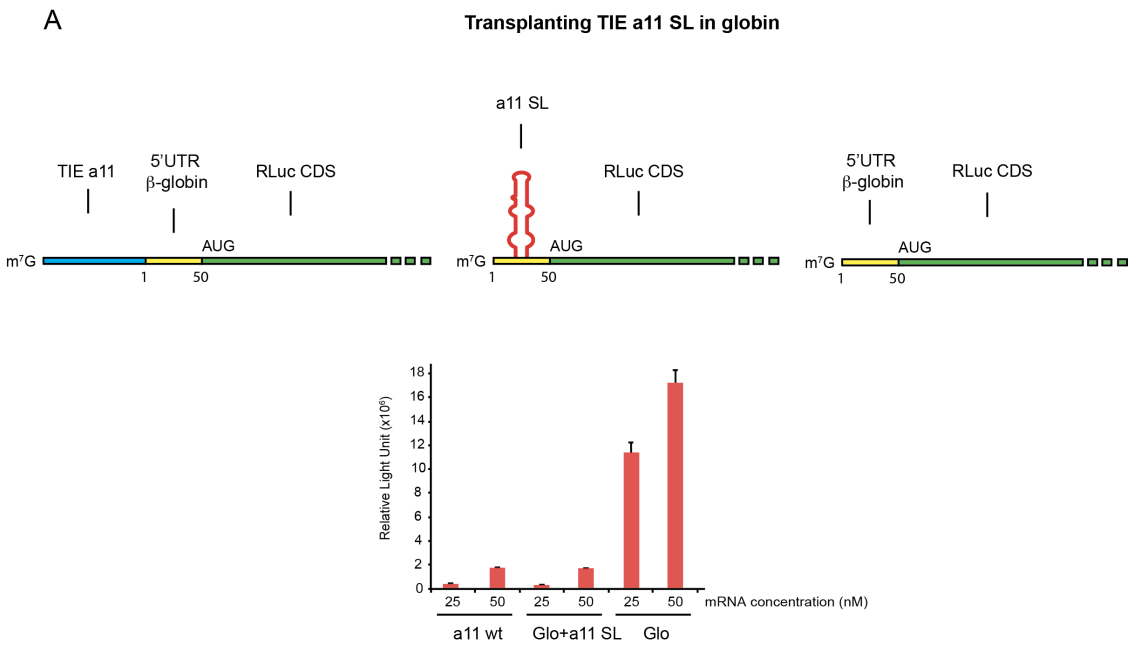
A



B

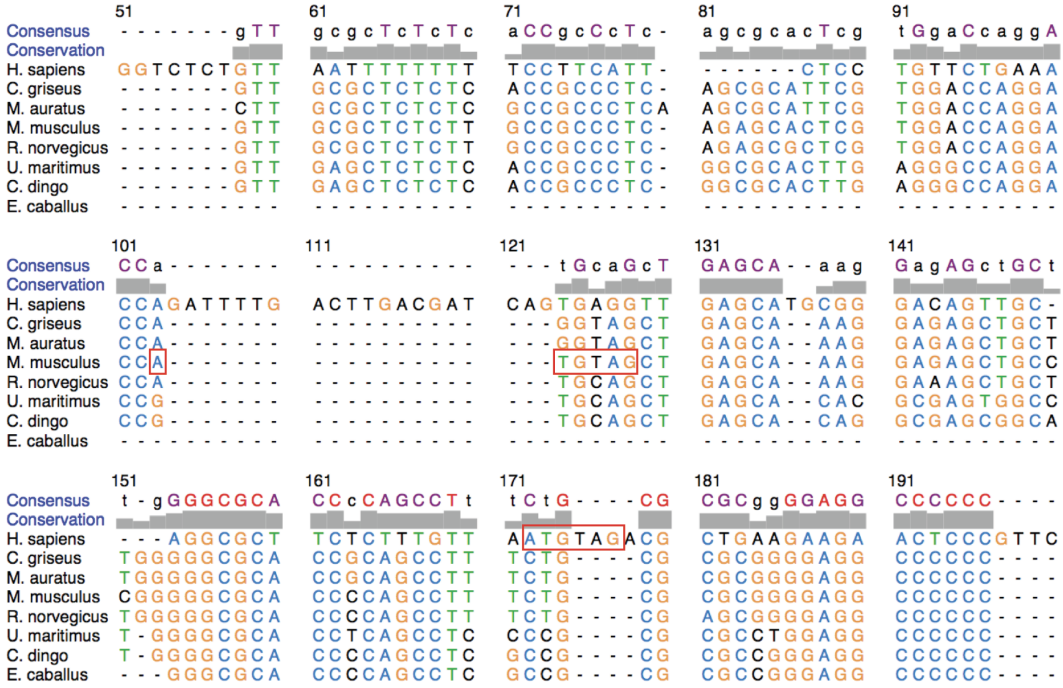


Supplementary figure S5

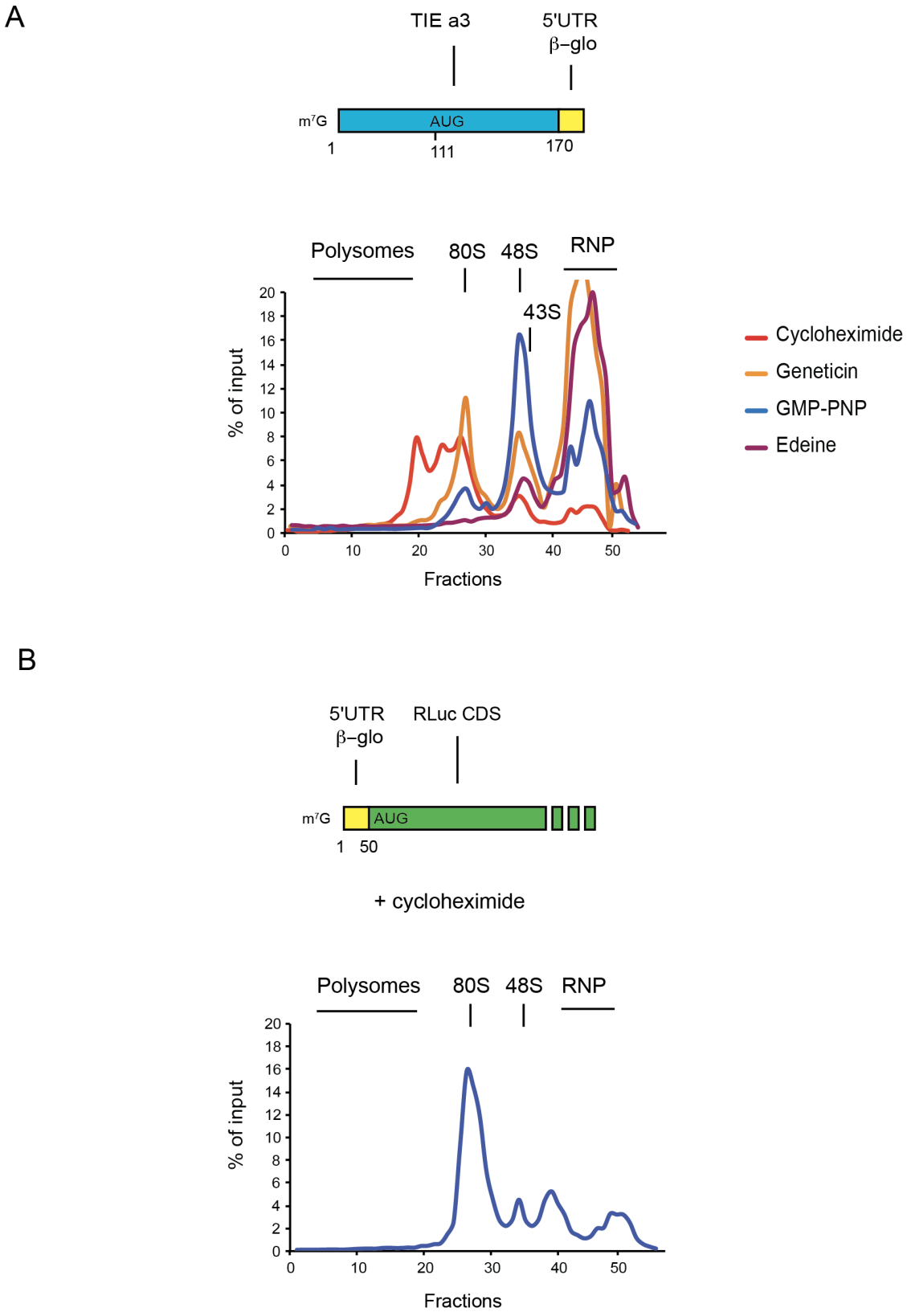


B

Conservation of uAUG-stop combination in TIE a3 among different species



Supplementary figure S6



3. Complementary result of TIE a11 study

Three-way junction in TIE a11 is not required for inhibition

In addition to the previously shown results, we performed series of complementary experiments on TIE elements. We focused our study on the full-length sequences and performed several mutations in TIE a3 and a11 to modify their structures. For TIE a11, the three-way junction (TWJ) was our main focus in the beginning of this study since it harbours the minimal TIE a11. Using site-directed mutagenesis, we introduced several mutations (**Fig. 42**). As previously explained (results, chapter 1), the upstream $(AUG/UAC)_{159\text{-}161}$ mutation had no effect on TIE a11 inhibitory function. We therefore speculated the involvement of other AUG-like codons that exist in the TWJ. As shown in figure 42, mutations of the AUG-like codons $(AUU/UAA)_{175\text{-}177}$, $(GUG/CAC)_{189\text{-}191}$, and $(GUG/CAC)_{202\text{-}204}$ have no effect on translation inhibition. The next step was to look into other structural motifs in the TWJ. We introduced mutations that affect the folding of stems or modify the sequences in the loops. Mutations of $(CAC/GUG)_{167\text{-}169}$, $(GGCG/UCGU)_{171\text{-}174}$, $(UGCU/GCGG)_{181\text{-}184}$, $(CGG/GCC)_{187\text{-}189}$, $(AGAA/UCUU)_{192\text{-}195}$ and $(CCG/GGC)_{198\text{-}200}$ had also no effect on TIE a11-mediated translation inhibition (**Fig. 42**). These results eliminated the possibility of TWJ requirement for TIE a11 inhibitory mechanism. It also raised the question of whether minimal TIE a11 indeed encompasses the main inhibitory element in TIE a11. These results were informative since they indicated that we should focus on other sequence motifs and structural elements located elsewhere in TIE a11. As previously explained, this shed the light on the role of start-stop combination and the GC-rich SL structural element in the mechanism of inhibition.

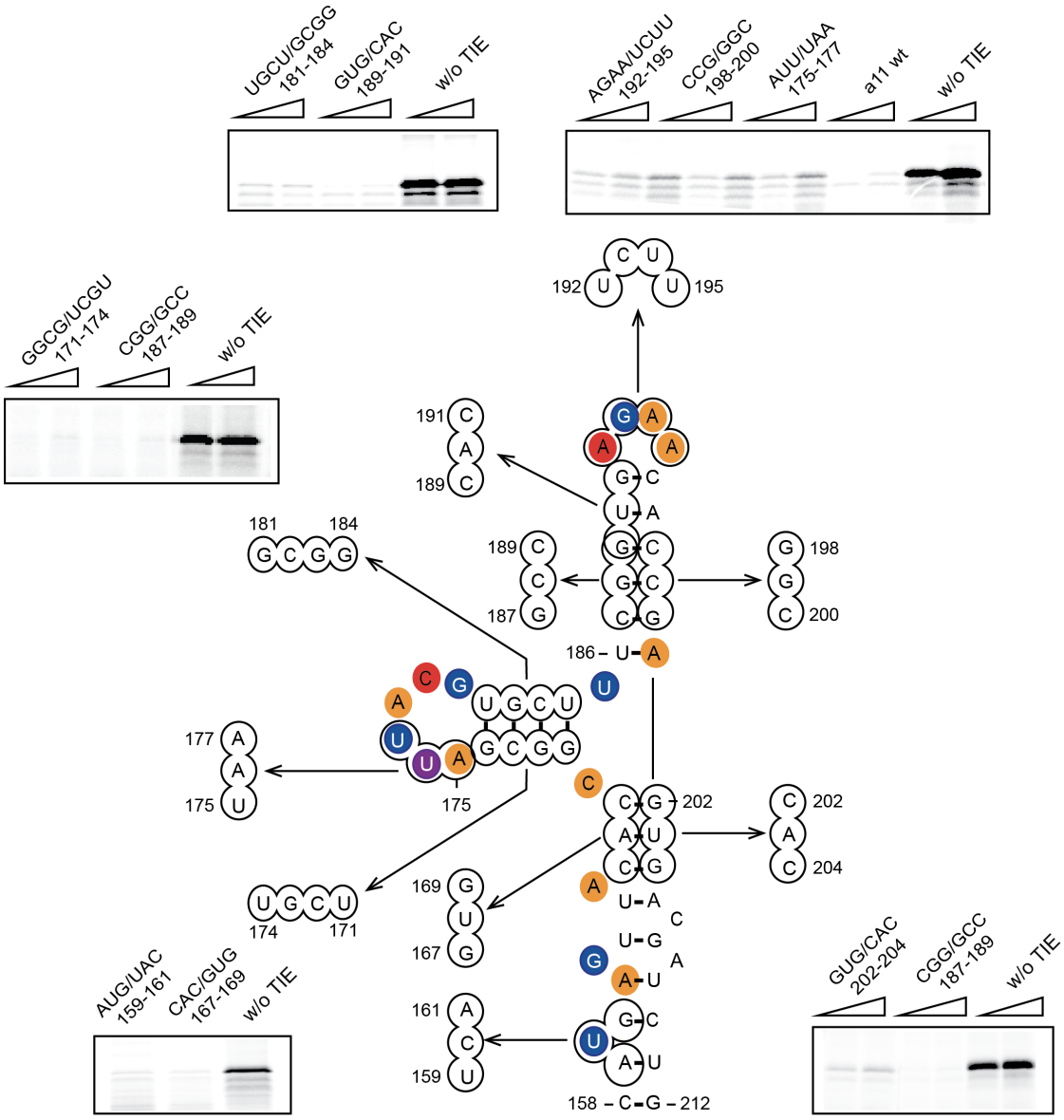


Figure 42: **Mutations in a11 TWJ show no effect on TIE-mediated inhibition.** Different mutations were performed in the stem and loop structures of TIE a11 TWJ. These constructs were obtained by site-directed mutagenesis and used for subsequent *in vitro* translation assay in RRL Translation products were loads on 10% SDS PAGE for analysis of RLuc protein expression.

Chapter 2: Study of Hox IRESes

In accordance with what was previously published (Xue et al., 2015), we conducted few additional experiments on Hox IRESes. Although, our main focus in that project was the TIE element, we were also interested in knowing Hox TIEs and IRESes act in synergy for Hox translation regulation. As previously explained (Introduction, section 4.7), cellular IRESes are poorly characterized due to the lack of sequence and structural similarities. According to experiments performed by Xue et al., 2015, IRES a3 is 163 nucleotide-long and IRES a11 is 280 nucleotide-long. We first performed structural studies on these two Hox IRESes by chemical probing and then functional studies using cell-free translation extracts.

1. Structural characterization of IRES a3

We performed chemical probing with DMS and CMCT using the same protocol as for isolated TIEs. We were first interested in determining whether TIE and IRES might interact with each other or whether they can fold independently. For that purpose, we performed probing of two RNA transcripts: first, the full-length 5'UTR Hoxa3 harbouring TIE and IRES and second a transcript containing only the IRES without TIE. We performed our probing experiments in triplicates and calculated the average of reactivities of each nucleotide (**Fig. 43**). We calculated the Pearson's correlation coefficient (R) to assess the folding of the IRES element in the presence of the TIE and when the IRES is alone. Our probing datasets indicated a correlation of R=0.8 for DMS and R=0.75 for CMCT suggesting a strong correlation of the nucleotide reactivities in both IRES transcripts. Therefore, we concluded that the two RNA regulons fold independently. The secondary structural model of IRES a3 that we established contains five stem-loop structures with base pairing of region 25-38 with 130-144 (**Fig. 44**). Altogether, this suggests that the Hox a3 TIE and IRES indeed fold as independent modules. For IRES a11, we performed similar probing experiments, however the interpretation of the probing datasets into a 2D structural model has not yet been completed in the time frame of my thesis. This will be accomplished in the next weeks.

Figure 43A

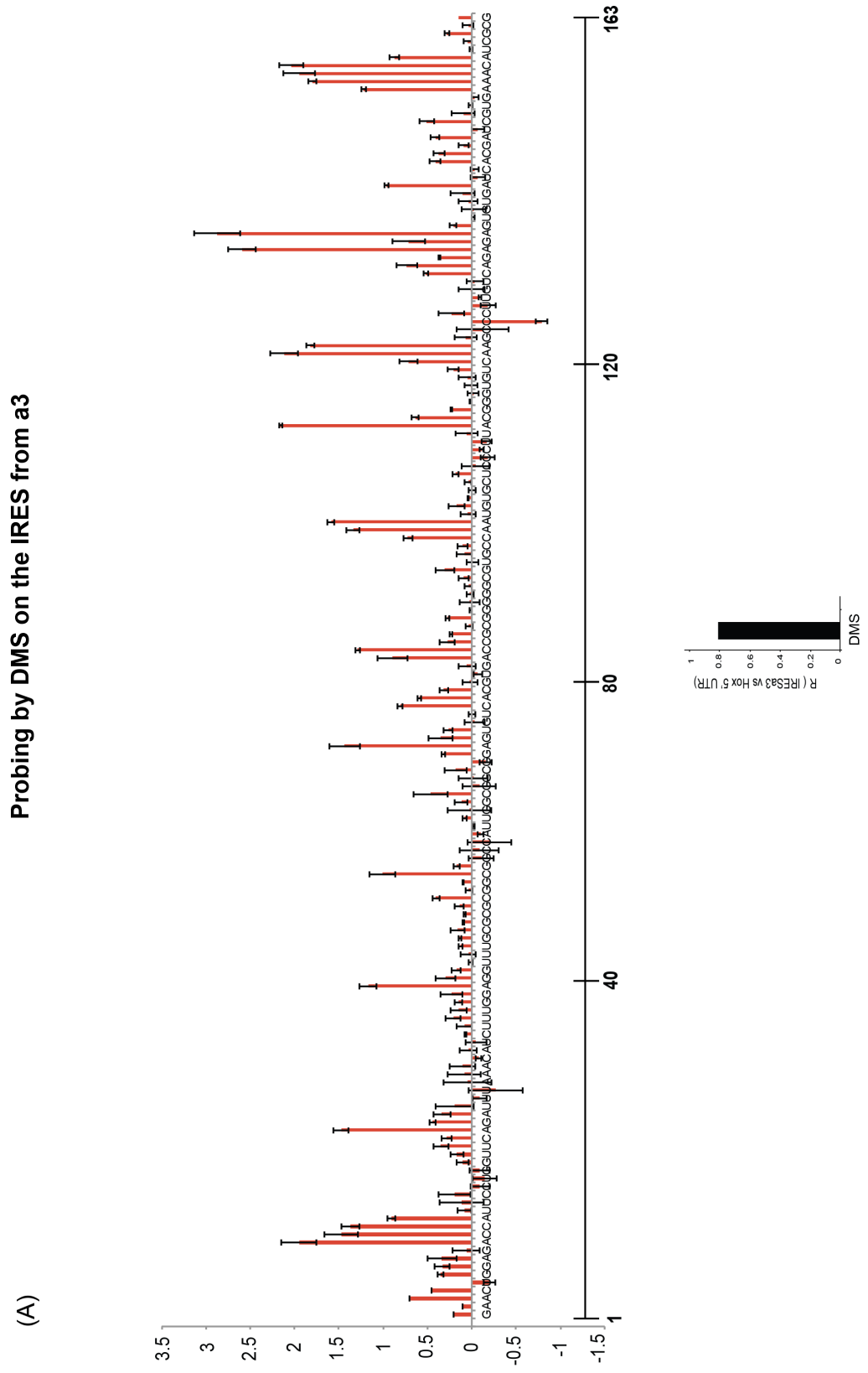

(A)
Probing by DMS on the IRES from a3
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
1
40
80
120
163
R (IRESa3 vs Hox 5' UTR)
DMS

Figure 43B

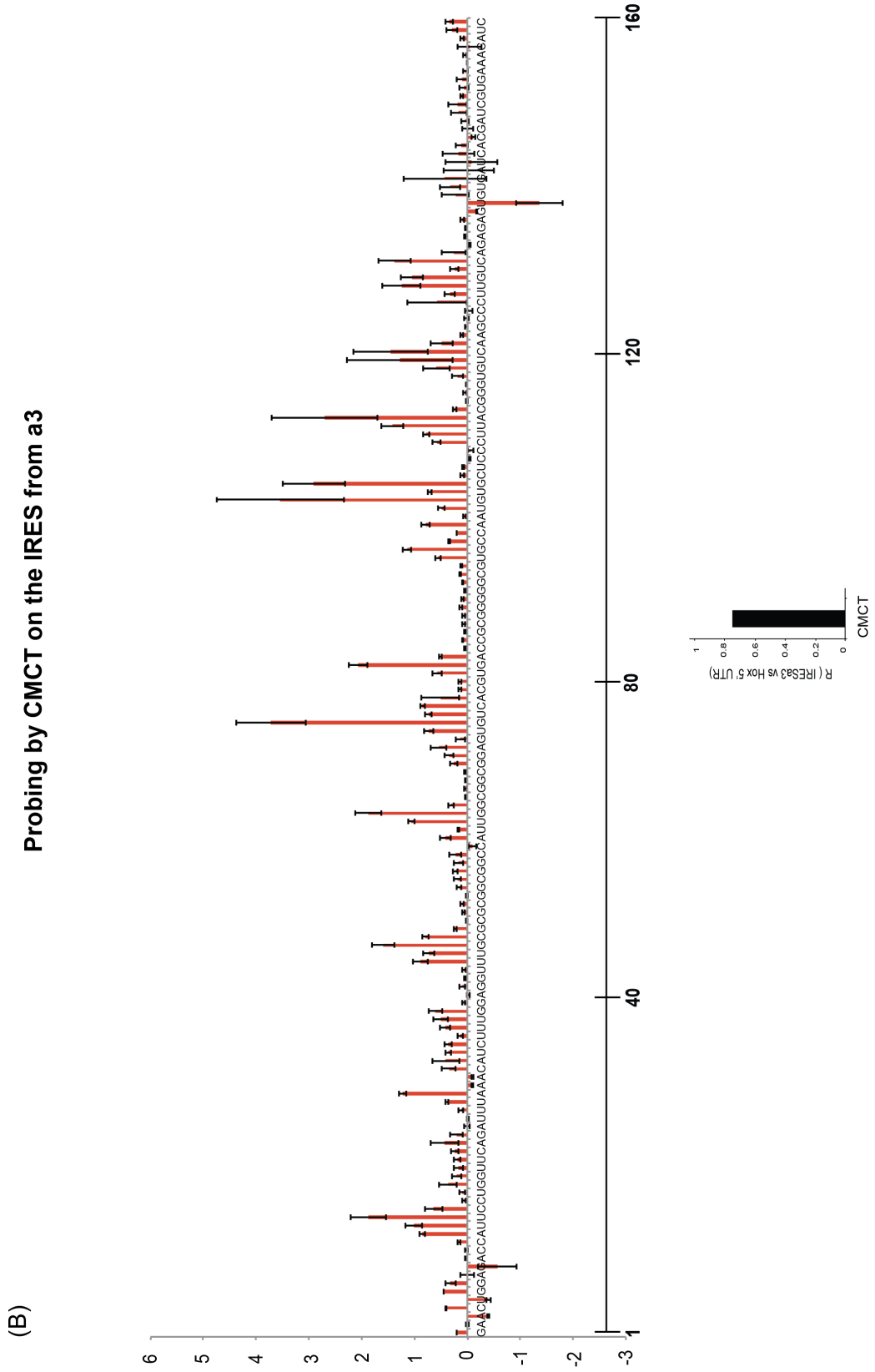


Figure 43: **Average of (A) DMS and (B) CMCT reactivities for IRES a3.** A scale of 40 nucleotides range was included to determine position of every nucleotide. Standard deviations and Pearson's correlation coefficient (R) are also shown.

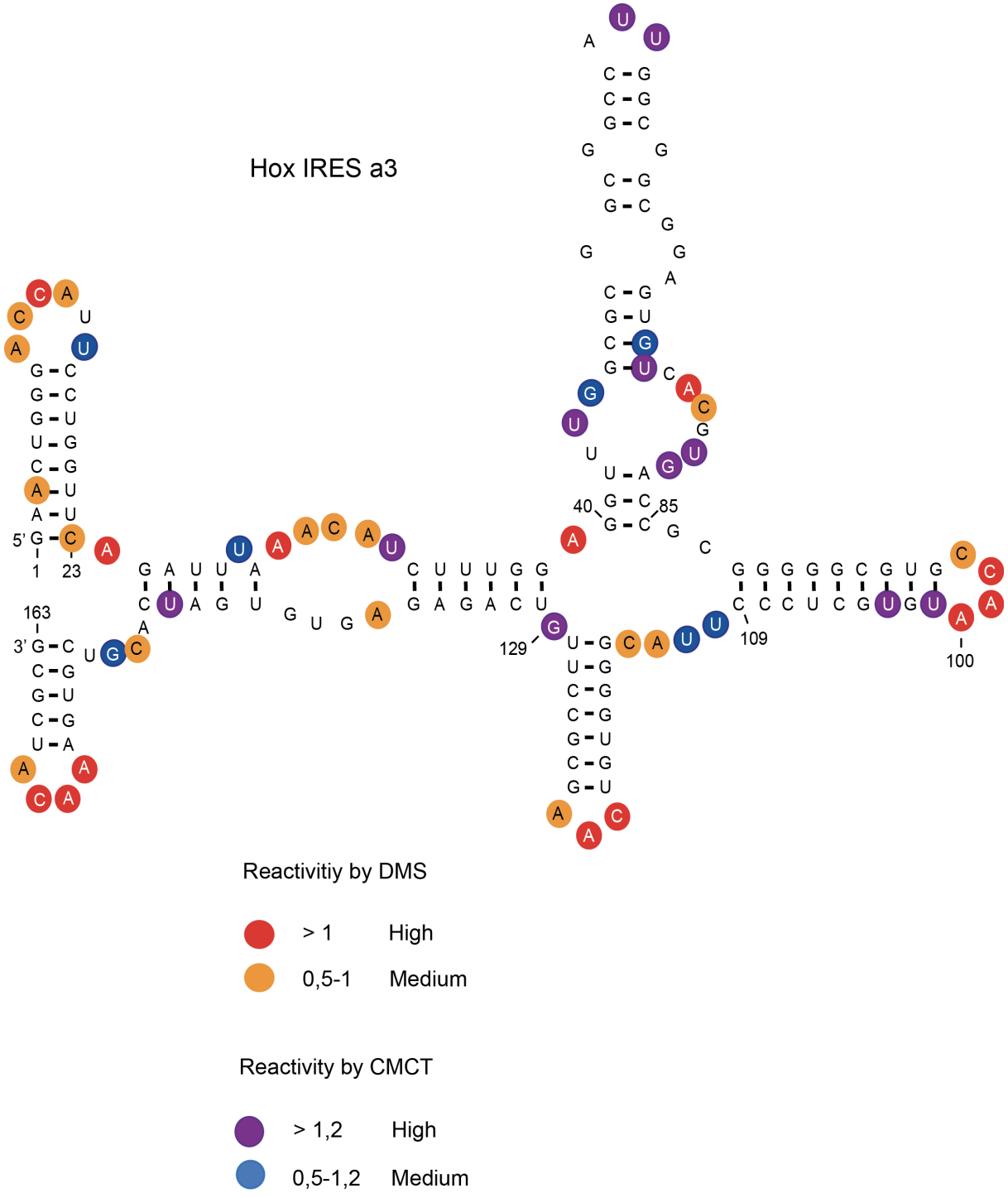


Figure 44: **Secondary structural model of IRES a3.** Probing of IREs a3 (163 nts) was obtained by chemical probing using base-specific reagents, DMS and CMCT. After modifications, reverse transcription was performed using fluorescently labelled primers to determine the position of modified nucleotides. These experiments were performed in triplicates, the reactivities are shown as average reactivity from three independent experiments.

