



HAL
open science

Etude de l'impact des facteurs eRF3 et Upf1 dans la traduction des ARN messagers porteurs d'uORF

Affaf Aliouat

► **To cite this version:**

Affaf Aliouat. Etude de l'impact des facteurs eRF3 et Upf1 dans la traduction des ARN messagers porteurs d'uORF. Biochimie, Biologie Moléculaire. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066183 . tel-02182941

HAL Id: tel-02182941

<https://theses.hal.science/tel-02182941>

Submitted on 14 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale 515 – Complexité du vivant

Laboratoire Adaptation biologique et vieillissement, CNRS – UPMC UMR8256

Equipe de recherche Traduction eucaryote

Etude de l'impact des facteurs eRF3 et Upf1 dans la traduction des ARN messagers porteurs d'uORF

Par Affaf ALIOUAT

Thèse de doctorat de Biochimie et Biologie moléculaire

Dirigée par

Dr. Olivier JEAN-JEAN et Dr. Samia SALHI

Présentée et soutenue publiquement le 12 juillet 2017

Devant un jury composé de :

Dr. Jean-jaques DIAZ

Dr. Stefania MILLEVOI

Pr. Ali LADRAM

Dr. Christine ALLMANG

Dr. Olivier JEAN-JEAN

Dr. Samia SALHI

Rapporteur

Rapporteur

Examineur

Examineur

Directeur de thèse

Co-directrice de thèse

Sommaire

Préambule : la régulation de l'expression des gènes	1
Introduction	4
I. l'ARN messenger eucaryote	4
1. Synthèse et maturation de l'ARN messenger	4
1.1. Maturation de l'extrémité 5' de l'ARN messenger	4
1.2. Epissage des ARN messagers	5
1.3. Maturation de l'extrémité 3' de l'ARN messenger	7
2. Structure de l'ARNm mature eucaryote	8
II. La traduction chez les eucaryotes	10
1. L'initiation	11
1.1. Les étapes de l'initiation de la traduction dépendante de la coiffe	13
1.1.1. Formation du complexe de pré-initiation 43S	13
1.1.1.1. La formation du complexe ternaire eIF2-GTP-Met-ARN _t ^{Met}	15
1.1.1.2. Les facteurs d'assemblage du complexe de pré-initiation 43S	15
1.1.2. Formation du complexe eIF4F	16
1.1.3. Formation du complexe d'initiation 48S	17
1.1.4. Balayage ou " <i>scanning</i> " de la région 5'UTR	17
1.1.5. Reconnaissance du codon initiateur	18
1.1.6. Les déterminants de la reconnaissance du codon initiateur	18
1.1.7. Arrivée de la sous-unité ribosomique 60S	19
1.2. Le cas du premier tour de traduction	20
2. L'élongation	21
3. La terminaison de la traduction et le recyclage des ribosomes chez les eucaryotes	22
3.1. Les principaux facteurs de terminaison et de recyclage	23
3.1.1. Le facteur eRF1	23
3.1.2. Le facteur eRF3	25
3.1.3. Le facteur Upf1	27
3.1.4. La protéine ABCE1	28
3.2. Le mécanisme de terminaison de la traduction	29
3.2.1. Reconnaissance des codons stop	30
3.2.2. Translecture des codons stop	31
3.2.3. Libération de la protéine néo-synthétisée	32
3.2.4. Dissociation du complexe de terminaison et recyclage des ribosomes	34
4. La réinitiation de la traduction chez les eucaryotes	36
III. Dégradation et surveillance des ARNm eucaryotes	37
1. Mécanisme général de dégradation cytoplasmique des ARNm	37
1.1. La désadénylation	39
1.2. Les voies exonucléasiques 5'-3' et 3'-5'	40
1.3. Coupures endonucléolytiques	40
1.4. Sites cytoplasmique de la dégradation des ARNm	40
2. La surveillance des ARNm et la dégradation des ARNm aberrants	41
2.1. Les acteurs majeurs de la voie NMD	42
2.1.1. Les protéines Upf	42
2.1.2. Les protéines SMG	44

2.2. Le complexe de Jonction Exon-Exon (EJC)	44
2.3. Mécanisme de dégradation des transcrits contenant des codons non-sens (NMD)	45
IV. Régulation de l'initiation de la traduction	48
1. Régulation par modulation de l'activité des facteurs d'initiation de traduction	48
1.1. Régulation par le facteur eIF2	48
1.2. Régulation par le facteur eIF4E	50
1.2.1. Régulation d'eIF4E par le facteur 4E-BP et la voie TOR	51
1.2.2. Régulation par phosphorylation d'eIF4E et la voie MAPK	53
2. Régulation par les éléments présents sur la région 5'UTR	55
2.1. Les mécanismes de régulation de la traduction par les uORF	57
2.2. Les uORF et la réponse au stress	59
2.3. Les uORF et la stabilité des ARNm	62
2.4. Les uORF et les pathologies humaines	63
3. Les Nouvelle technologie pour étudier les mécanismes de régulation traductionnelle	65
3.1. Qu'est-ce que le <i>ribosome profiling</i> ?	67
3.2. Limites du <i>ribosome profiling</i>	72
3.3. Les biais du <i>ribosome profiling</i>	72
3.3.1 Le choix de la nucléase	73
3.3.2. Les empreintes contaminantes	74
3.3.3. Profondeur de séquençage	75
Objectif de la thèse	76
Résultats	78
I. Analyse globale de la traduction des cellules déplétées en eRF3 ou Upf1	78
1. Stratégie d'étude	78
2. Mise au point des expériences de <i>ribosome profiling</i> et de <i>RNA-seq</i>	79
3. Analyse bioinformatique du <i>RNA-seq</i> et du <i>Ribo-seq</i>	82
4. Analyse d'une première série d'expériences de <i>Ribosome profiling</i>	84
5. Analyse d'une deuxième série d'expériences de <i>Ribosome profiling</i>	88
6. Validation des résultats du <i>Ribosome profiling</i>	89
7. Détection des transcrits exprimés dans les cellules HCT116	91
8. Identification des uORF fonctionnelles dans les cellules HCT116	93
9. Analyse de l'effet des facteurs eRF3 et Upf1 sur l'expression des gènes	97
Article 1 : <i>Involvement of translation termination factor eRF3 and nonsense-mediated mRNA decay factor Upf1 in the translational control of uORFs carrying mRNAs: a genome wide analysis.</i>	
II. Effet de la déplétion en Upf1 sur la formation du ribosome 80S	100
Discussion	105
1. Identification des uORF dans le traductome des cellules HCT116	106
2. Usage des codons d'initiation non canoniques par les uORF fonctionnelles	108
3. Rôles respectifs d'eRF3 et Upf1 dans le contrôle traductionnel par les uORF	109
4. Déplétion en eRF3 et translecture des codons stop	110
Bibliographie	113
Annexes	132
Article 2 : <i>Co-translational deadenylation of mRNAs in human cells: the example of MYC mRNA.</i>	

Préambule

La régulation de l'expression des gènes

L'acide désoxyribonucléique (ADN), support de l'information génétique, contient toutes les informations nécessaires à la formation d'un organisme vivant. Cette information génétique est organisée sous forme de gènes portés par les chromosomes dans le noyau des cellules eucaryotes. Ces gènes s'expriment selon un programme défini qui régule l'état de prolifération de la cellule. Ce programme dépend du profil d'expression spatio-temporelle de ces gènes qui doivent ainsi être exprimés avec le bon taux pour permettre le bon fonctionnement des cellules. Une variation même minime de l'expression des gènes suffit à l'apparition de pathologies ou à une prolifération cellulaire non contrôlée à l'origine des cancers.

L'expression d'un gène se fait par transcription de sa séquence ADN sous forme d'un ARN dit pré-messager (pré-ARNm) dans le noyau, cette étape est le siège d'une première régulation dite régulation transcriptionnelle. L'ARN transcrit est ensuite soumis à une maturation puis exporté vers le cytoplasme sous forme d'ARN messenger (ARNm). Au cours de ces processus, un second niveau de régulation contrôle la maturation et l'acheminement des ARNm. Une fois dans le cytoplasme, les ARN messagers sont pris en charge par les ribosomes et les facteurs de traduction qui assurent l'étape de traduction en protéines. Un troisième niveau de régulation dit post-transcriptionnel et/ou traductionnel intervient à cette étape, modifiant le profil d'expression de la protéine. L'expression des gènes dépend donc de nombreux facteurs qui interagissent pour favoriser ou réprimer leur action en fonction des besoins cellulaires. Des niveaux supplémentaires de régulations post-traductionnelles permettent d'affiner encore cette expression, ce sont par exemple, la dégradation des ARNm, les modifications post-traductionnelles et la dégradation des protéines. Toutes ces régulations fonctionnent sous forme de réseaux qui interagissent entre eux. Ainsi, l'expression d'un gène dans une cellule résulte de la régulation coordonnée de chacun de ces réseaux.

De façon schématique, l'expression des gènes est basée sur un système à trois niveaux : celui de l'ADN, support de l'information génétique, celui de l'ARNm messenger de cette information et celui des protéines qui assurent les fonctions cellulaires. Ces protéines peuvent avoir des fonctions diverses : protéines structurales permettant le maintien de l'architecture de la cellule, enzymes

nécessaires au métabolisme cellulaire, protéines régulatrices de l'expression des gènes. Ces protéines régulatrices sont des acteurs clés de la prolifération et de la différenciation cellulaire permettant le développement de l'organisme, l'élaboration des organes et le renouvellement des tissus.

La synthèse des protéines régulatrices est finement contrôlée au cours de la traduction par des éléments régulateurs portés par leur propre ARN messenger. Ce contrôle permet d'adapter rapidement le taux de ces protéines aux besoins de la cellule. Ce mécanisme de contrôle a une importance fondamentale parce qu'il concerne tous les types de cellules et affecte donc tous les tissus. Le défaut ou la perte du contrôle de traduction des ARN messagers codant les protéines régulatrices de l'expression des gènes peut aboutir à une modification du destin de la cellule et la faire passer d'un état différencié vers un état prolifératif menant au cancer. Si de nombreuses études sur les mécanismes contrôlant l'expression des gènes au niveau du noyau ont permis de mieux comprendre les processus mis en jeu au cours de la prolifération cellulaire, les mécanismes contrôlant la traduction des protéines régulatrices au niveau du cytoplasme sont encore peu décrits. Mon travail de thèse a été d'étudier l'enchaînement des événements qui gouvernent la traduction ou la destruction des ARN messagers, avec une attention particulière pour les ARNm des protéines régulatrices de l'expression des gènes.

Chez les Eucaryotes, la traduction des ARNm peut être régulée non seulement au niveau de la machinerie traductionnelle, ribosome et facteurs de traduction, mais aussi par la présence de motifs de séquence et de structures secondaires au sein même de l'ARNm. Les mécanismes de cette régulation traductionnelle portée par l'ARNm résultent essentiellement d'interactions entre les régions non traduites situées aux extrémités 5' et 3' de l'ARNm avec des protéines à effet en *trans* qui contrôlent spécifiquement la traduction et/ou la stabilité de l'ARNm.

Un de ces éléments régulateurs de la traduction portés par l'ARNm est constitué par un ou plusieurs petits cadres de lecture, ou uORF pour *upstream Open Reading Frame*, situés dans la région 5' non-traduite, en amont du cadre de lecture principal. Plusieurs analyses bioinformatiques ont révélé que plus de 40% des ARNm de mammifères portent de telles uORF régulatrices (Iacono et al., 2005; Timothy G. Johnstone et al., 2016; Raj et al., 2016). Ces uORF sont particulièrement fréquentes dans les transcrits codant des onco-protéines, des récepteurs d'hormones et des modulateurs transcriptionnels. Cependant, les mécanismes moléculaires par lesquels ces éléments

régulent l'expression des gènes ne sont pas complètement élucidés. Ces uORF semblent agir non seulement en contrôlant la traduction de l'ORF principale mais aussi en modulant la dégradation des ARNm par la voie de dégradation des ARNm aberrants ou NMD (*Nonsense Mediated mRNA Decay*).

Au cours de mon introduction, je reviendrai de façon sommaire sur la transcription de l'ARNm eucaryote, sa maturation et son transport du noyau vers le cytoplasme. Je décrirai plus en détail l'ARNm cytoplasmique, les différentes étapes de sa traduction, et la régulation de la traduction en insistant sur le rôle des uORF et leurs moyens d'études.

Introduction

I. L'ARN messenger eucaryote

1. Synthèse et maturation de l'ARN messenger

Dans le noyau de la cellule, la transcription de la séquence ADN sous forme d'un ARN pré-messenger débute par le recrutement ordonné de facteurs de transcription formant un complexe de pré-initiation. Chacun de ces facteurs est lui-même multiprotéiques et joue un rôle précis dans le processus d'initiation de la transcription : la reconnaissance des éléments du promoteur, l'orientation du complexe de pré-initiation, l'ouverture de la double hélice, le recrutement de l'ARN polymérase II (Pol II) et son activation. L'ARN polymérase II se déplace le long de l'ADN tout en synthétisant le brin d'ARN : le pré-messenger (pré-ARNm). Au cours de cette synthèse, des protéines se fixent à la molécule d'ARN et initient les processus de maturation post-transcriptionnelle : ajout d'une coiffe 7-méthylguanosine en 5', épissage des introns, polyadénylation en 3'. Ces deux processus, transcription et maturation des ARN se font de façon simultanée dans les cellules. La terminaison de la transcription est déterminée par plusieurs signaux dont celui de polyadénylation.

1.1. Maturation de l'extrémité 5' de l'ARN messenger

Une fois les 25 à 30 premiers nucléotides transcrits, la première étape de la maturation des ARNm consiste en l'ajout d'une coiffe (Shatkin, 1976). La coiffe est constituée d'une guanine méthylée en position N7, qui est ajoutée au premier nucléotide à l'extrémité 5' des pré-ARNm par une liaison 5'-5' triphosphate. Cette coiffe joue un rôle important dans le métabolisme des ARNm, elle participe en particulier à leur protection contre la dégradation dans le sens 5'-3' par les exonucléases (Lewis and Izaurralde, 1997). De plus, elle joue aussi un rôle important dans la poursuite des étapes ultérieures de maturation et notamment dans les réactions d'épissage, de polyadénylation nucléaire, de transport nucléocytoplasmique. Enfin, elle est un élément clé de l'initiation de la traduction.

Dans le noyau, la coiffe est liée à un complexe nucléaire CBC (*Cap Binding Complex*) constitué des protéines CBP80 et CBP20 (*Cap Binding Protein 80 et 20*). Ce complexe en plus de

protéger l'ARN contre la dégradation dans le sens 5'-3', participe au couplage de la transcription à la maturation. Grâce à différents partenaires d'interaction, le complexe de coiffe intervient dans la mise en place de l'épissage des introns, dans l'export des transcrits ainsi que dans le processus de maturation de la région 3' du pré-ARNm. Après l'export de l'ARNm mature dans le cytoplasme, le complexe CBC nucléaire, par interaction de CBP80 et du facteur d'initiation de la traduction eIF4G, permet l'initiation du premier tour de traduction durant lequel s'effectue le contrôle qualité des ARNm par la voie NMD (Kim et al., 2009; Maquat et al., 2010). Cette voie sera détaillée plus loin dans le chapitre concernant la traduction.

Une fois l'ARN messenger exporté dans le cytoplasme le complexe CBC nucléaire est remplacé par le facteur de coiffe cytoplasmique eIF4E qui fait partie du complexe d'initiation de la traduction eIF4F. Les différents rôles de ces facteurs, dans la circularisation de l'ARNm, dans le recrutement de la machinerie traductionnelle ainsi que dans les processus de régulation de la traduction seront détaillés plus loin dans le chapitre consacré à la traduction.

1.2. Epissage des ARN messagers

Les pré-ARNm sont constitués d'une alternance de séquences qui feront parties de l'ARNm mature, les exons, et de séquences excisées, les introns. Les introns sont en général dans les régions codantes, plus rarement dans la région 5' non traduite (5'UTR) de l'ARNm et exceptionnellement dans la région 3' non traduite (3'UTR). Au cours de leur maturation les pré-ARNm subissent une étape d'épissage, qui permet l'excision et l'élimination des introns, pour ne conserver que les exons et les lier entre eux (Jurica and Moore, 2003). La taille des introns varie de quelques dizaines à plusieurs centaines de milliers de nucléotides chez l'homme, avec une taille moyenne de 1900 nucléotides (nt) contre 125 nt pour les exons. On estime qu'il y a en moyenne 7,8 introns contre 8,8 exons par gène (Hoskins and Moore, 2012).

L'excision des introns et la ligation des exons sont réalisées en deux réactions de trans-estérifications successives par la machinerie d'épissage ou spliceosome (Jurica and Moore, 2003). Le spliceosome reconnaît les sites consensus d'épissage et catalyse les réactions de trans-estérifications par rapprochement des sites donneurs et accepteurs d'épissage situés de part et d'autre des introns (Will and Luhrmann, 2011). Le complexe de jonction d'exons EJC (*Exon*

Junction Complex) est alors déposé par le spliceosome, 24 nucléotides en amont de la jonction exon-exon (Le Hir, 2000; Le Hir et al., 2000). L'EJC, qui est transporté avec l'ARNm mature vers le cytoplasme, est impliqué dans l'export des ARNm hors du noyau. L'EJC participe aussi à la localisation subcellulaire des ARNm ainsi que dans la régulation de leur niveau de traduction (Giorgi and Moore, 2007). Enfin, l'EJC joue un rôle clé dans le contrôle de qualité des ARNm et sera dissocié à l'issue du premier cycle de traduction.

L'épissage des ARN messagers peut être alternatif. Dans ce cas, certains exons peuvent être exclus et, de la même façon, certains introns peuvent être maintenus dans la séquence de l'ARNm mature (**Figure 1**). Ainsi, à la fin de l'étape d'épissage, le transcrit d'un même gène donne plusieurs ARNm matures, qui codent pour différents variant protéiques. On a longtemps estimé qu'environ 60% des gènes subissent un processus d'épissage alternatif, permettant de générer 100 000 protéines différentes à partir des 24 000 gènes du génome humain (Nilsen and Graveley, 2010). Cependant, avec l'arrivée des nouvelles technologies de séquençage à haut débit, cette proportion semble être beaucoup plus importante et il semblerait qu'environ 90% des gènes sont le siège d'un épissage alternatif chez l'homme (Wang et al., 2008). L'épissage alternatif est un processus important de la régulation de l'expression des gènes.

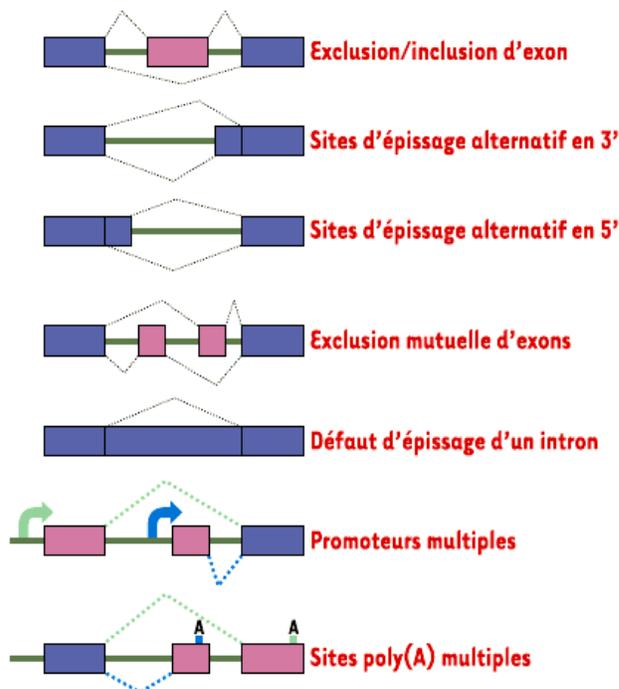


Figure 1 : Différents types d'épissage alternatif des ARN pré-messagers. En bleu les exons constitutifs, en violet les exons alternatifs, et en traits pleins les introns. Différentes combinaisons d'assemblage des exons et de rétention d'introns sont représentées. L'utilisation de multiples promoteurs peut aussi modifier l'épissage. Enfin, l'épissage alternatif peut conduire à l'utilisation de signaux de polyadénylation distincts (Corcos and Solier, 2005).

1.3. Maturation de l'extrémité 3' de l'ARN messenger

A la fin de la transcription, la maturation des pré-ARNm se termine par l'ajout d'une séquence poly-Adénosine ou poly(A) en 3' de la région 3'UTR. A l'exception des histones, tous les transcrits synthétisés par l'ARN polymérase II sont polyadénylés. Cette queue poly(A) peut avoir une taille variable selon le transcrit, allant de 200 à 300 résidus adénosine chez l'homme.

L'ajout de la queue poly(A) se fait en plusieurs étapes. Dans un premier temps, les pré-ARNm sont clivés au niveau d'une séquence poly(U) ou riche en doublets GU, appelée DSE pour *DownStream Element*, dans une région située à environ 10 à 30 nucléotides en aval du signal de polyadénylation, typiquement de forme AAUAAA. Une dizaine de résidus adénosine sont alors ajoutés à l'extrémité 3' par la Poly(A) Polymerase (PAP). La distance entre le signal AAUAAA et le DSE détermine donc la région dans laquelle va s'effectuer le clivage. Le site précis de cette coupure est de préférence en 3' d'un résidu A, le plus souvent après le di-nucléotide CA. L'efficacité du clivage semble être influencée par une séquence en amont, riche en résidus U appelée USE pour *UpStream Element*, par l'environnement nucléotidique du signal AAUAAA et probablement par la structure secondaire même de la 3'UTR (Tian and Graber, 2012). Cependant, les modalités du clivage ne sont que partiellement connues. Deux facteurs, le CPSF (*cleavage and polyadenylation specificity factor*) et le CstF (*cleavage stimulation factor*), qui se fixent respectivement sur le signal AAUAAA et sur le DSE, sont indispensables (Millevoi and Vagner, 2010). De même, la PAP semble intervenir dans le processus, mais la protéine qui porte l'activité endonucléase n'a pas encore été identifiée. Le clivage produit une extrémité 3'-OH sur les pré-ARNm, à laquelle la PAP ajoute une dizaine de résidus A, alors que le fragment présentant l'extrémité 5'-phosphate est rapidement dégradé. Dans un second temps, la séquence poly(A) est allongée, et cette extension dépend uniquement de la présence des 10 premiers résidus précédemment ajoutés et du CPSF. En effet, la séquence de 10 résidus A est nécessaire et suffisante pour que la protéine PABPN1 (*Poly(A) Binding Protein Nuclear 1*) s'y fixe et stimule, en coopération avec le CPSF, l'activité de la PAP en augmentant sa processivité (Kühn and Wahle, 2004) et permettant l'addition d'adénosines supplémentaires (Parent and Bisailon, 2006). La longueur initiale de la séquence poly(A) est d'environ 200 à 300 résidus adénosines chez les eucaryotes supérieurs et semble être contrôlée par la PABPN1 qui en changeant de conformation permet la dissociation de PAP et l'arrêt de la polymérisation. Cette queue poly(A) via son

interaction avec PABPN1, intervient dans le transport nucléo-cytoplasmique des ARNm. Une fois l'ARNm mature transporté dans le cytoplasme, la PABPN1 est remplacé par son homologue cytoplasmique PABPC1 qui le protège contre la dégradation par les ribonucléases agissant dans le sens 3' vers 5' et qui intervient dans l'initiation de leur traduction et dans le contrôle de leur stabilité.

2. Structure de l'ARNm mature eucaryote

L'ARN messenger eucaryote mature est une structure tripartite, la séquence codant la protéine se trouve au milieu du transcrit, au sein de la phase ouverte de lecture appelée ORF (*Open Reading Frame*), cette dernière commence par un codon d'initiation et se termine par un codon stop. Elle est constituée de triplets de nucléotides ou codons qui codent chacun un acide aminé. L'ARN messenger eucaryote ne peut être réduit à son cadre de lecture mais possède aussi en amont et en aval des séquences dites non traduites appelées 5' et 3' UTR (*Untranslated Transcribed Region*). Ces éléments sont impliqués dans la régulation de la traduction et dans la stabilité des ARN messagers.

En plus de la coiffe, la région 5'UTR d'un ARNm peut contenir des régions structurées telles que les IRES (*Internal Ribosome Entry Site*) qui peuvent recruter la machinerie de traduction indépendamment de la coiffe, et des uORF, qui, plus favorablement traduites que l'ORF principale, sous certaines conditions particulières (stress, carence, *etc.*) sont impliquées dans la régulation de la traduction ainsi que dans la stabilité des ARN messagers.

Dans sa région 3'UTR, l'ARNm peut contenir des régions structurées, des motifs de fixation de diverses protéines régulatrices liant l'ARN appelées de façon génériques RBP (*RNA-Binding Proteins*), des séquences spécifiques permettant l'appariement de petits ARN (miRNA) et enfin, la queue polyA (**Figure 2**). Si les séquences spécifiques assurant la liaison de protéines ou encore des miARN peuvent être présentes dans l'ensemble du messenger, elles se trouvent cependant de préférence dans les régions 3'UTR. Les éléments permettant l'adressage des ARNm et leur rétention au niveau des différents compartiments cytoplasmiques sont essentiellement situés dans les régions 3'UTR (St Johnston, 1995), sans doute parce qu'à l'inverse des 5'UTR, ces régions ne sont soumises à aucune contrainte, et fournissent un terrain propice à l'évolution des éléments de

régulation. Ainsi, la région 3'UTR est nécessaire et suffisante au contrôle de la distribution d'un ARNm réparti de façon non uniforme. Par son interaction avec des protéines régulatrices agissant en *trans*, l'ARNm va être dirigé et retenu dans le site où la protéine correspondante sera ultérieurement nécessaire.

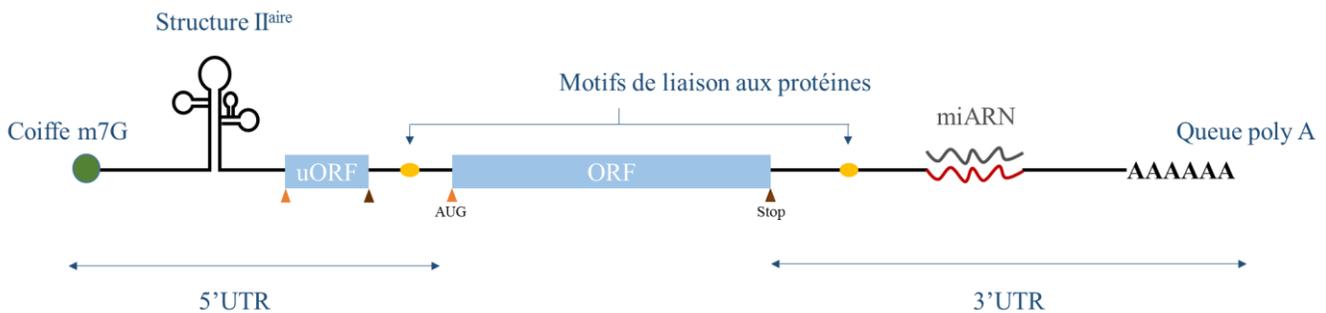


Figure 2 : Structure de l'ARNm mature. L'ARNm eucaryote porte une coiffe en 5', la région 5'UTR, le cadre ouvert de lecture ou ORF, la région 3'UTR et souvent une queue poly(A) en 3'. De plus, l'ARNm peut éventuellement contenir un IRES et des uORF en 5'UTR, et une séquence d'ancrage de miARN en 3'UTR. Les motifs cis-régulateurs de liaison aux protéines de liaison à l'ARN peuvent se trouver en 5' comme en 3'UTR.