2. Functional characterization of IRES a3

2.1. IRESes a3 and a11 are not functional in our cell-free translation assays

To get further insights into how Hox IRES function, we tested them in our cell-free translation extracts, Hox mRNA transcripts containing the full-length 5'UTRs from a3 and a11 mRNAs, upstream of the Renilla luciferase coding sequence. Our goal was to detect cap-independent translation guided by both the Hox IRESes. We tested different transcripts containing the complete 5'UTR and transcripts containing only the IRESes (**Fig. 45**). The mRNA transcripts were synthesized as uncapped mRNAs or capped with canonical 5' cap m^7G_{ppp} or A_{ppp}, a non-functional cap analog. Translation of A-capped mRNA allowed us to monitor cap-independent activity and we used m^7G-capped mRNA as a reference for cap-dependent mechanism. Unexpectedly, our *in vitro* translation assay in RRL with the full-length a3 transcripts did not show any significant luciferase expression whether uncapped, m^7G-capped or A-capped. In fact, with these mRNAs, the translation of luciferase was still inhibited. We also tested different cell-free translation systems like Wheat Germ Extract (WGE), drosophila S2 cell extracts, HeLa cell extracts and extracts prepared from HeLa cells that have been synchronised in G2/M (**Fig. 45**). The translation of RLuc in these four systems was not detectable significantly, meaning that the translational inhibition is still efficient possibly due to the presence of TIE regulons in the transcripts. We then tested the IRES a3 without the TIE either with m^7G cap or A-cap. With the transcripts lacking the TIE, we could detect a significant, although rather low luciferase expression for both transcripts in RRL, HeLa and G2M. The fact that the presence of the m^7G cap does not increase the translation efficiency detected with A-capped transcript indicates that the translation that is detected here with both constructs is cap-independent IRES activity. Altogether, the IRES activity of Hox a3 is rather weak in our cell-free translation extracts. Moreover, when a3 TIE element is present, the IRES a3 is not active at all. For IRES a11, results were quite similar. The same protocol was applied to IRES a11 (**Fig. 46**). Likewise a3, mRNA transcripts containing full-length a11 5'UTR are not active in all the cell-free translation systems tested and the addition of a functional m^7G cap or an A-cap does not change the IRES activity. In transcripts that contain only the a11 IRES, a slight IRES activity was detected with an A-capped transcripts in HeLa extracts suggesting that IRES a11 does not function very well in our cell-free translation assays (**Fig. 46**). Altogether, this suggests that Hox IRESes are not functional in our *in vitro* systems. A probable explanation is that these IRESes might require additional

specific factors (ITAFs) to enhance their activity that are actually missing in our cell-free translation systems.

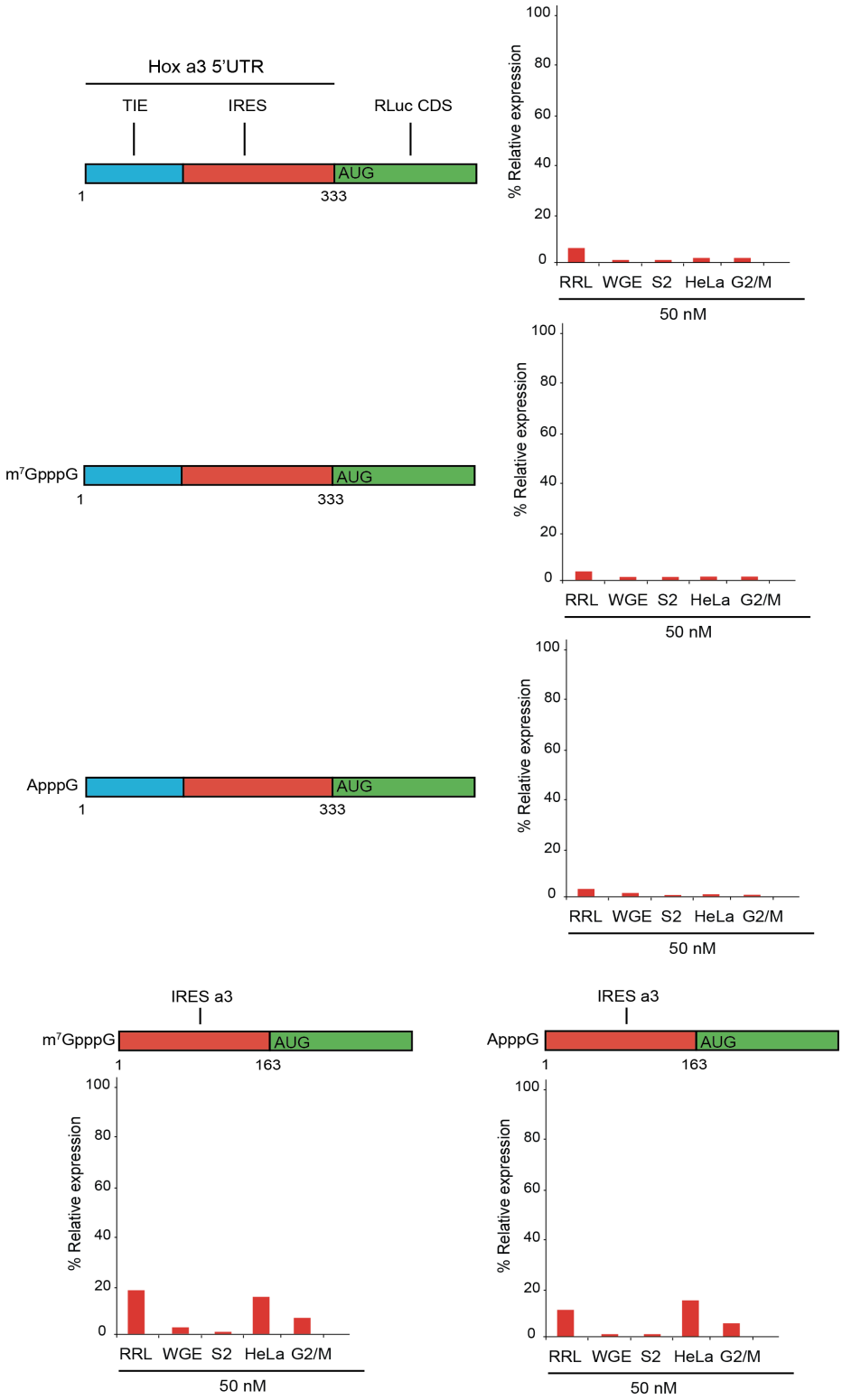


***Figure 45: In vitro* translation of IRES a3 constructs in 5 different systems.** Uncapped, m^7G or A-capped constructs of IRES a3 were translated *in vitro* in different systems. Results are indicated as relative expression of RLuc compared to control.

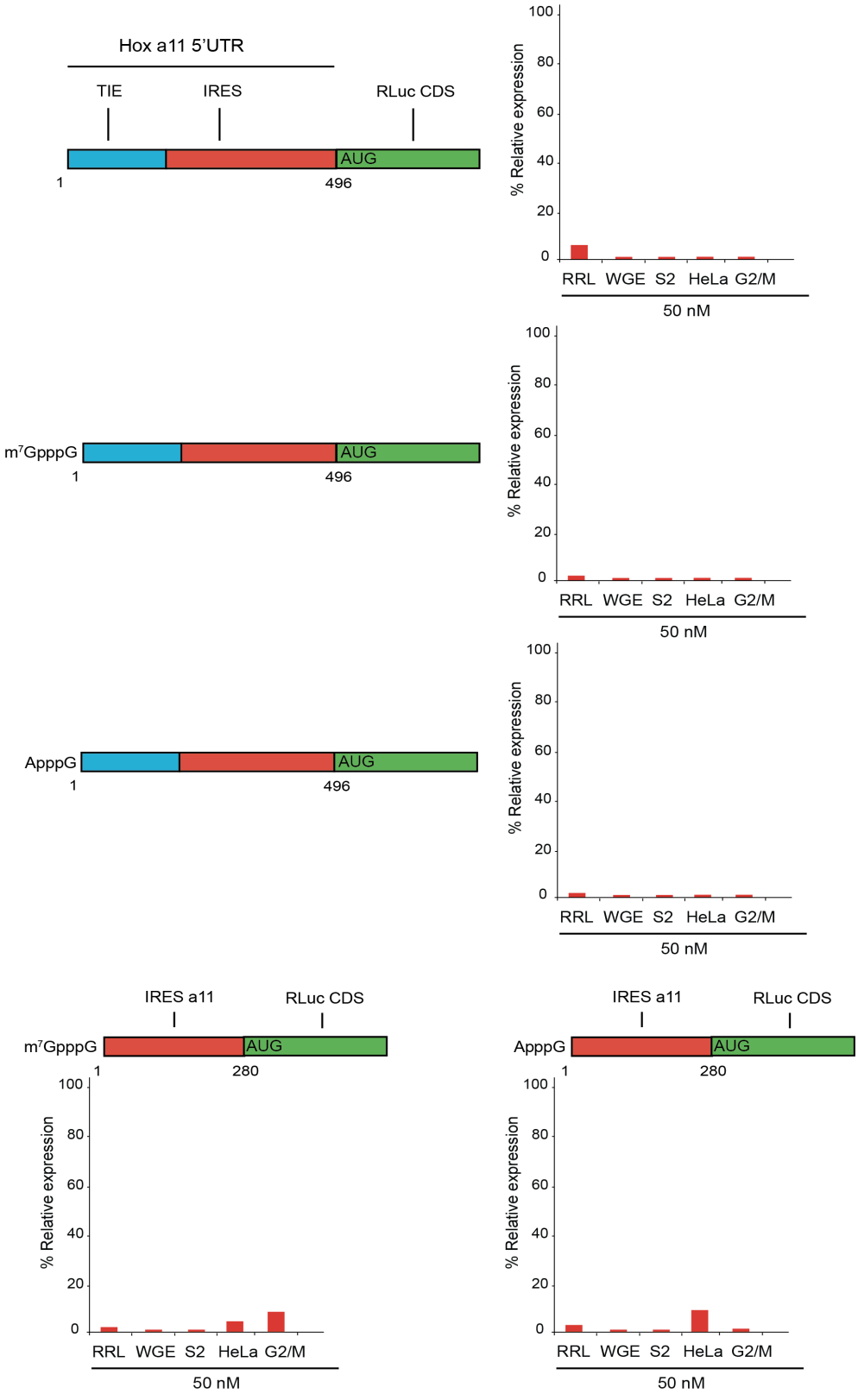


***Figure 46: In vitro* translation of IRES a11 constructs in 5 different systems.** Uncapped, m^7G or A-capped constructs of IRES a11 were translated *in vitro* in different systems. Results are indicated as relative expression of RLuc compared to control.

2.2. Sequential deletions of IRES a3 and IRES a11 reveal embedded inhibitory elements

Since we have not detected an *in vitro* IRES activity for a3 and a11, we predicted that this lack of activity might be due to the presence of inhibiting structures embedded in the IRESes. Therefore, we performed sequential 5' deletions of A-capped IRESes transcripts and tested *in vitro* translation with RRL. For IRES a3 (**Fig. 47**), the 5' truncated transcripts retrieved partial Luciferase expression when 102 nucleotides were deleted. This confirmed our prediction that IRES a3 contains an inhibitory element between regions 70-102 of IRES a3 that inhibits translation. Moreover, when the transcripts are capped with m^7G, cap-dependent translation is also inhibited. We noticed the presence of an $AUG_{100\text{-}102}$ that might also be involved in the inhibition of cap-dependent translation. To confirm it, we performed a frame-shifting mutation by deletion of G_{163}. If the $AUG_{100\text{-}102}$ is used for translation then with this deletion, an N-terminally extended luciferase would be produced. Indeed, our prediction was correct and we observed the longer luciferase (**Fig. 48**). This means that translation from this $uAUG_{100\text{-}102}$ was indeed responsible for the inhibition of cap-dependent translation we observed with the full-length IRES a3. Altogether, this suggests that the IRES a3 is not functional in RRL presumably because of its requirement of tissue specific ITAFs which are absent in our cell-free translation systems. Moreover, translation is inhibited in our cell-free translation extracts because the a3 IRES contains structured elements that prevent normal scanning. Finally, if normal scanning takes place, scanning ribosomes are trapped by an uAUG thereby resulting in the translation of a putative uORF that is embedded in the IRES.

For IRES a11, the same experiments were performed. We tested different 5' deletions of A-capped constructs (**Fig. 49A**). Interestingly, upon the deletion of region 134-221, luciferase expression increased significantly. Sequence analysis of this region also revealed the presence of a putative $uAUG_{156\text{-}158}$ that is embedded in IRES a11. Indeed, mutation of this uAUG significantly increased luciferase expression explaining why cap-dependent translation is inhibited by IRES a11 in our cell-free translation extracts (**Fig. 49B**).

Altogether, these results suggest that when using our cell-free translation extracts, Hox IRESes are not functional. In addition, they rather inhibit cap-dependent translation due to the presence of embedded inhibitory uORFs or structured elements in their sequence.

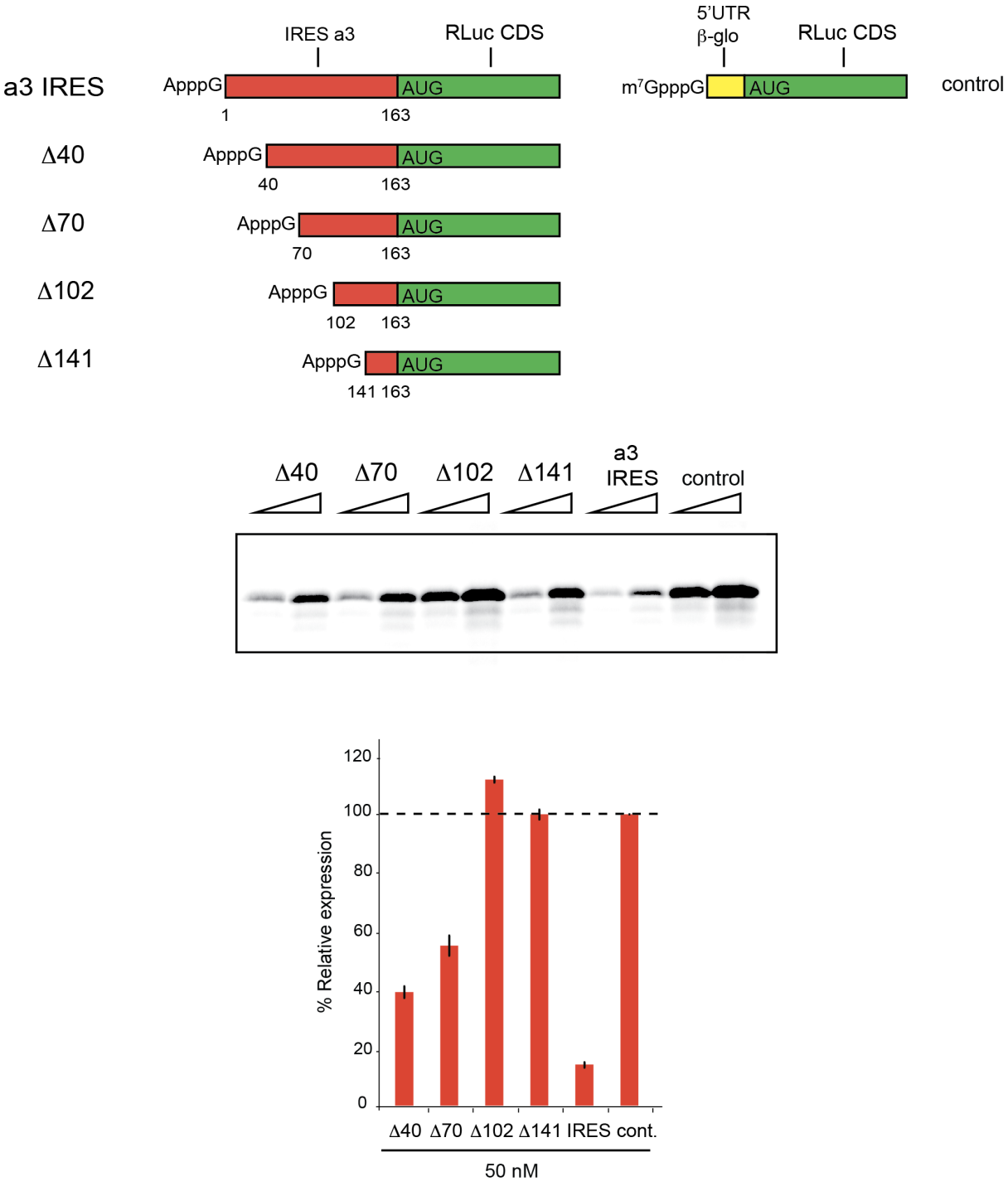


Figure 47: **Sequential 5' deletions in IRES a11.** Truncated IRES a3 constructs were translated *in vitro* in RRL. Values of translation expression were normalized to that of the control. Experiments were performed in triplicates. The percentages of RLuc expression is presented in histogram.

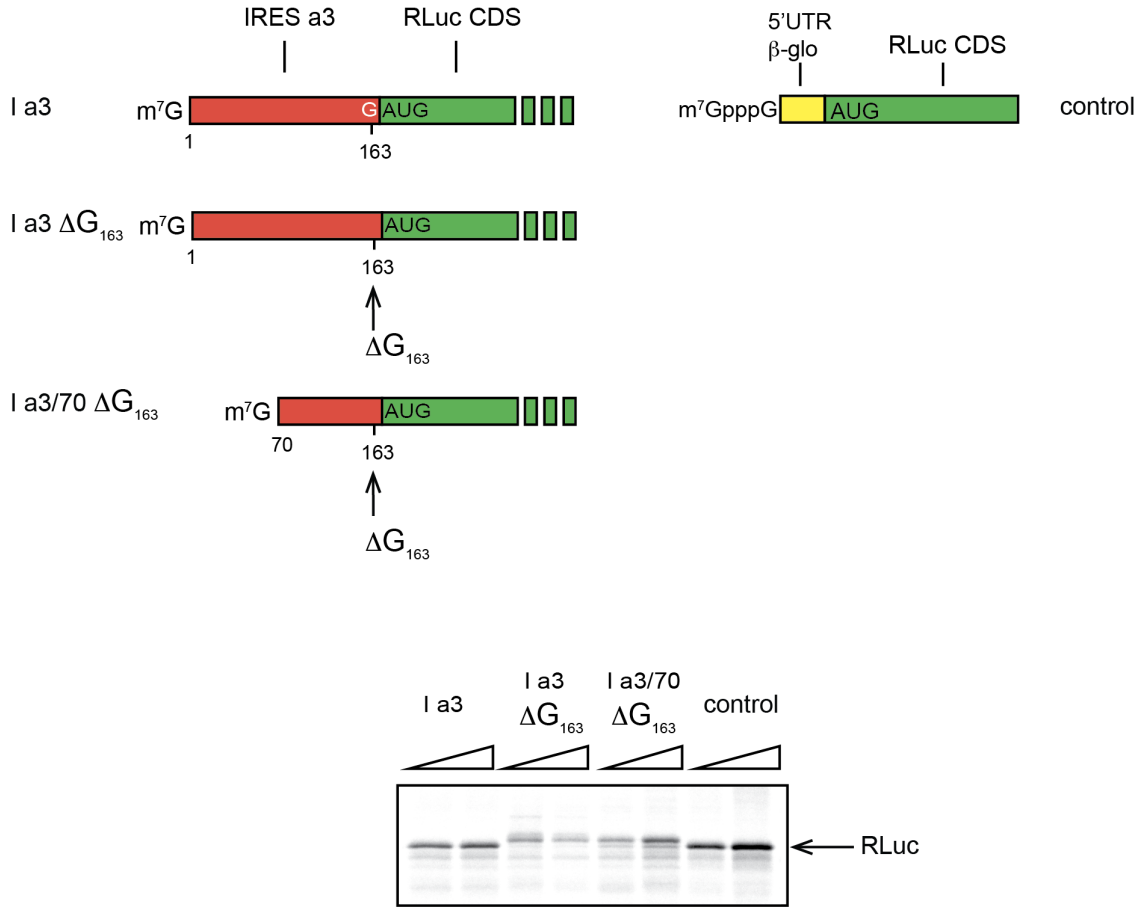
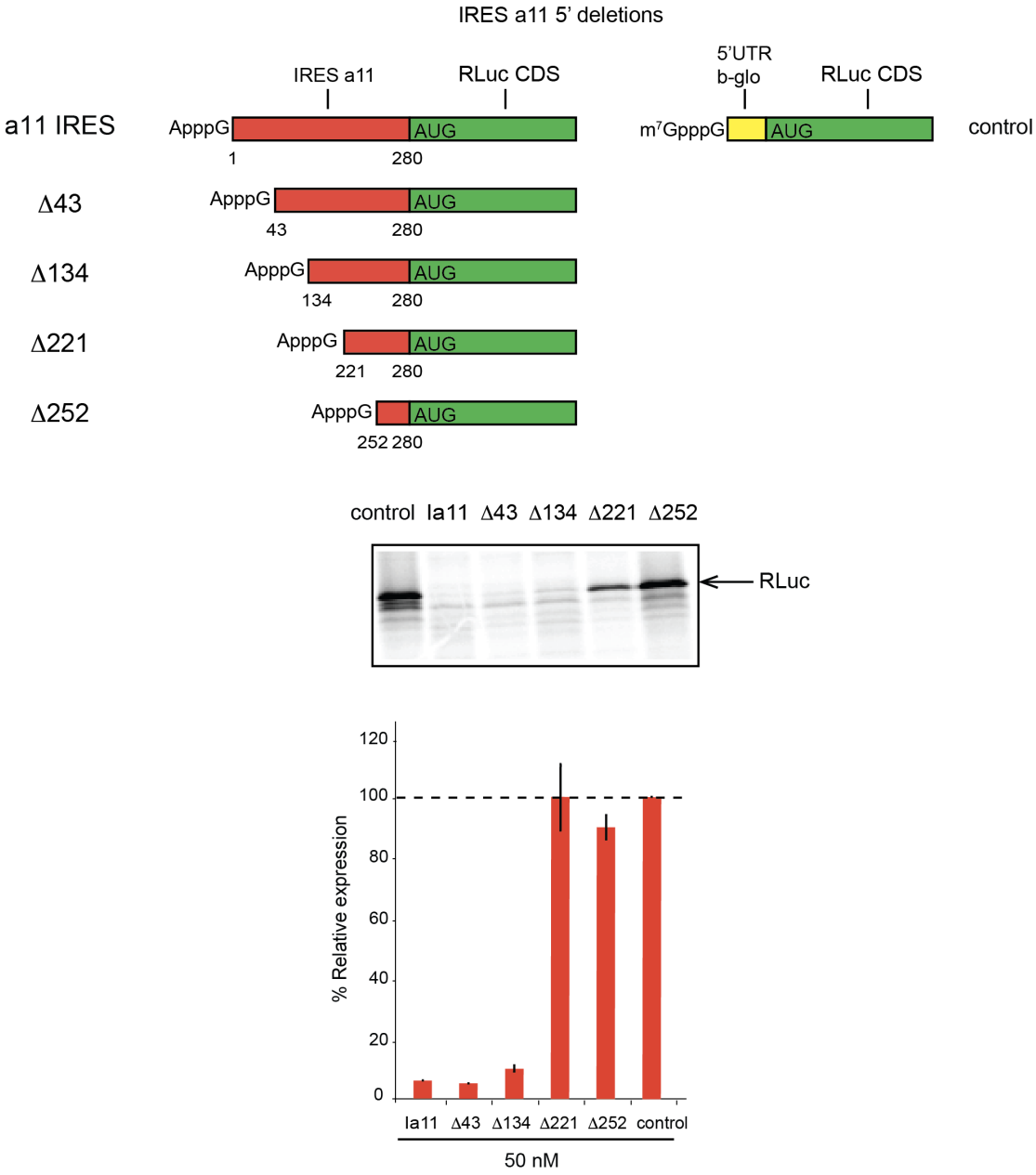


Figure 48: **Frame shift mutation in IRES a3 reveals embedded inhibitory element.** A deletion of G_{163} in IREs a3 was performed to study the translation of uORF in IRES a3 out of RLuc frame. Constructs were *in vitro* translated in RRL and products were loaded on 10% SDS PAGE.