De leur synthèse jusqu'à leur export dans le cytoplasme (**Figure 3**), les ARNm sont associés à de nombreuses protéines et forment différents complexes ribonucléoprotéiques qui subissent de nombreux remodelages. Dans le noyau, les ARNm coiffés, épissés et polyadénylés, sont associés à des protéines sous forme de complexes appelés hnRNP pour *heterogeneous nuclear RiboNucleoProtein*, avant d'être exportés vers le cytoplasme (Izaurralde and Mattaj, 1995; Nakielny and Dreyfuss, 1997). Certaines de ces protéines vont accompagner l'ARNm jusqu'au cytoplasme. Les protéines impliquées dans le transport nucléo-cytoplasmique des ARNm se fixent sur des séquences appelées NES (*nuclear export sequence*). De plus, les constituants du pore nucléaire franchi par l'ARNm participent aussi à ce transport actif. Après son acheminement vers le compartiment cytoplasmique approprié, l'ARNm est soit traduit, soit stocké sous forme de particules ribonucléoprotéiques appelés mRNP pour *messenger ribonucleoprotein* (Hershey et al., 1996). Ces mRNP constituent une sorte de réservoir dans lequel les ARNm sont sous forme inactive mais stable, puisqu'ils sont inaccessibles à l'appareil de traduction et indisponibles pour les réactions de désadénylation ou de dégradation cytoplasmiques (Moore and Proudfoot, 2009). Le

lieu de la dégradation des ARNm est encore mal connu. Il a longtemps été pensé que les ARNm étaient dégradés sous leur forme de stockage, au sein de structures cytoplasmiques non entourées de membranes appelées granules ou *P-bodies* pour *processing bodies*. Cette idée est néanmoins controversée et certains auteurs ont émis l'hypothèse que les ARNm étaient dégradés en majorité sous leur forme libre dans le cytosol (Weil, 2007). Il a été montré récemment chez la levure *Saccharomyces cerevisiae* que la dégradation des ARNm pouvait se faire au cours de la traduction (Pelechano et al., 2015). Chez les métazoaires, la dégradation co-translationnelle des ARNm pourrait avoir lieu seulement dans des conditions particulières comme au cours de l'embryogenèse ou au moment de la mitose (Park et al., 2016; Subtelny et al., 2014).

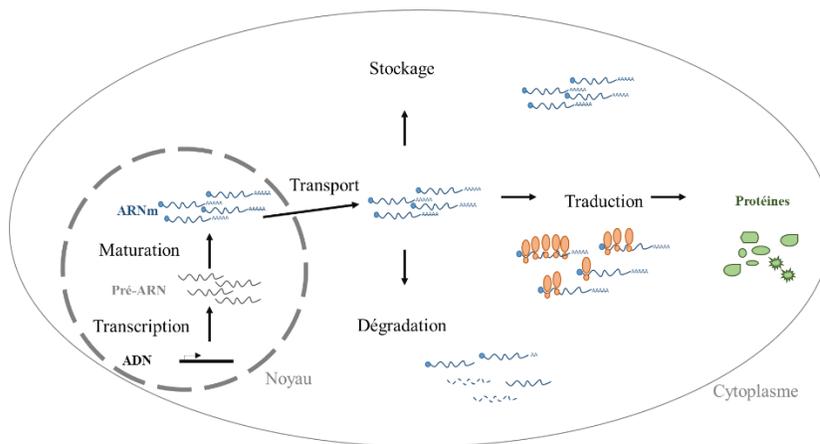


Figure 3 : Synthèse et devenir de l'ARNm eucaryote. Lorsque l'ARNm est transcrit, il est soumis à une maturation dans le noyau (ajout de la coiffe et de la queue poly(A)), puis il est exporté vers le cytoplasme où il peut être ensuite soit stocké, soit traduit, soit dégradé.

II. La traduction chez les eucaryotes

Arrivé à destination, l'ARNm mature sera traduit par un mécanisme découpé en trois étapes successives, l'initiation, l'élongation et la terminaison. L'initiation peut se dérouler principalement selon deux modèles, celui du balayage et celui de l'entrée interne des ribosomes. La phase d'initiation consiste dans le recrutement de la machinerie traductionnelle sur l'ARN messager, et le démarrage de la traduction. Lorsque la traduction est initiée, l'ARNm entre dans la phase d'élongation. Lors de cette étape, les acides aminés codés par l'ARNm sont ajoutés successivement au peptide en cours de synthèse. L'élongation se termine lorsque le ribosome rencontre un codon

stop. L'ARNm entre alors dans la phase de terminaison de la traduction. Elle est responsable de la dissociation des deux sous-unités ribosomiques fixées au messager depuis l'initiation, de la libération de la protéine néosynthétisée et du recyclage des sous-unités-ribosomiques. Les régulations traductionnelles ont lieu au cours des trois phases successives de la traduction et touchent à l'efficacité de la machinerie traductionnelle.

1. L'initiation

L'initiation permet à la machinerie traductionnelle de se mettre en place sur l'ARNm et la spécification de la phase de lecture. Il existe deux mécanismes principaux d'initiation de la traduction : l'initiation canonique dépendante de la coiffe utilisée par la grande majorité des ARNm et l'initiation dépendante des IRES pour *Internal Ribosome Entry Site*. Ce dernier mécanisme est indépendant de la coiffe et repose sur la présence de séquences particulières de l'ARNm formant des structures secondaires stables, qui, avec l'aide de protéines associées aux IRES, permettent le recrutement du ribosome directement sur le codon d'initiation. La traduction IRES dépendante permet la synthèse de protéines essentielles dans des conditions physiologiques particulières où les facteurs canoniques d'initiation sont indisponibles comme au cours des phases S et G2 du cycle cellulaire, en cas de stress cellulaires, au cours de l'apoptose ou dans des conditions pathologiques comme la prolifération cellulaire menant à certains cancers (Komar and Hatzoglou, 2015; Pestova et al., 2007). Il existe enfin d'autres mécanismes plus rares d'initiation de la traduction, notamment le mécanisme d'initiation hybride entre ces deux mécanismes, mis en évidence pour la traduction de l'histone H4 (Martin et al., 2011) et qui a été décrit pour les ARNm avec une très courte région 5' non traduite (Sinvani et al., 2015).

Cependant, chez les eucaryotes, l'essentiel de la traduction se fait de manière dépendante de la coiffe (von der Haar et al., 2004) et fait intervenir plus d'une douzaine de facteurs notés eIF pour *eukaryotic Initiation Factors*, ainsi qu'un facteur auxiliaire, la protéine de liaison à la queue polyA, PABP (**Tableau I**). Ils permettent le recrutement des deux sous-unités ribosomiques et l'association du codon initiateur, le plus souvent un codon AUG, avec l'anticodon porté par l'ARN de transfert (ARNt) initiateur (revue dans Sonenberg and Hinnebusch, 2009). Je vais décrire maintenant plus en détails les étapes de l'initiation dépendant de la coiffe afin de pouvoir discuter ultérieurement de la régulation de l'initiation de la traduction par l'intermédiaire des uORF.

Tableau I : Les facteurs d'initiation de la traduction eucaryotes (Jackson et al., 2010).

Nom	Nombre de sous-unités	Masse moléculaire (kDa)	Fonction
eIF1	1	12.7	Stimule l'interaction du complexe ternaire avec la sous-unité 40S du ribosome et prévient l'hydrolyse prématurée d'eIF2-GTP par eIF5. Contribue au scanning de l'ARNm et à la reconnaissance du codon initiateur.
eIF1A	1	16.5	Stimule la liaison du complexe ternaire avec la 40S. Contribue avec eIF1 à la reconnaissance du codon initiateur.
eIF2	3	36.1, 38.4, 51.1	Forme le complexe ternaire eIF2-GTP-Met-ARN ^{ti} qui lie la sous-unité ribosomique 40S.
eIF2B	5	33.7, 39, 50.2, 59.7, 80,3	GEF (<i>Guanosine nucleotide Exchange Factor</i>) impliqué dans l'échange du GDP en GTP sur eIF2.
eIF3	13	800	S'associe à la 40S, eIF1, eIF4G, eIF5 et stimule l'association du complexe ternaire à la 40S pour former le complexe 43S. Favorise l'association du complexe 43S avec l'ARNm et le scanning de l'ARNm. Prévient l'association prématurée des sous-unités ribosomique 40S et 60S.
eIF4A	1	46.1	ARN hélicase ATP dépendante de la famille des DEAD-box, impliquée dans le déroulement des structures secondaires. Elle favorise le balayage de l'ARNm.
eIF4B	1	69.3	RBP (<i>RNA binding protein</i>) qui favorise l'activité hélicase d'eIF4A.
eIF4E	1	24.5	Se lie à la coiffe m ⁷ -GTP située à l'extrémité 5' de l'ARNm.
eIF4G	1	175.5	Protéine d'échafaudage qui lie eIF4E, eIF4A, eIF3, PABP et l'ARNm. Stimule l'activité d'eIF4A.
eIF4H	1	27.4	RBP homologue à une région d'eIF4B qui favorise également l'activité hélicase d'eIF4A.
eIF5	1	49.2	GAP (<i>GTPase-Activating Protein</i>), favorise l'activité GTP-ase d'eIF2, induisant l'hydrolyse du GTP en GDP du facteur eIF2-GTP lors de la reconnaissance du codon initiateur.

eIF5B		138.9	GTP-ase qui favorise l'association des deux sous-unités ribosomiques.
eIF6	1	26.6	Lie la sous-unité 60S et empêche son association prématurée avec la sous-unité 40S.
PABP	1	70.7	Se lie à la queue poly(A) à l'extrémité 3' de l'ARNm et avec eIF4G. Permet ainsi la circularisation de l'ARNm ce qui facilite le recyclage de la sous-unité 40S en 5' de l'ARNm, après la terminaison. Se lie au facteur de terminaison eRF3.

1.1. Les étapes de l'initiation de la traduction dépendante de la coiffe

1.1.1. Formation du complexe de pré-initiation 43S

La formation du complexe de pré-initiation 43S est la première étape de l'initiation de la traduction coiffe dépendante. Cette étape nécessite au préalable la formation du complexe ternaire ainsi que l'assemblage des facteurs d'initiation eIF1, eIF1A, eIF3, eIF5 sur la petite sous unité ribosomique 40S (**Figure 4**).

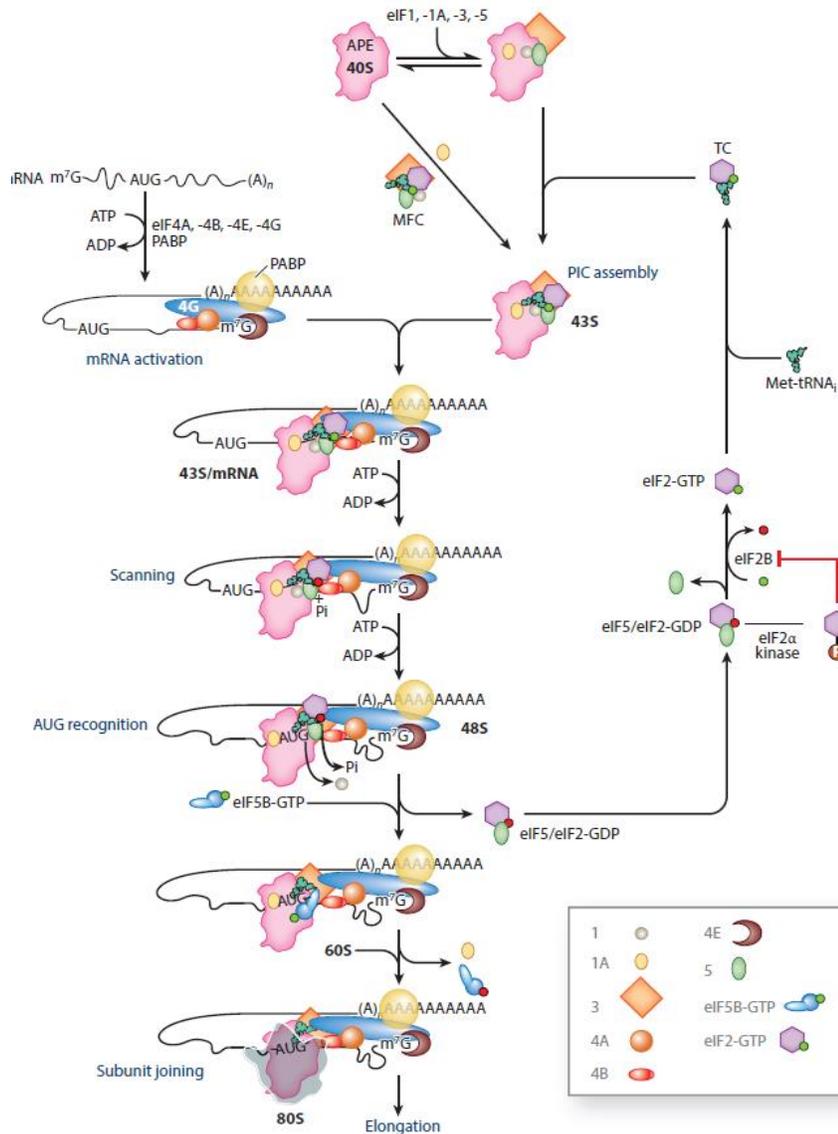


Figure 4 : Schéma général de l'initiation de la traduction coiffe dépendante chez les eucaryotes d'après Hinnebusch (2014). La formation du complexe ternaire (TC) qui consiste en l'association de l'ARNt initiateur chargé en méthionine et du facteur eIF2-GTP, est une étape préliminaire à l'initiation. Ensuite, la première étape de l'initiation est la formation du complexe 43S (PIC) qui se fait soit en une seule étape par association du complexe multi-factoriel MFC (TC+eIF5+eIF1+eIF1A+eIF3) avec la sous-unité 40S, soit en deux étapes par association du complexe ternaire et de la sous-unité 40S, portant déjà les facteurs eIF1, eIF1A, eIF3 et eIF5. En parallèle, le complexe eIF4F (eIF4A/eIF4E/eIF4G) avec ses modulateurs (eIF4B et PABP) se forme, par interaction de la protéine de liaison à la coiffe eIF4E liée à l'ARNm et de la protéine d'échafaudage eIF4G. La protéine PABP, liée à la queue poly(A) se lie à eIF4G et permet la circularisation de l'ARNm pour faciliter sa traduction. Il y a ensuite chargement du complexe 43S sur le complexe eIF4F lié à l'ARNm ce qui forme le complexe 48S. Ce dernier scanne la région 5'UTR de l'ARNm à la recherche du codon initiateur (AUG). La reconnaissance du codon initiateur induit la libération des facteurs d'initiation eIF5 et eIF2-GDP et le changement de conformation du PIC. Le recrutement de la grande sous-unité ribosomique (60S) au niveau du PIC par le facteur eIF5B-GTP permet de former le complexe d'initiation (80S). Cette étape est accompagnée par le relargage des facteurs eIF5B-GDP et eIF1A et signe la fin de l'initiation et le début de l'élongation de la traduction. Le recyclage du facteur eIF2-GDP est assuré par le facteur d'échange eIF2B.

1.1.1.1. La formation du complexe ternaire eIF2-GTP-Met-ARN_t^{Met}

Le facteur eIF2 est une GTP-ase qui se présente sous deux formes, eIF2-GTP et eIF2-GDP. Le recyclage de la forme eIF2-GDP en eIF2-GTP s'effectue grâce au facteur d'échange eIF2B. Lorsqu'eIF2 est sous sa forme active liée au GTP, il s'associe spontanément avec le Methionyl-ARN_t initiateur pour former le complexe ternaire eIF2-GTP-Met-ARN_t^{Met} (Sonenberg and Hinnebusch, 2009). Le Methionyl-ARN_t initiateur assure l'apport de la méthionine initiatrice qui formera l'extrémité *N*-terminale de la protéine en cours de synthèse (**Figure 4**). L'ARN_t initiateur a une grande affinité pour eIF2-GTP et n'en a pas pour eIF2-GDP. C'est cette différence d'affinité qui va permettre sa dissociation après hydrolyse du GTP lors de la reconnaissance du codon initiateur (Kapp and Lorsch, 2004).

1.1.1.2. Les facteurs d'assemblage du complexe de pré-initiation 43S

En parallèle à la formation du complexe ternaire, la petite sous unité ribosomique 40S s'associe aux facteurs eIF1, eIF1A, eIF3 et eIF5. Cette association favorise la fixation du complexe ternaire sur la 40S pour former le complexe de pré-initiation 43S (Majumdar et al., 2003). Les deux sous-unités ribosomiques sont maintenues dissociées grâce aux facteurs d'initiation qui se fixent sur la surface d'interaction de la 40S à la 60S, prévenant ainsi la formation de ribosome 80S inactif (**Figure 4**) :

- eIF1 se fixe près du site P et interagit avec eIF3,
- eIF1A se fixe au niveau du site A près du canal d'entrée de l'ARN_m et projette ses extrémités dans la cavité du site P, ce qui lui permet d'interagir avec les facteurs eIF3 et eIF2 et de stabiliser le complexe ternaire qui s'insère au niveau du site P,
- eIF5 interagit par ces extrémités N et C terminales avec eIF2 et eIF3.

Le facteur eIF3 est nécessaire à la liaison du complexe ternaire sur la petite sous-unité ribosomique 40S (Dong and Zhang, 2006), et il est aussi capable d'interagir directement avec l'ARN messager (Hinnebusch, 2006).

Les facteurs eIF1, eIF1A et eIF3 une fois liés à la petite sous-unité 40S facilitent la liaison des autres facteurs et permettent un changement de conformation de la sous-unité 40S du ribosome. Cette action ouvre le site de liaison à l'ARN_m (Liljas, 2013; Passmore et al., 2007; Sonenberg and

Hinnebusch, 2009). Le maintien d'une conformation ouverte de la sous-unité 40S faciliterait le balayage de l'ARNm par le complexe d'initiation jusqu'au codon initiateur.

1.1.2. Formation du complexe eIF4F

Le complexe 43S doit ensuite être recruté en 5' sur l'ARNm. Cette association n'est possible que si certains facteurs d'initiation sont déjà liés à la coiffe m⁷-GTP de l'ARNm. Ainsi, de façon concomitante à la formation du complexe de pré-initiation 43S, le complexe eIF4F se forme au niveau de la coiffe. Ce complexe est composé de trois facteurs d'initiation de la traduction eIF4A, eIF4E et eIF4G et de modulateurs eIF4B et/ou eIF4H et PABP. Le facteur eIF4E se lie directement à la coiffe de l'ARNm ce qui permettra le positionnement du complexe 43S à l'extrémité du transcrit (Sonenberg et al., 1979). Le facteur eIF4E recrute la protéine d'échafaudage eIF4G, l'interaction entre ces deux protéines permet d'ancrer ces protéines de façon stable sur l'ARNm grâce aux sites de liaison à l'ARNm d'eIF4G. Le facteur eIF4G se lie à l'ARN via des motifs RRM (*RNA recognition motif*) même en l'absence du facteur eIF4E (Mitchell et al., 2010). Ainsi la liaison d'eIF4E à la coiffe sert aussi à prévenir la formation de complexes aberrants sur l'ARNm. Sur le facteur eIF4G vient se fixer le facteur eIF4A, qui est une hélicase de l'ARN qui prévient la formation de structures secondaires dans la région proche de la coiffe de l'ARNm. La protéine PABP, qui fixe la queue poly(A) du messenger est également recrutée par la protéine d'échafaudage eIF4G. Les liaisons de PABP à eIF4G et à la queue poly(A) rapprochent les régions 5' et 3'UTR du messenger ce qui permet la circularisation de l'ARNm (Wells et al., 1998). Dans cette configuration, la petite sous-unité ribosomique 40S est liée par l'intermédiaire des facteurs d'initiation à l'extrémité 5' de l'ARNm ainsi qu'à son extrémité 3'. Selon ce modèle en boucle, PABP permet une synergie entre la coiffe et la queue poly(A) de l'ARN, ce qui augmente les efficacités de traduction et de recyclage des composantes du complexe d'initiation de la traduction (Amrani et al., 2008; Kahvejian et al., 2001; Rajkowitsch et al., 2004).

1.1.3. Formation du complexe d'initiation 48S

Le facteur eIF4F (eIF4A/eIF4E/eIF4G) a un rôle essentiel dans la mise en place du complexe d'initiation 48S, grâce au facteur eIF4G qui recrute le complexe de pré-initiation 43S via son interaction avec le facteur eIF3 (sous-unité eIF3e) porté par le complexe 43S. Des expériences menées *in vitro* ont montrés que le complexe de pré-initiation 43S se lie directement sur les ARNm dépourvus de structures secondaires (Pestova and Kolupaeva, 2002). Cependant les ARNm présentant des structures secondaires dans leur région 5'UTR sont déroulés au préalable au recrutement du complexe de pré-initiation 43S. L'activité hélicase ATP-dépendante du facteur eIF4A assure le dépliement de ces structures secondaires présentes sur la région 5'UTR de l'ARNm (Hinnebusch, 2014). L'efficacité de cette hélicase est potentialisée par les facteurs d'initiation se liant à l'ARN comme eIF4B et eIF4G (**Figure 4**). Cette efficacité peut être également stimulée par la protéine eIF4H homologue d'eIF4B (Rogers et al., 2001; Rozovsky et al., 2008).

1.1.4. Balayage ou «scanning» de la région 5'UTR

Une fois formé, le complexe d'initiation 48S doit retrouver le codon initiateur, qui peut se situer à plusieurs centaines de nucléotides en aval de la coiffe. Contrairement aux procaryotes où le codon initiateur est balisé par la séquence "Shine-Dalgarno", site de fixation du ribosome située à environ 6 à 12 nucléotides en amont du codon initiateur, chez les eucaryotes, la recherche du codon initiateur est assurée par le balayage linéaire de la région 5'UTR de l'ARNm. Ce balayage est permis encore une fois grâce à l'hélicase eIF4A, aidée par les facteurs eIF4G, eIF4B ou eIF4H. D'autres hélicases peuvent être requises pour le balayage des ARNm présentant des régions 5'UTR fortement structurées. Par exemple chez la levure, l'hélicase Ded1 intervient dans l'initiation dans le cas d'ARNm avec une longue région 5'UTR (von der Haar and McCarthy, 2002). Chez les mammifères, les hélicases DDX3, DHX29 et RHA interviennent dans l'initiation dans le cas d'ARNm avec une région 5'UTR fortement structurée (Soto-Rifo et al., 2012). L'initiation se fait lors de la rencontre du complexe d'initiation 48S et d'un codon initiateur dans un contexte favorable.

1.1.5. Reconnaissance du codon initiateur

L'identification du codon initiateur AUG est assurée principalement par antagonisme des facteurs eIF1 et eIF5. En effet, Le facteur eIF5 favorise la dissociation du facteur eIF1 du complexe de pré-initiation 43S et le passage de la sous-unité 40S à une conformation fermée. Une étude a montré que la surexpression d'eIF5 dans des cellules de mammifère altère la fidélité de reconnaissance des codons initiateurs tandis que ce phénomène est rétabli par la surexpression

concomitante des facteurs eIF1 et eIF5 (Loughran et al., 2012). Par ailleurs, une étude a montré que le facteur eIF1 est capable de dissocier les ribosomes ayant initié sur de mauvais codons initiateurs (Pisarev et al., 2006). Ces études montrent que le facteur eIF1 est en compétition avec l'appariement codon-anticodon. Une fois le codon initiateur reconnu, il y a accommodation de l'ARNt initiateur dans le site P du ribosome, hydrolyse du GTP en GDP du facteur eIF2 et déplacement des facteurs eIF1 et eIF1A par encombrement stérique. Le déplacement de ces facteurs est favorisé par le facteur eIF5 (**Figure 4**). Une fois le GTP hydrolysé, eIF2-GDP est relargué car il perd son affinité pour le complexe d'initiation. Il semblerait que le facteur eIF3 participe à la reconnaissance des codons initiateurs via ses sous-unités a, b et J (Chiu et al., 2010).

1.1.6. Les déterminants de la reconnaissance du codon initiateur

Chez les eucaryotes la nature des nucléotides environnants le codon initiateur ont un rôle primordial dans la reconnaissance du codon initiateur. L'importance de chacun des nucléotides entourant le codon initiateur a été déterminée par Marylin Kozak (Kozak, 1986). Selon ces règles, le codon AUG, codon initiateur de loin le plus fréquent, est dans un contexte très favorable à l'initiation lorsqu'il est entouré par la séquence nucléotidique 5'-RCCAAUGG-3', où l'adénosine du codon initiateur représente la position +1 de la phase de lecture et le résidu R en position -3 représente une purine. Cette séquence est souvent appelé "le contexte de Kozak". Il a été montré que les positions nucléotidiques -3 et +4 sont particulièrement importantes pour l'efficacité d'initiation de la traduction car ils stabilisent les changements conformationnels qui interviennent après l'appariement codon-anticodon. Ces résidus interagissent respectivement avec la sous-unité alpha du facteur eIF2 et l'ARN ribosomique 18S (ARNr) qui forme partiellement le site A et sont en compétition avec la liaison du facteur eIF1 (Pisarev *et al.*, 2006). En cas de contexte défavorable,

le complexe d'initiation de la traduction ignore l'AUG et continue le balayage. Eventuellement, l'initiation de la traduction se fera au prochain codon d'initiation AUG dans un contexte favorable. Ce mécanisme basé sur la force du codon d'initiation permet une diversification de l'extrémité N-terminale des protéines synthétisées et une modulation de l'efficacité d'initiation de la traduction.

L'initiation de la traduction peut toutefois s'effectuer au niveau d'autres codons que le codon AUG, à des codons initiateurs dits non-canoniques comme les codons CUG, ACG ou GUG. Cependant, si l'initiation sur ces codons non-canoniques est inefficace, elle est renforcée par la présence du contexte optimal de Kozak, en particulier par une purine dans la position -3 et par un G en position +4.

1.1.7. Arrivée de la sous-unité ribosomique 60S

Après reconnaissance du codon d'initiation, les facteurs d'initiation sont libérés à la suite de l'activité GTP-ase d'eIF5B qui stimule l'hydrolyse du GTP lié à eIF2 et le passage de la sous-unité 40S à une conformation fermée (Pestova et al., 2000). La dissociation des facteurs d'initiation eIF1, eIF2-GDP et eIF5 libère la surface d'interaction de la sous-unité 40S avec la sous-unité 60S. Après hydrolyse du GTP en GDP du facteur eIF5B, ce dernier perd son affinité pour le complexe d'initiation et se dissocie spontanément entraînant avec lui la dissociation du facteur eIF1A. Ce mécanisme permet ensuite la fixation de la grande sous-unité ribosomique 60S pour la constitution du ribosome 80S et le début de l'élongation (Nanda et al., 2013). Cependant, aucune étude ne montre précisément le devenir du facteur eIF3, qui semble rester fixé sur la sous-unité 40S (**Figure 4**).

Enfin, le facteur eIF2B qui est une GEF (*Guanosine nucleotide Exchange Factor*) assure l'échange GDP/GTP d'eIF2 afin de régénérer le facteur eIF2-GTP et former un complexe ternaire actif pour le cycle d'initiation suivant (Jackson et al., 2010; Sonenberg and Hinnebusch, 2009).

1.2. Le cas du premier tour de traduction

Chez les eucaryotes, l'épissage des introns peut être accompagné par des erreurs aboutissant à des ARNm contenant un cadre de lecture ouvert raccourci par un codon de terminaison prématuré (McGlinchy and Smith, 2008). Les codons de terminaison prématurés peuvent également être dus à une mutation non-sens dans la séquence du gène ou introduits par une erreur survenue au cours de la transcription.

La voie de dégradation des ARNm dite NMD est une voie de surveillance de l'ARNm traduction-dépendante qui détecte et élimine les transcrits contenant des codons de terminaison prématurés. Ces codons de terminaison prématurés ont le potentiel d'être délétères par les protéines tronquées qu'ils codent. Alors que l'initiation de la traduction est un mécanisme canonique qui est conduit grâce au facteur de coiffe eIF4E, le premier tour de traduction, premier ribosome qui traduit un ARNm donné, obéit à un mécanisme particulier qui permet à la machinerie de contrôle qualité NMD de se mettre en place pour vérifier l'intégrité de l'ARNm. Ce processus est mis en jeu immédiatement après l'export de l'ARNm dans le cytoplasme grâce au complexe de coiffe CBC (**Figure 5**). Il implique le chargement sur l'ARNm d'un ou plusieurs ribosomes, en fonction de l'efficacité de l'initiation de la traduction et de la longueur du cadre de lecture ouvert à traduire (Isken et al., 2008; Isken and Maquat, 2008). Une fois vérifié, l'ARNm subit des modifications dans le cytoplasme. Le CBC nucléaire est remplacé par le facteur cytoplasmique eIF4E. De plus, la protéine nucléaire (PABPN1) est associée à la queue poly(A) des ARNm liés au CBC. Les protéines PABPN1 sont majoritairement remplacées par PABPC1 sur les ARNm portant eIF4E (Schoenberg and Maquat, 2012). Ce changement est favorisé par le remplacement du CBC (Sato and Maquat, 2009). Bien que les ARNm liés aux CBC soient des précurseurs des ARNm liés à eIF4E, les deux particules ribonucléoprotéiques (mRNP) diffèrent de façon significative (**Figure 5**). Les ARNm liés aux CBC sont associés à un ou plusieurs complexes exon-jonction (EJC). Après remplacement du CBC par eIF4E, les EJC ne sont plus détectables, parce qu'ils sont situés dans la région codante des ARNm et sont donc déplacés par les ribosomes lors du premier tour de traduction (Gehring et al., 2009; Sato and Maquat, 2009). En dépit de ces différences, si l'ARNm n'est pas engagé dans une voie de dégradation, la suite des événements de traduction est la même pour le premier et les autres cycles de traduction. Par conséquent, les ARNm liés au CBC et ceux

liés à eIF4E utilisent les mêmes facteurs d'initiation de traduction : eIF4G, eIF3, eIF4B, eIF4A et eIF2 (Isken and Maquat, 2008).

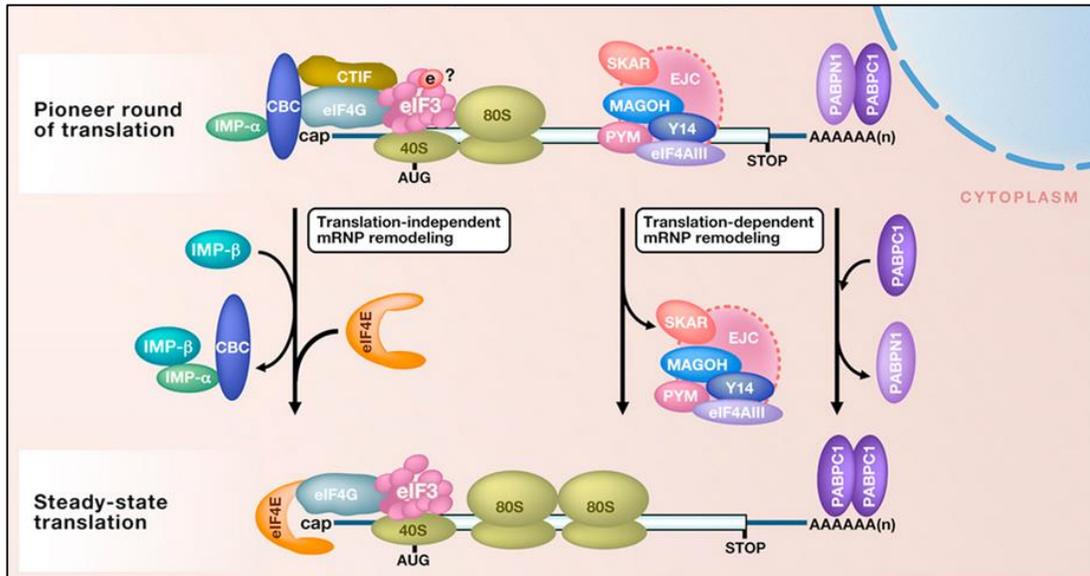


Figure 5 : Remodelage de l'ARNm au premier tour de traduction. Au premier tour de traduction le complexe CBC constitué des facteurs CBP80-CBP20 se lie directement à la coiffe de l'ARNm tandis qu'aux autres tours de traduction le facteur eIF4E est lié à la coiffe après remodelage de l'ARNm. Au premier tour de traduction l'ARNm porte au moins un complexe de jonction des exons (EJC), et PABPN1, ainsi que la sous unité eIF3-e qui ne fait pas partie du core du complexe eIF3, tandis qu'aux autres tours de traduction les EJC sont dégagés après passage du premier ribosome, PABPN1 sont remplacées par PABPC1 et le CBC est remplacé par eIF4E, cet échange est assuré par la protéine Importinβ (IMP-β) (Maquat et al., 2010).

2. L'élongation

L'élongation comprend l'ensemble des processus qui permettent à la machinerie traductionnelle la reconnaissance des codons, l'incorporation d'acides aminés et sa translocation sur l'ARNm. Chez les eucaryotes, l'élongation repose sur deux facteurs eEF pour *eukaryotic Elongation Factors* (Tableau II). Une fois le ribosome 80S formé et stabilisé sur le codon initiateur présent dans le site P, l'élongation de la traduction est déclenchée par incorporation d'ARNt chargés d'acides aminés sur le ribosome selon l'ordre des codons sur l'ARNm. Le facteur eIF1A assure l'apport de l'aminocyl-ARNt chargé correspondant au codon qui suit le codon initiateur au niveau du site A

du ribosome. Suite à l'appariement complet entre le codon de l'ARNm et l'anticodon de l'ARNt, le facteur eEF1A hydrolyse son GTP en GDP et se dissocie du ribosome. L'ARNt s'accommode dans le site A et provoque un changement de conformation du ribosome qui active le centre peptidyl-transférase de la sous-unité 60S. L'activité peptidyl-transférase du ribosome catalyse la formation de la liaison peptidique entre l'aminoacyl-ARNt présent dans le site A et la chaîne polypeptidique encore liée à l'ARNt présent dans le site P. Après formation de la liaison peptidique le ribosome avance de trois nucléotides du 5' vers le 3', cette translocation est favorisée par l'hydrolyse du GTP en GDP du facteur eEF2 (Kaul et al., 2011). De cette façon, le site A se retrouve libre et peut accueillir un autre aminoacyl-ARNt chargé complémentaire du codon suivant et l'ARNt désaminoacylé qui se trouve au niveau du site E est simultanément expulsé. Ainsi, le ribosome est prêt à un autre cycle d'élongation jusqu'à la rencontre d'un codon stop (Wilson and Nierhaus, 2006).

Tableau II : Les facteurs d'élongation de la traduction eucaryotes ainsi que leurs rôles.

Nom	Masse moléculaire (kDa)	Fonction
eEF1A	50,1	Apport le aminoacyl-ARNt, hydrolyse du GTP permettant l'accommodation de l'ARNt dans le site A du ribosome
eEF2	95,3	Translocation du ribosome favorisée par l'hydrolyse du GTP

3. La terminaison de la traduction et le recyclage des ribosomes chez les eucaryotes

Dès que le ribosome rencontre un codon stop, la terminaison de la traduction est déclenchée. La terminaison de la traduction est un évènement très important dans la synthèse protéique, elle permet non seulement la reconnaissance du codon stop et la libération du polypeptide synthétisé, mais aussi, la dissociation ainsi que le recyclage des sous-unités ribosomiques. On trouve trois codons stop chez les eucaryotes traditionnellement appelés ambre (UAG), ocre (UAA) ou opale (UGA). Le processus de terminaison de la traduction est principalement orchestré par deux facteurs eRF (*eucaryotic Release Factors*), mais deux autres protéines, ABCE1 (*ATP Binding Cassette E1*) et PABP ont un rôle important dans ce processus (Tableau III).

Tableau III : Les facteurs de terminaison et de recyclage eucaryotes.

Nom	Masse moléculaire (kDa)	Fonction
eRF1	49	Reconnaissance des codons stop. Activation du centre peptidyl-transférase du ribosome. Interagit avec eRF3. Recrute le facteur de recyclage ABCE1.
eRF3	60.5	GTP-ase qui potentialise l'activité du facteur de terminaison eRF1. Assure l'efficacité et la fidélité de la terminaison de la traduction.
PABP	70.7	Protège la queue poly(A) de l'ARNm des éventuels facteurs de désadénylation, interagit avec eRF3.
ABCE1	68	Recyclage des ribosomes.

3.1. Les principaux facteurs de terminaison et de recyclage

3.1.1. Le facteur eRF1

Le facteur de terminaison eRF1 est l'homologue eucaryote des facteurs procaryotes RF1 et RF2. Ce facteur est capable de reconnaître les trois codons stop (UAG, UAA et UGA) (Jackson et al., 2012). Au niveau structural, eRF1 est semblable à un ARNt, comprenant donc trois domaines fonctionnels (**Figure 6A**). Son domaine N-terminal est analogue au bras de l'anticodon de l'ARNt et il est capable d'entrer directement dans le site A du ribosome et d'interagir avec le codon stop (Dever and Green, 2012; Jackson et al., 2012). Des expériences de cross-linking d'eRF1 sur les codons stop (Bulygin et al., 2010; Chavatte et al., 2002) et des études de mutagenèse et d'analyse de séquences (Seit-Nebi et al., 2001) ont révélés que le domaine N-terminal d'eRF1 contient trois motifs NIKS, YxCxxF et une boucle GTS hautement conservés qui sont impliqués dans la reconnaissance des codons stop (**Figure 6B**). Il a été montré que la cavité proche du motif NIKS participe aussi à la reconnaissance du codon stop. Les résidus R65 et K109 sont importants pour le positionnement de l'ARNm dans cette cavité, tandis que les résidus S33 et S70 sont impliqués directement dans la reconnaissance du codon stop (Wada and Ito, 2014). Le domaine intermédiaire d'eRF1 est analogue au bras accepteur de l'ARNt. Ce domaine porte un motif GGQ qui est impliqué

dans l'activation du centre peptidyl-transférase du ribosome et permet la libération du polypeptide synthétisé (Seit-Nebi et al., 2001; Song et al., 2000). Enfin, le domaine C-terminal d'eRF1 est impliqué dans l'interaction avec ses partenaires, notamment, avec le facteur eRF3 (Cheng et al., 2009). Cette interaction favorise l'activité GTP-ase d'eRF3 (Kononenko et al., 2008). Le facteur eRF1 intervient aussi dans le recrutement du facteur du recyclage des ribosomes ABCE1.

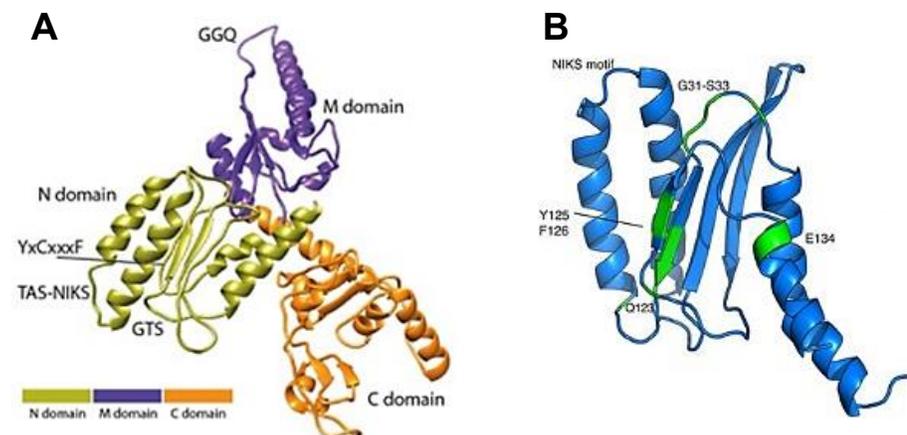


Figure 6 : Structure du facteur de terminaison eRF1. *A. Le facteur eRF1 comprend 3 domaines : en vert le domaine N-ter portant les motifs conservés NIKS, YxCxxxF ainsi que la boucle GTS qui sont impliqués dans la reconnaissance des codons stop, le domaine M en violet portant le motif GGQ et le domaine C-ter en orange. B. Structure du domaine N-ter du facteur eRF1 avec les résidus les plus touchés par les mutations (Matheisl et al., 2015; Palić et al., 2009).*

Ainsi, chez les eucaryotes, eRF1 a trois fonctions : il reconnaît les codons stop, stimule l'activité GTPase de eRF3 et active le centre peptidyl-transférase du ribosome qui catalyse l'hydrolyse du pont ester entre le peptide et l'ARNt du dernier peptidyl-ARNt.

3.1.2. Le facteur eRF3

Le facteur eRF3 est l'homologue eucaryote du facteur procaryote RF3 (Zhouravleva et al., 1995). Chez les mammifères, deux gènes nommés GSPT1 et GSPT2 pour *G to S transition factor* codent le facteur eRF3, ils donnent les protéines eRF3a et eRF3b, respectivement. Ces deux facteurs présentent 87% d'homologie de séquences et diffèrent par le domaine N-terminal. Notre équipe, a montré que seul le facteur eRF3a est essentiel à la terminaison de la traduction (Chauvin et al., 2005). Structuralement, eRF3 est constituée de trois domaines (**Figure 7**) et présente des homologies avec le facteur d'élongation eEF1A (Kong et al., 2004). Son domaine N-terminal est très variable selon les espèces. Chez les mammifères, ce domaine est riche en acides aminés proline, sérine et glycine et est très peu structuré.

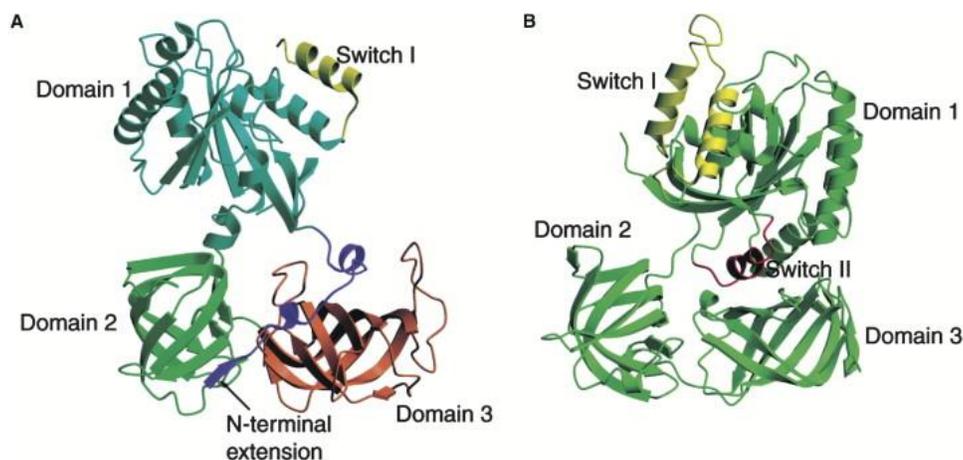


Figure 7 : Comparaison des structures du facteur de terminaison eRF3 et le facteur d'élongation eEF1A. *A. Structure du facteur eRF3 comprenant les trois domaines 1, 2 et 3; en violet l'extension N-terminale peu structurée, en vert domaines M et en rouge le domaine C-ter. B. Structure du facteur eEF1A avec ces domaines 1, 2, et 3 (Kong et al., 2004).*

Chez les mammifères, cette extrémité chez eRF3a (mais pas chez eRF3b) comprend une répétition stable de résidus glycine qui est codée par une répétition de codons GGC présente dans le premier exon du gène GSPT1. Une étude sur le polymorphisme de la répétition des trinuécléotides GGC sur une population portugaise a permis de détecter la présence de 5 allèles d'eRF3a : les allèles 7-GGC, 9-GGC, 10-GGC, 11-GGC et 12-GGC. Cette étude a permis de mettre en évidence

l'association entre la répétition de trinuécléotide polymorphe dans le facteur eRF3a et la susceptibilité au cancer (Brito et al., 2005). Une autre étude a révélé, que les personnes porteuses de l'allèle 12-GGC ont un risque de développer un cancer gastrique 20 fois plus élevé (Malta-Vacas et al., 2005). Les auteurs de cette étude suggèrent qu'une altération de la conformation d'eRF3a due à la présence des 12 répétitions du triplet GGC pourrait avoir un effet sur la fonction de la protéine et son rôle dans l'efficacité de la traduction. Le domaine N-terminal d'eRF3a porte en outre le site de fixation à la protéine de liaison des queues poly(A) des ARNm, PABP (Hoshino et al., 1999). En utilisant la résonance plasmonique de surface, mon équipe a étudié l'interaction entre PABP et eRF3, en présence et en absence de la queue poly(A) des ARNm. Ils ont ainsi montré que la répétition de glycine du domaine N-terminal d'eRF3 influence son affinité d'interaction avec PABP *in vitro* (Jerbi et al., 2016). Ils ont testé plusieurs formes alléliques d'eRF3a et ont montré que la forme impliquée dans le risque de cancer, celle codée par l'allèle 12-GGC, a une faible affinité pour PABP, comparée aux autres formes alléliques. Cette étude suggère aussi que la partie N-terminale d'eRF3 est cruciale pour la régulation de la desadénylation à travers son interaction avec PABP (Jerbi et al., 2016). De plus, il a été montré que le domaine C-terminal de PABP et son interaction avec le domaine N-terminal d'eRF3 jouent un rôle important dans la terminaison de la traduction. Par des expériences *in vitro*, Ivanov et ses collaborateurs (Ivanov et al., 2008) ont reconstitué un système de terminaison de traduction de mammifères et ont montré que PABP stimule la reconnaissance des codons stop de façon dépendante de son activité d'interaction avec l'ARN. La protéine PABP favorise le recrutement des facteurs de terminaison sur le ribosome augmentant ainsi l'efficacité de terminaison de la traduction. Cet effet de stimulation de la terminaison est due au positionnement optimal d'eRF3 sur le ribosome, qui favoriserait l'activité de reconnaissance des codon stop d'eRF1 (Ivanov et al., 2008).

Le domaine intermédiaire d'eRF3 est riche en acides aminés chargés, mais aucune fonction ne lui a été attribuée chez l'homme. Le seul rôle qui a été décrit à ce jour est celui du facteur de la levure chez qui il serait responsable de la modulation de la solubilité et le passage à l'état prion de la protéine (Liu et al., 2002).

La séquence en acides aminées du domaine C-terminal d'eRF3 est conservée chez tous les eucaryotes. Ce domaine contient les quatre motifs canoniques de liaison au GTP communs aux protéines de la superfamille des GTPases. Il a été montré grâce à des mutants des motifs de fixation

du GTP que l'hydrolyse du GTP par eRF3 facilite la reconnaissance des codons stop par eRF1 (Salas-Marco and Bedwell, 2004). Le domaine C-terminal d'eRF3 porte aussi les sites de liaison à différents partenaires, principalement eRF1 et le facteur Upf1 (*Up-Frameshift 1*).

3.1.3. Le facteur Upf1

Le facteur Upf1, appelé aussi RENT1 (*Regulator of nonsense transcripts 1*), est une ARN hélicase ATP dépendante appartenant à la superfamille des hélicases de type I. Ce facteur est relativement bien conservé chez les mammifères avec 48% d'identité de séquence avec la levure et 62% avec la drosophile (Culbertson and Leeds, 2003). Chez les mammifères, le facteur Upf1 interagit à la fois avec les facteurs de terminaison de la traduction eRF1 et eRF3 (Czaplinski et al., 1998), et forment le complexe dit de surveillance qui est impliqué dans la voie de dégradation NMD (*Nonsense-mediated mRNA decay*). Ce dernier permet la reconnaissance des ARNm contenant un codon stop prématuré et l'enclenchement du processus de dégradation. Ce processus sera plus approfondi dans le chapitre suivant. La région centrale d'Upf1, qui est la plus conservée, contient deux motifs en doigt de zinc riches en cystéine et sept motifs de la famille des hélicases de type I. L'activité hélicase est essentielle à la fonction d'Upf1. En effet, la mutation R844C, qui abolit son activité hélicase, inhibe la NMD dans les cellules de mammifères (Sun et al., 1998). Upf1 possède aussi une activité ATPase dépendante de l'ARN qui réside dans les motifs Ia et II du domaine hélicase (Bhattacharya et al., 2000). La mutation de deux résidus (DE636AA) dans le motif II non seulement supprime son activité ATPase mais aussi son activité hélicase (Bhattacharya et al., 2000). Plus récemment, il a été montré que l'activité ATPase d'Upf1 est nécessaire pour le relargage et le recyclage des facteurs du NMD qui sont fixés sur l'ARNm (Franks et al., 2010). En effet, chez les mutants DE636AA, les facteurs du NMD restent fixés sur des mRNP intermédiaires, partiellement dégradées en 3', qui s'accumulent dans les P-bodies (Franks et al., 2010). Les séquences N-terminale et C-terminale d'Upf1 contiennent de multiples résidus de serine qui peuvent être phosphorylés par la kinase SMG1 ; cette phosphorylation est une étape indispensable dans le mécanisme du NMD (Yamashita et al., 2001).

En plus de son rôle dans la voie NMD, Upf1 aurait un rôle dans l'efficacité de terminaison de la traduction et dans le recyclage des ribosomes aux codons stop prématurés. Chez la levure, grâce à des expériences étudiant la traduction par profilage de ribosome ou *ribosome profiling* dans

des cellules déplétées en Upf1p, Wu et ses collaborateurs ont constaté que la déplétion d'Upf1 conduit à une terminaison et à un recyclage défectueux au niveau des codons de terminaison normaux de plus de 1500 gènes (Wu et al., *Translational Control Meeting 2016, Cold Spring Harbor*). Une étude plus ancienne chez la levure a montré que la terminaison de la traduction était moins efficace en l'absence d'Upf1, que l'ARNm soit soumis à la NMD ou non (Keeling et al., 2004). Ainsi, chez la levure, Upf1 augmenterait l'efficacité de terminaison de la traduction indépendamment de la dégradation de l'ARNm. Contrairement à ce qui a été observé dans les cellules de mammifères où Upf1 aurait un rôle inhibiteur sur la terminaison de la traduction (Ivanov et al., 2008). Cette différence pourrait s'expliquer par le fait qu'il existe chez la levure deux homologues à Upf1 humain, Upf1p et Mtt1p (Czaplinski et al., 2000). C'est le facteur Mtt1p qui serait capable de fixer le domaine GTPase d'eRF3 et d'inhiber la terminaison de la traduction chez la levure. Upf1 humain aurait donc acquis au cours de l'évolution les caractéristiques de Mtt1p, ce qui conduirait à une terminaison de la traduction ou un recyclage du ribosome défectueux au niveau des codons de terminaison normaux.

3.1.4. La protéine ABCE1

La protéine ABCE1 (*ATP binding cassette E1*) est une protéine appartenant à la superfamille des *ATP binding cassette*. C'est une protéine essentielle et très conservée chez les eucaryotes et les archées (Pisarev et al., 2010). Structuralement, le domaine N-terminal d'ABCE1 contient deux domaines avec des groupements fer-soufre (Fe-S) suivis par deux domaines NBD pour *Nucleotide Binding Domain* (Becker et al., 2012; Karcher et al., 2008). Chez les eucaryotes, ABCE1 joue un rôle important dans différents processus cellulaires notamment la division cellulaire. Il a été montré que la suppression d'ABCE1 dans les cellules tumorales humaines HEK293 inhibe la prolifération cellulaire. Dans cette même étude, il a été montré que ABCE1 co-immunoprécipite avec les facteurs d'initiation de la traduction eIF2 et eIF5. Ces résultats suggèrent qu'ABCE1 serait impliqué dans l'assemblage du complexe d'initiation de la traduction (Chen et al., 2006). L'implication d'ABCE1 dans la terminaison de la traduction et le recyclage des ribosomes a été montrée par son interaction avec le facteur de terminaison de la traduction eRF1 (Barthelme et al., 2011). La déplétion du facteur ABCE1 induit une altération de la reconnaissance des codons stop (Khoshnevis et al., 2010). De plus, il a été montré que le facteur ABCE1 joue un rôle clé dans le

processus de recyclage des ribosomes (Pisarev et al., 2010) de façon ATP dépendante (Dever and Green, 2012; Jackson et al., 2012).

3.2. Le mécanisme de terminaison de la traduction

Chez les eucaryotes, la reconnaissance du codon stop est catalysée par les facteurs de terminaison eRF1 et eRF3 dans le site A du ribosome. Ces deux facteurs interagissent via leur domaine C-terminale et forment un complexe ternaire eRF1-eRF3-GTP dans lequel eRF3 est lié au GTP. Des expériences de déplétion en eRF3a dans différentes lignées de cellules humaines réalisées dans mon équipe, ont montrés qu'eRF1 est dégradé lorsqu'il n'est pas associé à eRF3 (Chauvin et al., 2005). De plus, en utilisant des formes d'eRF3a portant des mutations dans le site de liaison à eRF1, ils ont montrés que le facteur eRF3a était ubiquitinylé et dégradé par le protéasome lorsqu'il n'était pas associé à eRF1 (Chauvin and Jean-Jean, 2008). Ces résultats suggèrent que la formation du complexe de terminaison est contrôlée par des mécanismes de dégradation protéolytique qui permettent à la cellule d'adapter l'abondance de ces facteurs aux taux global de synthèse protéique qui surviennent lors du passage des cellules d'un état quiescent à un état prolifératif. Ces deux facteurs de terminaison fonctionnent de façon coopérative, la protéine eRF1 stabilise la fixation du GTP sur eRF3 et stimule son activité GTP-ase dépendante du ribosome (Mitkevich et al., 2006; Pisareva et al., 2006) et la protéine eRF3, potentialise l'activité de reconnaissance des codons stop (Salas-Marco and Bedwell, 2004) ainsi que l'activité de libération du polypeptide par eRF1 (Alkalaeva et al., 2006).

Ainsi, dès lors que le codon stop rentre dans le site A du ribosome, il recrute le complexe eRF1-eRF3-GTP pour former le complexe de pré-terminaison. Il a été montré que la liaison d'eRF3-GTP sur eRF1 entraîne un changement conformationnel des domaines intermédiaire et N-terminal d'eRF1 (Cheng *et al.*, 2009 ; Taylor *et al.*, 20012), ce changement de conformation entraîne une augmentation de l'affinité du complexe pour le ribosome se trouvant au niveau d'un codon stop (Fan-Minogue *et al.*, 2008). eRF1 reconnaît alors le codon stop et déclenche un changement de conformation du complexe de pré-terminaison, ce qui déclenche l'hydrolyse du GTP du facteur eRF3.