A


IRES a11 5' deletions
IRES a11
RLuc CDS
5'UTR b-glo
RLuc CDS
a11 IRES
ApppG
AUG
1
280
m7GpppG
control
Δ43
43
Δ134
134
Δ221
221
Δ252
252
control Ia11 Δ43 Δ134 Δ221 Δ252
RLuc
% Relative expression
120
100
80
60
40
20
0
Ia11 Δ43 Δ134 Δ221 Δ252 control
50 nM

B

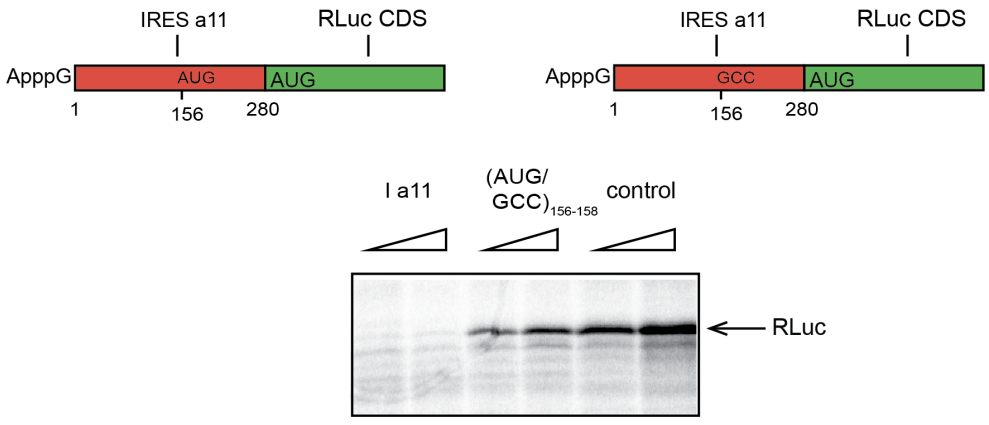

IRES a11
RLuc CDS
ApppG
AUG
1
156
280
GCC
I a11
(AUG/GCC)156-158
control
RLuc

Figure 49 (previous page): **Sequential 5' deletions in IRES a11 reveals embedded inhibitory element. (A)** Truncated IRES a3 constructs were translated *in vitro* in RRL. Values of translation expression were normalized to that of the control. Experiments were performed in triplicates. The percentages of RLuc expression is presented in histogram. **(B)** Mutation of $(AUG/GCC)_{156\text{-}158}$ in IRES a11. Constructs were translated *in vitro* in RRL and products were loaded on 10% SDS PAGE to monitor RLuc expression.

2.3. IRESes a3 and a11 are embryonic tissue-specific

To address why *in vitro* systems might not be the appropriate system to study cellular IRESes such as Hox IRESes, we performed an *in vivo* transfection assay with reporter plasmids containing IRESes a3 or a11 in their 5'UTR. We used two embryonic cell lines: kidney embryonic cell line HEK293T and the murine mesenchymal cell line C3H10T1/2 which are known to express Hox proteins (Phinney et al., 2005). We tested two transcripts: the full-length 5'UTR (FL) of Hox a3 and a11 that harbour both TIE and IRES elements and the IRES elements without the TIE (**Fig. 50**). The IRES activity was normalised by the expression of the Firefly luciferase which is also present on the plasmid and allows to monitor the transfection efficiency. Interestingly, in HEK293T cell line, the IRES-driven translation of both FL a3 and IRES a3 was 3.5 times stronger than the control confirming that the IRES is indeed functional in these cells. These results were very interesting considering our previous attempts to detect IRES activity *in vitro*. For FL a11 and IRES a11, we could also detect IRES activity but not as strong as with IRES a3. We also performed the same assay with mesenchymal cell line. Again, IRES activity was detected in FLa3 and FLa11 although less efficiently than in HEK293T cell line. We could detect a 30% of RLuc expression by IRES a3 and 22% by IRES a11 (**Fig. 50**).

Altogether, these results confirmed that indeed, embryonic cell lines express essential ITAFs that activate a3 and a11 IRESes and that they are not present in our *in vitro* systems. This proves that Hox IRESes are embryonic tissue-specific IRESes. In conclusion, deciphering the mechanism of Hox IRESes require a specific cell line where the IRESes are in fact functional.

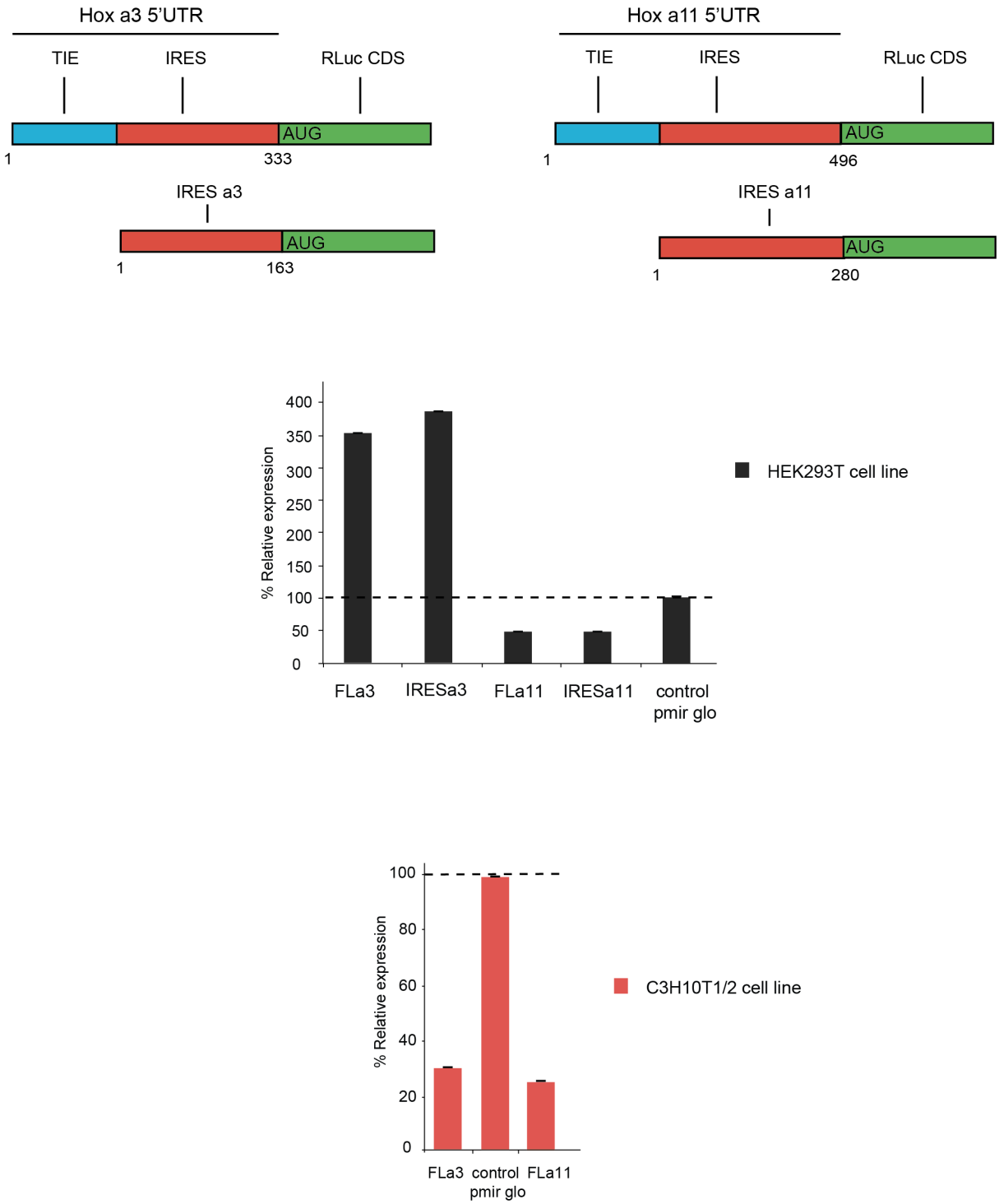


***Figure 50: In vivo* luciferase assays of Hox IRESes reveal activity in two embryonic cells lines**. Reporter constructs in pmirGlo containing FLa3, IRESa3, Fla11 and IRESa11 were transfected in the two cell lines HEK293FT and C3H10T1/2. Renilla luciferase expression was normalized to the control which was set to 100%. Experiments were performed in triplicates.

CONCLUSION AND PERSPECTIVES

CONCLUSION AND PERSPECTIVES

According to our results described in article (1), we show that there are two distinct models for TIE-mediated inhibition. Our findings indicate that TIE a3 harbours an uAUG, which is used to initiate translation through the full-length 5'UTR and leads to the synthesis of an 11 KDa-peptide. Interestingly, the uAUG is conserved among species but the peptide sequence is not suggesting it has no other function than inhibiting translation by avoiding the ribosome to reach the main ORF. Indeed, uORFs have been recognized as master regulators of translation. In the introduction, section (5.2.2.3), we highlighted the importance of regulatory uORFs under specific stress conditions such as the case of GCN4 mRNA, which contains four small uORFs in its 5'UTR. The translation of these uORFs is controlled by different nutritional conditions whereby under amino acid starvation, after translation of uORF1, leaky scanning through uORFs 2-4 leads to reinitiation on the main ORF (Hinnebusch, 2005). Nevertheless, uORFs can also contribute to IRES-mediated translation like in the case of CAT1 mRNA (Introduction, section 5.2.2.4). Under amino acid starvation, the translation of a uORF that is embedded in the IRES, induces structural remodelling to form the active state of the IRES and thereby induces CAT1 mRNA translation (Yaman et al., 2004). Regarding TIE a3, we show by *in vitro* and *in vivo* approaches that the uORF translation occurs. In order to test any putative *trans*-acting effect of the produced peptide, we supplemented our cell-free translation extracts with recombinant peptide produced *in vitro*. With this experiment we could not detect any activation on IRES a3-driven translation, which was still very low. At this point, we cannot rule out the fact that we did not have any IRES a3 activation simply because IRES a3 was not active *in vitro* since our cell-free translation extracts is not adapted to Hox IRES. Recently, we succeeded to measure for the first time IRES a3-driven translation *in vivo* with a reporter assay transfected in HEK cells (chapter 1). In the near future, we will set-up a protocol to prepare cell-free translation extract from HEK cells and test IRES-a3 activity with our reporter transcripts. With these extracts in our hands, we hope to recapitulate efficiently IRES-a3 activity *in vitro*. This will allow us to purify IRES-a3 programmed pre-initiation complexes. Then, by Mass Spectrometry, we should be able to identify the *ITAFs* that we think are absent from RRL to promote efficient IRES a3-mediated translation. Then, we will repeat the add-in experiment previously described with the uORF-encoded recombinant peptide in order to assess its influence on IRES a3 activity. If the peptide has no effect on the

IRES activity, it could also inhibit cap-dependent translation. To assess this possibility we will perform a simple experiment in RRL with our β-globin reporter mRNA and recombinant peptide added *in trans*. The next question would be to confirm first and then elucidate the putative role of eIF2D in TIE a3-mediated translation inhibition. We have shown that eIF2D is specifically present in a pre-initiation complex that has been programmed by TIE a3. In fact, eIF2D has been shown to bind at the same position on the 40S ribosomal subunit as eIF1 (Vaidya et al., 2017), suggesting structural and maybe functional homologies. Indeed, eIF2D has a homologous domain to eIF1 called SUI1 domain, which is involved in the scanning mechanism. As previously suggested, eIF2D might be involved in the scanning of subclasses of mRNAs that contain specific motifs such as A-repeats (Dmitriev et al., 2010). Interestingly, TIE a3 contains such A-rich motifs. The fact that our preliminary results show the importance of A-rich flanking sequence present upstream of uAUG in TIE a3-mediated inhibition brings support to our hypothesis that eIF2D is required in this process. Accordingly, our next step for the study of TIE a3 mechanism is to silence eIF2D in HEK293T cell line in order to confirm its involvement. We will co-transfect our TIE reporter plasmids with eIF2D siRNA and monitor the impact of this silencing on TIE inhibition. We have already established silencing controls; we have chosen eIF4E as our positive control and a non-target pool of siRNAs as a negative control. This experiment will inform us, at least initially, whether silencing eIF2D impacts TIE a3-mediated inhibition. Accordingly, we will study the molecular mechanism of eIF2D implication with TIE a3. One approach will be to use purified recombinant eIF2D. We will verify the specific interaction with TIE a3 by electrophoretic mobility shift assay or by cross-linking methods. We will also assess the role of eIF2D in the formation of TIE a3-programmed pre-initiation complex by sucrose gradient analysis and toe printing.

Concerning TIE a11, our current model suggests the importance of two regulatory elements: the start-stop combination and the GC-rich stem loop (SL). Both elements enable the sequestration of an 80S ribosome that is stalled on the start-stop (results, chapter 1). At the moment we believe that the distance between the AUG and SL is optimal (19 nucleotides) and critical for efficient ribosome stalling. Since the stalled ribosome contains the stop codon UAG in the ribosomal A-site, the release factor should be recruited and induce ribosome dissociation. One intriguing issue is to understand the mechanism that prevent release factor binding in order to enable ribosome stalling on the TIE a11. Our current hypothesis is that the SL is close enough to the ribosome to prevent the release factor recruitment. This hypothesis is reinforced by the fact that the distance between the SL and the AUG codon (which is an

AUG-like in some organisms) is conserved among species. Another compelling evidence is that the release factors are not present in the TIE a11-programmed pre-initiation complexes as shown by our Mass Spectrometry. In order to demonstrate the statement that the distance between the start codon and the SL is critical for translation inhibition, we will construct TIE a11 variants in which the distance between the start-stop and the SL has been increased by insertion of spacers of different lengths. These TIE a11 mRNA variants will be used to assemble pre-initiation complexes that will be analysed by sucrose gradients and later on by Mass Spectrometry. Another interesting feature of TIE a11 translation inhibition is the absence of eIF5B in the pre-initiation complexes (which was present with TIE a3). Factor eIF5B catalyses GTP hydrolysis, a step which triggers the release of eIFs and subsequent subunit joining to form the 80S complex (Fringer et al., 2007). Also, the presence of eIF3j subunit was unexpected. This subunit dissociates during early stage of initiation to allow mRNA entry (Fraser et al., 2007; Kolupaeva et al., 2005) and is therefore not expected to be present in initiation complexes. We will investigate further the role of eIF5B and eIF3j factors in TIE a11-mediated inhibition. We will also try to confirm the involvement of these factors by silencing as previously explained for eIF2D. Finally, Mass Spectrometry results show a number of non-canonical factors associated with TIE a11 initiation complexes that might also be interesting to investigate.

As previously mentioned, we would also like to investigate further the mode of action of the Hox IRESes. We have shown that they can fold independently of TIEs. This suggests that the two Hox regulons function as independent modules. We have shown that we can detect IRES activity with both a3 and a11 in HEK293T cell line. With cell-free extracts prepared from this cell line, we will attempt to decipher the mechanism of Hox IRESes. It would be interesting to perform a number of experiments such as pre-initiation complex assembly and purification for sucrose gradient analysis, Mass Spectrometry analysis and maybe cryo-Electron Microscopy. We expect to be able to classify the IRES in terms of ITAFs requirements and mode of action for ribosome recruitment. In our perspective, TIE elements in Hox cluster have distinct inhibitory mechanisms, the question remains: is this the case for Hox IRESes also? According to (Xue et al., 2015), IRES a11 is RpL38-dependent while IRES a3 is not. This suggests that Hox cluster mRNA might be differentially regulated during translation. The activation of the IRES most probably occurs in a tissue-specific manner. Since the translational control on Hox mRNAs is indispensable during embryonic development, TIEs inhibit cap-dependent translation in a constitutive manner. In contrast, the IRES activity should be tightly regulated. Therefore, the time- temporal- and tissue-specific expression of

Hox genes is solely governed by IRESes. In conclusion, after having studied the mode of action of TIEs, it is now crucial to increase our knowledge on the Hox IRESes to better understand the sequential expression of Hox proteins during development.

ANNEX

Article 2

Tracking the m^7G-cap during translation initiation by crosslinking methods

Gross Lauriane, Schaeffer Laure, Alghoul Fatima, Hayek Hassan, Allmang Christine, Eriani Gilbert, Martin Franck

Methods, 2018, Volume 15; issue 137 : 3-10. doi: 10.1016/j.ymeth.2017.12.019.

1. Tracking the m^{7}G-cap during translation initiation by crosslinking methods

1.1. Summary of article 2:

As part of my thesis, I participated in the work on histone H4 mRNA whose translation initiation is unconventional. The mechanism for initiating the translation of H4 mRNA is through a new mechanism called ribosome tethering (see Introduction section 4.1). It combines features of canonical initiation, such as the dependence on the cap, with features found in viral IRESes, such as the presence of structural elements in RNA that are involved in the recruitment of eIFs factors during recruitment of the translation machinery. In this article published in *Methods* in 2018, we followed the fate of the cap during the various stages of translation initiation using photoactivatable or chemical bridging methods of the m^{7}G cap. First, we studied the position of the cap on free H4 mRNA. To do this, the H4 mRNA was capped using cap analogues containing a photoactivatable thiol group. UV irradiation induces intra-molecular bridging, demonstrating that the cap is not accessible but instead sequestered in a pocket located in the coding phase of the RNA that we named Cap-Binding Pocket (CBP). We have thus confirmed the previous results from the laboratory, which indicate that the cap is not accessible before the initiation of the translation. Then, we followed the position of the cap during the different stages of cap-dependent initiation. The strategy was to use the radioactively labelled cap as a tracer to follow its position during the initiation process. The proteins located near the cap are visualized following a chemical crosslinking approach, which makes a covalent bond with the radiolabelled cap. This allowed us to show that the cap is recognized and still bound by the cap binding protein eIF4E protein at the 48S stage. On the other hand, when initiation reaches the 80S stage, the cap is not bound by eIF4E anymore and is rather located near the ribosomal proteins eS26 and eS28. These two proteins are positioned in the entry site of the mRNA on the 40S ribosomal subunit. In fact, H4 mRNA has a very small 5'UTR of only 9 nucleotides. Therefore, our study demonstrated that the eIF4E protein is required to disengage from the cap after 48S formation to allow assembly of the large 60S subunit. It would now be interesting to apply the experimental protocol developed to track the fate of the cap during the translation initiation of other mRNAs containing 5'UTR regions of varying lengths.

Methods 137 (2018) 3–10

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Tracking the m^{7}G-cap during translation initiation by crosslinking methods

Lauriane Gross, Laure Schaeffer, Fatima Alghoul, Hassan Hayek, Christine Allmang, Gilbert Eriani *, Franck Martin *

Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France

ARTICLE INFO

Article history:
Received 15 October 2017
Received in revised form 20 December 2017
Accepted 22 December 2017
Available online 4 January 2018

Keywords:
Cap-dependent translation
UV crosslinking
Chemical crosslinking
Ribosome
Histone H4 mRNA
eIF4E

ABSTRACT

In eukaryotes, cap-dependent translation initiation is a sophisticated process that requires numerous *trans*-acting factors, the eukaryotic Initiation Factors (eIFs). Their main function is to assist the ribosome for accurate AUG start codon recognition. The whole process requires a 5′-3′ scanning step and is therefore highly dynamic. Therefore translation requires a complex interplay between eIFs through assembly/release cycles. Here, we describe an original approach to assess the dynamic features of translation initiation. The principle is to use the m^{7}Gcap located at the 5′ extremity of mRNAs as a tracker to monitor RNA and protein components that are in its vicinity. Cap-binding molecules are trapped by chemical and UV crosslinking. The combination of cap crosslinking methods in cell-free translation systems with the use of specific translation inhibitors for different steps such as edeine, GMP-PNP or cycloheximide allowed assessing the cap fate during eukaryotic translation. Here, we followed the position of the cap in the histone H4 mRNA and the cap binding proteins during H4 mRNA translation.

1. Introduction

Translation initiation leads to the assembly of an elongation-competent 80S ribosome on the genuine AUG start codon. In eukaryotes, this is achieved through assembly of a 43S pre-initiation complex containing the small ribosomal subunit 40S on the 5′ m^{7}Gcap of mRNA. Next, the so-called 43S slides on the 5′UTR, by a 5′-3′ nucleotide-by-nucleotide inspection that is an ATP-consuming mechanism termed scanning, until the AUG start codon is found [1]. Then, the large ribosomal subunit 60S joins in order to assemble a complete 80S ribosome on the start codon. This sophisticated process is assisted by numerous *trans*-acting factors called eukaryotic Initiation Factors (eIF) [2]. The whole process is highly dynamic with many transient complexes that are difficult to trap. The interplay between the eIF is rather complex and methods to assess the numerous assembly/release cycles of individual players are rare.

Among these, a number of techniques for studying the structure and interaction of proteins with nucleic acids depend on methods of crosslinking. Crosslinking is the process of chemically joining two or more molecules by a covalent bond. It often implies chemical modification of one partner or incorporation of a reactive reagent such as photo reactive nucleotides that alter the reactivity of the original molecule. A great advantage of crosslinking approaches is to enable trapping of transient complexes by linking covalently the molecules that interact temporarily during dynamic processes such as translation initiation. Crosslinking methods provide a rapid mean of obtaining evidence for the proximity of functional groups in structurally complex RNAs and ribonucleoproteins. Accurate identification of the crosslinks is also a way to probe the conformation of the RNA of interest.

The position of the m^{7}G-cap during the whole translation initiation process is of special interest. Indeed, numerous studies have demonstrated that the cap first initiates the scanning process by recruiting the whole machinery at the 5′ extremity of mRNA. However, the positioning of the cap during the scanning process, after 80S formation on the AUG and later on during elongation is so far unexplored. Here, we describe two crosslinking methods using the cap as a bait to trap partner molecules in order to assess the cap positioning and its fate during the whole translation process. In this study, we used the mouse histone H4 mRNA as a model to follow the cap fate during the whole initiation process. This particular mRNA is part of the cell cycle-dependent histone mRNAs which are

Abbreviation: U, units.

* Corresponding authors.
E-mail addresses: g.eriani@ibmc-cnrs.unistra.fr (G. Eriani), f.martin@ibmc-cnrs.unistra.fr (F. Martin).

https://doi.org/10.1016/j.ymeth.2017.12.019
1046-2023/

not polyadenylated at their 3′ end [3]. They are massively and exclusively expressed during the S-phase of the cell cycle. They have short UTRs, H4 mRNA being the one with the shorter 5′ UTR of only 9 nucleotides [4]. The translation mechanism of histone H4 mRNA is non canonical. The 5′ m^7Gcap is not accessible but instead is sequestered by a cap binding pocket located in the coding region [5,6]. This enables the internal recruitment of eIF4F on a specific element called 4E-Sensitive Element that is also located in the coding region. Then, the ribosome is dropped on the start codon by a so-called 'ribosome tethering' mechanism. Since there is no scanning, the AUG start codon is positioned accurately in the P site of the ribosome by a direct interaction between an AGG triplet in the coding region of H4 mRNA and the loop of helix h16 of the 18S ribosomal RNA [7]. In order to better characterize the role of the cap in this sophisticated process, we studied its position during the distinct steps of translation initiation and elongation.

First, we monitored the position of the cap in the mRNA alone before translation initiation starts by UV-crosslinking with thio-modified caps. As previously mentioned, we confirmed that the cap is not accessible but rather located on an internal structure formed by H4 mRNA making a cap-binding pocket. Second, we investigated the fate of the cap during H4 mRNA translation initiation and elongation by chemical crosslinking. Thereby, we detected proteins that are positioned near the 5′-end of mRNA in initiation complexes formed during histone H4 mRNA translation.

2. Material and methods

2.1. In vitro transcription

Synthesis of RNAs containing thio-modified nucleosides can be performed with the T7 RNA polymerase system. The gene encoding mouse histone H4 (HIST1H4C) and the human β-globin mRNA have been previously cloned into pUC19 and YpGlo respectively. Transcription templates are synthesized by PCR amplification from the plasmids. The 5′ primers contain the T7 promoter sequence and the 3′ primers are designed to promote *in vitro* run-off transcription at the desired position. The PCR approach is also used to synthesize transcripts of H4 mRNA with truncated 5′ sequences. For that, new 5′ primers containing the T7 promoter sequence connected with the desired H4 sequences are designed to synthesize the PCR fragments with variable sizes. *In vitro* transcription is performed as follows. PCR DNA template (about 20 μg) is mixed with transcription buffer (40 mM Tris-HCl, pH 8.1, 20 mM $MgCl_2$, 5 mM DTT), 1 mM each ATP, CTP, GTP, UTP, 40 U of RNase inhibitor (RNasin-Promega), 0.2 mg/mL of T7 RNA polymerase, and milliQ water to 100 μL. The reaction mixture is incubated at 37 °C for 1 h. Then, 2 μL of pyrophosphatase (10108987001-Sigma) are added and the mix is further incubated for 30 min. Then, the DNA template is degraded by addition of 2 μL of DNase I RNase-free (04716728001-Sigma) and incubation for 1 h at 37 °C. Usually, to check RNA integrity, an aliquot is mixed with Formamide Dye and loaded on urea-denaturing 4% PAGE and visualized under UV light after ethidium bromide staining. To eliminate unincorporated nucleotides, the remaining RNA sample is loaded on a G25 column. Transcripts are further purified by separation on denaturing 4% PAGE and electro-elution from gel slices using a Biotrap apparatus (Schleicher and Schuell). Purified RNA samples are then phenol extracted to eliminate proteins from the synthesized transcripts. The concentration of purified RNA sample is determined by absorbance measurement at 260 nm. Before use, transcripts are folded in water by incubating at 80 °C for 2 min followed by slow cooling to 35 °C and kept on ice.

2.2. Capping of RNA transcripts

RNA transcripts can be m^7G-capped co-transcriptionally by T7 RNA polymerase in the presence of m^7GpppG cap analogue (New England Biolabs). More efficient capping is obtained with the vaccine-capping enzyme (VCE) using the ScriptCap™ m^7G Capping System (Epicentre Biotechnologies) which caps and methylates the 5′ end of mRNAs to nearly completion (Fig. 1A). The ScriptCap™ m^7G Capping System requires purified mRNA (purified by electro-elution, see 2.1.). When the GTP from the kit is substituted by s^6GTP (Jena Biosciences, 1 mM), a m^7s^6G-capped mRNA suitable for photo-crosslinking studies is synthesized by VCE after a 1-h incubation at 37 °C (Fig. 1B). After capping, enzymes used for the reaction are eliminated by phenol-chloroform extraction and the m^7s^6G-capped mRNAs are precipitated with ethanol in the presence of 250 mM NaCl. The capped mRNAs are recovered by centrifugation; the pellets are dried and then resuspended in autoclaved milliQ water.