3.2.1. Reconnaissance des codons stop

La caractéristique la plus importante dans ce processus est, tout d'abord, la compaction de l'ARNm et l'accommodation de quatre nucléotides dans le site A du ribosome. Ceci dépend du résidu A1825 de l'ARNr 18S qui interagit avec le nucléotide +2 du codon stop sur lequel vient s'empiler le nucléotide +3. Cette conformation permet aussi l'interaction du nucléotide +4 avec le résidu G626 de l'ARNr 18S qui stabilise l'ARNm (Brown et al., 1990; des Georges et al., 2014). Cette compaction de l'ARNm est très importante dans la reconnaissance du codon stop par l'ARNm (Kryuchkova et al., 2013; Shirokikh et al., 2010). Ensuite, les résidus Asparagine 61 et Lysine 63 du domaine NIKS interagissent avec l'Uracile (nucléotide +1) du codon stop. Il a été montré que l'hydroxylation de la Lysine 63 diminue la translecture et favorise la libération du polypeptide (Feng et al., 2014). L'interaction entre le résidu Gln 55 et le motif YxCxxxF avec les nucléotides +2 et +3 permet de discriminer un codon stop d'un codon sens. La chaîne principale de la Cystéine 127 du motif YxCxxxF forme deux liaisons hydrogènes avec le résidu A1825, ce qui le stabilise facilitant l'empilement des deux bases +2 et +3 du codon stop. Ensuite la Glutamine 55 ainsi que la Tyrosine 125 du motif YxCxxxF agissent ensemble pour positionner la chaîne latérale du Glutamate permettant son interaction avec le résidu Adénosine de la position +2 ou +3 des codons stop. Ces interactions ne sont possibles qu'avec des résidus purines des bases +2 ou +3 des codons stop, c'est ce qui permet donc la discrimination des autres codons sens (Brown et al., 2016). Le motif GTS situé près de la base +3 du codon stop, peut adopter deux conformations différentes qui sont interdépendantes du motif YxCxxxF. Dans le cas du codon stop UAG, la Thréonine 32 du motif GTS interagit avec la base G de la position +3 du codon stop (Wong et al., 2012). Cependant dans le cas du codon stop UGA, le motif YxCxxxF subit un petit mouvement pour permettre l'accommodation de la Guanosine en position +2 lui permettant de se mettre en face et d'interagir avec la Thréonine 32 du motif GTS (**Figure 8**).

Il a été montré chez la levure que les mutations des résidus Sérine 33 en Alanine du domaine GTS provoquent une diminution de l'efficacité de terminaison spécifiquement sur le codon UGA *in vivo*. De plus, les résidus Sérine 33 et 70 interagissent ensembles pour permettre la reconnaissance du codon stop UGA et cette interaction implique une conformation particulière modulé par le motif YxCxxxF. Enfin, il a été montré, dans cette même étude, que le phénotype des souches mutantes Arginine65 et Lysine109 en Alanine, présentent une augmentation de la

translecture au niveau des trois codons stop (Blanchet et al., 2015). La combinaison de toutes ces observations permet de dire que tous ces résidus du domaine N-terminal d'eRF1 sont importants dans la reconnaissance du codon stop et la moindre modification de ces résidus perturbe complètement le processus de terminaison (Cheng et al., 2009; Kolosov et al., 2005, 2005; Seit-Nebi, 2002).

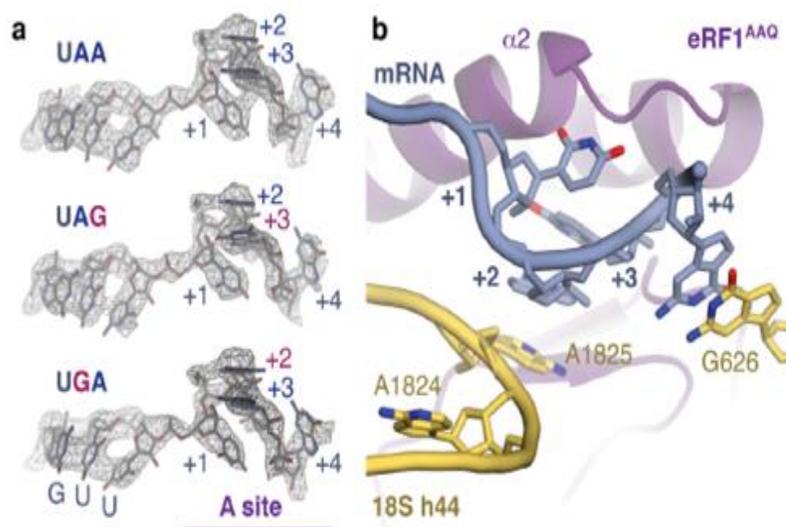


Figure 8 : Reconnaissance des codons stop. *a.* Cartes des densités de microscopie électronique de l'ARNm pour chacun des trois complexes de terminaison obtenus pour les trois codons stop. La conformation compacte des positions +2 et +3 est retrouvée dans les trois conditions. Le codon Valine (GUU) au site P et le codon stop (+1 à +3) et la base +4 au site A du ribosome sont indiqués. *b.* En violet eRF1 (AAQ) reconnaît quatre bases de l'ARNm de +1 à +4, dans le site A du ribosome. Les bases +2 et +3 sont empilés et interagissent avec A1825 de l'ARNr. La base en +4 interagit directement avec le résidu G626 de l'ARNr 18S (Brown et al., 2016).

3.2.2. Translecture des codons stop

Il est important de préciser que tous les codons stop ne déclenchent pas systématiquement la terminaison de la traduction. On peut avoir dans certains cas spécifiques et dans certaines conditions, une translecture du ribosome à travers certains codons stop. La traduction continue alors jusqu'à un codon stop situé en aval. La terminaison de la traduction peut être considérée comme un phénomène de compétition entre le complexe de terminaison eRF1-eRF3-GTP et un ARNt pouvant reconnaître de façon imparfaite (ARNt dit : "near-cognate") le codon stop (Druegeon et al., 1997; Le Goff et al., 1997). Cette compétition est différente d'un organisme à un autre, et est

influencée par l'identité du codon stop. Dans ce sens, plusieurs paramètres peuvent influencer l'efficacité de terminaison de la traduction notamment, le contexte nucléotidique qui entoure le codon stop (Bonetti et al., 1995; Namy et al., 2001), la nature des deux derniers acides aminés incorporés dans la chaîne polypeptidique naissante (Janzen et al., 2002), la présence d'éléments stimulateurs en aval du codon stop (Bertram *et al*, 2001) et le codon au site P de l'ARNm (Mottagui-Tabar et al., 1998). Ces éléments peuvent favoriser l'incorporation d'un aminoacyl-ARNt *near-cognate* au lieu du complexe de terminaison de la traduction, conduisant à la translecture du codon stop. Cette translecture conduit à une protéine aberrante avec une plus longue région C-terminal et parfois des fonctions biologiques différentes de la protéine originale. Récemment, des expériences de *ribosome profiling* ont permis de détecter plusieurs évènements de translecture chez la levure (Baudin-Baillieu et al., 2014), la drosophile ainsi que chez l'humain, ce qui suggère que c'est un mécanisme conservé chez les espèces (Dunn et al., 2013). Une étude menée chez la levure a montré qu'en plus des facteurs eRF1 et eRF3, la terminaison de la traduction est aussi régulée par le facteur d'initiation eIF3 (Beznosková et al., 2013). De plus, une autre étude par le même laboratoire montre que ce facteur est crucial dans les processus de translecture des codons stop (Beznosková et al., 2013). Ils proposent que le facteur eIF3 se lie au complexe de pré-terminaison sur le codon stop et interfère avec la reconnaissance de la 3^{ème} base du codon stop, lorsque ce dernier se trouve dans un contexte de terminaison défavorable. Ainsi, eIF3 favorise la translecture en permettant un mésappariement entre l'anti-codon de l'ARNt *near-cognate* avec la 3^{ème} base du codon stop. L'ARNt *near-cognate* décode le codon stop comme un codon sens et permet de poursuivre l'élongation. Cette même étude a permis de montrer qu'eIF3 agit spécifiquement sur des gènes qui contiennent des éléments stimulateurs de la translecture en aval du codon stop. Cette fonction d'eIF3 semble conservée au cours de l'évolution (Beznosková et al., 2013).

3.2.3. Libération de la protéine néo-synthétisée

Concomitamment à la reconnaissance du codon stop, eRF1 favorise l'hydrolyse du GTP du facteur eRF3. L'hydrolyse du GTP agit sur la conformation d'eRF1, qui subit un second changement conformationnel lui permettant de s'accommoder dans le site A du ribosome (Salas-Marco and

Bedwell, 2004). Cette accommodation lui permet de placer son motif catalytique GGQ dans le site P du ribosome près du centre peptidyl-transférase du ribosome (**Figure 9**).

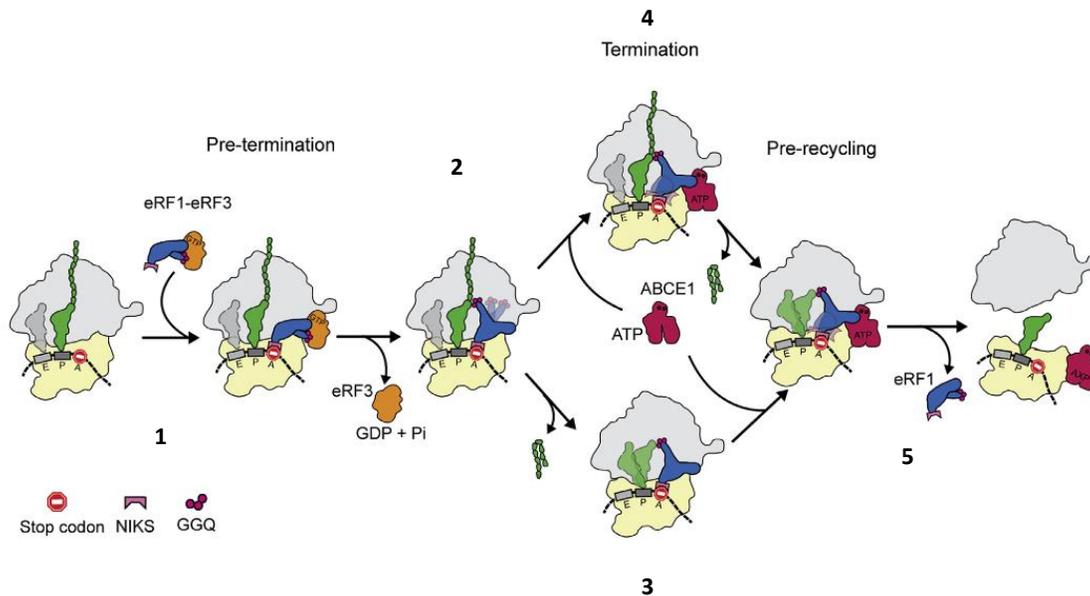


Figure 9 : Terminaison de la traduction et recyclage des ribosomes chez les eucaryotes. La terminaison de la traduction et le recyclage se décomposent en plusieurs étapes (1) la première étape est l'entrée du codon stop dans le site A du ribosome, par la suite il y a recrutement du complexe ternaire de terminaison de la traduction, et hydrolyse du GTP du facteur eRF3. (2) l'hydrolyse du GTP d'eRF3 favorise l'ancrage d'eRF1 dans le site A du ribosome, et le positionnement du motif GGQ d'eRF1 au niveau du centre peptidyl-transférase du ribosome. (3) cette interaction induit l'activation du centre peptidyl-transférase du ribosome et le relargage du polypeptide néosynthétisé, il en résulte un complexe constitué du 80S sur l'ARNm chargé d'un ARNt désaminoacylé dans le site P du ribosome ou entre les sites P/E. (4) parallèlement, cette conformation active d'eRF1 du complexe de pré-terminaison est stabilisée par son interaction avec ABCE1, ce qui stimule la libération du polypeptide néosynthétisé. Tandis que le domaine N-terminal d'eRF1 est délocalisé ce qui permet de découpler la reconnaissance de la libération du polypeptide. (5) indépendamment du processus de terminaison de la traduction, ABCE1 avec eRF1 fonctionnent ensemble pour dissocier la grande sous-unité ribosomique 60S, de la 40S encore associée à l'ARNm et l'ARNt désaminoacylé (Preis et al., 2014).

Après cette étape, le complexe de pré-terminaison passe à un état de complexe de terminaison ou de pré-recyclage. Cette conformation permet donc une attaque nucléophile d'une molécule d'eau, il en résulte une hydrolyse de la liaison entre l'ARNt et le polypeptide néosynthétisé et le relargage de la chaîne polypeptidique (Jin et al., 2010).

3.2.4. Dissociation du complexe de terminaison et recyclage des ribosomes

Dès lors que le polypeptide néosynthétisé est libéré, le processus de recyclage démarre. Au début de cette étape, le ribosome est encore fixé sur l'ARNm avec l'ARNt dans le site P et le facteur eRF1 dans le site A (**Figure 9**). Cette étape permet la libération des sous-unités ribosomiques ainsi que les facteurs de terminaison qui seront utilisés aux prochains tours de traduction.

Chez les procaryotes, le facteur de terminaison RF3 est impliqué dans la dissociation des facteurs de terminaison (Freistroffer et al., 1997) et dans le recyclage par l'intermédiaire de la protéine RRF pour *Ribosome Recycling Factor* (Janosi et al., 1996). Chez les eucaryotes, la dissociation et le recyclage sont assurés par un seul facteur ABCE1 (Barthelme et al., 2011). Une étude récente a montré que suite à la libération d'eRF3 du ribosome, le facteur eRF1 reste associé au ribosome et permet le recrutement d'ABCE1. Il a été montré que l'interaction ABCE1-eRF1 est requise pour la dissociation efficace des sous-unités ribosomiques (Pisarev et al., 2007). Le mécanisme par lequel ABCE1 agit dans la dissociation et le recyclage n'est pas encore bien élucidé. La protéine ABCE1 pourrait intervenir soit avant soit après le relargage du polypeptide (**Figure 9**). Après la libération du polypeptide, la protéine ABCE1 vient alors dissocier la sous-unité 60S qui est relarguée alors que la sous-unité ribosomique 40S et l'ARNt au site P restent fixés sur l'ARNm. ABCE1 fonctionne avec le facteur eIF6 pour dissocier les sous-unités ribosomiques (Pisarev et al., 2010). Le facteur d'initiation eIF6 se fixe sur la sous-unité-ribosomique 60S au niveau de son site de fixation à la 40S, et ainsi permet de maintenir les deux sous-unités ribosomiques dissociées (Miluzio et al., 2009; Sanvito et al., 1999). Des études récentes ont montré qu'ABCE1 interagit avec le ribosome via son domaine Fe-S (Barthelme et al., 2011; Becker et al., 2012; Preis et al., 2014; Shoemaker and Green, 2011). L'hydrolyse de l'ATP d'ABCE1 permet un changement de conformation du ribosome qui entraîne une déstabilisation de l'ARNt-désaminoacylé du site P (**Figure 10**), et la dissociation de la sous-unité 60S tandis que la sous-unité 40S et l'ARNt restent associés à l'ARNm.

Ce sont les facteurs d'initiation eIF3, par sa sous-unité eIF3j, eIF1 et eIF1A qui sont responsables de dissocier l'ARNt de la sous-unité 40S. La libération de l'ARNt est suivie par la dissociation de l'ARNm de la sous-unité 40S (Pisarev et al., 2010, 2007).

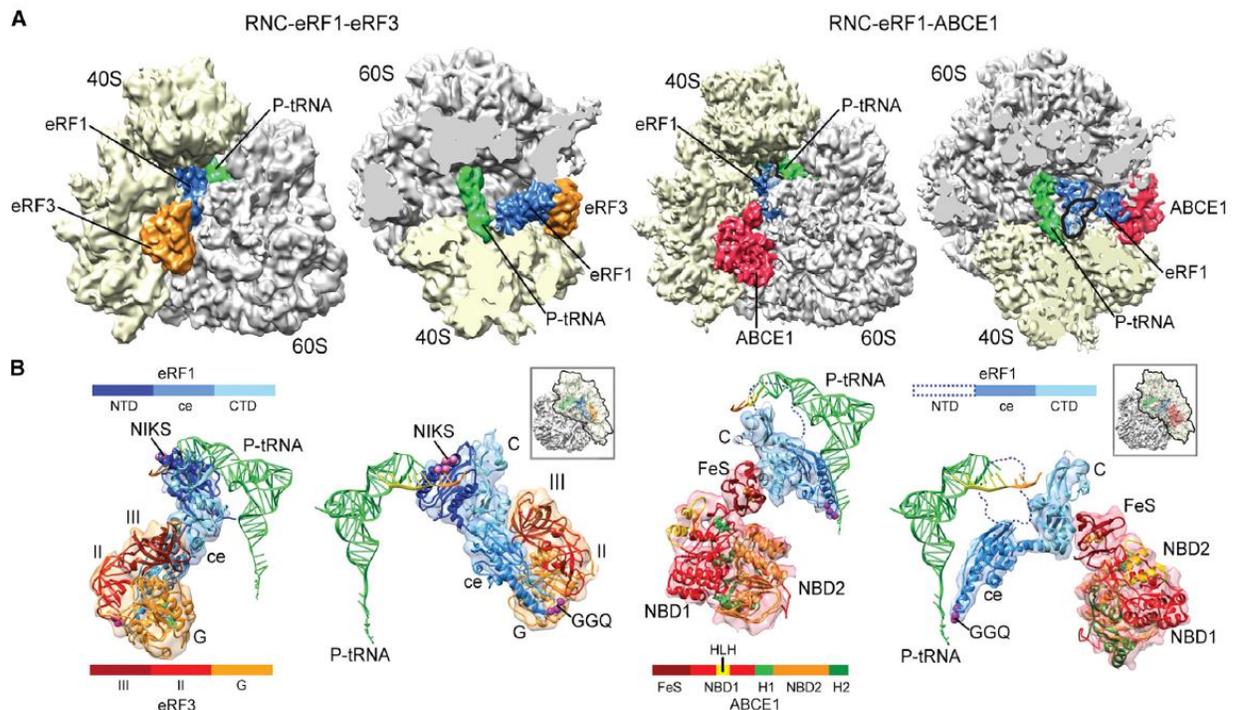


Figure 10 : Structure cryo-EM du complexe de terminaison eRF1-eRF3 et du complexe de recyclage eRF1-ABCE1 avec le peptidyl-ARNt. *A.* à gauche le ribosome 80S associé au complexe de terminaison de la traduction. A droite le ribosome associé au pré-complexe de terminaison/recyclage. *B.* Modèles moléculaires eucaryotes du peptidyl ARNt, eRF1, eRF3c et ABCE1 sur le ribosome. A gauche, le complexe de terminaison comprenant l'ARNt (vert), eRF1 (bleu) et eRF3c (orange-rouge). Les trois domaines de eRF1 et le domaine C-terminal de eRF3, constitué des domaines G, II et III. A droite, le complexe de pré- terminaison/recyclage est constitué de l'ARNt (vert), eRF1 (bleu) et ABCE1 (orange-rouge). Seuls les domaines central et C-terminal de eRF1 sont visibles. Les domaines fer-souffre (FeS), domaines de fixations à l'ARN (NBD1 et NBD2) et les charnières (H1 et H2) sont visibles. Le motif NIKS de eRF1 est positionné à proximité du codon stop (boules roses). Le domaine central de eRF1 contenant la boucle GGQ (boulles violettes) est situé contre eRF3 dans le complexe de terminaison et basculé vers le PTC dans le complexe de terminaison/recyclage (Preis et al., 2014).

Il a été suggéré qu'en plus d'ABCE1, la conformation circulaire de l'ARNm, favorisée par l'interaction entre PABP et eIFG, et la proximité entre les régions 5' et 3'UTR favorisent le recyclage des ribosomes (Jackson et al., 2010; Valásek, 2012). De plus, des études menées chez les levures et sur des cellules de mammifères ont montré qu'une fois la grande sous-unité

ribosomique dissociée, ABCE1 peut recruter le complexe multi-factoriel de traduction au niveau de la 40S pour permettre la réinitiation de la traduction (Dever and Green, 2012).

Ainsi, après terminaison de la traduction, deux processus sont possibles : soit la sous-unité 40S et l'ARNt sont relargués et recyclés pour un nouveau cycle de traduction, soit la sous-unité 40S reste associée à l'ARNm en attendant la reformation du complexe d'initiation pour réinitier la traduction au niveau d'un codon d'initiation situé en aval sur le même ARNm : c'est le cas des ARNm portant des petits cadres de lecture (Rajkowitsch et al., 2004).

4. La réinitiation de la traduction chez les eucaryotes

La réinitiation de la traduction est un mécanisme traductionnel qui nécessite la présence d'uORF en amont de la région codante de l'ARNm. La réinitiation après une courte séquence codante a été observée dès la découverte du modèle de balayage de l'ARNm par le ribosome (Kozak, 1987a). Dans cette étude, ils ont montré que l'insertion d'un codon stop quelques nucléotides en aval d'un codon d'initiation induit la réinitiation de traduction au niveau d'un codon d'initiation situé en aval. Environ la moitié des gènes codants des protéines portent des uORF dans la région 5'UTR de l'ARNm (Iacono et al., 2005; Timothy G. Johnstone et al., 2016). La présence de ces uORFs est corrélée avec une réduction de 30% à 80% de l'expression des protéines codées par l'ORF principale (Calvo et al., 2009) et ce en affectant l'efficacité de traduction de la région codante en aval. La présence d'une ou plusieurs uORF en amont de la phase codante de l'ARNm présente une situation spéciale pour le complexe d'initiation qui scanne le 5'UTR de l'ARNm à la recherche du codon initiateur. Lorsque le complexe d'initiation 48S qui scanne le 5'UTR de l'ARNm reconnaît, en amont du codon d'initiation du cadre de lecture principal, un ou plusieurs codons AUG dans un contexte favorable à l'initiation de la traduction, il initie la traduction à leur niveau. Ces codons AUG sont souvent suivis de près par un codon de terminaison et on observe alors la synthèse d'un court polypeptide. Après terminaison de la traduction de cette uORF, certaines sous-unités 40S du ribosome restent associées à l'ARNm, continuent à le scanner à la recherche d'un autre codon initiateur qui peut être ou pas celui du cadre de lecture principal de la protéine. La sous-unité 40S qui a reformé un complexe d'initiation peut donc réinitier la traduction au niveau du prochain codon initiateur se trouvant dans un contexte favorable. Récemment, il a été montré, qu'après terminaison de la traduction d'une petite uORF, l'association du facteur eIF3 peut persister sur la sous-unité

40S, le complexe d'initiation multi-factoriel (complexe ternaire, eIF1, eIF1A et eIF5) rejoint ce complexe et la réinitiation se produit (Asano, 2014). Cependant, l'efficacité de réinitiation dépend de plusieurs paramètres, notamment la taille de l'uORF, la distance qui sépare son codon stop du codon initiateur de l'ORF (Valásek, 2012), la séquence peptidique codée par l'uORF (Rahmani et al., 2009; Roy and von Arnim, 2013; von Arnim et al., 2014) et bien d'autres paramètres qui seront détaillés plus loin.

Récemment, Gunišová et ses collaborateurs ont soumis toutes les uORF de GCN4 à une analyse fine afin d'identifier toutes les séquences cis-régulatrices entourant ces uORF, qui contribuent de manière autonome ou en synergie à l'efficacité globale de réinitiation sur GCN4. Ils ont pu déterminer que, en plus des éléments présents en 5' de l'uORF1 et l'uORF2, les 12 premiers nucléotides des séquences 3' des uORF 1 et 3 contiennent un motif conservé AU1-2A/UUAU2 qui favorise la réinitiation de façon autonome. Ils ont également identifié des séquences inhibitrices de la réinitiation dans les régions 3', immédiatement après les séquences AU riche des uORF2 et uORF3. De plus, dans cette même étude ils ont souligné la contribution du domaine N-terminal d'eIF3 dans la réinitiation et ont pu constater qu'il y a une corrélation négative entre l'efficacité de réinitiation et l'efficacité de terminaison de la traduction (Gunišová et al., 2016). Les uORF et leurs mécanismes de régulation de la traduction seront plus approfondis dans le prochain chapitre.

III. Dégradation et surveillance des ARNm eucaryotes

1. Mécanisme général de dégradation cytoplasmique des ARNm

La dégradation des ARNm a longtemps été considérée comme une simple étape d'élimination des ARNm. Au cours des dernières années, cette vision a été totalement remise en cause. Les temps de demi-vie des ARNm, révélateurs de leur vitesse de dégradation, peuvent varier énormément. Ainsi chez les mammifères, ils varient de 15 minutes (ARNm *c-fos*) à 24 heures (ARNm de la globine), montrant que la dégradation des ARN est un processus spécifique et non un processus par défaut.

La dégradation des ARNm est essentielle. En l'absence d'une dégradation rapide des ARNm, une régulation rapide de la production de protéines par la seule induction transcriptionnelle serait

impossible (Camier and Séraphin, 2007). Les différentes voies de dégradations des ARNm "normaux" (par opposition à aberrants dont nous verrons le processus particulier de dégradation plus loin) sont décrites dans le schéma ci-dessous (**Figure 11**).

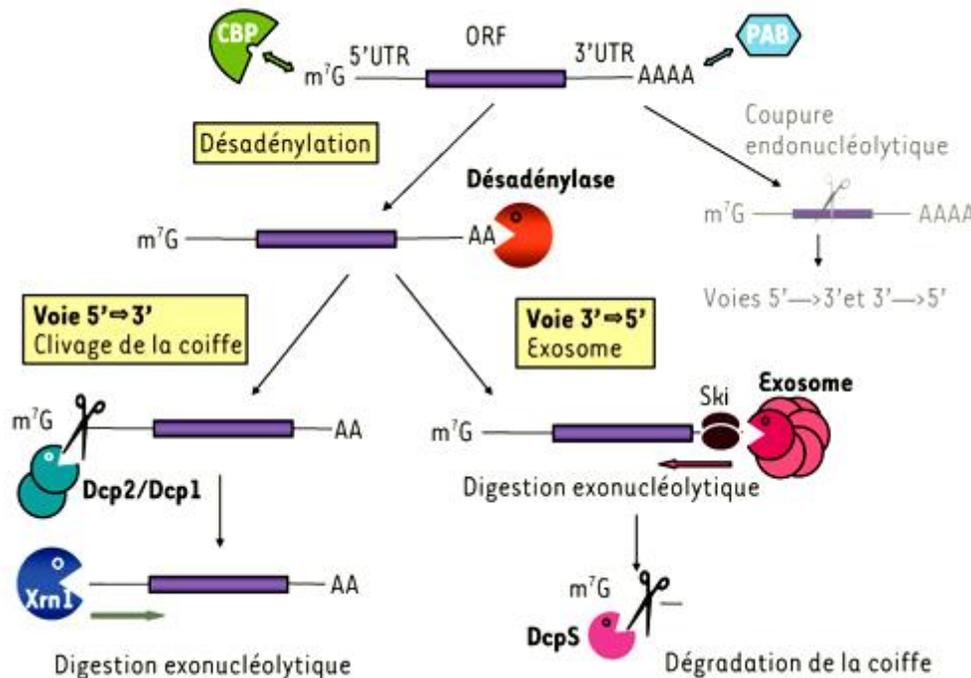


Figure 11 : Les principales voies cytoplasmiques de dégradation des ARNm normaux. L'événement initial le plus fréquent est une désadénylation. Le corps de l'ARNm est ensuite dégradé par une des 2 voies : 5'-3' ou 3'-5'. Dans le cas de la voie 5-3', la désadénylation est suivie du clivage de la coiffe par l'enzyme Dcp2/Dcp1 puis d'une digestion exonucléolytique par Xrn1. Dans le cas de la voie 3'-5', la désadénylation est suivie du recrutement du complexe Ski et d'une digestion exonucléolytique par l'exosome ; la coiffe est finalement dégradée par l'enzyme DcpS. En grisé est représenté une voie plus rare, amorcée par une coupure endonucléolytique de l'ARNm, suivie des voies classiques (Camier and Séraphin, 2007).

Que de précautions pour manipuler les ARNm au laboratoire afin qu'ils ne soient pas immédiatement dégradés par les ribonucléases ! Et pourtant, dans les cellules les ARN côtoient de nombreuses nucléases. Dans la cellule, plusieurs facteurs contribuent à la protection des ARNm contre l'action des ribonucléases. Lors de sa synthèse, l'ARNm subit des modifications à ses extrémités 5' et 3' : ajouts d'une coiffe et d'une queue poly(A). Ces marques, liées par des protéines spécifiques, les *cap-binding protein* (CBP) et *Poly(A)-binding protein* (PABP), le protègent de

l'action des exoribonucléases (Shatkin and Manley, 2000). L'ARNm est également associé à d'autres protéines au sein de larges particules ribonucléoprotéiques (mRNP) qui notamment le protègent des coupures endonucléolytiques non spécifiques. La dégradation d'un ARNm résulte d'un profond remaniement de la mRNP, déclenché par des signaux spécifiques conduisant au recrutement des enzymes et activateurs de la dégradation.

1.1. La désadénylation

La dégradation cytoplasmique des ARNm chez les eucaryotes débute, le plus souvent, par un raccourcissement de la queue poly(A), on parle de désadénylation. Cette étape détermine en grande partie la vitesse de dégradation de l'ARNm. Deux complexes protéiques cytoplasmiques (très conservés chez les eucaryotes) possèdent une activité de désadénylation :

- le complexe Pan2-Pan3, activé par la PABP associée à la queue poly(A). Il semble intervenir lors d'une étape initiale de désadénylation rapide (Brown and Sachs, 1998; Yamashita et al., 2005).
- le complexe Ccr4-Caf1-Not, majoritaire chez les mammifères, la drosophile et la levure, inhibé par la PABP. Il contient 2 exonucléases, Ccr4 et Caf1 (Bianchin et al., 2005; Daugeron et al., 2001; Finoux and Séraphin, 2006; Tucker et al., 2002).

L'étape de désadénylation n'inactive pas définitivement l'ARN, celui-ci peut éventuellement être polyadénylé à nouveau ou, dans certains cas, stabilisé par des protéines spécifiques (chez les métazoaires). Cependant le plus souvent, lorsqu'un certain nombre de résidus A restant sur l'ARNm est atteint, de l'ordre de 10 chez la levure ou 30-60 chez les mammifères, l'ARNm s'engage dans une dégradation rapide et irréversible.

Deux voies sont possibles pour la seconde étape : la voie 5'-3' ou la voie 3'-5', selon le sens de dégradation.

1.2. Les voies exonucléasiques 5'-3' et 3'-5'

Dans la voie 5'-3', après désadénylation, la mRNP subit un remodelage conduisant au recrutement et à l'activation de l'enzyme de clivage de la coiffe en 5' formée par le complexe Dcp1/Dcp2

(*decapping proteins 1et 2*), c'est une étape irréversible (Cougot et al., 2004; Simon et al., 2006). L'ARN clivé est ensuite dégradé de façon processive par l'exonucléase Xrn1 dans le sens 5'-3'.

Dans la voie 3'-5', la désadénylation est suivie du recrutement d'un complexe exonucléolytique appelé exosome, *via* l'association des protéines Ski. L'exosome est un complexe d'une dizaine de sous-unités catalysant une digestion de 3'-5'. Il possède un cœur de neuf exoribonucléases 3'-5' associées à d'autres protéines nécessaires à son activité nucléaire et cytoplasmique (Büttner et al., 2006; Dziembowski et al., 2007; Mitchell et al., 1997). La voie 3'-5' se termine avec l'hydrolyse de la coiffe par l'enzyme DcpS. Cette même enzyme hydrolyse également la coiffe libérée lors de la dégradation 5'-3' (Liu et al., 2002; van Dijk et al., 2003).

1.3. Coupures endonucléolytiques

Dans certains cas, la dégradation peut être amorcée par une coupure endonucléolytique, qui donne naissance à 2 fragments d'ARN : un fragment 5' coiffé et dépourvu de queue poly(A), pris en charge par l'exosome, et un fragment 3' sans coiffe pris en charge par l'exonucléase Xrn1. Les endonucléases impliquées restent cependant mal connues (pour revue Camier and Séraphin, 2007).

1.4. Sites cytoplasmique de la dégradation des ARNm

De nombreuses études ont mis en évidence l'existence de sites (ou *foci*) d'accumulation de protéines impliquées dans le métabolisme des ARNm dans le cytoplasme (Cougot et al., 2004; Eulalio et al., 2007; Parker and Sheth, 2007; van Dijk et al., 2003). Ces sites (appelés *P-bodies* pour « *processing bodies* », ou *Dcp-bodies* pour « *decapping bodies* ») contiennent en particulier des acteurs de la voie de dégradation 5'-3' (Dcp2, Dcp1, Xrn1,...). Les enzymes et activateurs de la dégradation ne sont pas uniquement localisés dans ces *foci*, on les trouve également diffus dans le cytoplasme. Certaines enzymes de dégradation, comme l'exosome ou les désadénylases, semblent être localisées de manière diffuse dans le cytoplasme et absentes des *foci* (Parker and Sheth, 2007). La proportion de dégradation totale s'effectuant dans les *foci* cytoplasmiques et la localisation intracellulaire des autres événements de dégradation restent à déterminer, ce qui pose la question de l'organisation spatiale de la dégradation des ARNm dans les cellules.

2. La surveillance des ARNm et la dégradation des ARNm aberrants

Comme nous l'avons vu plus haut, des mutations non-sens peuvent être présentes sur l'ADN suite à des erreurs lors des processus de réplication, ou suite à des stress génotoxiques, par exposition aux rayonnements UV par exemple. Ces mutations seront présentes à la fois sur l'ADN ainsi que sur l'ARNm. D'autres erreurs peuvent toucher uniquement l'ARNm, celles-ci surviennent lors de la transcription ou de l'épissage des introns et peuvent engendrer des mutations non-sens. La surveillance des ARN messagers regroupe l'ensemble des mécanismes qui visent à contrôler la qualité des ARN en éliminant les ARNm contenant des codons stop prématurés PTC pour *Premature Termination Codon* (Isken and Maquat, 2007). Des protéines tronquées qui pourraient avoir des effets délétères pour l'organisme, résultent de la traduction de ces transcrits aberrants. La dégradation de ces ARNm aberrants est donc indispensable aux cellules et implique des voies de dégradation qui se déroulent dans des compartiments spécialisés tels que les *P-bodies*.

L'activation des différents mécanismes de surveillance des ARN permet de discriminer les transcrits normaux de ceux contenant des PTC et d'empêcher leur traduction en déclenchant leur dégradation. La surveillance des ARNm peut agir directement sur les pré-ARNm par un processus d'inhibition d'épissage. Durant l'épissage, un phénomène appelé NAS pour *Nonsense-associated altered splicing* permet d'éliminer par épissage alternatif les exons contenant des PTC. Les transcrits matures eux sont ciblés par un mécanisme de dégradation des ARNm non-sens qui est la voie NMD (Isken and Maquat, 2007). Ces processus sont donc complémentaires et peuvent coexister au sein d'une même cellule pour contrôler efficacement la qualité des ARNm. Dans ce chapitre, je ne détaillerais que le mécanisme de la voie NMD.

Tableau IV : Les facteurs de la voie de surveillance et de dégradation des ARNm aberrants eucaryotes et leur rôle (Fatscher et al., 2015).

Nom	Masse moléculaire (kDa)	Fonction
Upf1	123	Protéine de liaison à l'ARN, possédant une activité hélicase. La phosphorylation de ces domaines N- et C-ter permettent le recrutement et la fixation des protéine SMG5, SMG7, PNRC2 et des facteurs de décoiffage Dcp.
Upf2	148	Régule l'activité hélicase d'Upf1, stimule l'activité kinase de SMG1, établie le lien entre les protéines Upf1 et Upf3.

Upf3a	55	Etablie le lien entre les protéines, Upf1 et Upf2 avec l'EJC (<i>Exon junction complex</i>).
Upf3b	56	Etablie le lien entre les protéines Upf1 et Upf2 avec l'EJC, dominant fonctionnel sur Upf3a.
SMG1	410	Phosphoryle les domaines N- et C-ter du facteur Upf1 sur plusieurs résidus SQ et TQ.
SMG5	114	Forme un complexe avec SMG7, recrute la protéine PP2A responsable de la déphosphorylation d'Upf1. Possède une plus grande affinité pour le facteur Upf1 phosphorylé.
SMG6	160	Se lie à la protéine Upf1 phosphorylée ou déphosphorylée, assure le clivage endonucléolytic des ARNm cible.
SMG7	122	Forme un complexe avec la protéine SMG5, nécessaire à la liaison des protéines SMG5 et SMG7 avec la protéine Upf1 phosphorylée, recrute la protéine POP2 pour la désadénylation de l'ARNm cible.
SMG8	110	Régule l'activité kinase la protéine SMG1, induit son changement de conformation ce qui l'inactive.
SMG9	58	Régule l'activité kinase de SMG1, nécessaire à la formation du complexe avec SMG1.

2.1. Les acteurs majeurs de la voie NMD

2.1.1. Les protéines Upf

Il existe quatre membres de la famille Upf chez les eucaryotes : UPF1, UPF2, UPF3a et UPF3b. Il a été montré que la déplétion de ces facteurs inhibe la voie NMD (Gehring et al., 2005, 2003, Mendell et al., 2004, 2002; Wittmann et al., 2006), et que la liaison artificielle des protéines Upf sur la région 3'UTR d'un ARNm dépourvu de PTC déclenche la NMD et conduit à leur dégradation (Gehring et al., 2005, 2003; Kunz et al., 2006; Lykke-Andersen et al., 2000).

Upf1 est le facteur le plus conservé. Il présente 98% d'homologie de séquence entre les protéines murine et humaine (Sun et al., 1998), cette dernière possède 48% d'identité avec la levure et 62% avec la drosophile (Culbertson and Leeds, 2003). Upf1 a une activité ATPase ARN/ADN dépendante et une activité hélicase à ARN 5'-3'. Son activité ATPase est nécessaire pour le relargage et le recyclage des facteurs de la NMD fixés sur l'ARNm (Franks et al., 2010). La

fonction hélicase est essentielle, la mutation R844C dans le domaine hélicase a un effet dominant négatif qui inhibe la NMD (Sun et al., 1998). Le domaine N-terminal d'Upf1 est riche en résidus proline/glycine et permet une interaction avec SMG5 (*Suppressor with Morphogenetic effect on Genitalia 5*), qui est nécessaire à la déphosphorylation d'Upf1 (Ohnishi et al., 2003). Les cycles de phosphorylation / déphosphorylation d'Upf1 sont indispensables au mécanisme de la NMD (Yamashita et al., 2001). Son domaine C-terminal contient de nombreux sites potentiels de phosphorylation. La protéine Upf1 est localisée principalement dans le cytoplasme, mais il a été montré qu'elle est également capable d'entrer dans le noyau (Mendell et al., 2002). L'inactivation du gène *upf1* est létal chez la souris, les embryons meurent entre 3,5 et 5,5 jours (Medghalchi et al., 2001), révélant ainsi le rôle essentiel de cette protéine dans de nombreuses fonctions biologiques. De plus, il a été reporté qu'Upf1 joue un rôle clé dans le maintien de l'intégrité du génome et la protection des séquences télomériques (Azzalin et al., 2007).

Upf2 sert d'adaptateur et permet de faire un lien entre les protéines Upf1 et Upf3, c'est une protéine de 1272 acides aminés avec un poids moléculaire de 148 kDa (Mendell et al., 2000; Serin et al., 2001). Dans sa région N-terminale, elle possède un signal de localisation nucléaire NLS (*Nuclear localisation signal*), et trois domaines MIF4G (*middle of eIF4G-like*) qui lui permettent d'interagir avec les facteurs d'initiation de la traduction eIF4A et eIF3 (Mendell et al., 2000). Le domaine C-terminal est impliqué dans l'interaction avec Upf1 (Mendell et al., 2000).

Les deux protéines Upf3a et Upf3b interagissent avec Upf2 par leur partie N-terminale. Elles présentent toute deux une localisation majoritairement nucléaire mais peuvent aller vers le cytoplasme (Lykke-Andersen et al., 2000; Serin et al., 2001). Malgré leurs similarités, leurs effets sur la NMD semblent être différents. En effet, la fixation artificielle d'Upf3b sur la région 3'UTR d'un ARNm dépourvu de PTC induit sa dégradation, alors que l'effet d'Upf3a est très faible (Lykke-Andersen et al., 2000). De plus, la déplétion d'Upf3b bloque la voie NMD mais pas celle d'Upf3a (Kunz et al., 2006).

2.1.2. Les protéines SMG

Les protéines SMG assurent la phosphorylation et déphosphorylation des protéines Upf. Dans cette famille composée de nombreux membres, 4 protéines jouent un rôle majeur dans la NMD : SMG1, SMG5, SMG6 et SMG7.

SMG1 phosphoryle Upf1 sur ses régions N- et C-terminales. Elle appartient à la famille des PI3KK (*Phosphatidylinositol 3-kinase related protein kinases*). La phosphorylation d'Upf1 par SMG1 est une des étapes clef de la NMD. En effet, les mutants de SMG1 dépourvus de l'activité kinase inhibe la voie NMD, alors que la surexpression de SMG1 augmente la phosphorylation d'Upf1 et donc l'efficacité de la voie NMD (Yamashita et al., 2001). SMG1 forme un complexe avec les protéines SMG8 et SMG9. Des expériences *in vivo* ont permis de mettre en évidence que SMG8 est nécessaire pour le recrutement de SMG1 sur le complexe de surveillance et régule l'activité kinase de SMG1 sur Upf1, alors que SMG9 est nécessaire pour l'interaction entre SMG1 et SMG8 (Yamashita et al., 2009).

Les protéines SMG5 et SMG7 participent à la déphosphorylation d'Upf1. Ce ne sont pas des phosphatases, mais elles assurent le lien entre la protéine phosphatase 2A (PP2A) et la forme phosphorylée d'Upf1. SMG7 possède une partie N-terminale, contenant un domaine 14-3-3 de fixation aux phosphosérines, impliqué dans l'interaction avec Upf1 phosphorylée. Une mutation au sein de ce domaine inhibe son interaction avec Upf1 phosphorylée *in vitro* (Fukuhara et al., 2005).

Cette séquence conservée 14-3-3, est également présente chez SMG5 et SMG6. La région C-terminale de ces protéines possède un domaine PIN (*Pilt N-terminal domain*) caractéristique des protéines phosphodiesterases avec une activité nucléasique. C'est SMG6 qui est responsable du clivage endonucléolytique de l'ARNm à proximité du PTC par l'intermédiaire de son domaine PIN (Eberle et al., 2009; Gatfield and Izaurralde, 2004; Huntzinger et al., 2008).

2.2. Le complexe de Jonction Exon-Exon (EJC)

L'EJC est un complexe protéique déposé lors de l'épissage sur l'ARNm, à 20-24 nucléotides en amont de chaque jonction exon-exon (Le Hir, 2000; Le Hir et al., 2000; Linde and Kerem, 2011). Le cœur de l'EJC est formé d'un hétérotétramère, composé de quatre protéines, qui reste associé

de façon stable à l'ARNm : eIF4A3, MLN51, Y14 et Magoh (Andersen et al., 2006; Ballut et al., 2005; Tange et al., 2005). Le complexe EJC joue le rôle de plateforme d'ancrage aux facteurs Upf2 et Upf3a et Upf3b (Jurica and Moore, 2002; Makarov et al., 2002). Chez les eucaryotes supérieurs, l'épissage et le dépôt de l'EJC jouent un rôle primordial dans le mécanisme de reconnaissance du PTC et dans la dégradation de l'ARNm. La présence de PTC sur un ARN monoexonique n'active pas la dégradation par la voie NMD (Maquat and Li, 2001).

2.3. Mécanisme de dégradation des transcrits contenant des codons non-sens (NMD)

Le mécanisme de la NMD a été mis en évidence, au cours de l'étude de pathologies humaines causées par des mutations non-sens ou de décalages du cadre de lecture, engendrant des codons stop prématurés PTC (Baserga and Benz, 1992; Urlaub et al., 1989). Il a été montré que les ARNm portant des PTC sont rapidement dégradés, empêchant ainsi leur traduction, et protégeant la cellule des mauvais effets liés à la production de protéines tronquées (Chang *et al.*, 2007). La voie NMD est activée lorsqu'un ribosome rencontre un codon stop prématuré. La reconnaissance d'un codon stop anormal et le déclenchement de la NMD est un processus complexe. Ce mécanisme est néanmoins bien élucidé et repose chez les mammifères, sur des interactions entre des ribonucléoprotéines, l'EJC et les protéines Upf (Bhuvanagiri et al., 2010; Rebbapragada and Lykke-Andersen, 2009; Silva and Romão, 2009).

La NMD a été longtemps considérée comme un processus de surveillance des ARNm ayant lieu uniquement au cours du premier tour de traduction. Comme décrit dans le chapitre I, les premières études menées sur cette voie avaient montré que les ARNm qui lient CBP80 sont dégradés par la NMD, ce qui suggérait que la NMD avait lieu uniquement au premier tour de traduction (Ishigaki et al., 2001). Cependant, il a été rapporté par la suite que la NMD pouvait aussi avoir lieu pour les ARNm qui ont lié eIF4E grâce à une interaction entre eIF4E et Upf1 chez les mammifères. De plus, de nombreux ARNm sont régulés par la NMD, même après le premier cycle de traduction et en absence d'EJC. Les complexes EJC ne seraient donc pas indispensables à l'activation de la voie NMD. Ces résultats laissent penser que la NMD n'est pas limitée au premier tour de traduction (Durand and Lykke-Andersen, 2013; Rufener and Mühlemann, 2013).

Lors de la terminaison de la traduction sur un codon stop normal, PABPC1 chez les mammifères interagirait avec eRF3 et cette interaction empêche l'activation de la voie NMD (**Figure 12**). Cependant, Upf1 peut également interagir avec eRF3 et cette interaction est en compétition avec l'interaction eRF3-PABPC1. Plusieurs études ont montré qu'Upf1 est capable de se lier à la région 3'UTR ou à la phase codante des ARNm régulés ou non par la NMD (pour revue Lykke-Andersen and Jensen, 2015). Upf1 se trouverait donc fixé aux ARNm avant la traduction puis serait détaché de l'ARNm par les ribosomes. Ainsi, lors d'une terminaison normale de la traduction sur un codon stop, le complexe eRF1-eRF3 serait recruté et il y aurait compétition entre Upf1 et PABPC1 pour interagir avec eRF3. En absence d'EJC, le recrutement d'Upf2 serait beaucoup plus lent et PABPC1 pourrait interagir avec eRF3 en dissociant Upf1, ce qui aurait pour conséquence d'inactiver la NMD et de favoriser la terminaison classique de la traduction. En revanche, en présence d'un EJC, le complexe Upf2-Upf3 s'associe au complexe EJC, et engendre l'activation de la voie NMD. Le changement de conformation provoqué par la fixation d'Upf2 à Upf1 permet de recruter SMG1 lié à SMG8 et SMG9. Avec l'aide d'Upf2, SMG8 et SMG9 se dissocient de SMG1, dont l'activité kinase est alors activée. SMG1 phosphoryle les résidus sérine et thréonine présents dans les régions N- et C-terminales d'Upf1, ce qui déclenche la dissociation des sous-unités ribosomiques, l'activité hélicase d'Upf1 et les mécanismes de dégradation de l'ARNm (**Figure 12**). La protéine Upf1 hyperphosphorylée va servir de plateforme pour la liaison des protéines SMG5, SMG6 et SMG7. La première voie de dégradation, majoritaire, est déclenchée par l'interaction de SMG6 avec Upf1 ; dans les cellules humaines SMG6 clive l'ARNm au niveau du codon stop prématuré. Les fragments 3' d'ARN sont alors rapidement dégradés par Xrn1 et les fragments 5' sont digérés par l'exosome. La deuxième voie de dégradation des ARNm, minoritaire, est déclenchée par le recrutement du complexe SMG5-SMG7 qui interagit avec les sérines phosphorylées d'Upf1. Ce complexe associé à la phosphatase PP2A déphosphoryle Upf1 puis s'associe au complexe Ccr4-Caf1-NOT pour désadényler la queue poly(A) et permettre à Xrn1 de dégrader l'ARNm.

Upf1 pourrait aussi s'associer au complexe de décoiffage Dcp1A-Dcp2-PNRC2 (*Pro-rich nuclear receptor co-activator 2*) sans être phosphorylée. Cette voie de dégradation serait indépendante de SMG5, SMG6 et SMG7, et constitue une troisième voie potentielle de dégradation des ARNm aberrants ((Lykke-Andersen and Jensen, 2015).

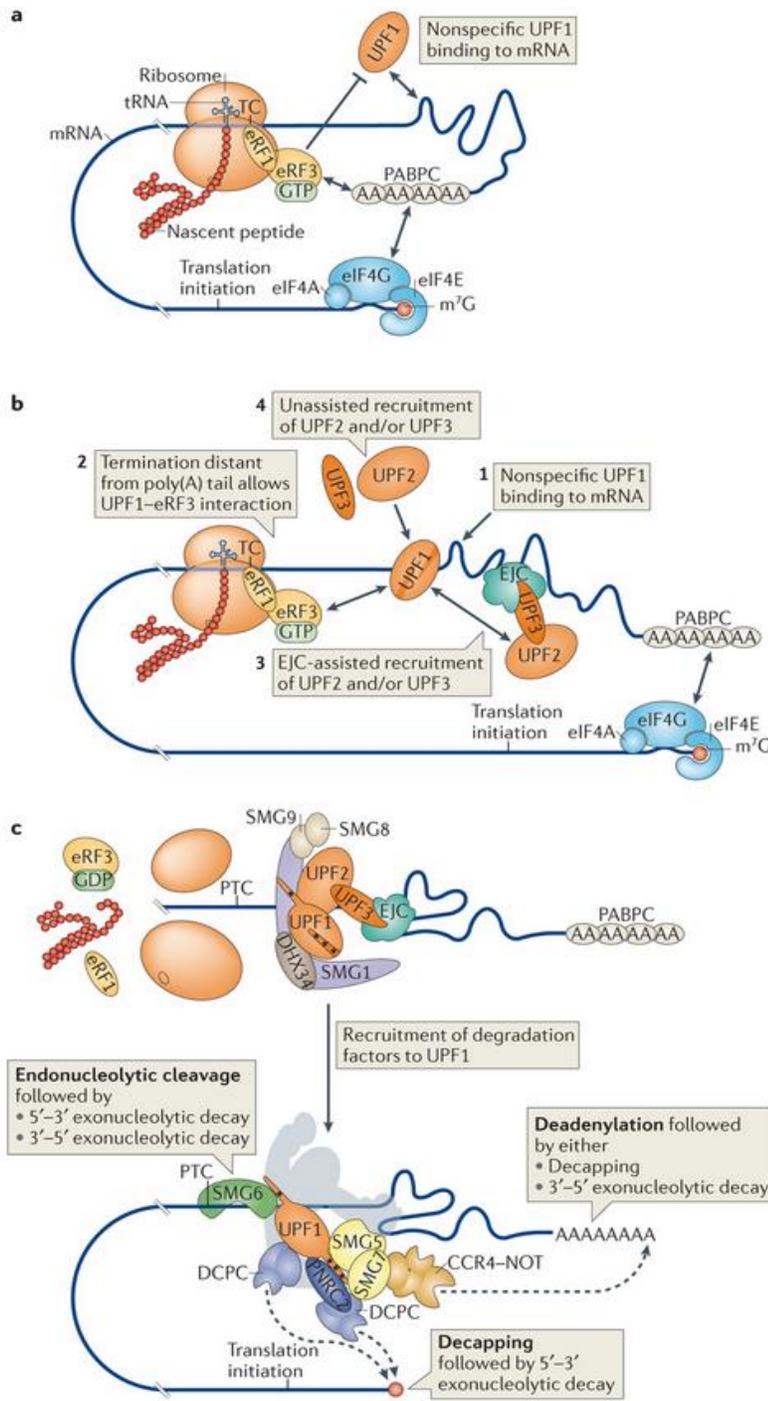


Figure 12 : Modèle universel de la voie NMD. *a.* Une terminaison de la traduction efficace empêche le déclenchement de la NMD. Qui est efficace lorsque le codon stop est proche de la queue poly(A) et de la coiffe. *b.* La terminaison anormale de la traduction se produit lorsque le codon de terminaison (TC) est éloigné de la queue poly(A). L'activation de la NMD nécessite l'interaction d'Upf1 et non PABP avec les facteurs de terminaison. Upf1 se lie de manière non spécifique à l'ARNm avant la traduction et interagit avec les facteurs de terminaison sur le ribosome positionné au niveau du TC. En absence d'EJC, Upf1 est activé par Upf2 et Upf3. La présence d'EJC facilite le recrutement d'Upf2-Upf3. A cette étape la terminaison de la traduction est arrêtée en faveur de l'activation de la voie NMD. *c.* En haut. La kinase SMG1 est activée lorsque SMG8 et SMG9 sont dissociés, qui est facilité par Upf2 et DHX34. SMG1 phosphoryle Upf1 sur ses résidus Ser et Thr en N- et C-terminale. L'activité hélicase d'Upf1 est alors activée et stimule les voies de dégradations de l'ARN. *c.* En bas. La phosphorylation d'Upf1 favorise la dégradation de l'ARNm par trois mécanismes distincts : le clivage endonucléolytique par SMG6. Le recrutement du complexe CCR4-NOT médiée par SMG7. Le

recrutement du complexe de désadénylation (DCPC) directement par Upf1 ou par l'intermédiaire de PNRC2. Les ARNm qui sont la cible de la voie NMD sont ensuite dégradés par les activités exonucléolytiques cellulaires (Lykke-Andersen and Jensen, 2015).

IV. Régulation de l'initiation de la traduction

Les régulations de la traduction touchent majoritairement l'étape d'initiation bien que les trois phases de la traduction soient régulées. La régulation peut être globale et affecter tous les ARNm, ou sélective et ne toucher qu'une catégorie d'ARNm. Je ne parlerai dans ce chapitre que de la régulation traductionnelle d'un ARNm qui se fait grâce à l'action de facteurs d'initiation et d'éléments présents dans sa région 5'UTR.

1. Régulation par modulation de l'activité des facteurs d'initiation de traduction

Deux facteurs d'initiations eIF2 et eIF4E contribuent de façon majeure dans la régulation de l'initiation de la traduction. Rappelons que le facteur eIF2 est responsable de la formation du complexe ternaire qui se fixe à la sous-unité 40S pour former le complexe de pré-initiation 43S, tandis que le facteur eIF4E participe à l'assemblage du complexe d'initiation 48S sur l'ARNm.