2.3. Short-range crosslinking using thio-containing 5′ cap

Two pmoles of *in vitro* synthesized and capped m^7s^6G-histone H4 mRNA are dissolved in water, refolded at 95 °C for 2 min and then incubated at 0 °C, then are irradiated with 312 nm for 30 min at a distance of 8 cm from the bulb using a Bio-Link BLX 312 on ice as previously described [6]. Importantly, previous experiments with histone H4 mRNA demonstrated that it has to be refolded in water to favour the optimal conformation to visualize the UV-crosslink of the cap in its cap binding pocket. However, for other mRNAs, it might be critical to first set-up an optimal refolding protocol that may require Mg^{2+} to reach the optimal conformation. Then, crosslinks are detected by primer extension using fluorescently labelled H4-specific primers by reverse transcriptase. Reverse transcription is performed in final volume of 20 μL. First, 2 pmol of cross-linked mRNA are mixed with 2 pmol of fluorescently (VIC or NED, from Integrated DNA Technologies) labelled primer (H4-rev160: 5′-TGAGGCCGGAGATGCGCTTC-3′) and denatured first at 95 °C for 2 min followed by primer annealing at 65 °C for 5 min and incubation on ice for 2 min. Then, 0.2 U AMV reverse transcriptase (Life Sciences Advanced Technologies Inc.) in its buffer as recommended by the manufacturer is added and supplemented with 1.5 mM of each dNTP. Reverse Transcription extension parameters are: 42 °C for 20 min, 50 °C for 30 min, 60 °C for 10 min. Sequencing reactions are performed in parallel in similar conditions, but containing 0.5 mM of each dNTP and 16.6 μM of ddCTP. Reactions are stopped by phenol-chloroform extraction and ethanol precipitation. Samples are resuspended in 10 μL deionized formamide and loaded on a 96-well plate for sequencing on an Applied Biosystems 3130xl genetic analyzer. The resulting electropherograms are analysed using QuSHAPE [8], which aligns signal within and across capillaries, as well as to the dideoxy references, and corrects for signal decay. Normalized reactivities range from 0 to ∼2, with 1.0–1.2 being the average reactivity for highly reactive positions. The average reactivity (∗.shape file) is calculated for each H4 mRNA transcript from three independent experiments with Pearson correlation coefficients ≥0.9.

2.4. 5′-end radioactive labelling of RNA transcripts

For chemical crosslinking studies, mRNA transcripts (80 pmol-10 μg) are first 5′-end-labelled. For that, 50 μCi of [α-^{32}P]-GTP (6000 Ci/mmol, 10 mCi/mL, Hartmann Analytic) is used instead of unlabelled GTP during the capping reaction by the ScriptCap™ m^7G Capping System (Fig. 1C). After radioactive capping, capped mRNA transcripts are separated from unincorporated GTP by denaturing PAGE (4%) and eluted overnight from gel slice at 4 °C in

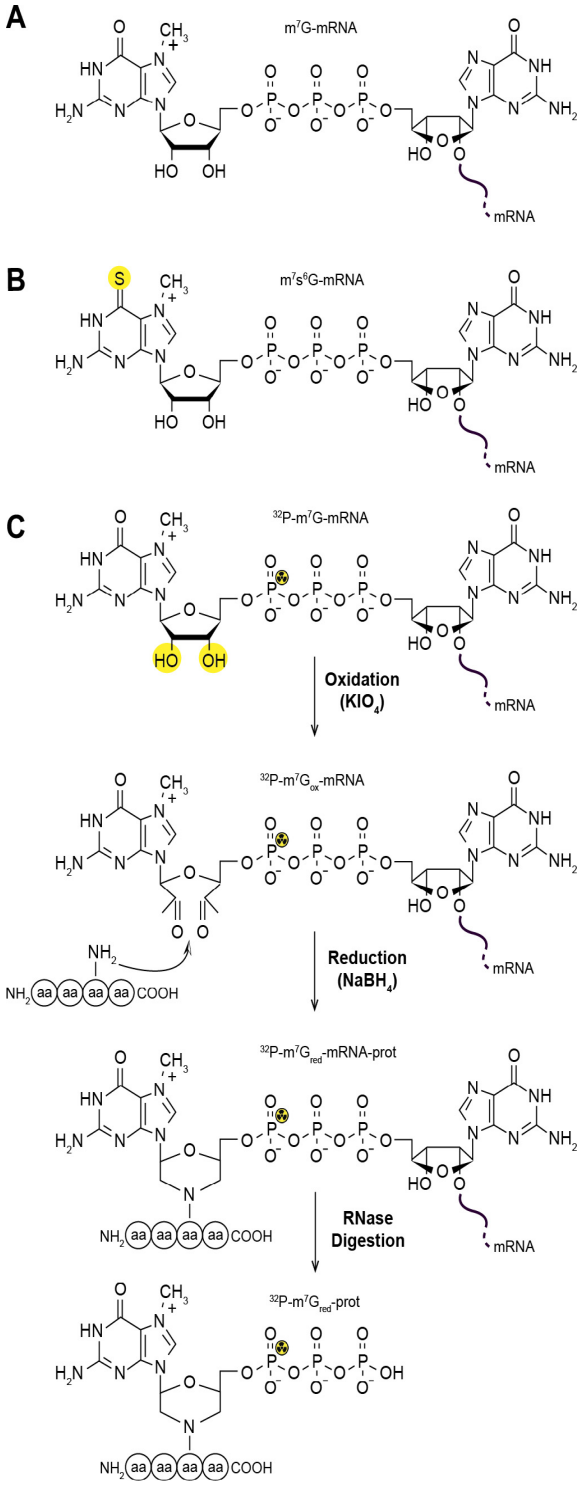


Fig. 1. Capped mRNAs used in this study. A. Structure of native m^7G-capped mRNA. B. Structure of m^7s^6G-mRNA that is site-selectively labelled with 6-thioguanosine (in yellow) within the cap structure. C. Structure of radiolabelled m^7G-capped mRNA that has been use for chemical crosslinking. The 2′ and 3′ OH (in yellow) are first oxidised by KIO_4 and then reduced by $NaBH_4$. This enables a covalent link between a NH_2 group from the lateral chain of a protein that is in a close proximity. The RNA moiety is then degraded by RNase treatment leaving a ^{32}P labelled protein that is further analysed by SDS-PAGE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

buffer (0.3 M NaCl, 0.5 mM EDTA, 10 mM Tris-HCl pH 7.5). After passive elution, the radioactive m^7G-capped mRNA are precipitated with ethanol in the presence of 250 mM NaCl. The radioactive capped mRNAs are recovered by centrifugation, the pellets are dried and then resuspended in autoclaved milliQ water. Global radioactivity of the mRNA samples are measured by Cerenkov counting and adjusted to 50,000 cpm per μL.

2.5. Chemical crosslinking of the 5′ cap to ribosome initiation complexes

Chemical crosslinking of the 5′ cap of mRNAs is a powerful approach to study RNA binding proteins such as initiation factors. The method was originally used to identify the cap-binding protein eIF4E [9]. The target mRNA containing a radioactively labelled cap, $m^7[^{32}P]G(5')pppG$, is oxidised with sodium periodate to convert the 2′,3′-cis-diol of the 5′-terminal m^7G to a reactive dialdehyde (it is worth noting here that the ribose at the 3′ end of the mRNA is similarly oxidised in a dialdehyde) (Fig. 1C). Then, the oxidised mRNA is incubated in cell-free protein-synthesis systems from rabbit reticulocyte lysate (RRL), or wheat germ extract (WGE) or alternatively with purified components. The resulting mRNA-ribosome initiation complexes are treated with $NaBH_4$ to reduce and stabilize the Schiff bases between mRNA 5′ termini and amino groups of neighbouring proteins, yielding to covalently linked protein-RNA conjugates. The mRNA is then further trimmed by RNase A (Roche Diagnostics) thereby leaving the radiolabelled cap covalently linked to its protein partner(s). Cell-free extracts or protein components are then separated by SDS-PAGE and radioactive crosslinked proteins are revealed by autoradiography. Usually, only polypeptides positioned near the 5′ end are crosslinked to the mRNA, making this approach very useful for studying the spatial relationships between molecules in ribosomal complexes or other ribonucleoprotein (RNP) complexes that are close to the cap. In the example of histone H4 mRNA shown here, the periodate-oxidised mRNA is assembled with initiation complexes in RRL in the presence of various translation inhibitors [10]. Edeine is binding in the mRNA tunnel of the small ribosomal subunit and interferes with the codon-anticodon interaction between initiator $Met\text{-}tRNA_i^{Met}$ and the AUG start codon in the P site. This means that edeine treatment impairs formation of the 48S complex. In other words, edeine treatment enable the accumulation of translation initiation complex prior AUG codon recognition. GMP-PNP is a non-hydrolysable GTP analogue that prevents ribosomal subunit joining by inhibiting the GTPase activity of eIF5B subunit-joining factor [11]. The use of GMP-PNP in cell-free extracts allows the accumulation of 48S particles (small ribosomal subunits with a $Met\text{-}tRNA_i^{Met}$ assembled on the AUG initiation codon). Finally, cycloheximide induces accumulation of 80S particles by blocking translation at the translocation step. In fact, cycloheximide allows one complete round of translocation to proceed before inhibiting further elongation. The inhibitory effect results from binding of both cycloheximide and deacylated tRNA to the E-site, which leads to the arrest of the ribosome on the second codon [12]. Cap-specific crosslinks are validated by competition with an excess of m^7GDP. Experimentally, 5 μg of purified and radiolabelled (250,000 cpm) capped H4 mRNA are incubated for 2–3 h at 0 °C in 250 μL of 100 mM sodium acetate, pH 5.3, 10 mM EDTA, 0.2 mM sodium periodate. Then, glycerol is added to 2% (v/v) final concentration for neutralisation. After 10 min incubation at room temperature, the mixture is phenol extracted twice and ethanol precipitated. The RNA pellet is dissolved in 10 μL of water. One μL of oxidised RNA is incubated with 2.5 μL of RRL, in 10 mM HEPES-KOH, pH7.6, 1 mM ATP, 75 mM KCl, 1 mM DTT, 1 mM $Mg(Ac)_2$ and 1 mg/mL cycloheximide, or 2 mM GMP-PNP or 10 μM edeine or 1 mM m^7GDP in a final volume of 10 μL. After 10 min incubation at 30 °C, 1 μL of 0.2 M $NaBH_4$ is added and incubation is pursued for 2–3 h at 0 °C. Then, RNA is digested by 1 μL of RNase A (10 mg/mL, ThermoFisher Scientific-EN0531) for 30 min at 37 °C and samples are fractionated by SDS-PAGE.

Radioactive bands correspond to proteins crosslinked to the 5′ m^7G-cap of the mRNA are revealed by autoradiography. The specificity of the crosslink is further confirmed by adding the cap analogue m^7GDP that competes with the m^7G-cap and clears the cross-link. Protein identification can be further performed by immuno-precipitation and/or mass spectrometry analysis.

2.6. Sucrose gradient analysis of crosslinked initiation complexes

The cap crosslinking capacity in RRL was further analysed after sucrose gradient fractionation. The protocol used to prepare the complexes with 5′-labelled capped mRNA was slightly modified compared to the one described in Section 2.5. Prior the formation of 80S/H4 initiation complexes, 30 μL of untreated RRL were incubated for 5 min at 30 °C and 20 min in ice in a buffer containing 1 μL of RNase inhibitor (40U)(RNasin, Promega-N2511), 75 mM KCl and 0.5 mM $MgCl_2$. In order to lock the ribosome at translation initiation, RRL were pre-incubated for 3 min at 30 °C and 20 min at 0 °C in the presence of 10 μM edeine or 2 mM GMP-PNP or 1 mg/mL cycloheximide or without inhibitor. Finally, formation of initiation complexes was obtained by adding the oxidised histone H4 mRNA to a final concentration of 500 nM and incubating for 5 min at 30 °C. Then $NaBH_4$ was added to a final concentration of 20 mM and reduction was performed at 0 °C for 1 h. The resulting RRL programed with H4 mRNA were separated on 7–47% linear sucrose gradients in buffer (25 mM Tris-HCl, pH 7.4, 75 mM KCl, 0.5 mM $MgCl_2$, 1 mM DTT, 1 mM cycloheximide). The reactions were loaded on the gradients and centrifuged (37,000 rpm for 2.5 h at 4 °C) in a SW41Ti rotor. Gradients are fractionated on a Biocomp gradient fractionator and fractions were analysed by Cerenkov counting in order to identify the fractions containing translation initiation complexes. Indeed, we demonstrated that the stability of translation initiation complexes is not affected by the 1 h-incubation at 0 °C after reduction as attested by the radioactive profiles. Then the fractions (300 μL) corresponding to RNPs, 80S and polysomes were RNase A treated as described in Section 2.5 for 30 min at 37 °C before being precipitated by TCA (12.5% final concentration) for 30 min at 0 °C. Precipitated proteins were recovered by 15 min centrifugation at 15000 × g, washed by 20 μL acetone, dried and resuspended in 10 μL of loading blue (400 mM Tris-HCl pH 8.8, 4% SDS, 4% βME, 30% glycerol, 0,05% bromophenol blue). Proteins were fractionated by SDS-PAGE and radioactive bands were detected by photostimulable phosphor plate (Fuji Imagine plates). For reproducibility, we noticed that the use of freshly oxidised mRNA is critical.

3. Results and discussion

3.1. UV crosslinking of the cap in its cap-binding pocket from H4 mRNA

Although a certain number of chemical and photoreagents are available for short-range RNA-RNA crosslinking studies, the

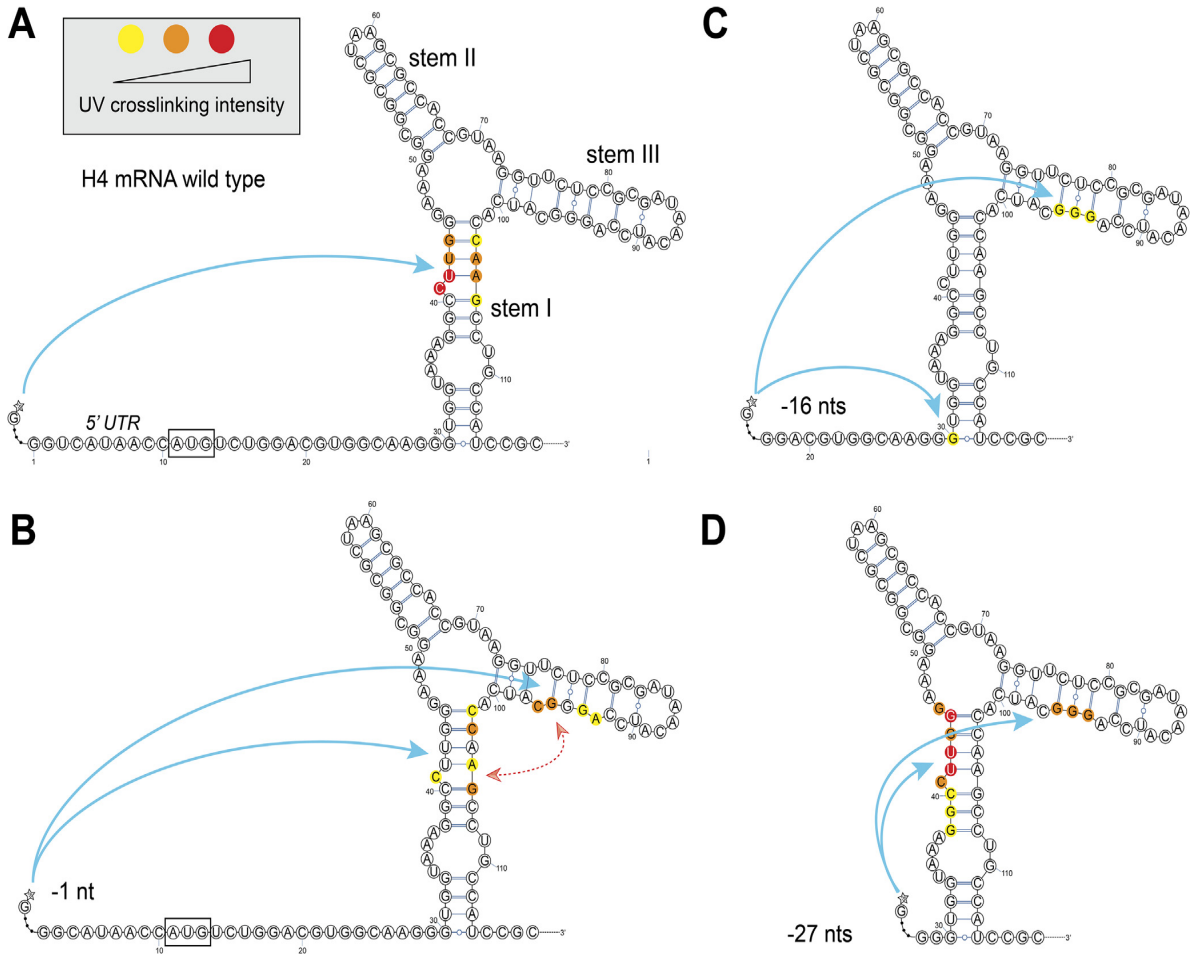


Fig. 2. UV-crosslinking of m^7s^6G-H4 mRNA. A. Secondary structure of the 114 first nts of histone H4 mRNA and interaction with the m^7s^6G cap structure. The structure contains three helical stems (I-III) connected by a three-way helix junction (TWJ). The initiation codon is boxed. Crosslink experiments on m^7s^6G-capped wild-type H4 mRNA. Crosslinks were performed at 360 nm and revealed by reverse transcription. Reverse transcription stops are indicated by colours according to the intensity scale shown in the inset. B-D. Similar experiments performed on mutants deleted of 1, 16 and 27 first nucleotides.

thionucleotides ^{6}sG and ^{4}sU are preferred due to their simple molecular structure, relative stability, and high reactivity. In addition, ^{6}sG and ^{4}sU only differ from the corresponding G and C nucleotides by a single atomic substitution, the replacement of a nucleobase oxygen by sulphur, thus reducing the risk for perturbation of the RNA structure. Exposure of the thionucleotides to UV light produces a sulphur radical that can react very efficiently with functional groups that are in proximity (in proteins or nucleic acids). To increase crosslinking specificity, low energy (360 nm) irradiation is preferred, yielding "zero-length" crosslinks [13] that represent direct contacts (in a 2-Å sphere) of the nucleotide of interest and protein or nucleic acid. Thionucleotides can be incorporated by either random incorporation or site-specific substitution. *In vitro* transcription is a convenient way to incorporate statistically and randomly ^{6}sG and ^{4}sU. This way of incorporation allows a rapid screening of the potential interactions over the entire RNA transcript. Site-specific incorporation by transcription or RNA ligation eliminates the ambiguity of the origin of crosslinking and generally improves the experimental signal since all the RNA molecules are containing the photo agent at a single or very few positions. It allows accurate probing of any position, and generally increases the experimental signal since all the RNA molecules contain a photo agent at a single position rather than spread throughout the molecule, however it is time- and labour-consuming since it requires the synthesis of site-specifically modified RNA. A simplified variant of this approach consists in the synthesis of short RNA fragments containing a single G or U residue that is substituted by ^{6}sG or ^{4}sU during *in vitro* transcription. For instance a set of mRNAs that contained single ^{4}sU at unique positions between −26 and +11, relative to the A of the initiation codon, were synthesized to explore mRNA contacts in the ribosome mRNA channel during translation initiation [14]. Similarly, TISU mRNA fragments containing single unique ^{4}sU between nts −4 and +8 were synthesized to identify ribosomal proteins interacting with TISU sequence [15].

Here we describe another method of site-specific incorporation of photo agent in the m^{7}G cap at the 5′ end of any mRNA. The thionucleotide ^{6}sG has been used for mapping the cap-binding pocket in the mouse histone H4 mRNA [5]. Indeed, the m^{7}G-cap of eukaryotic mRNAs is a modified guanine nucleotide connected to the mRNA via an unusual 5′ to 5′ triphosphate linkage. *In vivo*, it is added co-transcriptionally to the mRNA. *In vitro*, the modified cap is added post-transcriptionally by the Vaccine Capping Enzyme (VCE) using the ScriptCap™ m^{7}G Capping System (Epicentre Biotechnologies) (see Section 2.2). During this process, ^{6}sGTP is substituted to the usual GTP, it is subsequently modified to form a m^{7}s^{6}G-cap structure. In that form, the m^{7}s^{6}G-cap structure is an effective structural probe that combines high crosslinking efficiency, site-specific incorporation with high level of incorporation. Then, the modified mRNA is subjected to UV at 312 nm exposure and detection of the crosslinks is subsequently performed by primer extension with [^{32}P] labelled oligonucleotides and polyacrylamide gel separation as described previously [6,5] or by a an alternative protocol that we are here describing in Section 2.3.

On the secondary structure representation of the three-way helix junction (TWJ) domain of H4 mRNA we observed that capped-wild-type H4 mRNA exhibits strong crosslinks in the stem I of the TWJ, involving, in order of decreasing intensity, nts 41, 42, 43, 44 in addition to nts 103, 104, 105 and 106 on the other strand of the stem (Fig. 2). This strongly supports the hypothesis that these nts form the core of the cap-binding pocket (CBP) as previously described [5]. Next, we performed 8 progressive deletions

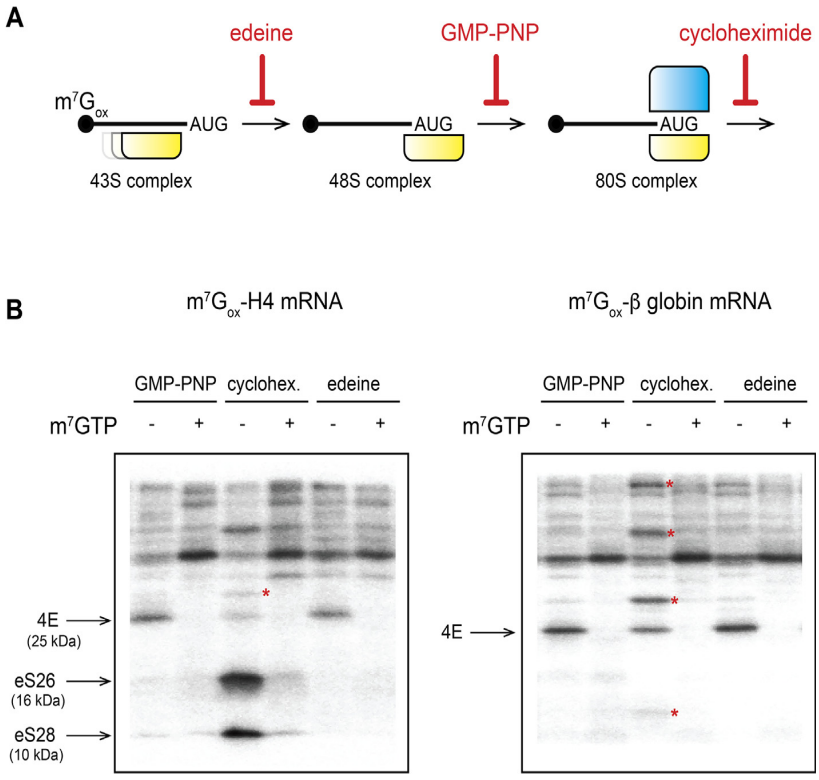


Fig. 3. Chemical crosslinking of oxidised histone H4 and β-globin mRNA. A. Cap-dependent translation initiation is inhibited by specific inhibitors (shown in red). Edeine prevents codon-anticodon interaction of Met-tRNA$_{i}^{Met}$ on AUG codon and thereby leads to 43S accumulation. GMP-PNP, is a non-hydrolysable GTP analogue that blocks eIF5B GTPase activity, it leads to 48S accumulation on the AUG start codon. Cycloheximide inhibits translocation after 60S joining and triggers 80S formation on the AUG start codon. B. Chemical crosslinking of ^{32}P-labelled oxidised histone H4 mRNA and β-globin mRNA in RRL in the presence of GMP-PNP, cycloheximide, edeine with and without m^{7}GDP. Characterized cap-specific crosslinks are indicated by arrows. Uncharacterized cap-specific crosslinks are indicated by red asterisks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the 5′ UTR and checked whether this changed the cap-crosslinking capacity. The truncated mRNA fragments were synthesized starting from shortened truncated H4 genes and capped as the wild-type mRNA with a m^7s^6G-cap structure using the ScriptCap™ m^7G Capping System. One nt deletion induced a severe effect with a near disappearance of the crosslinks of nts 41–44 and new crosslinks at nts 93–94, 96–97 whereas crosslinks at nts 102–103 and 105–106 are maintained. This suggests that nts 93–97 are spatially close of the stem I in the TWJ and can both react with mutant −1 nt of H4 mRNA. Deletions −2 to −8 nts show that the strong crosslink at positions 41–42 is definitively lost and confirm the vicinity of nts 93–97 and nts 103–106 (Supplementary Fig. 1).

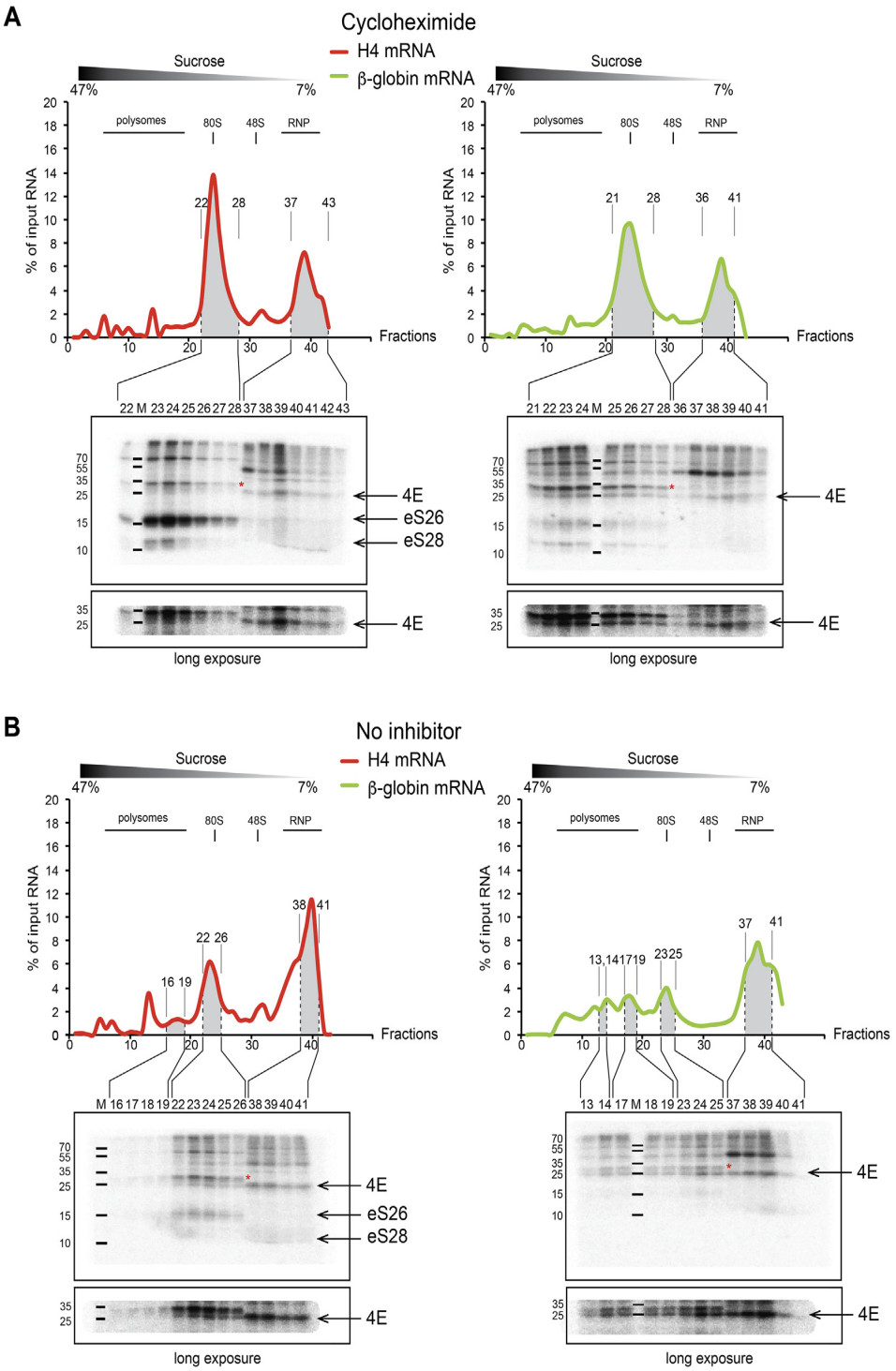


Fig. 4. Chemical crosslinking of oxidised histone H4 and β-globin mRNA and sucrose fractionation. A. Sucrose gradient fractionation of cycloheximide-blocked translation in RRL. The percentage of input mRNA is measured by Cerenkov counting of the sucrose gradient fractions. Proteins contained in fractions that are shown in grey are precipitated and analysed by SDS-PAGE. Characterized cap-specific crosslinks are indicated by arrows; red asterisks indicate uncharacterized cap-specific crosslinks. B. Sucrose gradient fractionation of translation in RRL. The percentage of input mRNA is measured by Cerenkov counting of the sucrose gradient fractions. Proteins contained in fractions that are shown in grey are precipitated and analysed by SDS-PAGE. Characterized cap-specific crosslinks are indicated by black arrows, uncharacterized cap-specific crosslinks are indicated by red asterisks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Most of these mutants exhibit strong crosslink reactivity around nts 104–105. After a 3 nt-deletion, the mutants exhibit different reactivities spread over nts 13 and 30 that indicate that short UTRs can react with closer regions of the mRNA structure (Supplementary Fig. 1). Following these progressive deletions of the UTR, we performed six additional deletions that brought the reactive cap at positions 16, 22, 27, 29, 31 and 37 of H4 mRNA. Except deletion 27, all the mutants lose their ability to react with the original cap-binding site in the pocket 41–44 and 104–106. With these mutants, new positions are crosslinked, providing new data on their position versus the modified cap. Surprisingly, mutant −27 which has the cap at the bottom of stem I, recovered the strong crosslinking activity of nts 41–46 but no reactivity on the other strand of the stem. This indicates that the cap of mutant −27 does not bind the CBP in the native manner since it does not react as the wild-type mRNA does with nts 103–106. The low energy crosslinking experiments might be useful to detect close interactions such as the interaction of the m^7Gcap in the cap-binding pocket of H4 mRNA. In the H4 case, the specificity of the binding pocket was demonstrated by performing a competition assay with m^7GpppG cap analogue [5]. Deletion experiments of nts of the 5′ UTR revealed the spatial proximity of the CBP of stem I and stem III where recurrent crosslinks were found. One major drawback of the deletion approach may concern either the alteration of the folding of the mRNA structure or the severe reduction of reactivity observed with several mutants. For instance, we repetitively observed low reactivity of several independent preparations of mutants −5, −16 and −22 nts. Whether this is due to an unusual high flexibility of the RNA sequence harbouring the 5′ cap resulting in a higher exposure to the solvent side remains to be explored.

3.2. Chemical crosslinking of the cap in RRL

The next issue we addressed was the position of the cap in H4 mRNA during the translation initiation process. For that purpose, we performed chemical cross-linking experiments using mRNA that contain a radiolabelled m^7Gcap that has been oxidised. This histone H4 mRNA transcript was then incubated in RRL in the presence of edeine, GMP-PNP or cycloheximide and then further reduced to induce covalent bonds with proteins that are in contact with the cap. As a control, we performed the same experiment with radioactive oxidised capped β-globin mRNA. Crosslinking experiments in the presence of an excess of m^7GDP as a competitor allowed us to discriminate cap-specific crosslinks that are abrogated versus unspecific bonds that are still present with the competitor (Fig. 3). In the case of H4 mRNA, we detected specific crosslinks of eIF4E, eS26 and eS28 that were identified with specific antibodies as previously described [7]. The crosslinks of eIF4E are detected with edeine and GMP-PNP, but are drastically decreased with cycloheximide. This suggests that eIF4E is bound on the cap on the 43S and 48S complexes but is released from the 80S complex. On the contrary, eIF4E is bound on the cap on β-globin mRNA and remained bound after 80S assembly. Concerning eS26 and eS28, no specific crosslinks are detected on β-globin mRNA. Altogether, these results indicate that in the case of H4 mRNA translation, eIF4E is bound on the cap prior and after 48S complex formation, then eIF4E is released from the cap and is replaced by eS26 and eS28 in the 80S complex. In contrast, eIF4E stayed bound on the cap of β-globin mRNA during the whole translation initiation process. However, the major drawback of this experiment is that the whole input oxidised mRNA is most probably not fully assembled in translation initiation complexes suggesting that a part of the oxidised mRNA is in fact crosslinked with soluble proteins that are also visualised here as attested by the radioactive count profiles (Fig. 4). In other words, the detected crosslinks could also be due to RNPs components in contact with the cap rather than ribosomal complex proteins.

3.3. Translation initiation step-resolved chemical crosslinking of the cap

In order to unambiguously assign the crosslinked proteins to translation initiation complexes, we coupled the chemical crosslinking approach with sucrose gradient sedimentation. The reduced proteins contained in each sucrose gradient fraction are precipitated and analysed by SDS-PAGE (Fig. 4). When translation was blocked with cycloheximide, the eS26 and eS28 crosslinks with H4 mRNA co-sedimented with 80S fractions confirming that the cap is indeed crosslinked to ribosomal proteins that are embedded in the fully assembled ribosome (Fig. 4A). In contrast, eIF4E crosslink was found in the soluble RNP fractions at the top of the gradient. Interestingly, with β-globin mRNA, eIF4E crosslinks were found both in the soluble RNP fractions and in the 80S fractions. We also performed the same experiments without any inhibitor. As expected, we obtained polysomes indicating ongoing translation (Fig. 4B). eIF4E-crosslinks on β-globin mRNA were still found in soluble RNP fractions and in 80S fractions. However, the results are different for histone H4 mRNA, when translation is active, eIF4E-crosslinks are also detected in 80S fractions. All together, these results allowed us to propose a dynamic model for the cap and eIF4E fates during translation initiation and elongation (Fig. 5). In the case of β-globin mRNA, eIF4E is bound to the cap

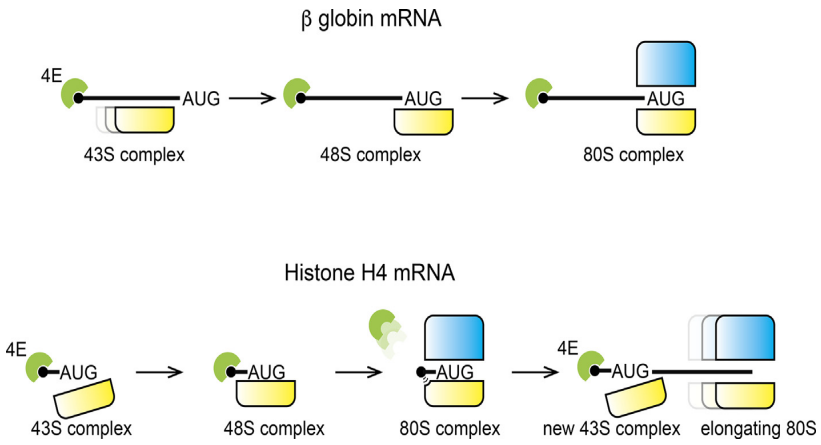


Fig. 5. Model of the cap fate during translation of histone H4 mRNA and β-globin mRNA. During β-globin mRNA translation, eIF4E remains bound on the cap during the whole process. On the contrary, H4 mRNA translation requires eIF4E binding/release cycles to allow efficient ribosome assembly.

for efficient 43S recruitment to promote 5′-3′ scanning. After 48S assembly and 80S formation on the AUG start codon, eIF4E stays on the cap in order to promote further recruitment of the next 43S particle. Histone H4 mRNA is translated by a rather different mechanism. First, eIF4E is bound to the 5′ cap for 43S tethering on the AUG start codon without any scanning [7,5]. At the 48S stage, eIF4E is still bound on the cap, but 80S formation requires the release of eIF4E to allow specific contacts with eS28 and eS26 on the small ribosomal subunit. This is correlated with the fact that the 5′ UTR of H4 mRNA is only 9-nt long. Indeed, such a short 5′ UTR is seemingly incompatible with eIF4E bound to the cap and an assembled 80S because the large subunit would clash with eIF4E. Indeed, many functional studies by toe- and footprinting experiments have demonstrated that the 80S covers ~14 to 19 nucleotides upstream of the AUG start codon. This is probably why eIF4E has to be released from the cap in order to allow 60S subunit joining for histone H4 translation initiation. However, during translation elongation, eIF4E-crosslinks are also detected in the 80S fractions. This suggests that once the 80S ribosome has proceeded to elongation, eIF4E binds to the cap that is now accessible again to enable translation initiation for the next ribosome by tethering. In conclusion our approach allowed us to monitor the 'breathing' of eIF4E which undergo successive cycles of binding/release during the translation process.

4. Conclusion

By combining chemical and UV crosslinking approaches of the cap, we have been able to track the proteins bound to cap that is present on an mRNA of interest during the translation process. By identifying the proteins that are in proximity to the cap, these methods give interesting insights on the dynamic interplay between eIF, the ribosome and mRNA. It will be of interest to further use these approaches on other mRNAs in order for example to assess the influence of 5′UTR sequence. The accurate position of the cap and cap-binding factors such as eIF4F during scanning is still poorly understood. Crosslinking of the cap is a method of choice to address such fundamental aspects of translation initiation.

Acknowledgments

We are grateful to Liz Grayhack and Jeff Coller for their kind invitation to write this article. We are also grateful to Bruno Sargueil for YpGlo plasmid. This work was supported by CNRS, Université de Strasbourg and Agence Nationale pour la Recherche (ANR-11-SVSE802501; ANR-17-CE11-0024-01; ANR-17-CE12-0025-01).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ymeth.2017.12.019.

References

[1] A.G. Hinnebusch, The scanning mechanism of eukaryotic translation initiation, Annu. Rev. Biochem. 83 (2014) 779–812, https://doi.org/10.1146/annurev-biochem-060713-035802.
[2] A.G. Hinnebusch, Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation, Trends Biochem. Sci. 42 (2017) 589–611, https://doi.org/10.1016/j.tibs.2017.03.004.
[3] S. Jaeger, S. Barends, R. Giegé, G. Eriani, F. Martin, Expression of metazoan replication-dependent histone genes, Biochimie 87 (2005) 827–834, https://doi.org/10.1016/j.biochi.2005.03.012.
[4] V.S. Meier, R. Bohni, D. Schumperli, Nucleotide sequence of two mouse histone H4 genes, Nucl. Acids Res. 17 (1989) 795.
[5] F. Martin, S. Barends, S. Jaeger, L. Schaeffer, L. Prongidi-Fix, G. Eriani, Cap-assisted internal initiation of translation of histone H4, Mol. Cell. 41 (2011) 197–209.
[6] M. Nowakowska, J. Kowalska, F. Martin, A. D'Orchymont, J. Zuberek, M. Lukaszewicz, E. Darzynkiewicz, J. Jemielity, Cap analogs containing 6-thioguanosine – reagents for the synthesis of mRNAs selectively photo-crosslinkable with cap-binding biomolecules, Org. Biomol. Chem. 12 (2014) 4841–4847, https://doi.org/10.1039/c4ob00059e.
[7] F. Martin, J.-F. Ménétret, A. Simonetti, A.G. Myasnikov, Q. Vicens, L. Prongidi-Fix, S.K. Natchiar, B.P. Klaholz, G. Eriani, Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation, Nat. Commun. 7 (2016) 12622, https://doi.org/10.1038/ncomms12622.
[8] F. Karabiber, J.L. McGinnis, O.V. Favorov, K.M. Weeks, QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis, RNA 19 (2013) 63–73, https://doi.org/10.1261/rna.036327.112.
[9] N. Sonenberg, A.J. Shatkin, Reovirus mRNA can be covalently crosslinked via the 5' cap to proteins in initiation complexes, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 4288–4292.
[10] N. Garreau de Loubresse, I. Prokhorova, W. Holtkamp, M.V. Rodnina, G. Yusupova, M. Yusupov, Structural basis for the inhibition of the eukaryotic ribosome, Nature 513 (2014) 517–522, https://doi.org/10.1038/nature13737.
[11] N.K. Gray, M.W. Hentze, Regulation of protein synthesis by mRNA structure, Mol. Biol. Rep. 19 (1994) 195–200.
[12] T. Schneider-Poetsch, J. Ju, D.E. Eyler, Y. Dang, S. Bhat, W.C. Merrick, R. Green, B. Shen, J.O. Liu, Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin, Nat. Chem. Biol. 6 (2010) 209–217, https://doi.org/10.1038/nchembio.304.
[13] A. Favre, G. Moreno, M.O. Blondel, J. Kliber, F. Vinzens, C. Salet, 4-thiouridine photosensitized RNA-protein crosslinking in mammalian cells, Biochem. Biophys. Res. Commun. 141 (1986) 847–854, https://doi.org/10.1016/S0006-291X(86)80250-9.
[14] A.V. Pisarev, V.G. Kolupaeva, M.M. Yusupov, C.U.T. Hellen, T.V. Pestova, Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes, EMBO J. 27 (2008) 1609–1621, https://doi.org/10.1038/emboj.2008.90.
[15] O. Haimov, H. Sinvani, F. Martin, I. Ulitsky, R. Emmanuel, A. Tamarkin-Ben-Harush, A. Vardy, R. Dikstein, Efficient and accurate translation initiation directed by TISU involves RPS3 and RPS10e binding and differential eukaryotic initiation factor 1A regulation, Mol. Cell. Biol. 37 (2017) e00150–e217, https://doi.org/10.1128/MCB.00150-17.

2. Study of the structure of the Dicistrovirus C (DCV) 5'UTR by chemical probing methods

2.1. Summary of my contribution to a collaboration

The expertise in RNA structure probing that was acquired during my thesis allowed me to participate in collaboration with the team of Dr. Jean-Luc Imler (IBMC-UPR9022). The goal of this collaboration was to establish a secondary structural model of the DCV Cloverleaf (CL) that is present in the 5'UTR of the genomic RNA. The CL is a structural element that is believed to be essential for viral replication and protection against host 5'-3' exonuclease degradation during immune response (Barton et al., 2001).

- **Structural characterization of the DCV cloverleaf**

Chemical probing of a DCV RNA fragment containing the longer fragment containing the full 5'UTR (CL+IRES) (**Fig. 51**) and CL only (short 5'UTR) (**Fig. 52**) were performed in triplicates to determine the average reactivity value of each nucleotide. These data were superposed with mfold predictions (Mathews et al., 2016) in order to build a 2D model of the DCV cloverleaf. First, we were interested in determining whether the CL and the IRES interact or rather fold independently. So, we performed chemical probing using DMS and CMCT with both RNA constructs. The average of reactivity values was calculated for each nucleotide. Then, we calculated Pearson correlation coefficients (R) for DMS and CMCT profiles which revealed strong correlation between the short and long RNAs ($R > 0.7$). This showed that the DCV cloverleaf and the IRES fold independently.

The secondary structural model of DCV cloverleaf consists of three stem loops (**Fig. 51** and **52**):

- The first stem-loop is predicted to be short and tightly folded. It spans nucleotides (2-42) in the DCV 5'UTR. This stem loop encompasses a protected 5' extremity of the viral genome (ΔG = -16.10 kcal/mol).
- The second predicted stem-loop is longer (43-129) and has few bulging nucleotides (ΔG = -17.80 kcal/mol). A possible intra loop interaction with U_{86} and U_{89} with A_{78} might be an explanation for the absence of reactivity in the nucleotides U_{76} and G_{77} which are located in a loop in our model.

- The third stem-loop spans a long region (123-300) and has several internal bulges with AU-rich base pairs in the stem structure.

Overall, the 2D model shows structured stem-loops with few single-stranded regions. According to this model, we can predict a stable 5' proximal structure that could eventually, in addition to the VpG, prevent 5'-3' exonuclease degradation. Our collaborators are currently working on this hypothesis. This structure shows accordance with the previously proposed protective role of the poliovirus cloverleaf (Barton et al., 2001). Results from this study will be presented in a manuscript that is currently in preparation and my contribution to this work will enable me to be a co-author of this publication.

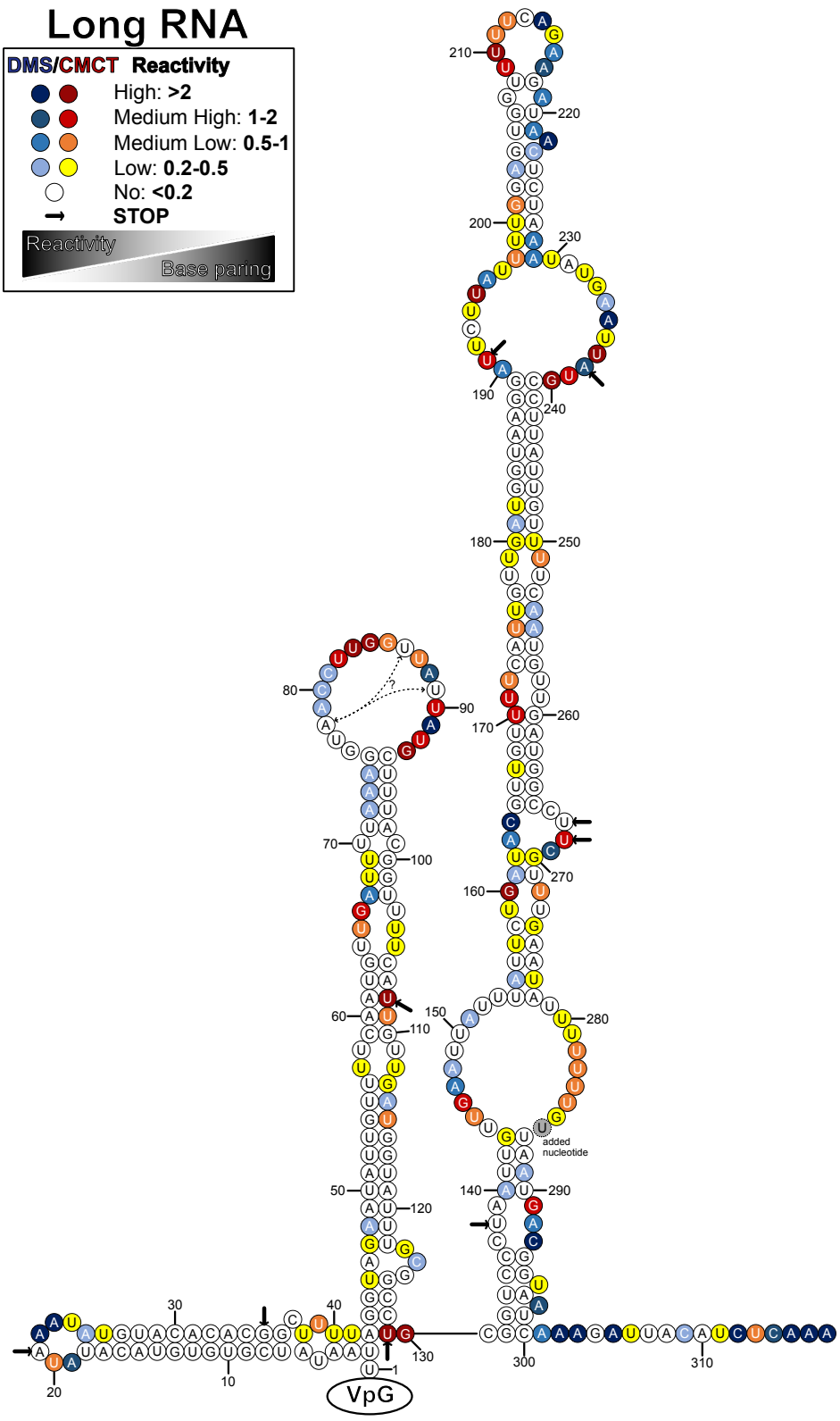


Figure 51: **secondary structural model of DCV cloverleaf determined by chemical probing using DMS and CMCT.** Structure was determined using full length DCV 5'UTR (long RNA). The reactivities are shown as average reactivity calculated from three independent experiments. RT stops from probing are determined by arrows on specific nucleotides.

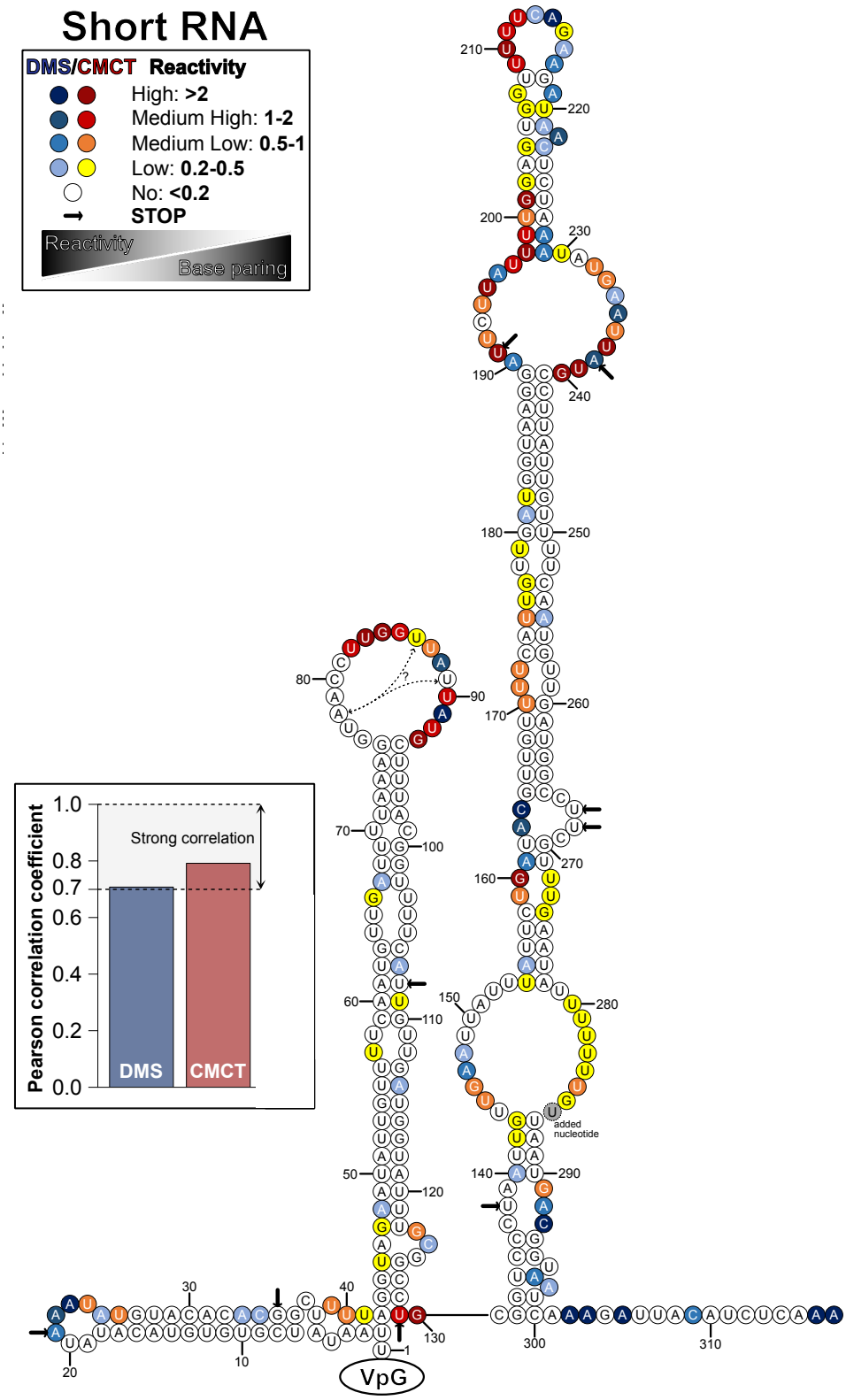


Figure 52: **secondary structural model of DCV cloverleaf determined by chemical probing using DMS and CMCT.** Structure was determined using DCV cloverleaf (short RNA). The reactivities are shown as average reactivity calculated from three independent experiments. On the left is shown the Pearson's correlation coefficient (R) between long and short DCV 5'UTR RNA. RT stops from probing are determined by arrows on specific nucleotides.

Résumé de thèse en français

Inhibition de la traduction des ARN messagers Hox par les éléments de type TIE

Chez les eucaryotes, les ARN messagers sont protégés à leur extrémité 5' par une structure m^7G appelée coiffe et par une extension polyadénylée à l'autre extrémité 3'. Ces protections garantissent la stabilité des ARN messagers dans le cytoplasme en bloquant l'accès aux nombreuses ribonucléases qui coexistent dans ce compartiment cellulaire. Outre cette fonction protectrice, ces modifications sont également utilisées pour d'autres processus cellulaires majeurs. Ainsi, la majorité des ARN messagers cellulaires sont traduits par un mécanisme dit canonique qui utilise ces modifications. En effet, la coiffe m^7G guide l'assemblage de la machinerie traductionnelle sur l'extrémité 5' de l'ARN messager. On appelle ces assemblages, complexes de pré-initiation, ils comprennent la petite sous-unité du ribosome 40S ainsi que des facteurs protéiques auxiliaires appelés « eukaryotic Initiation Factors » ou eIFs. Le complexe ainsi formé va ensuite glisser sur l'extrémité 5' non traduite, du 5' vers le 3', cette étape est appelée « scanning », à la recherche du codon AUG de démarrage de la traduction. Lorsque celui-ci est atteint, les eIFs se décrochent permettant ainsi l'arrimage de la grande sous-unité 60S du ribosome. Le complexe formé est maintenant une particule ribosomique 80S complète apte à démarrer sa fonction de traduction. L'ensemble de ce processus est appelé l'étape d'initiation de la traduction. Etant donné qu'il est amorcé par la coiffe m^7G, on qualifie ce type de mécanisme de traduction de coiffe-dépendante. Il existe cependant d'autres mécanismes alternatifs qui ne nécessitent pas la présence de la coiffe. Dans ce cas, le ribosome est recruté non pas à l'extrémité 5' mais plutôt à l'intérieur d'un ARN messager, grâce à la présence de structures ribonucléiques plus ou moins complexes qui attirent la machinerie traductionnelle. Ces structures sont appelées « Internal Ribosome Entry Sites » ou IRES. Ces éléments IRES recrutent le ribosome sur l'ARNm, ils ne nécessitent donc pas de coiffe en 5' et peuvent s'affranchir de la plupart des eIFs. Puisque ce type de mécanisme ne nécessite pas la présence de la coiffe, on qualifie cette traduction guidée par les IRES de coiffe-indépendante. Initialement découverts dans des génomes de virus à ARN de polarité positive, les IRESs sont également retrouvés dans certains ARN messagers cellulaires particuliers.