1.1. Régulation par le facteur eIF2

Le couplage d'eIF2 au GTP est l'une des principales régulations de l'initiation de la traduction (**Figure 13**). Le facteur eIF2, qui forme le complexe ternaire avec le GTP et le Met-ARNt^{Met}, est un complexe hétérotrimérique constitué de trois sous-unités : α , β et γ . eIF2 γ porte deux sites de liaison, l'un pour le GTP, le second pour le Met-ARNt^{Met}. La sous-unité eIF2 β est capable de lier l'ARNm. Elle est essentielle dans la reconnaissance du Met-ARNt^{Met} et sa liaison au site P de la petite sous-unité ribosomique 40S (Valásek, 2012). Les deux sous-unités γ et β interagissent avec eIF2B (échangeur GDP-GTP) et eIF5 (activateur de GTPase). Elles forment la partie effectrice du facteur eIF2. Suite à la reconnaissance du codon initiateur, le facteur eIF5 interagit avec la sous-unité β du facteur eIF2 ce qui provoque un changement conformationnel de ce dernier et permet l'activation de l'activité GTPase de la sous-unité eIF2 α . A la fin de l'initiation, eIF2 est relargué sous sa forme inactive d'eIF2-GDP. Pour être réactivé pour de nouveaux cycles de traduction, eIF2 doit être rechargé en GTP grâce au facteur d'échange eIF2B (Das and Maitra, 2001; Kimball, 1999;

Pain, 1996). La sous-unité eIF2 α est la principale sous-unité régulatrice de ce complexe car elle contient à la fois un site de phosphorylation et des sites de liaison à l'ARN.

En réponse aux stress, eIF2 α est phosphorylée sur la sérine 51, elle séquestre le facteur eIF2B, qui est limitant par rapport à eIF2, et bloque la réaction d'échange du GDP en GTP, inhibant ainsi la formation du complexe ternaire.

Chez les mammifères, on compte quatre protéines kinases capables de phosphoryler eIF2 α (**Figure 13**). Ces protéines kinases qui diffèrent par leur mode d'activation, permettent à la machinerie traductionnelle d'adapter la synthèse protéique aux changements des conditions physiologiques dans lesquelles se trouve la cellule (Holcik and Sonenberg, 2005):

- la kinase GCN2 (*General Control Nonderepressible 2*) est activée par la carence en acides aminés (conduisant à la présence d'ARNt déchargés), mais également par l'inhibition du protéasome ou les radiations UV,
- la kinase HRI (*Heme Regulated Inhibitor kinase*) est activée par le stress oxydatif, par le choc thermique ou par une carence en hème,
- la kinase PKR (*Protein Kinase R*) est activée par la présence d'ARN double brins amenés par, des infections virales,
- la kinase PERK (*PKR-like Endoplasmic Reticulum Kinase*) est activée en cas de stress lié au réticulum endoplasmique, dû à des protéines mal repliées, ou encore en cas d'hypoxie.

En phosphorylant eIF2 α , ces kinases limitent la disponibilité en complexe ternaire. Il en résulte une atténuation de la traduction globale, en faveur de la traduction de certains ARNm, en particulier ceux portant des uORF comme par exemple ceux d'ATF4, ATF5, CHOP et GADD34 (Hinnebusch, 2011; Palam et al., 2011). Le mécanisme général de ce processus est détaillé plus bas.

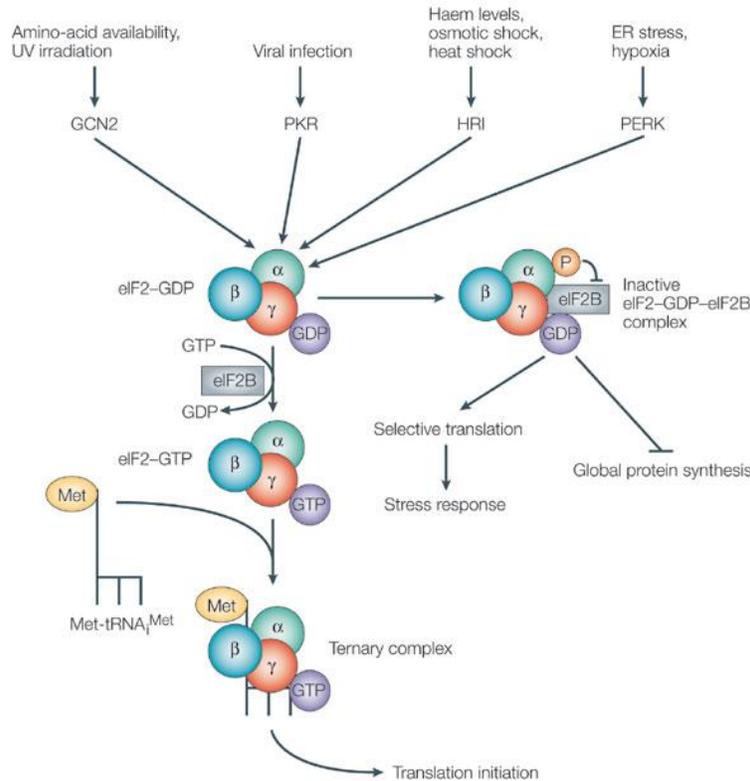


Figure 13 : Régulation de la formation du complexe ternaire. En cas de stress, les kinases qui phosphorylent eIF2 sont activées. La phosphorylation d'eIF2 sur sa sous-unité α empêche l'échange de GDP en GTP et la formation du complexe ternaire. Il en résulte une diminution de la traduction globale en faveur des ARNm impliqués dans la réponse au stress. La déphosphorylation eIF2 α permet à eIF2B de recharger eIF2 en GTP, ce qui est nécessaire à la formation du complexe ternaire (Holcik and Sonenberg, 2005).

1.2. Régulation par le facteur eIF4E

La formation du complexe eIF4F qui se lie à l'ARNm et permet l'arrimage de la petite sous unité ribosomique 40S sur l'ARNm, dépend de la disponibilité du facteur d'initiation eIF4E. La disponibilité de eIF4E dans les cellules impacte particulièrement la traduction des messagers présentant une région 5'UTR structurée, tels que certains proto-oncogènes impliqués dans le cycle cellulaire (cycline D1), la prolifération cellulaire (c-Myc) ou la survie cellulaire (Bcl2) (Mamane et al., 2007). Il a été montré que la traduction de ces ARNm est favorisée par une surexpression du facteur eIF4E (De Benedetti and Harris, 1999) et que sa surexpression transforme les cellules saines en cellules tumorales (Lazaris-Karatzas et al., 1990). De ce fait eIF4E est considéré comme un proto-oncogène (Furic et al., 2010; Jia et al., 2012).

Les régulations majeures de l'initiation de la traduction font intervenir de nombreuses protéines kinases qui agissent en cascades. Deux voies de signalisation sont fortement impliquées dans la régulation traductionnelle et en particulier au niveau d'eIF4E : la voie TOR (*Target Of Rapamycin*) ou mTOR chez les mammifères, et la voie MAPK (*Mitogen-Activated Protein Kinases*).

1.2.1. Régulation d'eIF4E par le facteur 4E-BP et la voie TOR

Le facteur eIF4E peut être séquestré par les facteurs 4E-BP (*eIF4E-Binding Protein*). Il existe trois différents facteurs 4E-BP dans les cellules de mammifères, 4E-BP1, 4E-BP2 et 4E-BP3, qui sont tous capables de se lier au facteur eIF4E (Joshi et al., 2004). La liaison eIF4E-4E-BP inhibe l'initiation de la traduction coiffe dépendante. La protéine 4E-BP est capable de se lier au facteur eIF4E par ses domaines N- et C-terminaux (Gosselin et al., 2011; Peter et al., 2015). 4E-BP porte également quatre sites très conservés de phosphorylation sur des résidus sérine ou thréonine (Ser65, Thr37, Thr46 et Thr70). La phosphorylation de 4E-BP affecte sa liaison à eIF4E. Les sites de phosphorylation sont les cibles de la voie de signalisation mTOR (von Manteuffel et al., 1996). L'action de cette voie sur le niveau de phosphorylation de 4E-BP diffère selon l'état de la cellule et le stress qui l'affecte. A l'inverse d'eIF4E qui est considéré comme un proto-oncogène, le facteur 4E-BP est considéré comme un anti-oncogène, sa surexpression est capable d'inverser le phénotype cancéreux induit par la surexpression d'eIF4E (Rousseau et al., 1996).

La voie mTOR permet la régulation de la prolifération et de la croissance cellulaire et elle est conservée chez tous les eucaryotes. Elle est activée par la disponibilité en facteurs de croissance, en hormones et en acides aminés, et inhibée par le stress. Sa dérégulation est intimement liée à l'apparition de maladies tel que le diabète ou le cancer. La voie mTOR agit sur de nombreux processus transcriptionnels et traductionnels, je ne parlerai ici que de la régulation traductionnelle dépendante de mTOR. La voie de signalisation mTOR met en place tout une cascade de kinases, en amont et en aval de la protéine kinase mTOR (**Figure 14**).

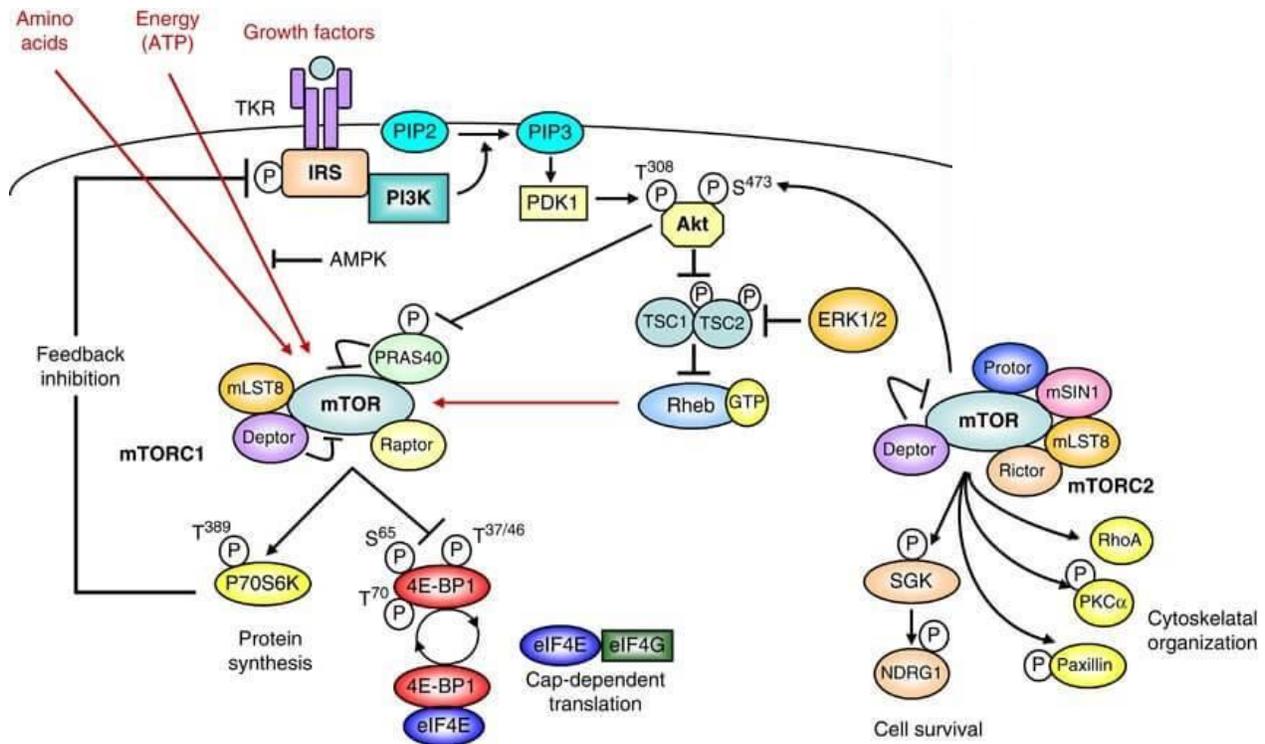


Figure 14 : Voie de signalisation mTOR et contrôle traductionnel, (d'après Le Bacquer *et al.*, 2006). Il existe deux complexes impliquant mTOR : mTORC1 et mTORC2. Trois signaux majeurs contrôlent mTORC1 : les facteurs de croissance, les nutriments (amino acides en particulier) et le statut bioénergétique de la cellule. La principale voie d'activation de mTORC1 est celle des facteurs de croissance qui passe par la voie de signalisation de la PI3kinase/Akt. Cette voie est activée par des signaux provenant de la surface cellulaire tels que les facteurs de croissance, des cytokines et des hormones (en particulier l'insuline) qui agissent en se fixant à des récepteurs transmembranaires à activité tyrosine kinase (TKR). La fixation du ligand spécifique sur la partie extracellulaire du récepteur (IRS) entraîne sa dimérisation et son autophosphorylation par son activité tyrosine kinase intracellulaire. Cette activité tyrosine kinase phosphoryle la phosphatidyl-inositol-3-kinase (PI3K). La PI3K catalyse alors la conversion du phosphatidylinositol-4,5-bisphosphate (PIP2) en phosphatidylinositol-3,4,5-trisphosphate (PIP3). Le PIP3 active la 3-Phosphoinositide-Dependent-Kinase 1 (PDK1), qui a son tour phosphoryle la sérine-thréonine kinase Akt sur la Thréonine 308. La kinase AKT phosphoryle différents substrats dont le complexe hétérodimérique TSC1/TSC2 sur la sous-unité TSC2. En plus des signaux passant par la kinase AKT, les signaux activant les kinases ERK (ERK1/2) passent par le complexe TSC1/TSC2. Par sa sous-unité TSC2, le complexe TSC1/TSC2 possède une activité GAP (GTPase activating protéine) pour la GTPase Rheb. Cette activité induit la conversion de Rheb-GTP en Rheb-GDP. La forme Rheb-GTP interagit directement avec le complexe mTORC1 et active sa kinase mTOR. La kinase mTOR possède de nombreuses cibles. Cependant deux cibles directes principales de mTOR sont impliquées dans la régulation traductionnelle : la kinase de la protéine ribosomique S6, P70S6K, et la protéine inhibitrice du facteur d'initiation de la traduction eIF4E, 4E-BP1. Le complexe mTORC2 constitué entre autre de Rictor (rapamycin-insensitive companion of mTOR), mLST8 et de mTOR est impliqué dans l'organisation du cytosquelette et dans la survie cellulaire. Il active Akt par phosphorylation de la Serine 473.

Ces kinases sont responsables du contrôle de la machinerie traductionnelle, notamment par le biais du facteur eIF4E via 4E-BP, et permettent un lien serré entre la synthèse protéique et la prolifération cellulaire (Zoncu *et al.*, 2011). Selon ses partenaires d'interaction, la protéine kinase

mTOR peut former les deux complexes mTORC1 (*mTOR Complex 1*) quand elle est associée à Raptor pour *Regulatory associated protein of mTOR* et mTORC2, quand elle est associée à Rictor pour *Rapamycin insensitive companion of mTOR*.

L'activation du complexe mTORC1 va activer ses cibles régulatrices de la traduction (**Figure 14**). Il s'agit principalement de la protéine inhibitrice 4E-BP (von Manteuffel et al., 1996), qui régule la traduction coiffe dépendante, et de la protéine kinase S6 (*S6K*). L'activation de la S6K phosphoryle le facteur d'initiation eIF4B et permet son activation. S6K phosphoryle également la protéine ribosomique RPS6 (*Ribosomal protein small 6*), impliquée dans la biogenèse des ribosomes (Hannan et al., 2003). Le complexe mTORC2 assure un rétrocontrôle positif sur le complexe mTORC1 (**Figure 14**). En effet, une fois activé le complexe mTORC2 stimule l'activité de la protéine kinase Akt (Zinzalla et al., 2011). Ce dernier se trouve juste en amont de mTORC1 au niveau de la cascade de phosphorylation ce qui permet l'activation du complexe mTORC1.

1.2.2. Régulation par phosphorylation d'eIF4E et la voie MAPK

Lorsque eIF4E n'est pas séquestré par 4E-BP et sous l'effet de différents stress, le facteur eIF4E peut être phosphorylé sur sa sérine 209 par la kinase Mnk (*MAPK interacting kinases*) qui est recrutée par eIF4G (Pyronnet et al., 1999). La kinase Mnk est sous le contrôle de la voie MAPK pour *Mitogen Activated Protein Kinase* (Wang et al., 1998). La phosphorylation d'eIF4E est corrélée avec le niveau de traduction et la prolifération cellulaire. En revanche, sa déphosphorylation a lieu en cas de carence en acides aminés, d'infection virale ou d'arrêt du cycle cellulaire en phase G0 (Scheper and Proud, 2002). Il semblerait que la phosphorylation d'eIF4E, permet la dissociation d'eIF4E de la coiffe des ARNm pour se fixer préférentiellement sur la coiffe d'une catégorie spécifique d'ARNm et permettre leur traduction. Cependant, le rôle de cette phosphorylation sur la synthèse protéique reste controversé (Scheper et al., 2002). Une étude analysant le taux de traduction en fonction de la phosphorylation d'eIF4E a montré que la phosphorylation de la Ser209 d'eIF4E favorise la traduction de proto-oncogènes, et ce sans variation de la quantité d'eIF4E (Furic et al., 2010).

La voie de signalisation MAPK comprend plusieurs cascades de kinases conservées chez les eucaryotes. Comme la voie mTOR, cette voie de signalisation est doublement impliquée dans la prolifération cellulaire : de façon directe en activant des facteurs de transcription, et indirectement en facilitant l'initiation de la traduction à travers la phosphorylation d'eIF4E (Furic et al., 2010). La voie MAPK est activée par des facteurs de croissance, des hormones et par la disponibilité en acides aminés, qui sont, à l'instar de la voie mTOR, captés par des récepteurs TKR. Par la suite, la voie se décompose en quatre niveaux successifs de kinases (**Figure 15**), qui aboutissent à l'activation des kinases Mnk1 et Mnk2.

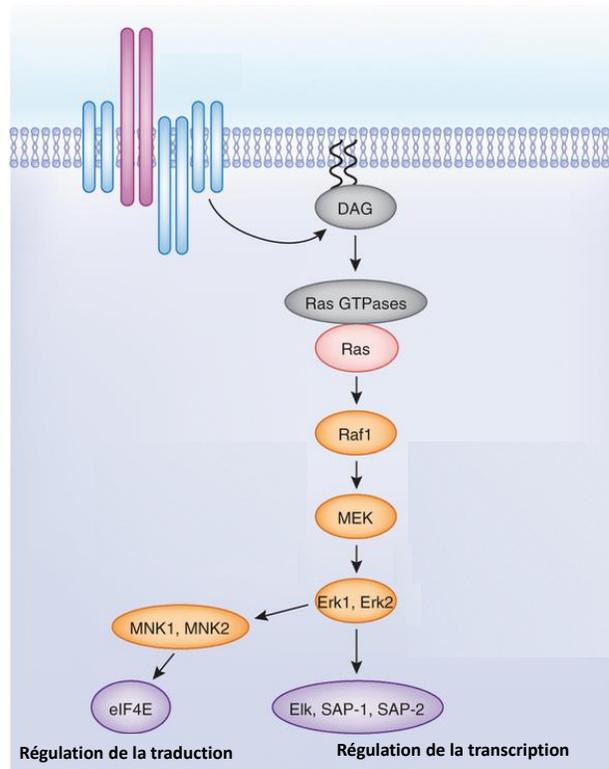


Figure 15 : Voie de signalisation MAPK et contrôle traductionnel. La stimulation des récepteurs à activité tyrosine kinase mène à l'activation de petites protéines appelées Ras GTPases qui à leur tour activent les MAP kinase kinase kinase (MAPKKK) telle que la protéine kinase Raf1. Raf1 active les MAP kinase kinase (MAPKK) comme MEK qui vont activer à leur tour les MAP kinases (MAPK) telles que Erk1/2. Suite à leur activation, les protéines kinases Erk1 et Erk2 peuvent soit rester dans le cytoplasme et activer les kinases MNK1 et MNK2 (MAPK interacting kinases) qui régulent la synthèse protéique en phosphorylant le facteur eIF4E, soit migrer dans le noyau et phosphoryler plusieurs facteurs de transcription tels que Elk1, SAP-1, SAP-2. Ces facteurs de transcription sont alors capables d'activer la transcription de gènes qui interviennent dans le contrôle de la prolifération cellulaire.

Les modes d'activation et les rôles des voies mTOR et MAPK sont très similaires. En effet, dans les deux cas, la présence de facteurs de croissance, d'hormones, d'acides aminés permet l'activation des récepteurs TKR membranaires ou de RCPG (*Récepteurs Couplés à une Protéine G*). D'autre part, ces voies de signalisation sont toutes les deux impliquées dans la régulation du métabolisme cellulaire en général, et dans la régulation de la traduction en particulier. Les deux voies de signalisation sont soumises à des boucles de rétrocontrôle par des kinases. Dans la voie mTOR, la S6K est responsable de deux rétrocontrôles négatifs, l'un agissant sur l'activation de la PI3K (**Figure 14**), et l'autre sur l'activation du complexe mTORC2. De plus, mTORC1 inhibe également la PI3K, tandis que mTORC2 active Akt. Dans la voie MAPK, c'est Erk qui inhibe l'activation de Ras, de Raf et de MEK. Ces différents rétrocontrôles permettent une autorégulation pour un meilleur maintien de l'homéostasie cellulaire.

2. Régulation par les éléments présents sur la région 5'UTR

La région 5'UTR joue un rôle crucial dans la régulation de la traduction des ARNm car elle est en relation étroite avec la sous-unité ribosomique 40S pendant l'étape de l'initiation de la traduction. Tous les éléments de séquence et de structure de la région 5'UTR jouent un rôle plus ou moins important dans la régulation traductionnelle : la coiffe, les structures secondaires et la présence d'uORF dont je parlerai de façon détaillée.

- **La coiffe** : un ARNm coiffé est toujours traduit de façon plus efficace que le même ARNm non coiffé. Cependant, certains ARNm sont traduits par un mécanisme d'initiation interne qui n'implique pas la coiffe.

- **Les structures secondaires** : la présence de structures secondaires et de tiges boucles stables dans la 5'UTR des ARNm, diminue l'efficacité de l'initiation de la traduction, du fait que les facteurs d'initiation à activité hélicase ont des difficultés à dérouler de telles structures (Sonenberg and Hinnebusch, 2009). Cet effet dépend de la position de ces structures, car il est d'autant plus prononcé quand elles sont situées à proximité de la coiffe. L'assemblage du complexe de pré-initiation 43S au niveau de la coiffe est donc gêné par la présence de telles structures secondaires, ainsi que la progression de la sous-unité 40S le long de l'ARNm. La fixation du complexe de pré-initiation 43S sur l'ARNm et le déplacement de la sous-unité 40S le long du 5'UTR peuvent être

modulés par des interactions entre des structures secondaires spécifiques et des protéines régulatrices de l'initiation de la traduction, qui sont généralement des facteurs répresseurs. L'exemple le plus connu est celui du messager de la ferritine (revue dans Pantopoulos, 2004). En l'absence de fer dans la cellule, le facteur IRP1 (*iron responsive protein-1*) bloque l'initiation de la traduction au niveau de la coiffe de l'ARNm de la ferritine en se fixant sur une séquence très structurée, appelée IRE (*iron responsive element*), qui est située à 40 nucléotides de la coiffe. Par conséquent, la ferritine n'est synthétisée que dans les cellules correctement approvisionnées en fer, et c'est sa concentration qui règle la conversion entre les formes de haute et de faible affinité du facteur IRP1 pour l'élément IRE. A l'inverse, dans les cellules ravitaillées en fer, le facteur IRP1 est rapidement dégradé par le protéasome.

- **Les uORF** : Par définition une uORF est comprise entre un codon d'initiation et un codon stop dans la même phase présente dans la région 5'UTR des ARN messagers. Les uORF sont donc des petits cadres de lecture ouverts de tailles très variables allant de 2 codons jusqu'à environ 150 codons. La disposition des uORF sur les transcrits est aussi très variables : elles peuvent être en amont du cadre de lecture principal (le codon stop de l'uORF se trouve en amont du codon d'initiation de l'ORF principale), ou chevauchantes, (le codon stop de l'uORF se trouve en aval du codon d'initiation de l'ORF principale). L'uORF peut aussi être comprise dans l'ORF principale avec son codon d'initiation et son codon stop situés après le codon initiateur de l'ORF principale mais bien sûr dans une autre phase de lecture (**Figure 16**). De plus les uORF peuvent être multiples et dans des situations diverses dans le 5'UTR (**Figure 16**).

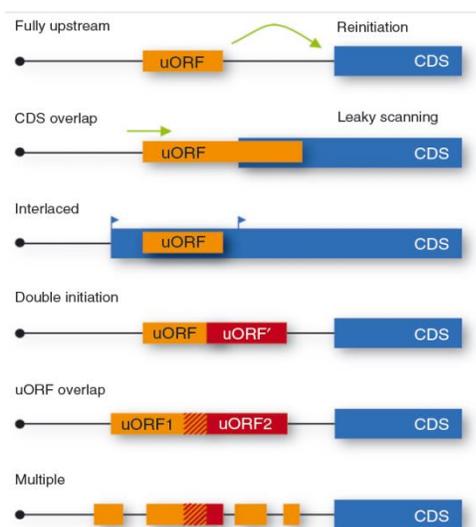


Figure 16 : Les différents arrangements des uORF dans la région 5'UTR. La région 5'UTR (ligne noire) peut contenir des uORF (rectangle orange) entre la coiffe et l'ORF principale (CDS, rectangle bleu). Les uORF peuvent être localisées entièrement en amont de l'ORF principale, chevaucher l'ORF principale, ou être comprises dans l'ORF principale entre deux codons d'initiation en phase. Les flèches vertes indiquent les deux voies possibles de traduction de l'ORF principale en présence d'uORF : la réinitiation ou le leaky scanning quand le codon initiateur de l'uORF n'est pas dans un contexte favorable. Les 3 schéma du bas représentent plusieurs cas d'uORF multiples, chevauchantes, dans la même phase ou non (Wethmar, 2014).

2. 1. Les mécanismes de régulation de la traduction par les uORF

Le rôle de l'uORF est principalement régulateur et tend à diminuer l'efficacité de traduction de l'ORF principale. Le mécanisme général de ce processus est simple. Le complexe d'initiation qui balaye la région 5'UTR d'un ARNm va se lier au premier codon d'initiation AUG dans un contexte favorable, ici le codon AUG de l'uORF, et induire la traduction de l'uORF. Après terminaison de la traduction de l'uORF, la sous-unité 40S dépourvue du complexe ternaire eIF2-GTP-Met-ARNt_i^{Met} reprend sa migration le long de l'ARN jusqu'à être rechargé en complexe ternaire. Le complexe d'initiation, une fois reformé, va initier la traduction à un autre AUG en aval. Ce peut être l'AUG de l'ORF principale ou un autre AUG dans un contexte favorable comme celui d'une seconde uORF et la même séquence d'évènements se reproduit : traduction de l'uORF, terminaison, migration le long de l'ARNm et réinitiation à un AUG en aval.

Ce mécanisme est modulé par de nombreuses variables : la distance entre la coiffe et l'uORF, le contexte entourant le codon d'initiation de l'uORF, la distance entre les différents codons d'initiation ou plus précisément entre les uORF dans le cas d'uORF multiples, le contexte entourant le codon de terminaison de l'uORF, la distance entre le codon de terminaison de l'uORF et le codon d'initiation suivant, mais aussi, la longueur des uORF et la présence de structure secondaire (**Figure 17**).