Récemment, des éléments IRES ont été trouvés dans la région 5’ non traduite (5'UTR) d’une famille d'ARNm cellulaires Homeobox plus communément appelés Hox (Xue et al., 2015). Les gènes Hox sont très bien conservés parmi les espèces, ils sont classés en quatre groupes (A-D) sur différents chromosomes. Les ARNm Hox codent pour des facteurs transcriptionnels qui contrôlent l’élaboration du squelette dans le corps embryonnaire le long de l'axe tête-queue. De nombreuses études se sont concentrées sur la régulation de l'expression des gènes Hox à la fois au niveau de leur transcription et de leur traduction. Outre un élément IRES, un autre élément régulateur d'ARN appelé Élément Inhibiteur de la Traduction (« Translation Inhibitory Element » ou TIE) a été décrit dans la région 5'UTR des ARNm de Hox (Xue et al., 2015). Ce nouveau type d’élément régulateur inhibe, par un mécanisme jusqu'ici inconnu, la traduction canonique dépendante de la coiffe. L'élément TIE agit en synergie avec l'IRES pour permettre que la traduction de l'ARNm Hox soit finement contrôlée et uniquement pilotée par son mécanisme coiffe indépendant guidé par l’IRES. En effet, l’action conjuguée de ces deux éléments empêche la superposition de deux mécanismes de traduction différents ce qui rendrait la régulation de la traduction complexe et de ce fait moins précise. Le fait que seul le mécanisme guidé par l’IRES soit utilisé permet d’allumer ou d’éteindre l’expression des ARN messager Hox à souhait à des stades précis du développement embryonnaire simplement en activant ou réprimant l’IRES.

L'objectif de mon projet de doctorat était de comprendre le mode d'action de deux éléments TIE a3 et a11 localisés respectivement dans les ARNm Hox a3 et a11. Pour commencer, nous avons mis au point un système de traduction *in vitro* acellulaire à l'aide de lysats de réticulocytes de lapin pour tester l'inhibition de la traduction orchestrée par le TIE avec un gène rapporteur qui exprime la luciférase Renilla. L'inhibition par le TIE a été efficacement reproduite *in vitro* à l’aide de ce test (**Fig. 1A**). Ceci nous également permis de déterminer le domaine minimal requis pour l’inhibition pour les TIE a3 et a11 (**Fig. 1B**).

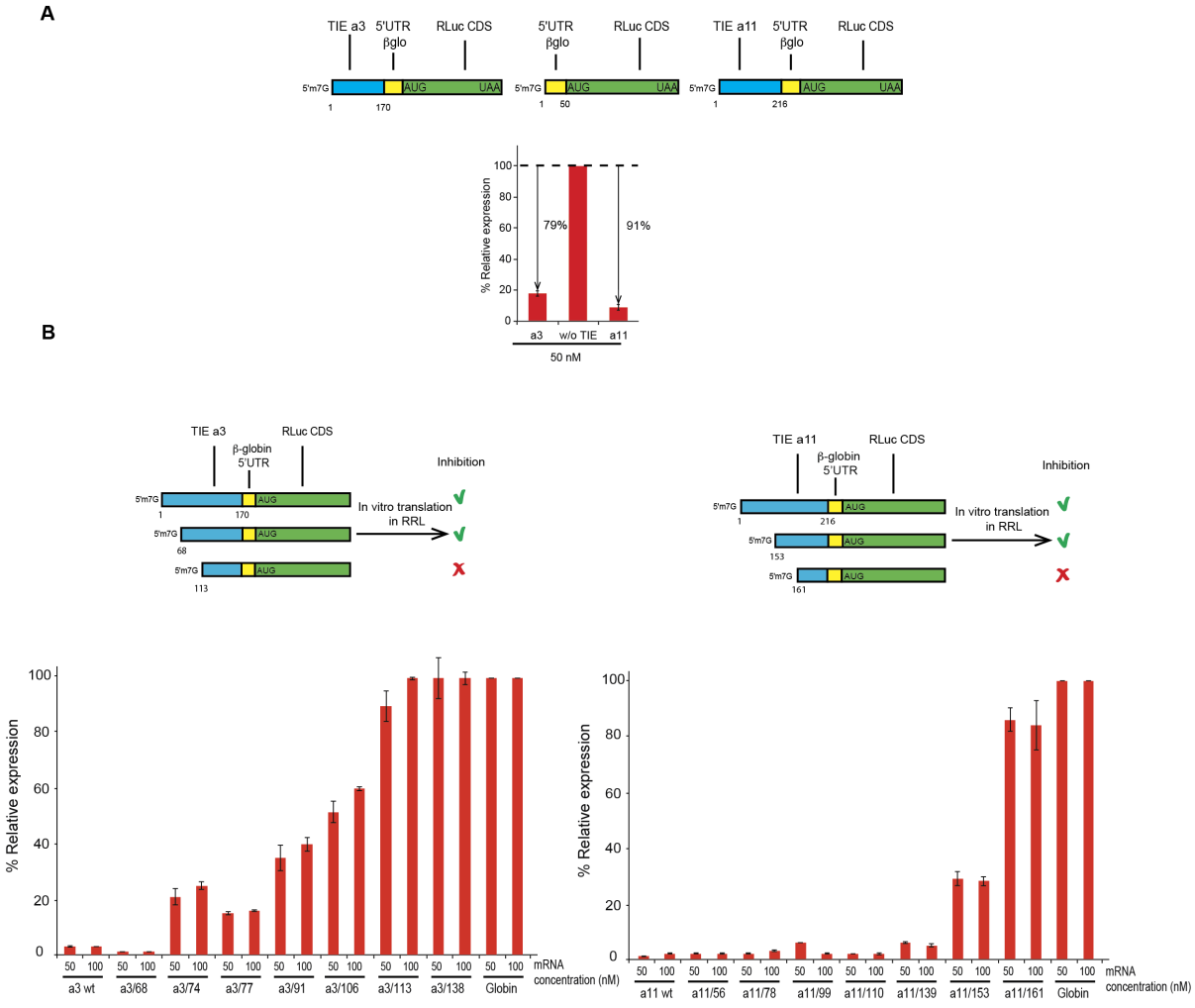


Figure 1: **(A)** étude de l'inhibition de la traduction médiée par les éléments TIE a3 et a11 à l'aide d'extraits de traduction acellulaires. **(B)** Cartographie du domaine minimal requis pour l'inhibition de la traduction par a3 (à gauche) et a11 (à droite).

Ensuite, nous nous sommes intéressées aux structures secondaires supposées des deux éléments TIE. La méthode que nous avons utilisée est la méthode de sondage chimique à l'aide des réactifs DMS et CMCT. Ces réactifs modifient les nucléotides présents dans les régions monocaténaires et non appariées de l'ARN. Le DMS réagit avec les résidus 'A' et 'C' tandis que le CMCT modifie les résidues 'G' et 'U'. Ces expériences nous ont permis d'établir un modèle de la structure secondaire des éléments TIE de a3 (**Fig. 2A**) et a11 (**Fig. 2B**).

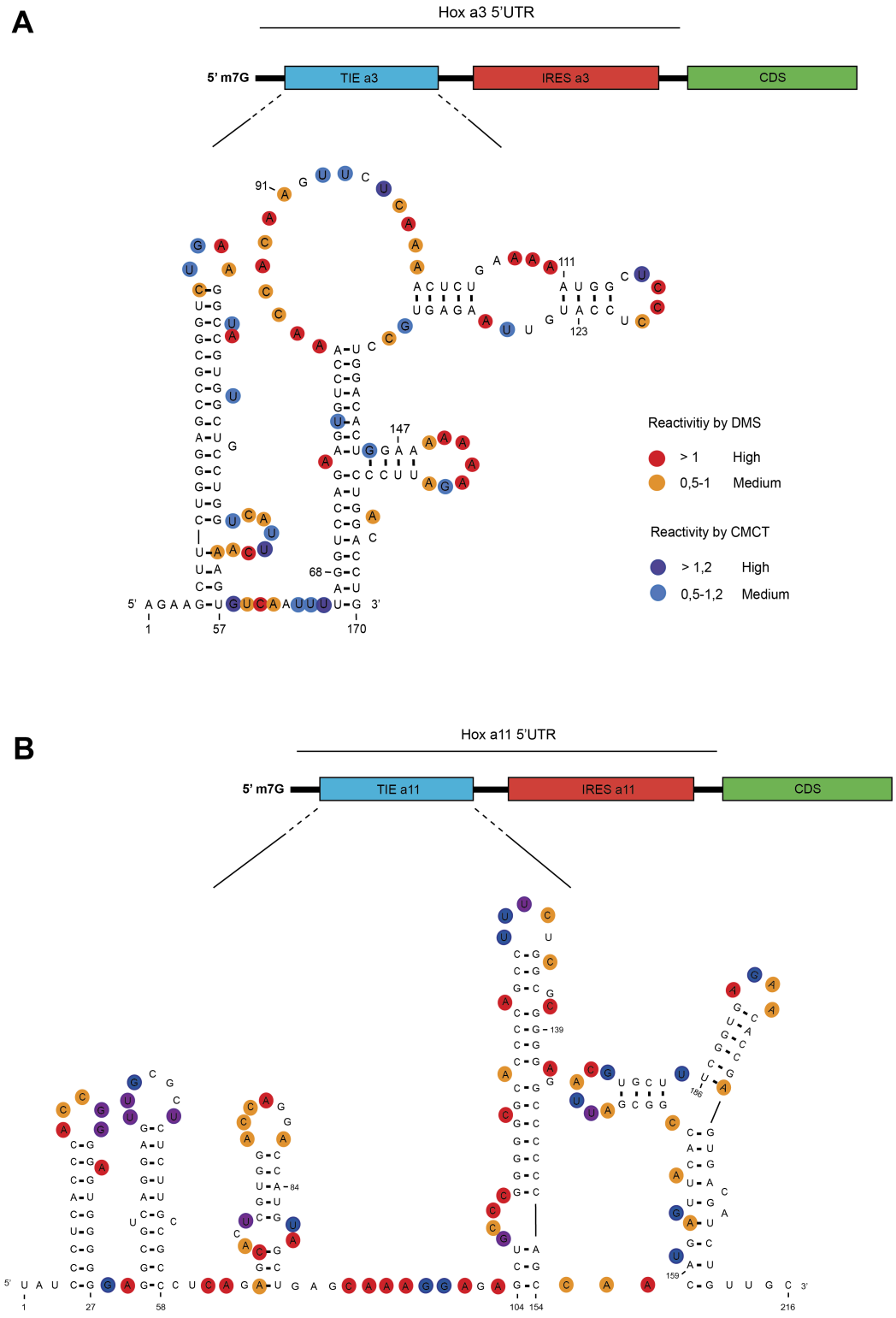


Figure 2: modèles de la structure secondaire des éléments TIE a3 **(A)** et a11 **(B)** construits à l'aide des expériences de sondages chimiques réalisées avec les réactifs DMS et CMCT.

Puis, par fractionnement sur gradients de sucrose, nous avons étudié l'assemblage du ribosome sur les deux éléments TIE a3 et a11. Grâce à ces études, nous avons pu mettre au

point l'assemblage *in vitro* de complexe de preinitiation assemblés sur les TIE a3 et a11 à partir de transcrits et d'extraits de traduction acellulaires.

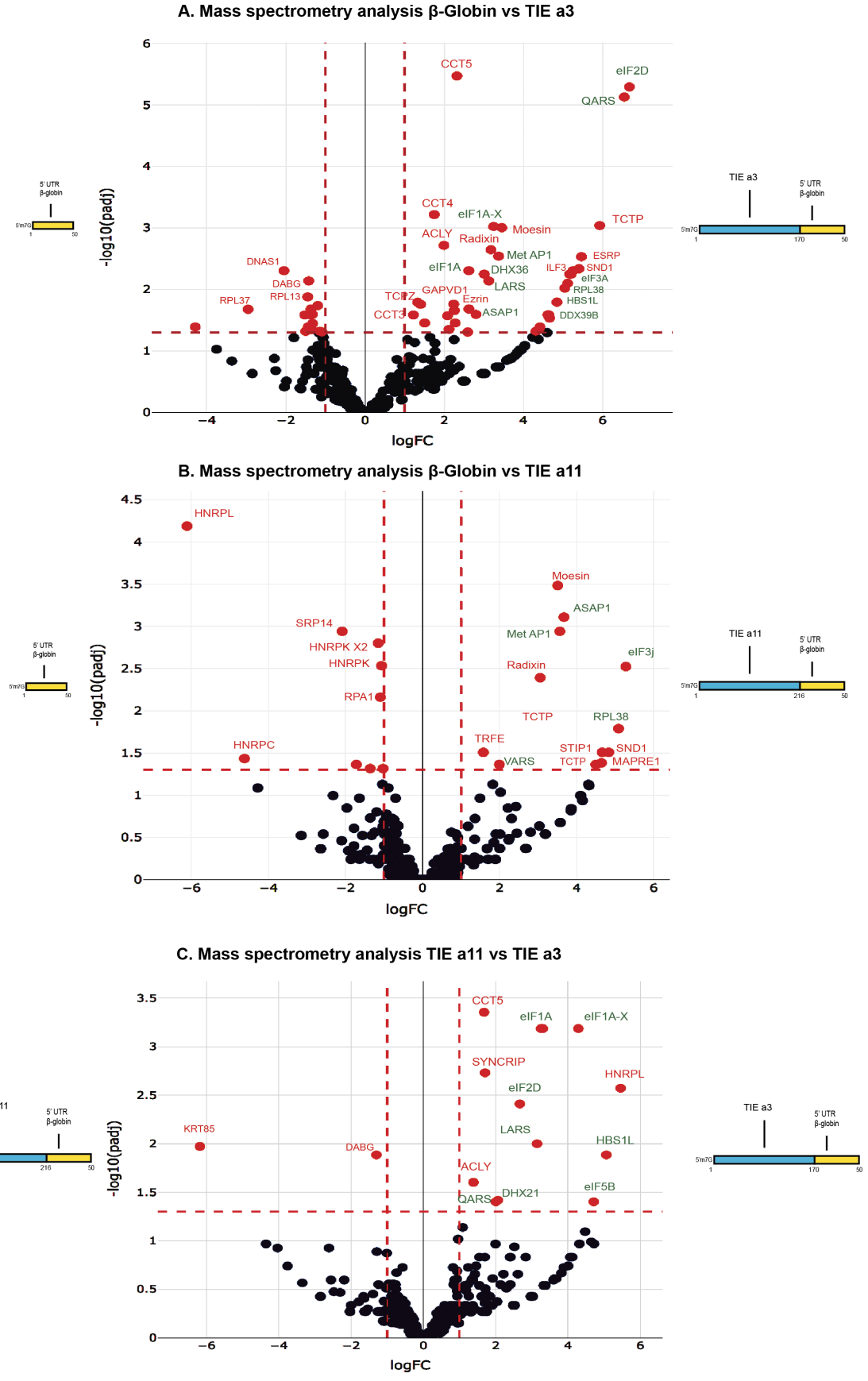


Figure 3 (page précédente): identification des protéines des complexes de pré-initiation assemblés sur les TIE a3 et a11 et sur un ARN contrôle de référence contenant la 5'UTR de la β-globine. Les protéines sont identifées par spectrométrie de masse et les résultats sont présentés sous forme de 'volcano plots' en comparant la composition des complexes deux à deux : **(A)** a3 *versus* 5'UTR β-globine, **(B)** a11 *versus* 5'UTR β-globine et **(C)** a3 *versus* a11.

Nous avons ensuite entrepris la purification des complexes d'initiation de la traduction pour identifier les facteurs impliqués en utilisant une approche développée dans notre laboratoire pour l'analyse de leur composition par spectrométrie de masse (**Fig. 3**). Les études ont été réalisées en comparant la composition des complexes deux à deux. Nous avons comparé les complexes assemblés sur a3 par rapport à notre référence de contrôle à savoir l'extrémité 5'UTR de l'ARN messager de la β-globine (**Fig. 3A**), les complexes sur a11 par rapport à la β-globine (**Fig. 3B**) et enfin nous avons également comparé les deux éléments TIE a3 et a11 entre eux (**Fig. 3C**).

Ces études par spectrométrie de masse nous ont permis de comprendre que les facteurs requis pour l'inhibition de la traduction sont différents pour chaque TIE. Ces données semblent indiquer que l'inhibition de la traduction par les TIE se fait en piégeant le ribosome pendant l'étape de scanning sur des séquences localisées dans les TIE, empêchant de ce fait le ribosome de poursuivre sa course jusqu'au codon d'initiation de la séquence codante des ARN messagers Hox. Nous avons également localisé les sites d'initiation qui permettent de stopper le ribosome en utilisant la technique dite de « toe-printing » qui est basée sur la détection d'arrêts prématurés de la reverse transcriptase pendant l'extension d'une amorce marquée radioactivement, ces arrêts sont induits par la présence du ribosome sur l'ARNm et de ce fait permettent de localiser précisément le ribosome sur l'ARNm. Cette technique nous a permis de mettre en évidence des codons de démarrage de la traduction localisés dans les TIE a3 et a11 et qui marquent le début de ce qu'on appelle des « upstream Open Reading Frame » ou uORFs. Bien que fonctionnant différemment, les deux éléments TIE a3 et a11 utilisent pourtant de manière similaire des séquences de type uORF. Cependant, nous avons démontré que a3 et a11 inhibent la traduction par deux mécanismes moléculaires radicalement différents. L'ensemble de nos données nous ont permis d'établir un modèle du mécanisme d'action de l'inhibition de la traduction des TIE a3 et a11 (**Fig. 4**).

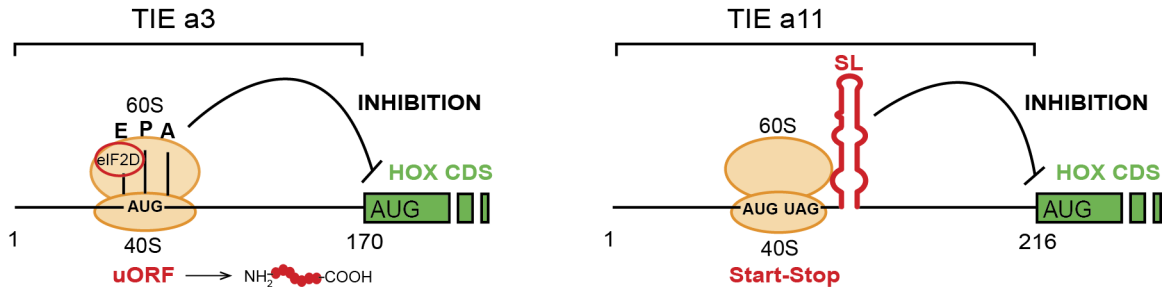


Figure 4: modèle du mécanisme moléculaire de l'inhibition de la traduction cap-dépendante contrôlée par l'élément TIE a3 (à gauche) et a11 (à droite).

Alors que a3 recrute des facteurs supplémentaires pour inhiber la traduction tel que eIF2D, a11 quant à lui entrave la progression d'un complexe ribosomique 80S grâce à la présence de structures secondaires (SL) très stables. Le TIE a3 inhibe la traduction en raison de la présence d'un cadre de lecture ouvert (uORF) en amont du codon AUG à la position 111. Cette uORF est localisée et traduite grâce au facteur eIF2D, un facteur d'initiation non canonique GTP-indépendant. Nous avons également pu détecter le peptide d'une taille de 9 KDa produit à partir de la traduction de l'uORF. En revanche, le TIE a11 fonctionne différemment, il séquestre le ribosome 80S en bloquant sa progression grâce à une structure SL en épingle à cheveux riche en GC très stable (ΔG = -25,00 kcal/mol). Cette structure peut également empêcher le « scanning » des ribosomes lorsqu'elle est transplantée dans la région 5'UTR d'un autre ARNm (ARNm codant pour la β-globine). Dans le TIE a11, le ribosome 80S est bloqué sur une séquence 'START-STOP' composée d'un codon de démarrage AUG suivi d'un codon STOP. L'action combinée de cette séquence minimale et de la structure SL localisée à proximité bloque le ribosome 80, l'empêche d'avancer et de se dissocier, interdisant ainsi le passage d'autres complexes de pré-initiation. C'est de cette manière que la traduction de la phase codante principale Hox est inhibée.

Ces résultats ont été confirmés *in vivo* dans deux lignées cellulaires : des cellules humaines HEK293T et des cellules murines embryonnaires C3H10T1/2. Pour cela, nous avons utilisé un plasmide rapporteur codant pour la luciférase. Nous y avons inséré dans l'extrémité 5'UTR de ce gène rapporteur, les TIE a3 et a11. Ces constructions nous ont permis de confirmer les résultats que nous avions obtenus avec nos systèmes de traduction acellulaires. Par exemple, nous avons vérifié l'implication des uORFs dans le mode de fonctionnement des TIE a3 et a11. A l'avenir, nous projetons également d'utiliser ce même test pour confirmer l'implication du facteur eIF2D dans le mode d'action du TIE a3.

En conclusion, nous avons déchiffré les mécanismes d'inhibition traductionnelle par les éléments TIE a3 et a11. La TIE a3 inhibe grâce à la présence d'une uORF, elle est utilisée par le ribosome grâce au facteur eIF2D. En revanche, le TIE a11 bloque la progression 80S grâce à une structure en épingle à cheveux très stable. Mon travail de thèse ouvre de nombreuses perspectives. Parmi les nombreuses questions ouvertes, il serait notamment intéressant de comprendre comment les éléments IRES sont-ils régulés pendant le développement embryonnaire pendant lequel les protéines Hox sont si déterminantes. D'autre part, une autre question fondamentale serait focalisée sur la recherche d'autres éléments de type TIE dans les autres ARN messagers Hox ou encore dans d'autres régions 5'UTR provenant d'ARNm cellulaires autre que les Hox.

Mots-clefs: traduction, ribosome, ARNm de Hox, TIE, IRES

Liste d'articles:

Accepté:

- Tracking the m^7G-cap during translation initiation by crosslinking methods.
Lauriane Gross, Laure Schaeffer, **Fatima Alghoul**, Hassan Hayek, Christine Allmang, Gilbert Eriani, Franck Martin
Methods 2018 DOI: 10.1016/j.ymeth.2017.12.019. Epub 2018 Jan 4.

Chapitre d'ouvrage:

- Immunoprecipitation methods to isolate messenger ribonucleoprotein complexes (RNPs).
Hassan Hayek, Lauriane Gross, **Fatima Alghoul**, Franck Martin, Gilbert Eriani, Christine Allmang
Chapter 2, Volume 2, Advances in Experimental Medicine and Biology Book.

En préparation:

- Translation Inhibitory Elements from Hox a3 and a11 mRNAs use upstream ORFs to repress cap-dependent translation
Fatima Alghoul, Laure Schaeffer, Gilbert Eriani, Franck Martin

- Secondary structure of DCV 5'UTR by chemical probing
Loïc Talide, **Fatima Alghoul**, Franck Martin, Jean-Luc Imler, Carine Meignin

- RNA structural probing with DMS and CMCT chemical reagents using fluorescent probes
Fatima Alghoul, Christine Allmang, Gilbert Eriani, Franck Martin

Invitation pour Methods in Molecular Biology- Mars 2020

REFERENCES

A

Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol. Cell *72*, 727-738.

Advani, V.M., and Dinman, J.D. (2016). Reprogramming the genetic code: The emerging role of ribosomal frameshifting in regulating cellular gene expression. BioEssays *38*, 21–26.

Akulich, K.A., Andreev, D.E., Terenin, I.M., Smirnova, V. V., Anisimova, A.S., Makeeva, D.S., Arkhipova, V.I., Stolboushkina, E.A., Garber, M.B., Prokofjeva, M.M., et al. (2016). Four translation initiation pathways employed by the leaderless mRNA in eukaryotes. Sci. Rep. *6*, 37905.

Alasti, F., Sadeghi, A., Sanati, M.H., Farhadi, M., Stollar, E., Somers, T., and Van Camp, G. (2008). A Mutation in HOXA2 Is Responsible for Autosomal-Recessive Microtia in an Iranian Family. Am. J. Hum. Genet. *82*, 982–991.

Alexander, T., Nolte, C., and Krumlauf, R. (2009). Hox genes and segmentation of the hindbrain and axial skeleton. Annu. Rev. Cell Dev. Biol. *25*, 431–456.

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

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

Andino, R., Rieckhof, G.E., and Baltimore, D. (1990). A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. Cell *63*, 369–380.

Andreev, D.E., Fernandez-Miragall, O., Ramajo, J., Dmitriev, S.E., Terenin, I.M., Martinez-Salas, E., and Shatsky, I.N. (2007). Differential factor requirement to assemble translation initiation complexes at the alternative start codons of foot-and-mouth disease virus RNA. RNA *13*, 1366–1374.

Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., and Penalva, L.O.F. (2012). Before it gets started: Regulating translation at the 5′UTR. Comp. Funct. Genomics *2012*, 1–8.

Asano, K., Vornlocher, H.P., Richter-Cook, N.J., Merrick, W.C., Hinnebusch, A.G., and Hershey, J.W.B. (1997). Structure of cDNAs encoding human eukaryotic initiation factor 3 subunits. Possible roles in rna binding and macromolecular assembly. J. Biol. Chem. *272*, 27042–27052.

Atkinson, G.C., Baldauf, S.L., and Hauryliuk, V. (2008). Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. BMC Evol. Biol. *8*, 290.

B

Babendure, J.R., Babendure, J.L., Ding, J.-H., and Tsien, R.Y. (2006). Control of mammalian translation by mRNA structure near caps. RNA *12*, 851–861.

Balasubramanian, S., and Neidle, S. (2009). G-quadruplex nucleic acids as therapeutic targets. Curr. Opin. Chem. Biol. *13*, 345–353.

Barbosa, C., Peixeiro, I., and Romão, L. (2013). Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. PLoS Genet. *9*, e1003529.

Barrett, S.P., and Salzman, J. (2016). Circular RNAs: analysis, expression and potential functions. Development *143*, 1838–1847.

Barton, D.J., O'Donnell, B.J., and Flanegan, J.B. (2001). 5′ cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. EMBO J. *20*, 1439–1448.

Bastide, A., Karaa, Z., Bornes, S., Hieblot, C., Lacazette, E., Prats, H., and Touriol, C. (2008). An upstream open reading frame within an IRES controls expression of a specific VEGF-A isoform. Nucleic Acids Res. *36*, 2434–2445.

van Batenburg, F.H.D. (2001). PseudoBase: structural information on RNA pseudoknots. Nucleic Acids Res. *29*, 194–195.

Becker, M., Kuhse, J., and Kirsch, J. (2013). Effects of two elongation factor 1A isoforms on the formation of gephyrin clusters at inhibitory synapses in hippocampal neurons. Histochem. Cell Biol. *140*, 603–609.

Ben-Asouli, Y., Banai, Y., Pel-Or, Y., Shir, A., and Kaempfer, R. (2002). Human interferon-γ mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. Cell *108*, 221–232.

Berry, K.E., Waghray, S., Mortimer, S.A., Bai, Y., and Doudna, J.A. (2011). Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. Structure *19*, 1456–1466.

Bhardwaj, U., Powell, P., and Goss, D.J. (2019). Eukaryotic initiation factor (eIF) 3 mediates Barley Yellow Dwarf Viral mRNA 3′–5′ UTR interactions and 40S ribosomal subunit binding to facilitate cap-independent translation. Nucleic Acids Res. *47*, 6225–6235.

Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M.J., and Namy, O. (2015). New insights into stop codon recognition by eRF1. Nucleic Acids Res. *43*, 3298–3308.

Błaszczyk, L., and Ciesiołka, J. (2011). Secondary structure and the role in translation initiation of the 5'-terminal region of p53 mRNA. Biochemistry *50*, 7080–7092.

Blau, L., Knirsh, R., Ben-Dror, I., Oren, S., Kuphal, S., Hau, P., Proescholdt, M., Bosserhoff, A.-K., and Vardimon, L. (2012). Aberrant expression of c-Jun in glioblastoma by internal ribosome entry site (IRES)-mediated translational activation. Proc. Natl. Acad. Sci. *109*, E2875–E2884.

Boeras, I., Song, Z., Moran, A., Franklin, J., Brown, W.C., Johnson, M., Boris-Lawrie, K., and Heng, X. (2016). DHX9/RHA Binding to the PBS-Segment of the Genomic RNA during HIV-1 Assembly Bolsters Virion Infectivity. J. Mol. Biol. *428*, 2418–2429.

Boncinelli, E. (1997). Homeobox genes and disease. Curr. Opin. Genet. Dev. *7*, 331–337.

Bonnal, S., Boutonnet, C., Prado-Lourence, L., and Vagner, S. (2003). IRESdb: The internal ribosome entry site database. Nucleic Acids Res. *31*, 427–428.

Bornes, S., Prado-Lourenco, L., Bastide, A., Zanibellato, C., Iacovoni, J.S., Lacazette, E., Prats, A.C., Touriol, C., and Prats, H. (2007). Translational induction of VEGF internal ribosome entry site elements during the early response to ischemic stress. Circ. Res. *100*, 305–308.

Bromleigh, V.C., and Freedman, L.P. (2000). p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells. Genes Dev. *14*, 2581–2586.

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature *524*, 493–496.

Brown, C.M., Stockwell, P.A., Trotman, C.N.A., and Tate, W.P. (1990). Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. Nucleic Acids Res. *18*, 6339–6345.

Browning, K.S., Gallie, D.R., Hershey, J.W., Hinnebusch, A.G., Maitra, U., Merrick, W.C., and Norbury, C. (2001). Unified nomenclature for the subunits of eukaryotic initiation

factor 3. Trends Biochem. Sci. *26*, 284.

Bugaut, A., and Balasubramanian, S. (2012). 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. Nucleic Acids Res. *40*, 4727–4741.

Buratti, E., Tisminetzky, S., Zotti, M., and Baralle, F.E. (1998). Functional analysis of the interaction between HCV 5'UTR and putative subunits of eukaryotic translation initiation factor eIF3. Nucleic Acids Res. *26*, 3179–3187.

Bürglin, T.R. (1998). The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? Dev. Genes Evol. *208*, 113–116.

Bürglin, T.R., and Ruvkun, G. (1992). New motif in PBX genes. Nat. Genet. *1*, 319–320.

Butcher, S.E., and Jan, E. (2016). tRNA-mimicry in IRES-mediated translation and recoding. RNA Biol. *13*, 1068–1074.

Byrd, M.P., Zamora, M., and Lloyd, R.E. (2005). Translation of Eukaryotic Translation Initiation Factor 4GI (eIF4GI) Proceeds from Multiple mRNAs Containing a Novel Cap-dependent Internal Ribosome Entry Site (IRES) That Is Active during Poliovirus Infection. J. Biol. Chem. *280*, 18610–18622.

C

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl. Acad. Sci. U. S. A. *106*, 7507–7512.

Candeias, M.M., Powell, D.J., Roubalova, E., Apcher, S., Bourougaa, K., Vojtesek, B., Bruzzoni-Giovanelli, H., and Fåhraeus, R. (2006). Expression of p53 and p53/47 are controlled by alternative mechanisms of messenger RNA translation initiation. Oncogene *25*, 6936–6947.

Cao, Y., Portela, M., Janikiewicz, J., Doig, J., and Abbott, C.M. (2014). Characterisation of translation elongation factor eEF1B subunit expression in mammalian cells and tissues and co-localisation with eEF1A2. PLoS One *9*, e114117.

Castets, M., Schaeffer, C., Bechara, E., Schenck, A., Khandjian, E.W., Luche, S., Moine, H., Rabilloud, T., Mandel, J.L., and Bardoni, B. (2005). FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. Hum. Mol. Genet. *14*, 835–844.

Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007). The Nonsense-Mediated Decay RNA Surveillance Pathway. Annu. Rev. Biochem. *76*, 51–74.

Chard, L.S., Kaku, Y., Jones, B., Nayak, A., and Belsham, G.J. (2006). Functional analyses of

RNA structures shared between the internal ribosome entry sites of hepatitis C virus and the picornavirus porcine teschovirus 1 Talfan. J. Virol. *80*, 1271–1279.

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

Chekulaeva, M., and Rajewsky, N. (2019). Roles of Long Noncoding RNAs and Circular RNAs in Translation. Cold Spring Harb. Perspect. Biol. *11*, a032680.

Chen, S.-J., Lin, G., Chang, K.-J., Yeh, L.-S., and Wang, C.-C. (2008). Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. J. Biol. Chem. *283*, 3173–3180.

Cheung, Y.-N., Maag, D., Mitchell, S.F., Fekete, C.A., Algire, M.A., Takacs, J.E., Shirokikh, N., Pestova, T., Lorsch, J.R., and Hinnebusch, A.G. (2007). Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. Genes Dev. *21*, 1217–1230.

Chicher, J., Simonetti, A., Kuhn, L., Schaeffer, L., Hammann, P., Eriani, G., and Martin, F. (2015). Purification of mRNA-programmed translation initiation complexes suitable for mass spectrometry analysis. Proteomics *15*, 2417–2425.

Choi, J., Ieong, K.W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., O'Leary, S.E., Dominissini, D., Rechavi, G., Soltis, S.M., et al. (2016). N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. Nat. Struct. Mol. Biol. *23*, 110–115.

Chojnowski, J.L., Masuda, K., Trau, H.A., Thomas, K., Capecchi, M., and Manley, N.R. (2014). Multiple roles for HOXA3 in regulating thymus and parathyroid differentiation and morphogenesis in mouse. Development *141*, 3697–3708.

Cohen-Chalamish, S., Hasson, A., Weinberg, D., Namer, L.S., Banai, Y., Osman, F., and Kaempfer, R. (2009). Dynamic refolding of IFN-γ mRNA enables it to function as PKR activator and translation template. Nat. Chem. Biol. *5*, 896–903.

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

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

Comelli, L., Marchetti, L., Arosio, D., Riva, S., Abdurashidova, G., Beltram, F., and Falaschi,

A. (2009). The homeotic protein HOXC13 is a member of human DNA replication complexes. Cell Cycle *8*, 454–459.

Connolly, E., Braunstein, S., Formenti, S., and Schneider, R.J. (2006). Hypoxia inhibits protein synthesis through a 4E-BP1 and elongation factor 2 kinase pathway controlled by mTOR and uncoupled in breast cancer cells. Mol. Cell. Biol. *26*, 3955–3965.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17–37.

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

Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M., and Borden, K.L.B. (2006). eIF4E is a central node of an RNA regulon that governs cellular proliferation. J. Cell Biol. *175*, 415–426.

D

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

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

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development *132*, 2931–2942.

Dever, T.E., Kinzy, T.G., and Pavitt, G.D. (2016). Mechanism and Regulation of Protein Synthesis in *Saccharomyces cerevisiae*. Genetics *203*, 65–107.

Dibrov, S.M., Johnston-Cox, H., Weng, Y.-H., and Hermann, T. (2007). Functional Architecture of HCV IRES Domain II Stabilized by Divalent Metal Ions in the Crystal and in Solution. Angew. Chemie Int. Ed. *46*, 226–229.

Dmitriev, S.E., Terenin, I.M., Andreev, D.E., Ivanov, P.A., Dunaevsky, J.E., Merrick, W.C., and Shatsky, I.N. (2010). GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor. J. Biol. Chem. *285*, 26779–26787.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

E

Echeverría, N. (2015). Hepatitis C virus genetic variability and evolution. World J. Hepatol. *7*, 831–845.

Van Eden, M.E., Byrd, M.P., Sherrill, K.W., and Lloyd, R.E. (2004). Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. RNA *10*, 720–730.

Eisenstein, R.S., and Munro, H.N. (1990). Translational regulation of ferritin synthesis by iron. Enzyme *44*, 42–58.

Elfakess, R., and Dikstein, R. (2008). A translation initiation element specific to mRNAs with very short 5′UTR that also regulates transcription. PLoS One *3*, e3094.

Elfakess, R., Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., and Dikstein, R. (2011). Unique translation initiation of mRNAs-containing TISU element. Nucleic Acids Res. *39*, 7598–7609.

Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L.P., and Sonenberg, N. (2001). The major mRNA-associated protein YB-1 is a potent 5′ cap-dependent mRNA stabilizer. EMBO J. *20*, 5491–5502.

F

Farabaugh, P.J. (2002). Programmed Translational Frameshifting. Annu. Rev. Genet. *30*, 507–528.

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

Fernandez, J., Yaman, I., Merrick, W.C., Koromilas, A., Wek, R.C., Sood, R., Hensold, J., and Hatzoglou, M. (2002). Regulation of internal ribosome entry site-mediated translation by eukaryotic initiation factor-2α phosphorylation and translation of a small upstream open reading frame. J. Biol. Chem. *277*, 2050–2058.

Ferrara, N. (2002). VEGF and the quest for tumour angiogenesis factors. Nat. Rev. Cancer *2*, 795–803.

Filbin, M.E., and Kieft, J.S. (2009). Toward a structural understanding of IRES RNA function. Curr. Opin. Struct. Biol. *19*, 267–276.

Flather, D., and Semler, B.L. (2015). Picornaviruses and nuclear functions: targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus.

Front. Microbiol. *6*, 594.

Fraser, C.S. (2015). Quantitative studies of mRNA recruitment to the eukaryotic ribosome. Biochimie *114*, 58–71.

Fraser, C.S., Berry, K.E., Hershey, J.W.B., and Doudna, J.A. (2007). eIF3j Is Located in the Decoding Center of the Human 40S Ribosomal Subunit. Mol. Cell *26*, 811–819.

Fraser, D.J., Phillips, A.O., Zhang, X., Van Roeyen, C.R., Muehlenberg, P., En-Nia, A., and Mertens, P.R. (2008). Y-box protein-1 controls transforming growth factor-β1 translation in proximal tubular cells. Kidney Int. *73*, 724–732.

Fringer, J.M., Acker, M.G., Fekete, C.A., Lorsch, J.R., and Dever, T.E. (2007). Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining. Mol. Cell. Biol. *27*, 2384–2397.

G

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

Gabellini, D., Colaluca, I.N., Vodermaier, H.C., Biamonti, G., Giacca, M., Falaschi, A., Riva, S., and Peverali, F.A. (2003). Early mitotic degradation of the homeoprotein HOXC10 is potentially linked to cell cycle progression. EMBO J. *22*, 3715–3724.

García-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. Ciba Found. Symp. *0*, 161–182.

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. *6*, 881–892.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211–223.

Giaccia, A.J., and Kastan, M.B. (1998). The complexity of p53 modulation: Emerging patterns from divergent signals. Genes Dev. *12*, 2973–2983.

Gingras, A.C., Svitkin, Y., Belsham, G.J., Pause, A., and Sonenberg, N. (1996). Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. Proc. Natl. Acad. Sci. U. S. A. *93*, 5578–5583.

Golob-Schwarzl, N., Schweiger, C., Koller, C., Krassnig, S., Gogg-Kamerer, M., Gantenbein, N., Toeglhofer, A.M., Wodlej, C., Bergler, H., Pertschy, B., et al. (2017). Separation of low and high grade colon and rectum carcinoma by eukaryotic translation initiation

factors 1, 5 and 6. Oncotarget *8*, 101224–101243.

Graber, T.E., and Holcik, M. (2007). Cap-independent regulation of gene expression in apoptosis. Mol. Biosyst. *3*, 825.

Gray, N.K., and Hentze, M.W. (1994). Regulation of protein synthesis by mRNA structure. Mol. Biol. Rep. *19*, 195–200.

Grover, R., Ray, P.S., and Das, S. (2008). Polypyrimidine tract binding protein regulates IRES-mediated translation of p53 isoforms. Cell Cycle *7*, 2189–2198.

Grover, R., Candeias, M.M., Fhraeus, R., and Das, S. (2009). P53 and little brother p53/47: Linking IRES activities with protein functions. Oncogene *28*, 2766–2772.

Grover, R., Sharathchandra, A., Ponnuswamy, A., Khan, D., and Das, S. (2011). Effect of mutations on the p53 IRES RNA structure: Implications for de-regulation of the synthesis of p53 isoforms. RNA Biol. *8*, 132–142.

Guenther, U.P., Weinberg, D.E., Zubradt, M.M., Tedeschi, F.A., Stawicki, B.N., Zagore, L.L., Brar, G.A., Licatalosi, D.D., Bartel, D.P., Weissman, J.S., et al. (2018). The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. Nature *559*, 130–134.

H

Halaby, M.-J., Harris, B.R.E., Miskimins, W.K., Cleary, M.P., and Yang, D.-Q. (2015). Deregulation of Internal Ribosome Entry Site-Mediated p53 Translation in Cancer Cells with Defective p53 Response to DNA Damage. Mol. Cell. Biol. *35*, 4006–4017.

Han, L., Witmer, P.D., Casey, E., Valle, D., and Sukumar, S. (2007). DNA methylation regulates microRNA expression. Cancer Biol. Ther. *6*, 1284–1288.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Pestova, T. V, Hellen, C.U.T., and Frank, J. (2013). Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. Nature *503*, 539–543.

Hentze, M.W., and Kuhn, L.C. (2002). Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc. Natl. Acad. Sci. *93*, 8175–8182.

Hilliker, A., Gao, Z., Jankowsky, E., and Parker, R. (2011). The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex.

Mol. Cell *43*, 962–972.

Hinnebusch, A.G. (2005). TRANSLATIONAL REGULATION OF GCN4 AND THE GENERAL AMINO ACID CONTROL OF YEAST. Annu. Rev. Microbiol. *59*, 407–450.

Hinnebusch, A.G. (2014). The Scanning Mechanism of Eukaryotic Translation Initiation. Annu. Rev. Biochem. *83*, 779–812.

Honda, M., Ping, L.H., Rijnbrand, R.C., Amphlett, E., Clarke, B., Rowlands, D., and Lemon, S.M. (1996). Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. Virology *222*, 31–42.

Huez, I., Créancier, L., Audigier, S., Gensac, M.-C., Prats, A.-C., and Prats, H. (1998). Two Independent Internal Ribosome Entry Sites Are Involved in Translation Initiation of Vascular Endothelial Growth Factor mRNA. Mol. Cell. Biol. *18*, 6178–6190.

Huez, I., Bornes, S., Bresson, D., Créancier, L., and Prats, H. (2014). New Vascular Endothelial Growth Factor Isoform Generated by Internal Ribosome Entry Site-Driven CUG Translation Initiation. Mol. Endocrinol. *15*, 2197–2210.

Hussain, T., Llácer, J.L., Fernández, I.S., Munoz, A., Martin-Marcos, P., Savva, C.G., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2014). Structural changes enable start codon recognition by the eukaryotic translation initiation complex. Cell *159*, 597–607.

I

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

Iimura, T., Denans, N., and Pourquié, O. (2009). Chapter 7 Establishment of Hox Vertebral Identities in the Embryonic Spine Precursors. Curr. Top. Dev. Biol. *88*, 201–234.

Iizuka, N., Kohara, M., Hagino-Yamagishi, K., Abe, S., Komatsu, T., Tago, K., Arita, M., and Nomoto, A. (1989). Construction of less neurovirulent polioviruses by introducing deletions into the 5’ noncoding sequence of the genome. J. Virol. *63*, 5354–5363.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Iwasaki, S., Floor, S.N., and Ingolia, N.T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. Nature *534*, 558–561.

J

Jackson, R.J., Hellen, C.U.T., and Pestova, T. V (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jackson, R.J., Hellen, C.U.T., and Pestova, T. V. (2012). Termination and post-termination events in eukaryotic translation. In Advances in Protein Chemistry and Structural Biology, pp. 45–93.

Jaeger, S., Martin, F., Rudinger-Thirion, J., Giegé, R., and Eriani, G. (2006). Binding of human SLBP on the 3′-UTR of histone precursor H4-12 mRNA induces structural rearrangements that enable U7 snRNA anchoring. Nucleic Acids Res. *34*, 4987–4995.

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

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

Jenkins, R.H., Bennagi, R., Martin, J., Phillips, A.O., Redman, J.E., and Fraser, D.J. (2010). A conserved stem loop motif in the 5'untranslated region regulates transforming growth factor-β(1) translation. PLoS One *5*, e12283.

Johnson, A.G., Grosely, R., Petrov, A.N., and Puglisi, J.D. (2017). Dynamics of IRES-mediated translation. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160177.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. Cell *131*, 530–543.

K

Kafasla, P., Morgner, N., Pöyry, T.A.A., Curry, S., Robinson, C. V., and Jackson, R.J. (2009). Polypyrimidine Tract Binding Protein Stabilizes the Encephalomyocarditis Virus IRES Structure via Binding Multiple Sites in a Unique Orientation. Mol. Cell *34*, 556–568.

Kafasla, P., Morgner, N., Robinson, C. V, and Jackson, R.J. (2010). Polypyrimidine tract-binding protein stimulates the poliovirus IRES by modulating eIF4G binding. EMBO J. *29*, 3710–3722.

Kanamori, Y., and Nakashima, N. (2001). A tertiary structure model of the internal ribosome

entry site (IRES) for methionine-independent initiation of translation. RNA *7*, 266–274.

Karabiber, F., McGinnis, J.L., Favorov, O. V., and Weeks, K.M. (2013). QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. RNA *19*, 63–73.

Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster. Adv. Genet. *27*, 309–362.

Kearse, M.G., and Wilusz, J.E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. *31*, 1717–1731.

Kempf, B.J., and Barton, D.J. (2008). Poliovirus 2APro Increases Viral mRNA and Polysome Stability Coordinately in Time with Cleavage of eIF4G. J. Virol. *82*, 5847–5859.

Khan, D., Sharathchandra, A., Ponnuswamy, A., Grover, R., and Das, S. (2013). Effect of a natural mutation in the 5′ untranslated region on the translational control of p53 mRNA. Oncogene *32*, 4148–4159.

Khawaja, A., Vopalensky, V., and Pospisek, M. (2015). Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. Wiley Interdiscip. Rev. RNA *6*, 211–224.

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

Kim, G.-W., Lee, S.-H., Cho, H., Kim, M., Shin, E.-C., and Oh, J.-W. (2016). Hepatitis C Virus Core Protein Promotes miR-122 Destabilization by Inhibiting GLD-2. PLOS Pathog. *12*, e1005714.

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science *301*, 331–333.

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

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

Komar, A.A., and Hatzoglou, M. (2011). Cellular IRES-mediated translation. Cell Cycle *10*, 229–240.

Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Xue, S., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA

translation and vertebrate tissue patterning. Cell *145*, 383–397.

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

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

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. Nucleic Acids Res. *29*, 5226–5232.

Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S., and Wilusz, J.E. (2015). Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. Genes Dev. *29*, 2168–2182.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat. Chem. Biol. *3*, 218–221.

de la Cruz, J., Iost, I., Kressler, D., and Linder, P. (2002). The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. *94*, 5201–5206.

L

Laing, C., and Schlick, T. (2009). Analysis of Four-Way Junctions in RNA Structures. J. Mol. Biol. *390*, 547–559.

Lamaa, A., Le Bras, M., Skuli, N., Britton, S., Frit, P., Calsou, P., Prats, H., Cammas, A., and Millevoi, S. (2016). A novel cytoprotective function for the DNA repair protein Ku in regulating p53 mRNA translation and function. EMBO Rep. *17*, 508–518.

Lambert, B., Vandeputte, J., Remacle, S., Bergiers, I., Simonis, N., Twizere, J.C., Vidal, M., and Rezsohazy, R. (2012). Protein interactions of the transcription factor Hoxa1. BMC Dev. Biol. *12*, 29.

Laporte, J., Malet, I., Andrieu, T., Thibault, V., Toulme, J.-J., Wychowski, C., Pawlotsky, J.-M., Huraux, J.-M., Agut, H., and Cahour, A. (2000). Comparative Analysis of Translation Efficiencies of Hepatitis C Virus 5' Untranslated Regions among Intraindividual Quasispecies Present in Chronic Infection: Opposite Behaviors Depending on Cell Type. J. Virol. *74*, 10827–10833.

Lappin, T.R.J., Grier, D.G., Thompson, A., and Halliday, H.L. (2006). HOX genes: seductive science, mysterious mechanisms. Ulster Med. J. *75*, 23–31.

LaRonde-LeBlanc, N.A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to

DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev. *17*, 2060–2072.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lee, A.S.Y., Kranzusch, P.J., and Cate, J.H.D. (2015). EIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature *522*, 111–114.

Lee, A.S.Y., Kranzusch, P.J., Doudna, J.A., and Cate, J.H.D. (2016). EIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. Nature *536*, 96–99.

Lee, K.-M., Chen, C.-J., and Shih, S.-R. (2017). Regulation Mechanisms of Viral IRES-Driven Translation. Trends Microbiol. *25*, 546–561.

Lee, V.M.-Y., Goedert, M., and Trojanowski, J.Q. (2001). Neurodegenerative Tauopathies. Annu. Rev. Neurosci. *24*, 1121–1159.

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. *46*, D708–D717.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol. Cell *66*, 22-37.

Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. Science *313*, 1918–1922.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. *19*, 158–174.

Lewis, E.B. (2007). A gene complex controlling segmentation in Drosophila. In Genes, Development, and Cancer: The Life and Work of Edward B. Lewis, pp. 229–242.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121–122.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Livingstone, M., Atas, E., Meller, A., and Sonenberg, N. (2010). Mechanisms governing the control of mRNA translation. Phys. Biol. *7*, 021001.

Locker, N., Easton, L.E., and Lukavsky, P.J. (2007). HCV and CSFV IRES domain II

mediate eIF2 release during 80S ribosome assembly. EMBO J. *26*, 795–805.

Locker, N., Chamond, N., and Sargueil, B. (2011). A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. Nucleic Acids Res. *39*, 2367–2377.

Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., and Blasi, F. (2014). Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev. Dyn. *243*, 59–75.

López, M.D., and Samuelsson, T. (2008). Early evolution of histone mRNA 3′ end processing. RNA *14*, 1–10.

López de Quinto, S., and Martínez-Salas, E. (1997). Conserved structural motifs located in distal loops of aphthovirus internal ribosome entry site domain 3 are required for internal initiation of translation. J. Virol. *71*, 4171–4175.

López de Quinto, S., and Martínez-Salas, E. (2000). Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo. RNA *6*, 1380–1392.

Lozano, G., and Martínez-Salas, E. (2015). Structural insights into viral IRES-dependent translation mechanisms. Curr. Opin. Virol. *12*, 113–120.

Lukavsky, P.J. (2009). Structure and function of HCV IRES domains. Virus Res. *139*, 166–171.

Luukkonen, B.G., Tan, W., and Schwartz, S. (1995). Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. J. Virol. *69*, 4086–4094.

M

Maeda, R.K., and Karch, F. (2009). Chapter 1 The Bithorax Complex of Drosophila. In Current Topics in Developmental Biology, pp. 1–33.

Mahmoudi, S., Henriksson, S., Corcoran, M., Méndez-Vidal, C., Wiman, K.G., and Farnebo, M. (2009). Wrap53, a Natural p53 Antisense Transcript Required for p53 Induction upon DNA Damage. Mol. Cell *33*, 462–471.

Mahmoudi, S., Henriksson, S., Farnebo, L., Roberg, K., and Farnebo, M. (2011). WRAP53 promotes cancer cell survival and is a potential target for cancer therapy. Cell Death Dis. *2*, e114–e114.

Mailliot, J., and Martin, F. (2018). Viral internal ribosomal entry sites: four classes for one goal. Wiley Interdiscip. Rev. RNA *9*, e1458.

Malik, S., Girisha, K.M., Wajid, M., Roy, A.K., Phadke, S.R., Haque, S., Ahmad, W., Koch, M.C., and Grzeschik, K.H. (2007). Synpolydactyly and HOXD13 polyalanine repeat: Addition of 2 alanine residues is without clinical consequences. BMC Med. Genet. *8*, 78.

Mallo, M., and Alonso, C.R. (2013). The regulation of Hox gene expression during animal development. Development *140*, 3951–3963.

Marcel, V., Van Long, F.N., and Diaz, J.J. (2018). 40 years of research put p53 in translation. Cancers (Basel). *10*.

Marintchev, A. (2013). Roles of helicases in translation initiation: A mechanistic view. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 799–809.

Martin, F., Barends, S., Jaeger, S., Schaeffer, L., Prongidi-Fix, L., and Eriani, G. (2011). Cap-assisted internal initiation of translation of histone H4. Mol. Cell *41*, 197–209.

Martin, F., Ménétret, J.-F., Simonetti, A., Myasnikov, A.G., Vicens, Q., Prongidi-Fix, L., Natchiar, S.K., Klaholz, B.P., and Eriani, G. (2016). Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. Nat. Commun. *7*, 12622.