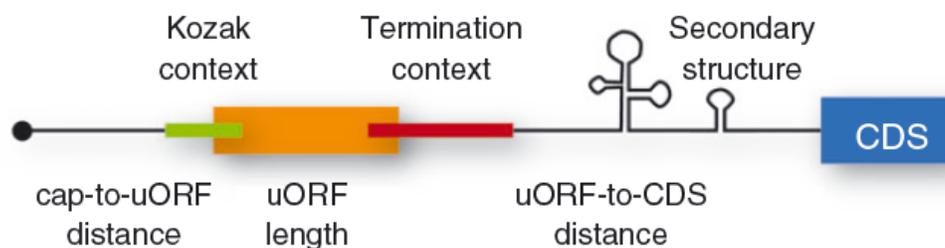


Figure 17 : Les différents éléments de l'ARNm qui modulent l'effet régulateur des uORF (Wethmar, 2014).

Parmi ces variables, certaines ont une importance particulière. Les codons AUG des uORF ne sont pas toujours forcément reconnus et il a été montré que le contexte nucléotidique de Kozak **RCCAUGG** (Kozak, 1986) était très important dans la reconnaissance des codons d'initiation des uORF. En présence de contexte nucléotidique non favorable à l'initiation de la traduction le

complexe d'initiation dépasse le codon d'initiation par un mécanisme appelé *leaky scanning* et initie à l'AUG en aval que ce soit celui de l'ORF principale ou celui d'une autre uORF. Comme nous le verrons dans la partie résultats, l'initiation de la traduction d'une uORF peut se faire à un codon non-AUG et dans ce cas le contexte nucléotidique entourant ce codon d'initiation non-canonique est particulièrement important. Dans tous les cas, la force du codon d'initiation de l'uORF est une variable clé de l'effet régulateur de l'uORF.

Dans le cas où l'uORF est traduite, plusieurs possibilités s'offrent aux ribosomes. Au codon stop de l'uORF, le complexe de traduction peut se dissocier, en particulier si l'uORF est suivie d'un stretch de nucléotides GC qui favorise la dissociation des deux sous-unités ribosomiques et empêche la réinitiation à l'ORF principale. Les ribosomes en cours d'élongation peuvent aussi rester bloqués au niveau de l'uORF, ce qui bloque l'accès à l'ORF principale. Ce blocage peut être peptide dépendant (Geballe and Morris, 1994). Il existe des exemples de régulation par le peptide codé par l'uORF qui provoque une diminution massive de l'expression de l'ORF principale, c'est le cas du messenger de la tyrosine phosphatase PTPRJ et de la S-adenosylmethionine decarboxylase (Karagyozov et al., 2008; Raney et al., 2002). Cependant, le peu d'exemples décrits chez les mammifères, n'ont pas permis d'identifier une séquence peptidique particulière qui serait responsable de ce phénomène. Une étude réalisée chez différentes espèces de plantes afin d'analyser la conservation des peptides codés par des uORF, a révélé que les uORF riches en sérine, thréonine et tyrosine, qui sont potentiellement phosphorylables, peuvent bloquer les ribosomes (Hayden and Jorgensen, 2007). De plus, le polypeptide codé par l'uORF peut avoir un effet répresseur sur la traduction de l'ARNm en interagissant avec un composant de la machinerie de traduction ou en réduisant la stabilité de l'ARNm en réponse à des facteurs régulateurs tels que le sucre, l'arginine ou les polyamines (Wethmar et al., 2010b).

Les ribosomes, ou plus exactement sa petite sous-unité 40S, peuvent aussi rester accrochés à l'ARNm après traduction de l'uORF. Dans ce cas, la petite sous-unité 40S va reprendre le balayage de la région 5'UTR et réinitier la traduction au codon d'initiation suivant. Cependant, le redémarrage du balayage et la réinitiation dépendent de plusieurs paramètres. Le redémarrage du balayage dépend de la taille de l'uORF (plus celle-ci est courte, plus le redémarrage est efficace) et du temps nécessaire à la traduction de l'uORF qui est lié à la vitesse d'élongation (plus cette vitesse est grande et plus le redémarrage est efficace). Dans ces circonstances, plusieurs facteurs d'initiation

indispensables à la réinitiation, particulièrement les facteurs eIF3 et eIF4G, restent accrochés aux ribosomes pendant l'élongation et même après terminaison de la traduction de l'uORF. Ces facteurs seraient emmenés par la petite sous-unité du ribosome au redémarrage du balayage (Poyry, 2004; Unbehaun et al., 2004). Ainsi, plus l'uORF est longue, plus les ribosomes en cours de traduction ont tendance à perdre de façon stochastique les facteurs d'initiation qui y sont restés accrochés (Kozak, 1987b). Le facteur eIF3, en particulier sa sous-unité h, qui permet de recruter le complexe ternaire (eIF2-GTP-Met-tRNA_i^{Met}) est indispensable au processus de réinitiation (Munzarová et al., 2011; Roy et al., 2010). La réinitiation, en elle-même, dépend de la distance entre l'uORF et le codon d'initiation suivant (Kozak, 2002, 2001) ainsi que de la disponibilité en complexe ternaire. Ces deux paramètres sont liés : la distance entre l'uORF et le codon d'initiation suivant permettent à la petite sous-unité du ribosome de se recharger en complexe ternaire. La disponibilité en complexe ternaire dépend comme nous l'avons vu plus haut de l'état de phosphorylation du facteur eIF2 α qui varie en fonction des conditions de vie cellulaire et en particulier selon leur état de stress.

2.2. Les uORF et la réponse au stress

En réponse à des stimuli internes ou externes, les cellules modifient rapidement leur programme d'expression des gènes. Un des mécanismes le plus répondu pour inhiber la traduction globale passe par la phosphorylation du facteur d'initiation de la traduction eIF2 (Hinnebusch et al., 2007). Le mécanisme le mieux étudié de la réponse aux stress est celui induit par une carence en acides aminés. Lorsque il y a carence en acides aminés, on observe une diminution rapide de l'association des ribosomes aux ARNm et l'inhibition générale de la traduction des ARNm (Maurin et al., 2005). On observe aussi une augmentation de la phosphorylation de la kinase GCN2 qui induit une augmentation de la phosphorylation d'eIF2 α . Ces processus ont été conservés au cours de l'évolution, et ont été particulièrement bien étudiés chez la levure et chez les mammifères. Chez les mammifères, la kinase GCN2 bloque la traduction de la plupart des gènes à l'exception de quelques ARNm particuliers, notamment celui qui code pour l'activateur de la transcription ATF4. La kinase GCN2 possède en C-terminal un domaine homologue de l'histidyl-ARNt synthétase qui lui permet de se lier aux ARN de transferts non chargés qui s'accumulent dans la cellule en cas de déficit en acides aminés, particulièrement en acides aminés indispensables. Il résulte de cette liaison un changement conformationnel de GCN2 permissif pour son activité kinase conduisant à

l'autoactivation de GCN2 (Dong et al., 2000). Le seul substrat de GCN2 connu à l'heure actuelle est la sous-unité α du facteur d'initiation de la traduction eIF2. La phosphorylation de eIF2 α va entraîner une diminution du complexe ternaire et une augmentation de la traduction de l'ARNm d'ATF4 qui passe par des mécanismes similaires à ceux décrits aussi pour le gène GCN4 de levure (Lu et al., 2004). L'augmentation de la traduction d'ATF4 peut alors induire la transcription de gènes cibles possédant dans leur promoteur des éléments de réponse aux acides aminés, AARE pour *Amino Acid Response Element*.

ATF4 est un activateur de transcription, appartenant à la famille des ATF/CREB (activating transcription factor/cyclic AMP response element binding protein). Il possède un domaine bZip (basic region-leucine zipper) qui lui permet de former des homo- ou hétéro-dimères. ATF4 régule l'expression des gènes impliqués dans le stress oxydatif, la synthèse des acides aminés, la différenciation cellulaire et l'angiogenèse (Ameri and Harris, 2008).

La régulation de la traduction de l'ARNm d'ATF4 est basée sur la capacité des ribosomes de scanner les ARNm et de ré-initier la traduction après des petits cadres de lecture. L'ARN messenger ATF4 porte dans sa région 5'UTR deux uORF, la première, uORF1, codant 3 acides aminés est située à environ 50 nucléotides de l'extrémité 5' de l'ARNm d'ATF4, la deuxième est située à environ 90 nucléotides après le codon stop de l'uORF1 et chevauche l'ORF d'ATF4 (**Figure 18**). Cette situation des deux uORF d'ATF4 est remarquablement conservée chez les métazoaires.

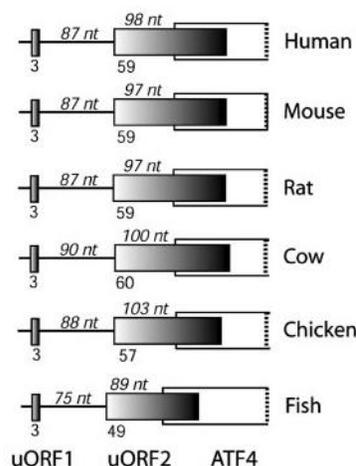


Figure 18 : Conservation des deux uORF présentes dans la région 5' non codante de l'ARNm d'ATF4 chez les métazoaires. Pour chaque espèce, la distance entre les ORF est notée en italique au-dessus du schéma et la longueur des uORF est notée sous le schéma en codons (Vattem and Wek, 2004).

Comme pour le gène homologue GNC4 de levure, le mécanisme de régulation de l'expression d'ATF4 repose sur la disponibilité en complexe ternaire.

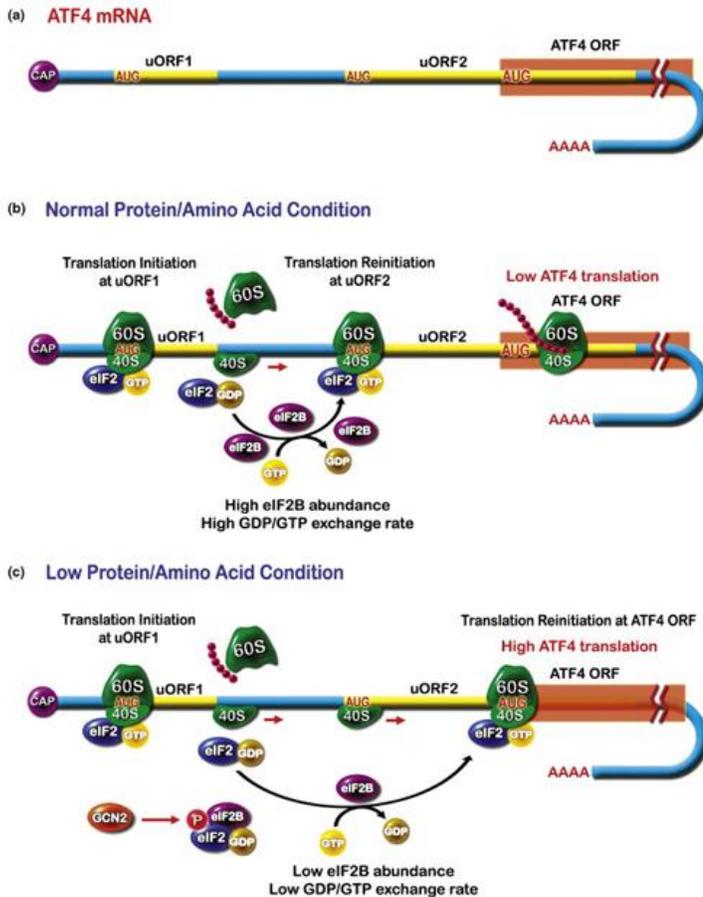


Figure 19 : Régulation traductionnelle de l'ARNm d'ATF4. (a) Structure de l'ARNm d'ATF4. (b) En conditions normale, après traduction de l'uORF1 d'ATF4, la disponibilité en complexe ternaire favorise la réinitiation à l'uORF2 d'ATF4. L'uORF2 chevauche l'ORF principale et la réinitiation au niveau de cette ORF principale ne se fait pas ou peu. (c) En cas de stress, eIF2 α est phosphorylée, ceci a pour conséquence la diminution de la quantité de complexes ternaires disponibles. La probabilité pour qu'il y ait réinitiation à l'uORF2 est donc diminuée, favorisant l'initiation à l'ORF principale d'ATF4 (Kilberg et al., 2009).

En conditions normales en acides aminés, il y a peu de eIF2 α phosphorylé, et donc une grande disponibilité en complexe ternaire. Le premier uORF1 est traduit puis la distance séparant l'uORF1 du codon d'initiation de l'uORF2 étant optimale, il y a réinitiation de la traduction au niveau de l'uORF2. Il faut noter que de façon tout aussi remarquable que la conservation des distances entre les uORF, le contexte nucléotidique des codons d'initiation de ces uORF est lui aussi très conservé et permet une initiation forte de la traduction. L'uORF2 capte ainsi tous les ribosomes et empêche l'initiation à l'AUG d'ATF4. Puisque ces deux derniers le permettent donc le deuxième petit cadre de lecture va jouer un rôle inhibiteur de la traduction, de la région codant ATF4 (**Figure 19**).

Dans des conditions de stress comme la carence en acides aminés, eIF2 α est en partie phosphorylé et séquestré par le facteur d'échange eIF2B. Il y a donc peu de complexes ternaires disponibles, la réinitiation se fait alors plus lentement, et la petite sous-unité 40S du ribosome va passer l'AUG de l'uORF2 avant d'être rechargée en complexe ternaire. Cependant, les quelques 100 nucléotides supplémentaires qui séparent l'AUG de l'uORF2 de celui de l'ORF d'ATF4 sont suffisant pour recharger un nouveau complexe ternaire sur la sous-unité 40S qui démarre la traduction à l'AUG d'ATF4. La traduction d'ATF4 permet un retour à l'état d'équilibre après un stress (Blais et al., 2004; Kilberg et al., 2009). Le facteur transcriptionnel ATF5 obéit au même processus de régulation traductionnel (Watatani et al., 2008). Tout comme ATF4, la région 5'UTR de l'ARNm d'ATF5 porte deux uORF conservées qui sont impliquées dans la réponse cellulaire à la carence en acides aminés ou au stress à l'arsénite, à travers la phosphorylation du facteur eIF2 α (Watatani et al., 2008). Plus récemment, des régulations par uORF ont été mises en évidence lors du contrôle traductionnel de plusieurs protéines, CHOP (*C/EBP (CCAAT/enhancer binding protein) homologous protein*) (Palam et al., 2011), VEGFA, CEBP et bien d'autres.

2.3. Les uORF et la stabilité des ARNm

Comme nous l'avons vu plus haut, les codons stop des uORF jouent le rôle de codon stop prématuré. En effet, le premier passage des ribosomes sur l'uORF ne permet pas le décrochage des éventuels complexes EJC présent en aval de cette dernière. Il semble aussi que le codon de terminaison de l'uORF est reconnu comme prématuré parce que la région 3' en aval de l'uORF est grande par rapport à celle des codons stop conventionnels des ORF principales (McGlinchy et al., 2010; Sachs, 2006). Ainsi, les uORF peuvent déclencher la dégradation de l'ARNm par la voie NMD (Ruiz-Echevarria et al., 1996). Plusieurs études ont montrés l'implication la NMD dans la régulation physiologique des ARNm portant des uORF (Mendell et al., 2004; Rehwinkel et al., 2006; Wittmann et al., 2006; Yepiskoposyan et al., 2011). Cependant, les transcrits porteurs d'uORF ne sont pas tous cibles de la voie de dégradation NMD. Il existe plusieurs exemples de transcrits où l'uORF diminue l'expression protéique en réprimant la traduction de l'ORF principale indépendamment de la dégradation de l'ARNm (McGlinchy et al., 2010; Sachs, 2006). Il semble que la taille de l'uORF est fortement impliquée dans la dégradation par la NMD. La présence de longues uORF sur l'ARNm a été corrélée avec le déclenchement de la NMD tandis que la présence

de petites uORF ne le permet pas (Silva and Romão, 2009). Cet effet de la taille de l'uORF serait dû à la proximité du complexe de terminaison de la traduction et de la protéine PABP, du fait de la circularisation de l'ARNm pendant la traduction (Peixeiro et al., 2012; Silva et al., 2008). Chez les mammifères, il est difficile de déterminer la taille minimale de l'uORF pour que cette dernière soit capable de déclencher la NMD (Mendell et al., 2004). Cependant, une étude réalisée chez les plantes a permis de définir la taille minimale pour qu'une uORF déclenche la NMD : la réduction de la taille de l'uORF en dessous de 36 nucléotides abolit sa capacité à déclencher la NMD (Nyikó et al., 2009). Chez les mammifères, les ARNm porteurs de longues uORF et dont l'expression est régulée par la NMD en conditions normales, deviennent résistant à la NMD en conditions de stress (Gardner, 2008; Zhao et al., 2010). L'ARNm du régulateur transcriptionnel IFRD1 (*Interferon-related developmental regulator1*), qui porte une seule longue uORF de 52 codons, en est l'exemple type : il échappe à la NMD et n'est plus dégradé en conditions de stress (Zhao et al., 2010). L'inhibition de la NMD en réponse à la phosphorylation d'eIF2 α peut être expliquée par le fait que la diminution de disponibilité en complexe ternaire, diminue la traduction de l'uORF d'IFRD1 par un mécanisme de *leaky scanning*, qui favorise la traduction de l'ORF principale d'IFRD1. En cas de stress et de diminution de la disponibilité en complexe ternaire, le codon d'initiation de l'uORF dans un contexte peu favorable est moins bien reconnu par les complexes d'initiation balayant la région 5'UTR. L'uORF d'IFRD1 n'est donc pas traduite, son codon stop n'est pas reconnu comme codon stop prématuré et ne déclenche pas la NMD. L'ARNm d'IFRD1 est donc stabilisé par le stress.

2.4. Les uORF et les pathologies humaines

Il est clairement établi que les uORF sont impliquées dans la régulation de l'expression des gènes à deux niveaux : au niveau traductionnel, dans ce cas l'uORF empêche la traduction de l'ORF principale et par conséquent diminue la synthèse protéique, et en déclenchant la dégradation de l'ARNm par la NMD. On peut donc imaginer que les mutations qui touchent aux uORF pourraient perturber ces processus de régulation de la synthèse protéique et auront de lourdes conséquences sur la régulation de l'expression des gènes et le destin de la cellule. En effet, Calvo et ses collaborateurs, ont étudié les 12 millions de cas de SNP qui affectent les uORF présents dans les bases de données répertoriant les SNP (Calvo et al., 2009; Sherry et al., 2001). Cette étude leur a

permis d'identifier 509 gènes avec des polymorphismes qui permettent de créer ou d'éliminer des uORF. Parmi ces gènes, 366 portent plusieurs uORF et 146 une seule uORF (Calvo et al., 2009). En comparant l'expression protéique des variants avec des uORF "normales" et ceux avec des uORF touchées par des SNP, ces auteurs ont montré que ces uORF étaient responsables de 30 à 60% de diminution de la synthèse protéique. Chez l'homme, il a été décrit une SNP qui permet de créer une uORF dans la région 5'UTR de l'ARNm du facteur FXII (*clotting factor XII*). La présence de cette SNP a été associée à des troubles thromboemboliques, qui résultaient d'une modification du taux de FXII présent dans le sang des patients atteints de cette pathologie (Bersano et al., 2008). Cette SNP localisée au niveau du nucléotide 24 de la région 5'UTR du transcrit FXII, permet simultanément de créer une très petite uORF de deux codons et altère le contexte de Kozak entourant l'AUG de l'ORF principale du facteur FXII.

L'effet régulateur exercé par les uORF répressives de l'expression des gènes peut aussi être modifié par des mutations qui les suppriment. Chez l'homme, l'exemple le plus décrit est celui qui touche le gène HR (*Hairless homolog*) ; son ARNm porte une uORF de 34 codons. Une mutation qui altère le codon d'initiation, supprime son uORF et augmente la synthèse de la protéine HR, a été associée à une pathologie autosomique dominante induisant la perte de cheveux (Baek et al., 2009; Wen et al., 2009).

Comme décrit plus haut dans la partie qui traite de la régulation traductionnelle par les uORF, dans certains cas, les uORF peuvent réprimer la traduction de façon dépendante des peptides qu'ils codent par ces uORF ; le peptide agirait par un mécanisme de blocage des ribosomes en cours de traduction (Lovett and Rogers, 1996). C'est le cas pour l'ARNm du récepteur DRD3 humain (*Dopamine D3 receptor*) qui porte une seule uORF de 36 codons. En effet, il a été montré que le polymorphisme de l'uORF entraînant la substitution de la Lysine 9 par un acide Aspartique augmentait les chances de prédisposition à la schizophrénie. La présence de l'Asp9 dans le peptide de l'uORF diminue l'efficacité de blocage des ribosomes en cours de la traduction et augmente la traduction de la protéine DRD3 (Sivagnanasundaram et al., 2000).

Un autre mécanisme de régulation originale par les uORF a été décrit pour les transcrits des gènes C/EBP (*CCAAT/enhancer binding protein*) qui portent des codons d'initiation alternatifs dans leur ORF principal. L'exemple le plus représentatif est celui de C/EBP β , un facteur transcriptionnel impliqué dans la différenciation de divers types cellulaires : adipocytes,

ostéoclastes, ostéoblastes, etc. (Huber et al., 2012). L'ORF principale de C/EBP β contient trois codons d'initiation en phase : le premier est séparé de 23 nucléotides du deuxième et de 198 nucléotides du 3ème. L'ARNm de C/EBP β contient aussi une uORF unique, codant 11 acides aminés, et se trouvant à l'intérieur de l'ORF principale entre les deux premiers codons d'initiation. Le codon stop de cette uORF est séparé du 3ème codon d'initiation AUG de la plus petite isoforme de l'ORF de C/EBP β par 530 nucléotides. *In vivo*, la perte de l'expression de C/EBP β dans les cellules satellites induit une augmentation de la largeur des fibres musculaires et une augmentation de la réparation des muscles après lésion (Marchildon et al., 2012), alors que la perte de l'expression de l'uORF de C/EBP β , modifie l'équilibre entre les trois isoformes, et aura pour conséquence une régénération tissulaire déficiente des muscles après lésion et une altération de la différenciation des ostéoclastes (Wethmar et al., 2010a).

Tous ces exemples de modifications physiopathologiques liées aux uORF, mettent en évidence l'importance des uORF dans les mécanismes de régulation de l'expression des gènes ainsi que l'effet des mutations pouvant modifier ces régulations. Les progrès récents dans les technologies de séquençage de dernière génération permettent une étude beaucoup plus précise et plus fine de ces uORF et de leur rôle régulateur. Ces avancées permettront sûrement la reconnaissance des uORF comme de possibles cibles thérapeutiques.

3. Les Nouvelles technologies pour étudier les mécanismes de régulation traductionnelle

Les efforts visant à connaître l'expression globale des gènes ont historiquement mis l'accent sur la mesure des niveaux d'ARNm, par exemple, en utilisant des puces d'expression ou *microarrays* (Brown and Botstein, 1999) ou le séquençage à haut débit des ARN ou *RNA-seq* (Wang et al., 2009). Jusqu'à récemment, la connaissance exacte de la traduction globale était beaucoup plus difficile que la mesure des niveaux d'ARNm. Cela a changé avec le développement de l'approche du profilage des ribosomes ou *ribosome profiling*, qui a permis de combler le vide en faisant le lien entre les études du transcriptome et du protéome, qui jusque-là ont été considérés comme étant faiblement corrélés (Vogel and Marcotte, 2012; Wang et al., 2009). L'étude de l'état de traduction des ARNm ou traductome permet maintenant de mieux établir les corrélations entre les résultats des études de transcriptomique et celles de protéomique.

L'état de traduction d'un ARNm est corrélé à son association avec les ribosomes à un temps donné, dans une condition donnée et notre capacité à cartographier la position des ribosomes sur les ARNm a été essentielle pour l'étude du traductome et la compréhension des mécanismes de traduction. Le *ribosome profiling* «footprinting» décrit pour la première fois en 2009 par Nick Ingolia et ses collaborateurs (Ingolia et al., 2009) a depuis connu un succès important et de nombreux laboratoires dans le monde entier utilisent de nos jours cette technique pour étudier la dynamique de la traduction *in vivo*.

Historiquement, les origines de cette approche remontent à la découverte des polysomes il y a plus de 50 ans et à la conclusion selon laquelle le traitement par les endoribonucléases peuvent convertir ces polysomes en monosomes liés chacun à un court fragment d'ARNm protégé. Cette technique dite de «*toeprinting*» a d'abord été exploitée par Joan Steitz pour identifier et séquencer les sites d'initiation de la traduction de l'ARN du bactériophage procaryote R17 (Steitz, 1969). En raison des limites des méthodes de séquençage disponibles à cette époque, les expériences de *toeprinting* ont principalement utilisé des systèmes de traduction *ex vivo* avec une seule ou un très petit nombre d'espèces d'ARNm. Grâce à l'avènement des technologies modernes de séquençage haut débit et les méthodes associées d'amplification de fragments courts d'ARNm pour générer des banques d'ADNc, il est aujourd'hui possible de cartographier à l'échelle du génome les positions des ribosomes sur les ARNm des cellules. En outre, la détermination des positions des fragments protégés permet de connaître empiriquement l'identité des produits de traduction, entre le codon d'initiation et le stop des cadres de lecture. Cela a conduit à la découverte de nombreux produits protéiques nouveaux ou alternatifs. La répartition des empreintes des ribosomes peut aussi donner un aperçu des mécanismes du contrôle traductionnel, il peut être utilisé pour identifier les phénomènes des pauses de traduction, et enfin, il permet d'identifier les petits cadres de lecture (uORF) en amont ou même en aval de l'ORF principale. Enfin, les nouvelles adaptations de l'approche du profilage des ribosomes permettent de suivre la traduction médiée par des sous-groupes de ribosomes sur la base de leur emplacement physique dans la cellule ou de leurs partenaires d'interactions.

L'étude du traductome présente plusieurs avantages, et notamment de s'assurer de la réelle utilisation fonctionnelle des ARNm. Contrairement à l'étude du protéome, elle inclut la notion de dynamique de la traduction. Au niveau des séquences, le *ribosome profiling* permet d'identifier précisément la position des codons initiateurs, et donc permet l'identification des isoformes

protéiques. D'un point de vue mécanistique, il révèle la dynamique traductionnelle telle que la vitesse d'élongation ou la présence de sites de pauses traductionnelles.

Depuis son développement, cette technique a été utilisée pour comprendre deux phénomènes intéressants de la régulation traductionnelle : l'utilisation de codons d'initiation non canoniques ou non-AUG, et l'utilisation des uORF dans les régions 5' non traduites des ARNm (Brar and Weissman, 2015). Le *ribosome profiling* a aussi permis de caractériser finement les changements globaux de traduction qui se produisent en cas de stress cellulaire, comme au cours d'un choc thermique (Shalgi et al., 2013), d'un stress oxydatif (Gerashchenko et al., 2012) ou d'une infection virale (Noam Stern-Ginossar et al., 2012).