Martínez-Salas, E., Pacheco, A., Serrano, P., and Fernandez, N. (2008). New insights into internal ribosome entry site elements relevant for viral gene expression. J. Gen. Virol. *89*, 611–626.

Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G., and Diaz-Toledano, R. (2015). Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res. *206*, 62–73.

Mathews, D.H., Turner, D.H., and Watson, R.M. (2016). RNA secondary structure prediction. Curr. Protoc. Nucleic Acid Chem. *2016*, 11.2.1-11.2.19.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V, Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.-J., Chen, Q., et al. (2017). Reversible methylation of m6Am in the 5' cap controls mRNA stability. Nature *541*, 371–375.

Meier, V.S., Böhni, R., and Schumperli, D. (1989). Nucleotide sequence of two mouse histone H4 genes. Nucleic Acids Res. *17*, 795–795.

MEIJER, H.A., and THOMAS, A.A.M. (2002). Control of eukaryotic protein synthesis by upstream open reading frames in the 5′-untranslated region of an mRNA. Biochem. J. *367*, 1–11.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat. Genet. *36*, 1073–1078.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T. V, Qian, S.-B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Miller, W.A., Wang, Z., and Treder, K. (2007). The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. Biochem. Soc. Trans. *35*, 1629–1633.

Miras, M., Miller, W.A., Truniger, V., and Aranda, M.A. (2017). Non-canonical Translation in Plant RNA Viruses. Front. Plant Sci. *8*, 494.

Mitchell, S.F., and Lorsch, J.R. (2008). Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. J. Biol. Chem. *283*, 27345–27349.

Mlynarczyk, C., and Fåhraeus, R. (2014). Endoplasmic reticulum stress sensitizes cells to DNA damage-induced apoptosis through p53-dependent suppression of p21 CDKN1A. Nat. Commun. *5*, 5067.

Modelska, A., Turro, E., Russell, R., Beaton, J., Sbarrato, T., Spriggs, K., Miller, J., Gräf, S., Provenzano, E., Blows, F., et al. (2015). The malignant phenotype in breast cancer is driven by eIf4A1-mediated changes in the translational landscape. Cell Death Dis. *6*, e1603–e1603.

Mokrejš, M., Mašek, T., Vopálenský, V., Hlubuček, P., Delbos, P., and Pospíšek, M. (2010). IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. Nucleic Acids Res. *38*, D131–D136.

Morris-Desbois, C., Réty, S., Ferro, M., Garin, J., and Jalinot, P. (2001). The Human Protein HSPC021 Interacts with Int-6 and Is Associated with Eukaryotic Translation Initiation Factor 3. J. Biol. Chem. *276*, 45988–45995.

Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of mRNA translation. Mol. Cell. Biol. *20*, 8635–8642.

Mosenkis, J., Daniels-McQueen, S., Janovec, S., Duncan, R., Hershey, J.W., Grifo, J.A., Merrick, W.C., and Thach, R.E. (1985). Shutoff of host translation by encephalomyocarditis virus infection does not involve cleavage of the eucaryotic initiation factor 4F polypeptide that accompanies poliovirus infection. J. Virol. *54*, 643–645.

Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell *45*, 201–207.

Mühlemann, O. (2008). Recognition of nonsense mRNA: towards a unified model. Biochem. Soc. Trans. *36*, 497–501.

Murray, J., Savva, C.G., Shin, B.-S., Dever, T.E., Ramakrishnan, V., and Fernández, I.S. (2016). Structural characterization of ribosome recruitment and translocation by type IV IRES. Elife *5*.

N

Nekrasov, M.P., Ivshina, M.P., Chernov, K.G., Kovrigina, E.A., Evdokimova, V.M., Thomas, A.A.M., Hershey, J.W.B., and Ovchinnikov, L.P. (2003). The mRNA-binding protein YB-1 (p50) prevents association of the eukaryotic initiation factor eIF4G with mRNA and inhibits protein synthesis at the initiation stage. J. Biol. Chem. *278*, 13936–13943.

Nicholson, B.L., Wu, B., Chevtchenko, I., and White, K.A. (2010). Tombusvirus recruitment of host translational machinery via the 3' UTR. RNA *16*, 1402–1419.

Nunes, F.D., Almeida, F.C.S. de, Tucci, R., and Sousa, S.C.O.M. de (2007). Homeobox genes: a molecular link between development and cancer. Pesqui. Odontológica Bras. *17*, 94–98.

Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol. Biol. *71*, 367–378.

O

Ofir-Rosenfeld, Y., Boggs, K., Michael, D., Kastan, M.B., and Oren, M. (2008). Mdm2 Regulates p53 mRNA Translation through Inhibitory Interactions with Ribosomal Protein L26. Mol. Cell *32*, 180–189.

Ogram, S.A., Spear, A., Sharma, N., and Flanegan, J.B. (2010). The 5'CL-PCBP RNP complex, 3' poly(A) tail and 2A(pro) are required for optimal translation of poliovirus RNA. Virology *397*, 14–22.

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

Otto, G.A., and Puglisi, J.D. (2004). The Pathway of HCV IRES-Mediated Translation Initiation. Cell *119*, 369–380.

P

Paek, K.Y., Hong, K.Y., Ryu, I., Park, S.M., Keum, S.J., Kwon, O.S., and Jang, S.K. (2015). Translation initiation mediated by RNA looping. Proc. Natl. Acad. Sci. U. S. A. *112*, 1041–1046.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol. Cell *66*, 9-21.

Pan, H., Zhao, X., Zhang, X., Abouelsoud, M., Sun, J., April, C., Amleh, A., Fan, J.-B., Hu, Y., and Li, R. (2015). Translational Initiation at a Non-AUG Start Codon for Human and Mouse Negative Elongation Factor-B. PLoS One *10*, e0127422.

Parsyan, A., Shahbazian, D., Martineau, Y., Petroulakis, E., Alain, T., Larsson, O., Mathonnet, G., Tettweiler, G., Hellen, C.U., Pestova, T. V, et al. (2009). The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *106*, 22217–22222.

Parsyan, A., Svitkin, Y., Shahbazian, D., Gkogkas, C., Lasko, P., Merrick, W.C., and Sonenberg, N. (2011). MRNA helicases: The tacticians of translational control. Nat. Rev. Mol. Cell Biol. *12*, 235–245.

Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., and Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. Mol. Cell *26*, 41–50.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Pehar, M., Ko, M.H., Li, M., Scrable, H., and Puglielli, L. (2014). P44, the "longevity-assurance" isoform of P53, regulates tau phosphorylation and is activated in an age-dependent fashion. Aging Cell *13*, 449–456.

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

Pérard, J., Leyrat, C., Baudin, F., Drouet, E., and Jamin, M. (2013). Structure of the full-length HCV IRES in solution. Nat. Commun. *4*, 1612.

Peselis, A., and Serganov, A. (2014). Structure and function of pseudoknots involved in gene expression control. Wiley Interdiscip. Rev. RNA *5*, 803–822.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., and Liuni, S. (2001). Structural and

functional features of eukaryotic mRNA untranslated regions. In Gene, pp. 73–81.

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

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

Pestova, T. V, Lomakin, I.B., and Hellen, C.U.T. (2004). Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO Rep. *5*, 906–913.

Pfingsten, J.S., and Kieft, J.S. (2008). RNA structure-based ribosome recruitment: lessons from the Dicistroviridae intergenic region IRESes. RNA *14*, 1255–1263.

Phelan, M.L., Rambaldi, I., and Featherstone, M.S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Mol. Cell. Biol. *15*, 3989–3997.

Phinney, D.G., Gray, A.J., Hill, K., and Pandey, A. (2005). Murine mesenchymal and embryonic stem cells express a similar Hox gene profile. Biochem. Biophys. Res. Commun. *338*, 1759–1765.

Pickering, B.M., Mitchell, S.A., Spriggs, K.A., Stoneley, M., and Willis, A.E. (2004). Bag-1 Internal Ribosome Entry Segment Activity Is Promoted by Structural Changes Mediated by Poly(rC) Binding Protein 1 and Recruitment of Polypyrimidine Tract Binding Protein 1. Mol. Cell. Biol. *24*, 5595–5605.

Pilipenko, E. V, Gmyl, A.P., Maslova, S. V, Svitkin, Y. V, Sinyakov, A.N., and Agol, V.I. (1992). Prokaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. Cell *68*, 119–131.

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

Piñeiro, D., Ramajo, J., Bradrick, S.S., and Martínez-Salas, E. (2012). Gemin5 proteolysis reveals a novel motif to identify L protease targets. Nucleic Acids Res. *40*, 4942–4953.

Piñeiro, D., Fernández, N., Ramajo, J., and Martínez-Salas, E. (2013). Gemin5 promotes IRES interaction and translation control through its C-terminal region. Nucleic Acids Res. *41*, 1017–1028.

Pisarev, A. V, Kolupaeva, V.G., Yusupov, M.M., Hellen, C.U.T., and Pestova, T. V (2008). Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. EMBO J. *27*, 1609–1621.

Pisareva, V.P., Pisarev, A. V, Hellen, C.U.T., Rodnina, M. V, and Pestova, T. V (2006). Kinetic analysis of interaction of eukaryotic release factor 3 with guanine nucleotides. J. Biol. Chem. *281*, 40224–40235.

Pisareva, V.P., Pisarev, A. V, Komar, A.A., Hellen, C.U.T., and Pestova, T. V (2008). Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. Cell *135*, 1237–1250.

Pisareva, V.P., Pisarev, A. V, and Fernández, I.S. (2018). Dual tRNA mimicry in the cricket paralysis virus IRES uncovers an unexpected similarity with the hepatitis C virus IRES. Elife *7*.

Pleij, C.W.A. (1994). RNA pseudoknots. Curr. Opin. Struct. Biol. *4*, 337–344.

Prongidi-Fix, L., Schaeffer, L., Simonetti, A., Barends, S., Ménétret, J.-F., Klaholz, B.P., Eriani, G., and Martin, F. (2013). Rapid purification of ribosomal particles assembled on histone H4 mRNA: a new method based on mRNA-DNA chimaeras. Biochem. J. *449*, 719–728.

Q

Quinonez, S.C., and Innis, J.W. (2014). Human HOX gene disorders. Mol. Genet. Metab. *111*, 4–15.

R

Ragan, C., Goodall, G.J., Shirokikh, N.E., and Preiss, T. (2019). Insights into the biogenesis and potential functions of exonic circular RNA. Sci. Rep. *9*, 2048.

Ray, P.S., Grover, R., and Das, S. (2006). Two internal ribosome entry sites mediate the translation of p53 isoforms. EMBO Rep. *7*, 404–410.

Rehwinkel, J., Raes, J., and Izaurralde, E. (2006). Nonsense-mediated mRNA decay: target genes and functional diversification of effectors. Trends Biochem. Sci. *31*, 639–646.

Reinert, L.S., Shi, B., Nandi, S., Mazan-Mamczarz, K., Vitolo, M., Bachman, K.E., He, H., and Gartenhaus, R.B. (2006). MCT-1 protein interacts with the cap complex and modulates messenger RNA translational profiles. Cancer Res. *66*, 8994–9001.

Reynolds, N., O'Shaughnessy, A., and Hendrich, B. (2013). Transcriptional repressors: multifaceted regulators of gene expression. Development *140*, 505–512.

Rezsohazy, R. (2014). Non-transcriptional interactions of Hox proteins: Inventory, facts, and future directions. Dev. Dyn. *243*, 117–131.

Rezsohazy, R., Saurin, A.J., Maurel-Zaffran, C., and Graba, Y. (2015). Cellular and molecular insights into Hox protein action. Development *142*, 1212–1227.

Rijnbrand, R., van der Straaten, T., van Rijn, P.A., Spaan, W.J., and Bredenbeek, P.J. (1997). Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. J. Virol. *71*, 451–457.

Riley, A., Jordan, L.E., and Holcik, M. (2010). Distinct 5' UTRs regulate XIAP expression under normal growth conditions and during cellular stress. Nucleic Acids Res. *38*, 4665–4674.

Roberts, A.P.E., Lewis, A.P., and Jopling, C.L. (2011). miR-122 activates hepatitis C virus translation by a specialized mechanism requiring particular RNA components. Nucleic Acids Res. *39*, 7716–7729.

Romeo, D.S., Park, K., Roberts, A.B., Sporn, M.B., and Kim, S.J. (2014). An element of the transforming growth factor-beta 1 5'-untranslated region represses translation and specifically binds a cytosolic factor. Mol. Endocrinol. *7*, 759–766.

Roobol, A., Carden, M.J., Newsam, R.J., and Smales, C.M. (2009). Biochemical insights into the mechanisms central to the response of mammalian cells to cold stress and subsequent rewarming. FEBS J. *276*, 286–302.

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. J. Am. Chem. Soc. *137*, 2107–2115.

Roy, B., Vaughn, J.N., Kim, B.H., Zhou, F., Gilchrist, M.A., and Von Arnim, A.G. (2010). The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. RNA *16*, 748–761.

Rubin, E., Wu, X., Zhu, T., Cheung, J.C.Y., Chen, H., Lorincz, A., Pandita, R.K., Sharma, G.G., Hyo, C.H., Gasson, J., et al. (2007). A role for the HOXB7 homeodomain protein in DNA repair. Cancer Res. *67*, 1527–1535.

Rubio, C.A., Weisburd, B., Holderfield, M., Arias, C., Fang, E., DeRisi, J.L., and Fanidi, A. (2014). Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. Genome Biol. *15*, 476.

S

Salsi, V., Ferrari, S., Ferraresi, R., Cossarizza, A., Grande, A., and Zappavigna, V. (2009). HOXD13 Binds DNA Replication Origins To Promote Origin Licensing and Is Inhibited by Geminin. Mol. Cell. Biol. *29*, 5775–5788.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One *7*, e30733.

Sánchez-Herrero, E. (2013). Hox Targets and Cellular Functions. Scientifica (Cairo). *2013*, 1–26.

Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. EMBO J. *20*, 4803–4813.

Schleich, S., Strassburger, K., Janiesch, P.C., Koledachkina, T., Miller, K.K., Haneke, K., Cheng, Y.S., Küchler, K., Stoecklin, G., Duncan, K.E., et al. (2014). DENR-MCT-1 promotes translation re-initiation downstream of uORFs to control tissue growth. Nature *512*, 208–212.

Schleich, S., Acevedo, J.M., Clemm Von Hohenberg, K., and Teleman, A.A. (2017). Identification of transcripts with short stuORFs as targets for DENR•MCTS1-dependent translation in human cells. Sci. Rep. *7*, 3722.

Schmeing, T.M., Voorhees, R.M., Kelley, A.C., and Ramakrishnan, V. (2011). How mutations in tRNA distant from the anticodon affect the fidelity of decoding. Nat. Struct. Mol. Biol. *18*, 432–436.

Schüler, M., Connell, S.R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P.A., Westhof, E., and Spahn, C.M.T. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nat. Struct. Mol. Biol. *13*, 1092–1096.

Schult, P., Roth, H., Adams, R.L., Mas, C., Imbert, L., Orlik, C., Ruggieri, A., Pyle, A.M., and Lohmann, V. (2018). microRNA-122 amplifies hepatitis C virus translation by shaping the structure of the internal ribosomal entry site. Nat. Commun. *9*, 2613.

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

Shou, S., Carlson, H.L., Perez, W.D., and Stadler, H.S. (2013). HOXA13 regulates Aldh1a2

expression in the autopod to facilitate interdigital programmed cell death. Dev. Dyn. *242*, 687–698.

Silva, A.L., and Romão, L. (2009). The mammalian nonsense-mediated mRNA decay pathway: To decay or not to decay! Which players make the decision? FEBS Lett. *583*, 499–505.

Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Molecular biology: Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science (80-.). *310*, 1513–1515.

Sizova, D. V, Kolupaeva, V.G., Pestova, T. V, Shatsky, I.N., and Hellen, C.U. (1998). Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. J. Virol. *72*, 4775–4782.

Skabkin, M.A., Skabkina, O. V., Dhote, V., Komar, A.A., Hellen, C.U.T., and Pestova, T. V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. Genes Dev. *24*, 1787–1801.

Sokabe, M., Fraser, C.S., and Hershey, J.W.B. (2012). The human translation initiation multi-factor complex promotes methionyl-tRNAi binding to the 40S ribosomal subunit. Nucleic Acids Res. *40*, 905–913.

Somers, J., Pöyry, T., and Willis, A.E. (2013). A perspective on mammalian upstream open reading frame function. Int. J. Biochem. Cell Biol. *45*, 1690–1700.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731–745.

Song, Y., Tzima, E., Ochs, K., Bassili, G., Trusheim, H., Linder, M., Preissner, K.T., and Niepmann, M. (2005). Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation. RNA *11*, 1809–1824.

Soto-Rifo, R., Rubilar, P.S., Limousin, T., de Breyne, S., Décimo, D., and Ohlmann, T. (2012). DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. EMBO J. *31*, 3745–3756.

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

Spear, A., Sharma, N., and Flanegan, J.B. (2008). Protein-RNA tethering: the role of poly(C) binding protein 2 in poliovirus RNA replication. Virology *374*, 280–291.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene

expression during conditions of cell stress. Mol. Cell *40*, 228–237.

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

Sun, C., Querol-Audí, J., Mortimer, S.A., Arias-Palomo, E., Doudna, J.A., Nogales, E., and Cate, J.H.D. (2013). Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. Nucleic Acids Res. *41*, 7512–7521.

Susorov, D., Zakharov, N., Shuvalova, E., Ivanov, A., Egorova, T., Shuvalov, A., Shatsky, I.N., and Alkalaeva, E. (2018). Eukaryotic translation elongation factor 2 (eEF2) catalyzes reverse translocation of the eukaryotic ribosome. J. Biol. Chem. *293*, 5220–5229.

Sweeney, T.R., Abaeva, I.S., Pestova, T. V, and Hellen, C.U.T. (2014). The mechanism of translation initiation on Type 1 picornavirus IRESs. EMBO J. *33*, 76–92.

Swiatkowska, A., Zydowicz, P., Gorska, A., Suchacka, J., Dutkiewicz, M., and Ciesiołka, J. (2015). The role of structural elements of the 5′-terminal region of p53 mRNA in translation under stress conditions assayed by the antisense oligonucleotide approach. PLoS One *10*, e0141676.

T

Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C.Z., Lin, C.C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., et al. (2018). CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat. Commun. *9*, 152.

Takagi, M., Absalon, M.J., McLure, K.G., and Kastan, M.B. (2005). Regulation of p53 Translation and Induction after DNA Damage by Ribosomal Protein L26 and Nucleolin. Cell *123*, 49–63.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Thompson, A.A., and Nguyen, L.T. (2000). Amegakaryocytic thrombocytopenia and radio-ulnar synostosis are associated with HOXA11 mutation. Nat. Genet. *26*, 397–398.

Tiret, L., Le Mouellic, H., Maury, M., and Brûlet, P. (1998). Increased apoptosis of motoneurons and altered somatotopic maps in the brachial spinal cord of Hoxc-8-deficient mice. Development *125*, 279–291.

Topisirovic, I., Kentsis, A., Perez, J.M., Guzman, M.L., Jordan, C.T., and Borden, K.L.B.

(2005). Eukaryotic Translation Initiation Factor 4E Activity Is Modulated by HOXA9 at Multiple Levels. Mol. Cell. Biol. *25*, 1100–1112.

Toyoda, H., Franco, D., Fujita, K., Paul, A. V., and Wimmer, E. (2007). Replication of Poliovirus Requires Binding of the Poly(rC) Binding Protein to the Cloverleaf as Well as to the Adjacent C-Rich Spacer Sequence between the Cloverleaf and the Internal Ribosomal Entry Site. J. Virol. *81*, 10017–10028.

U

Unbehaun, A., Borukhov, S.I., Hellen, C.U.T., and Pestova, T. V (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. Genes Dev. *18*, 3078–3093.

Underhill, M.F., Smales, C.M., Naylor, L.H., Birch, J.R., and James, D.C. (2007). Transient Gene Expression Levels from Multigene Expression Vectors. Biotechnol. Prog. *23*, 435–443.

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

V

Vaidya, A.T., Lomakin, I.B., Joseph, N.N., Dmitriev, S.E., and Steitz, T.A. (2017). Crystal Structure of the C-terminal Domain of Human eIF2D and Its Implications on Eukaryotic Translation Initiation. J. Mol. Biol. *429*, 2765–2771.

Valášek, L.S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M.P., Hronová, V., Herrmannová, A., Hashem, Y., and Gunišová, S. (2017). Survey and summary: Embraced by eIF3: Structural and functional insights into the roles of eIF3 across the translation cycle. Nucleic Acids Res. *45*, 10948–10968.

W

Wang, X., and Proud, C.G. (2009). Nutrient control of TORC1, a cell-cycle regulator. Trends Cell Biol. *19*, 260–267.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N 6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. Cell *161*, 1388–1399.

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

Wang, Z., Treder, K., and Miller, W.A. (2009). Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. J. Biol. Chem. *284*, 14189–14202.

Wang, Z., Parisien, M., Scheets, K., and Miller, W.A. (2011). The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. Structure *19*, 868–880.

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.

Weingarten-Gabbay, S., Khan, D., Liberman, N., Yoffe, Y., Bialik, S., Das, S., Oren, M., and Kimchi, A. (2014). The translation initiation factor DAP5 promotes IRES-driven translation of p53 mRNA. Oncogene *33*, 611–618.

Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A.A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., and Segal, E. (2016). Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. Science *351*.

Weisser, M., Schäfer, T., Leibundgut, M., Böhringer, D., Aylett, C.H.S., and Ban, N. (2017). Structural and Functional Insights into Human Re-initiation Complexes. Mol. Cell *67*, 447-456.

Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., and Haeusler, A. (2019). Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. EMBO Mol. Med. *11*, e9423.

Westerheide, S.D., and Morimoto, R.I. (2005). Heat shock response modulators as therapeutic tools for diseases of protein conformation. J. Biol. Chem. *280*, 33097–33100.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature *513*, 65–70.

Wu, C., Wei, J., Lin, P.J., Tu, L., Deutsch, C., Johnson, A.E., and Sachs, M.S. (2012).

Arginine changes the conformation of the arginine attenuator peptide relative to the ribosome tunnel. J. Mol. Biol. *416*, 518–533.

X

Xue, S., and Barna, M. (2015). Cis-regulatory RNA elements that regulate specialized ribosome activity. RNA Biol. *12*, 1083–1087.

Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015). RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. Nature *517*, 33–38.

Y

Yamamoto, H., Collier, M., Loerke, J., Ismer, J., Schmidt, A., Hilal, T., Sprink, T., Yamamoto, K., Mielke, T., Bürger, J., et al. (2015). Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. EMBO J. *34*, 3042–3058.

Yamamoto, H., Unbehaun, A., and Spahn, C.M.T. (2017). Ribosomal Chamber Music: Toward an Understanding of IRES Mechanisms. Trends Biochem. Sci. *42*, 655–668.

Yaman, I., Fernandez, J., Liu, H., Caprara, M., Komar, A.A., Koromilas, A.E., Zhou, L., Snider, M.D., Scheuner, D., Kaufman, R.J., et al. (2004). The Zipper Model of Translational Control. Cell *113*, 519–531.

Yang, D.Q., Halaby, M.J., and Zhang, Y. (2006). The identification of an internal ribosomal entry site in the 5′-untranslated region of p53 mRNA provides a novel mechanism for the regulation of its translation following DNA damage. Oncogene *25*, 4613–4619.

Ye, X., Fong, P., Iizuka, N., Choate, D., and Cavener, D.R. (1997). Ultrabithorax and Antennapedia 5' untranslated regions promote developmentally regulated internal translation initiation. Mol. Cell. Biol. *17*, 1714–1721.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-Directed Cleavage of HOXB8 mRNA. Science (80-.). *304*, 594–596.

Young, D.J., and Guydosh, N.R. (2019). Hcr1/eIF3j Is a 60S Ribosomal Subunit Recycling Accessory Factor In Vivo. Cell Rep. *28*, 39-50.e4.

Yu, C.H., Teulade-Fichou, M.P., and Olsthoorn, R.C.L. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. Nucleic Acids Res. *42*, 1887–1892.

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

Z

Zhao, C., Datta, S., Mandal, P., Xu, S., and Hamilton, T. (2010). Stress-sensitive regulation of IFRD1 mRNA decay is mediated by an upstream open reading frame. J. Biol. Chem. *285*, 8552–8562.

Zhao, X., Yang, Y., Sun, B.F., Shi, Y., Yang, X., Xiao, W., Hao, Y.J., Ping, X.L., Chen, Y.S., Wang, W.J., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. *24*, 1403–1419.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.B. (2015). Dynamic m6 A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Abstract

Mechanism of inhibition of cap-dependent translation by the Translation Inhibitory Elements (TIE) a3 and a11 in Hox mRNAs

In eukaryotes, most cellular mRNAs undergo the canonical cap-dependent translation because the initiation is guided by the 5' m^7G cap. This process requires several factors termed eukaryotic Initiation Factors (eIFs). Homeobox (Hox) genes are evolutionarily conserved genes that encode for transcriptional factors controlling embryonic body plan along the head-tail axis. Recent reports suggest that subsets of HoxA mRNAs are translated in a cap-independent manner due to the presence of 5'UTR RNA regulon termed Translation Inhibitory Element (TIE) that inhibits cap-dependent translation.

The objective of my PhD project is to decipher the inhibitory mechanism of two TIE elements: TIE a3 and TIE a11 in Hoxa3 and Hoxa11 mRNAs respectively. For that, we established an *in vitro* cell-free translation system with Rabbit Reticulocytes Lysate that faithfully recapitulates TIE-mediated translation inhibition. We were able to map the minimal functional domain of each TIE element. We also established a model of the secondary structure for each TIE element using chemical probing methods with DMS and CMCT reagents. By sucrose gradient fractionation, we analysed the mode of action of each TIE element. Further experiments were conducted by 'Toe Printing' assay and site-directed mutageneses to better understand TIE-mediated inhibition. We also purified translation initiation complexes to identify the *trans*-acting factors using an approach developed in our laboratory for mass spectrometry analysis. Interestingly, the two TIE elements function by radically distinct mechanisms. Our current model for TIE a3 suggests that translation inhibition is due to an upstream Open Reading Frame (uORF) which is translated through the 5'UTR with the involvement of eIF2D, a non-canonical GTP-independent initiation factor. In contrast, TIE a11 sequesters 80S ribosome due to a start-stop codons combination located nineteen nucleotides upstream of a GC-rich stem-loop which blocks the ribosome at the stop codon. In addition, we show that the Hox TIE and Hox IRESes function as independent modules. Interestingly, Hox IRESes are not functional *in vitro* and their activity requires specific embryonic cell lines that harbour specific ITAFs to activate them. Future studies will be carried out to get further insights into how Hox mRNAs are regulated by this TIE/IRES interplay.

Key words: Hox mRNA, TIE, IRES, uORF