3.1. Qu'est-ce que le *ribosome profiling* ?

Le *ribosome profiling* est basé sur la méthode classique de l'empreinte des ribosomes *in vivo* (Steitz, 1969; Wolin and Walter, 1988), dans laquelle ces ARNm en cours de traduction sont traités avec des endonucléases pour détruire les régions qui ne sont pas protégées par les ribosomes. Il nécessite la collecte d'un échantillon de taille relativement importante par rapport au *RNA-seq*. Il nécessite aussi d'inhiber la traduction pour immobiliser les ribosomes et révéler leur position exacte sur les transcrits. Pour cette raison, deux variantes majeures de la technique ont été développées : l'une pour suivre le ribosome en cours d'élongation, l'autre pour identifier les sites d'initiation (**Figure 20**). Chaque approche présente des avantages. L'initiation de la traduction étant plus lente que l'élongation, du fait de l'existence de sites de blocage des ribosomes (pauses de traduction) qui biaisent l'information, le taux d'initiation de la traduction est un meilleur prédicteur du taux de traduction que ne l'est la mesure de la densité des ribosomes en cours d'élongation. En ce qui concerne la caractérisation des produits protéiques, il est également avantageux de bloquer les sites d'initiation car les données sur les emplacements des codons d'initiation peuvent être facilement interprétées pour prédire les différentes isoformes de protéines traduites à partir de différents codons d'initiation. Enfin, les informations sur l'initiation permettent la détection des sites internes d'initiation de la traduction (IRES). Cependant seules les informations sur les ribosomes en cours d'élongation peuvent être utilisées pour la détection des phénomènes tels que le décalage de cadre de lecture (*frameshifting*) ou la translecture des codons stop (**Figure 20**). Une étude récente a montré que ces deux approches se complètent et peuvent être très puissantes si elles sont utilisées en parallèle (N. Stern-Ginossar et al., 2012).

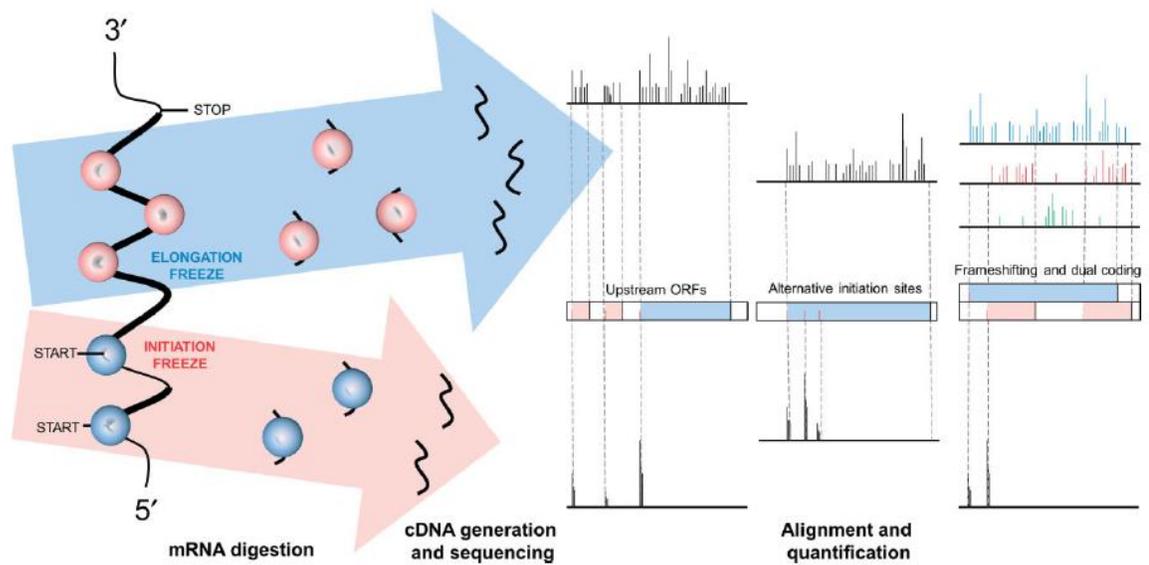


Figure 20 : Représentation schématique des deux principales stratégies de *ribosome profiling*. En bleu, l'immobilisation des ribosomes en cours d'élongation permet d'identifier les événements de décalage du cadre de lecture. En rose, l'immobilisation des ribosomes sur les sites d'initiation permet l'identification des uORF, des isoformes de protéines avec des régions N-ter alternatives (Michel and Baranov, 2013).

L'immobilisation des ribosomes en cours d'élongation peut être obtenue avec des drogues qui inhibent soit la translocation comme la cycloheximide et l'emetine, le transfert de peptidyle comme le chloramphenicol, ou encore par congélation rapide (Michel and Baranov, 2013). Tandis que l'immobilisation des ribosomes initiateurs peut être obtenue soit par le blocage direct des ribosomes avec des drogues spécifiques comme l'harringtonine et la lactimidomycine (Michel and Baranov, 2013), soit par l'enrichissement des ribosomes en cours d'élongation près des sites d'initiation en bloquant les ribosomes avec de la cycloheximide après prétraitement avec la puromycine qui provoque une terminaison prématurée (Fritsch et al., 2012).

Une fois les ribosomes en cours de traduction immobilisés, les ARNm sont traités avec des endonucléases pour détruire les régions qui ne sont pas protégées par les ribosomes (Ingolia et al., 2009). Un tel traitement laisse des "empreintes" de taille d'environ 30 nt, qui sont converties en une banque d'ADNc qui est soumise à un séquençage haut débit (**Figure 21a**).

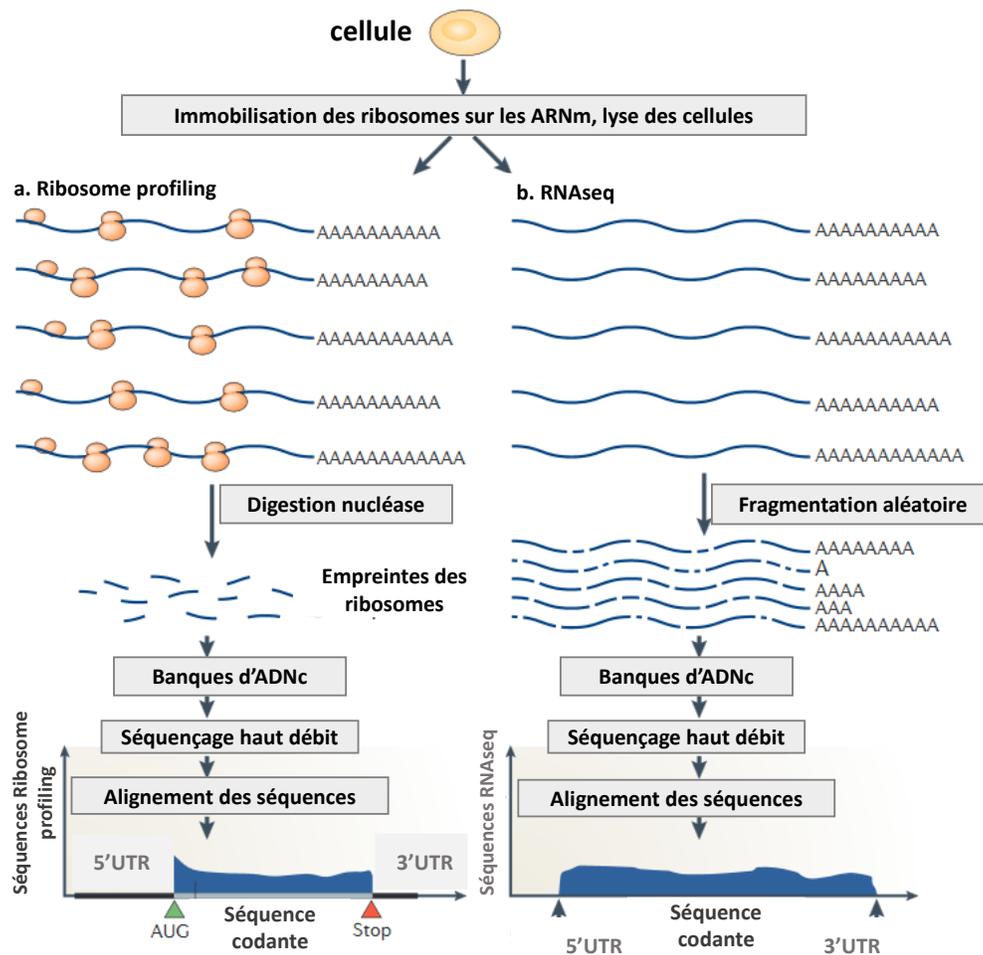


Figure 21 : Schéma comparatif du ribosome profiling et du RNAseq. *a. Ribosome profiling* : les ARNm sont digérés avec une endonucléase après immobilisation des ribosomes, il en résulte des petits fragments d'empreintes des ribosomes d'environ 30 nt qui sont purifiés et convertis en banques d'ADNc qui sont séquencés par la suite. Les séquences issues des empreintes des ribosomes s'alignent principalement au niveau des phases codantes des gènes. *b.* En parallèle au traductome on réalise le séquençage des ARNm ou RNA-seq, qui capture des fragments aléatoires recouvrant la totalité du transcrit (Brar and Weissman, 2015).

Les séquences issues des empreintes de ribosomes sont ensuite cartographiées au niveau du génome de référence approprié, pour définir l'emplacement exact des ribosomes en cours de traduction. Parallèlement au profil de ribosome est généralement effectué une banque d'ADNc pour mesurer l'abondance des ARNm par RNA-seq (**Figure 21b**). La quantification du signal issue des empreintes de ribosomes avec celui du RNA-seq, permettra par la suite la comparaison entre

les taux de synthèse protéique et l'abondance des ARNm ce qui permettra de déterminer l'efficacité de traduction pour chaque ARNm (Ingolia et al., 2009).

Le *ribosome profiling* permet donc de cartographier et de mesurer l'intégralité des empreintes de ribosome *in vivo* afin de quantifier la synthèse protéique et d'annoter les régions codantes (Brar et al., 2012; Ingolia et al., 2011a, 2009). Les progrès extraordinaires réalisés dans la technologie de séquençage permettent maintenant de mesurer l'activité de l'intégralité des ribosomes en cours de traduction dans des cellules de mammifères, qui codent pour environ 20.000 protéines différentes avec une région codante moyenne d'environ 500 nt (van Dijk et al., 2014). Ainsi, la digestion par les nucléases de toutes les empreintes des cellules donne autour de 10 millions d'empreintes possibles.

Grace aux propriétés biophysiques communes des ribosomes des différentes espèces et l'absence de manipulation génétique pour la préparation des échantillons, cette approche devient relativement facile et très adaptable aux cellules ou aux tissus de tout organisme, avec des modifications modestes. La traduction d'un large éventail d'organismes, bactéries, levures, parasites protozoaires, poissons zèbres, drosophile, nématodes, souris, rats, plantes, virus, jusqu'à la traduction mitochondriale dans les cellules humaines, a été étudiée par *ribosome profiling* (Andreev et al., 2015; Brar et al., 2012; Ingolia et al., 2011, 2009; Oh et al., 2011; Schafer et al., 2015) (Rooijers et al., 2013). De plus, une approche similaire a été utilisée dans le cas des chloroplastes dans les cellules végétales (Zoschke et al., 2013).

L'une des innovations les plus importantes de cette méthode a été de générer des fragments d'ARNm protégés à l'aide de la RNase I d'*Escherichia coli*. Cette RNase a la particularité d'avoir peu voire pas du tout de spécificité de séquence et d'hydrolyser l'ARNm de façon aléatoire, contrairement aux nucléases précédemment utilisées tel que la RNase A, la RNase T1 et la Nucléase de microcoque. Une telle précision d'hydrolyse par la RNase I, en particulier au niveau de l'extrémité 5', permet de déterminer les positions des ribosomes le long des phases codantes sur l'ARNm avec une précision au nucléotide près et permet aussi de retrouver une périodicité tri-nucléotide claire qui correspond à la signature des ribosomes (**Figure 22**). Cette périodicité permet de distinguer les empreintes résultant de la traduction (fragments d'ARN protégé par des ribosomes) par rapport aux empreintes de fragments d'ARN qui sont protégés par d'autres complexes, pour toutes autres raisons.

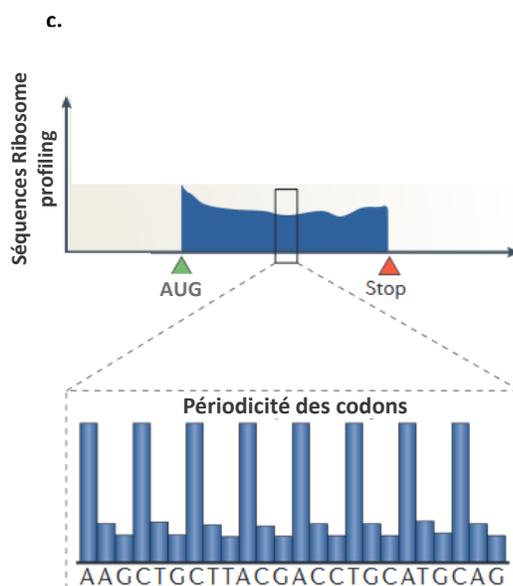


Figure 22 : Les séquences issues de ribosome profiling présentent une périodicité des codons due au mouvement trois à trois du ribosome (Brar and Weissman, 2015).

L'analyse de la répartition de la taille des empreintes est aussi un élément important du *ribosome profiling*. La répartition de la taille des fragments d'ARNm protégés par les ribosomes présente un pic symétrique avec une médiane sur 28-29 nt chez *Saccharomyces cerevisiae* ou 30-31 nt chez les cellules de mammifères. Ce changement reflète vraisemblablement la plus grande taille de la grande sous unité ribosomique 60S des mammifères par rapport aux levures. Cependant, des fragments protégés présentant des tailles d'environ 16 nt, 21 nt et entre 40 à 65 nt ont également été obtenus chez *S. cerevisiae* (Guydosh and Green, 2014). Les fragments d'environ 16 nt proviennent de ribosomes 80S bloqués sur l'extrémité 3' d'ARNm partiellement dégradés alors que les fragments de 40 à 65 nt ont été attribués à un empilement étroit des ribosomes 80S accumulés lorsqu'un ribosome est bloqué. Les fragments de 21 nt sont plus intrigants et semblent provenir d'une conformation alternative du ribosome 80S en cours de traduction, cette forme est stabilisée par l'anisomycine, tandis que le cycloheximide stabilise les espèces de ~ 30 nt.

3.2. Limites du *ribosome profiling*

Le *ribosome profiling* présente un certain nombre de biais qu'il ne faut pas négliger lors de l'analyse des résultats. Une des limitations majeures réside dans le fait que cette technique ne permet pas d'obtenir une mesure directe de l'efficacité de traduction mais uniquement de mesurer la densité de ribosomes sur un ARNm donné. En effet, la présence de ribosomes sur un ARNm n'est pas nécessairement synonyme de synthèse protéique active. De plus, dans le cas où les ribosomes seraient ralentis en phase d'élongation ou restent bloqués sur l'ARNm, leur densité pourrait croître sans pour autant conduire à une augmentation de la synthèse protéique. Il est donc important de valider les résultats de *ribosome profiling*, en utilisant des méthodes alternatives permettant de suivre les modifications de l'expression protéique telles que l'analyse des protéines par *western blot* ou l'utilisation de gènes rapporteurs.

Cependant, une approche bioinformatique, appelée RSS (*Ribosome release score*), basée sur l'hypothèse que, sur les transcrits réellement codants, le nombre de ribosomes associés en aval du codon stop décroît de façon significative par rapport à un ARN non-codants où l'association des ribosomes reste homogène le long de toute la séquence, permet de discriminer, parmi les empreintes de ribosomes, ceux qui proviennent d'ARNm codant de ceux qui proviennent d'ARN non-codants (Guttman et al., 2013).

3.3. Les biais du *ribosome profiling*

En général, chaque étape expérimentale, de la récolte cellulaire à la digestion par la nucléase, à la génération des banques d'ADNc, peut introduire des biais au niveau des résultats. Ainsi, la traduction et l'occupation des ribosomes sur un ARNm peuvent changer en quelques secondes après un stress, alors que la synthèse d'un nouvel ARNm nécessite plusieurs minutes (Andreev et al., 2015; Gerashchenko et al., 2012). De plus, historiquement, les polysomes sont stabilisés en traitant les cellules avec des inhibiteurs d'élongation de la traduction tels que la cycloheximide peu avant la lyse. Le principal défi technique du *ribosome profiling* a été d'inhiber rapidement la traduction pour figer instantanément des ribosomes en cours de traduction. La fiabilité de cette étape est particulièrement importante pour l'analyse. En effet, l'utilisation d'un inhibiteur de l'élongation de la traduction comme la cycloheximide peut modifier la distribution des ribosomes sur les régions

codantes en introduisant un biais non négligeable (Hussmann et al., 2015). L'utilisation de la cycloheximide aurait tendance à enrichir les empreintes de ribosomes situées à proximité du site d'initiation de la traduction ce qui ne serait pas forcément représentatif de la situation réelle dans la cellule avant l'ajout de la drogue (Guydosh and Green, 2014; Ingolia et al., 2012, 2009; Lee et al., 2012). Bien que cela ne semble pas interférer avec les mesures globales de la densité des ribosomes sur un ARNm, qui sont utilisés pour déterminer les taux de synthèse protéiques, l'utilisation de la cycloheximide peut provoquer des pics parasites d'accumulation de ribosomes sur des sites particuliers (Ingolia et al., 2011; Weinberg et al., 2016). D'où l'idée que la cycloheximide ne crée, ni ne détruit des empreintes de ribosome au milieu d'un cadre de lecture, mais il les redistribue (Hussmann et al., 2015). Pour éviter cette redistribution, des variantes du protocole n'utilisant pas d'inhibiteurs de la traduction (Lareau et al., 2014; Weinberg et al., 2016) ou au contraire utilisant un large excès d'inhibiteurs ont été développées (Gerashchenko and Gladyshev, 2014).

3.3.1 Le choix de la nucléase

Le choix de la nucléase utilisée pour obtenir les empreintes ribosomiques peut aussi introduire des biais significatifs dans la distribution des séquences protégées (Miettinen and Björklund, 2015). La plupart des nucléases utilisées dans les protocoles de *ribosome profiling*, telles que la RNase I, la RNase A, la nucléase micrococciale (MNase) et la RNase T1, agissent préférentiellement sur les ARN simple brin. Certaines nucléases ont une préférence pour digérer l'ARN en 3' ou en 5' de certains nucléotides comme la RNase A qui ne clive l'ARN qu'en 3' des nucléotides C et U. Par manque de spécificité de séquence, la RNase I fournit une empreinte robuste dans de nombreux systèmes eucaryotes. Cependant, la RNase I a tendance à enrichir les fragments protégés au niveau du site d'initiation de la traduction (Ingolia et al., 2012) et à induire une dégradation non spécifique des ribosomes, particulièrement dans les échantillons provenant de mammifères (Cenik et al., 2015). La MNase semble épargner l'ARNr mieux que la RNase I (Dunn et al., 2013) mais révèle la présence de sites protégés dans la région 3'UTR des transcrits. D'autres nucléases ont également été utilisées pour générer des empreintes ribosomiques dans les cellules humaines comme le cocktail RNase A-RNase T1 (Cenik et al., 2015). Il est donc évident que le choix de la nucléase introduit des biais non négligeables dans le profil des séquences d'empreintes ribosomiques. Ces biais,

associés à d'autres comme la présence de structures secondaires sur l'ARNm ou les biais dus à la préparation des banques d'ADNc conduisent à une distribution hétérogène des séquences obtenues par *ribosome profiling* le long des régions codantes. Cette hétérogénéité de distribution rend difficile voire impossible certaines analyses comme la détection de sites de pause des ribosomes lors de l'élongation (Irigoyen et al., 2016).

3.3.2. Les empreintes contaminantes

Le *ribosome profiling* est également sensible à la présence de contaminants qui peuvent créer des faux-positifs. Les fragments d'ARN ribosomique (ARNr) constituent 90% des contaminations. Ils résultent de l'étape de digestion des ARNm par les nucléases et peuvent avoir de lourdes conséquences sur les résultats en diminuant considérablement la profondeur de séquençage des empreintes ribosomiques (Ingolia et al., 2009), en particulier dans des conditions dans lesquelles les niveaux de traduction globale sont faibles. La déplétion des fragments d'ARNr contaminants est fortement recommandée et peut se faire de façon très efficace.

De plus, malgré la purification des ribosomes par ultracentrifugation sur gradient ou coussin de sucrose avant la récupération des fragments d'ARNm qui leurs sont associés, des complexes ribonucléiques de haut poids moléculaire peuvent co-sédimer avec les ribosomes et générer des séquences qui vont biaiser les résultats. Pour les éviter, il est possible de purifier les ribosomes par chromatographie d'affinité en utilisant des anticorps dirigés contre une protéine de la sous-unité ribosomique 60S (Ingolia et al., 2014), ou en étiquetant une des protéines de la sous-unité 60S (Sanz et al., 2009). De plus, une approche bioinformatique basée sur le filtrage des séquences en fonction de la taille attendue des fragments protégés par les ribosomes, appelée FLOSS pour *fragment length organization similarity score*, permet d'identifier de tels fragments et de les supprimer *in silico* (Ingolia et al., 2014). Cette approche permet d'améliorer la qualité des données en minimisant la présence de contaminants. La répartition typique des tailles d'empreintes de ribosomes utilisée dans l'analyse FLOSS est mesurée empiriquement pour chaque expérience en examinant la taille des empreintes ribosomiques de régions codant des protéines connues. Ceci permet d'identifier les fragments contaminants. Néanmoins, il existe des exemples dans lesquels les empreintes ribosomiques authentiques ne sont pas conformes au modèle de taille typique déterminé par FLOSS. Deux laboratoires ont analysé les empreintes de ribosome de taille

alternative et ont démontrés que ce n'était pas des fragments contaminants mais bien des fragments issus de la protection de ribosomes avec des conformations alternatives (Lareau et al., 2014) ou d'ARNm ayant des conformations particulières (Guydosh and Green, 2014).

Il est également possible de préparer des échantillons contrôles en utilisant des drogues permettant de d'empêcher l'association des ribosomes aux ARNm, afin de vérifier si le signal observé dans les échantillons d'intérêt provient réellement d'empreintes ribosomiques ou est dû à des contaminants. Pour ces expériences, la patéamine A est utilisée comme inhibiteur des étapes précoces de l'initiation de la traduction (Noam Stern-Ginossar et al., 2012).

3.3.3. Profondeur de séquençage

Même après déplétion des échantillons en ARNr les fragments obtenus, après la purification des empreintes ribosomiques, restent extrêmement contaminés en ARNr et ces derniers peuvent diminuer considérablement la profondeur de séquençage des empreintes ribosomiques. Parallèlement, dans le cas de séquençage haut débit tous les transcrits sont en compétition pour être séquencés. Ainsi, un transcrit abondant va être sur-représenté en termes de séquences comparé à un transcrit peu abondant qui sera plus faible.

La profondeur de séquençage est donc considérée comme l'étape limitante dans l'analyse des résultats de *ribosome profiling*. Ainsi, il est très important d'adapter le nombre de séquences au niveau d'expression des transcrits d'intérêt et aux questions biologiques posées. Par exemple, une étude qui s'appuie sur la phase de lecture, comme l'étude des uORF, nécessite une profondeur de séquençage plus importante qu'une analyse d'expression différentielle.

Objectifs de la thèse

Mon projet de thèse était l'étude de la régulation traductionnelle dans des cellules humaines présentant soit un défaut de terminaison de la traduction par déplétion du facteur eRF3, soit un défaut de la NMD par déplétion du facteur Upf1. Un point particulièrement important de ce projet était la régulation traductionnelle par les uORF.

Le point de départ de mon projet était une analyse, faite par notre équipe, de la régulation de l'expression des gènes dans des cellules humaines déplétées en facteur de terminaison de la traduction eRF3. Cette analyse par puces à ADN a révélé une augmentation des transcrits de l'activateur transcriptionnel ATF4 et de ces cibles par un mécanisme indépendant de la réponse aux stress (Ait-Ghezala *et al.*, 2012). L'ARNm d'ATF4 porte deux uORF régulatrices dans sa région 5'UTR. Les résultats de cette étude montrent que la déplétion en eRF3, en inhibant la terminaison de la traduction à la première uORF, favorise la traduction de l'ORF principale d'ATF4. Cette augmentation de la traduction de l'ORF principale d'ATF4 avait été mesurée par la technique du *Dual luciferase assay* à l'aide d'un gène rapporteur F-Luc (codant la luciférase de luciole, *Firefly luciferase*) placé en aval de la région 5'UTR d'ATF4 et d'un gène contrôle R-Luc (codant la *Renilla luciferase*). Sur la base de l'augmentation de l'activité *Firefly luciferase* par rapport à celle de la *Renilla luciferase* lors de la déplétion en eRF3, mais aussi de la grande conservation de séquence de la région 5'UTR d'ATF4 chez les mammifères et finalement de l'absence d'augmentation de la phosphorylation du facteur eIF2 témoignant d'une abondance normale en complexe ternaire eIF2-GTP-Met-tRNA_i^{Met}, il était suggéré que l'augmentation de traduction de l'ORF d'ATF4 était due à un mécanisme de translecture du codon stop de la première uORF d'ATF4 permettant la réinitiation non pas à l'AUG de la seconde uORF d'ATF4, mais à l'AUG de son ORF principale.

De plus, il était également montré que l'augmentation de l'expression d'ATF4 était en partie liée à une augmentation de la stabilité de son ARNm qui pouvait résulter d'un défaut de sa dégradation par la NMD, le codon stop de la première uORF d'ATF4 n'étant plus reconnu comme codon stop prématuré.

Dans le cas d'ATF4, ses deux uORF semblent donc agir non seulement en contrôlant la traduction de l'ORF principale mais aussi en modulant la stabilité des ARNm probablement en introduisant un codon stop prématuré activant, ainsi, la dégradation par la voie NMD. Cet exemple montrait clairement que ces deux mécanismes de régulation post-transcriptionnelle peuvent être associés.

L'analyse fine des résultats de cette étude montrait en outre, qu'en plus d'ATF4, plusieurs ARNm portant des uORF étaient aussi augmentés dans les cellules déplétées en eRF3. C'était le cas pour les ARNm des régulateurs transcriptionnels C/EBP β et IFRD1 qui étaient augmenté, respectivement, d'un facteur 1,98 et 2,21 fois.

Les questions posées par ces résultats étaient :

- Est-ce que toutes les uORF sont régulatrices ?
- Quels sont les caractéristiques des uORFs régulatrices ?
- Comment la terminaison de la traduction et la NMD interviennent dans cette régulation, et en particulier, dans le mécanisme de réinitiation de la traduction post-uORF ?

Seule une analyse de la traduction à l'échelle du génome pouvait permettre de tenter de répondre à ces questions. Après avoir confirmé l'augmentation de la traduction des gènes ATF4, IFRD1 et CEBP β par *Western blot* dans les cellules déplétées en Upf1 comme en eRF3, j'ai entrepris de réaliser une étude globale de la traduction par *ribosome profiling* dans des conditions de déplétion en eRF3 et en Upf1.

Résultats

Les travaux que j'ai menés au cours de ma thèse peuvent être divisés en deux parties. Mon premier projet a porté sur l'étude de l'effet des facteurs eRF3 et Upf1 sur l'expression des gènes portant des uORF. Le deuxième projet a été d'étudier l'effet du facteur Upf1 sur l'assemblage du ribosome. Les résultats des expériences de ces deux projets menés en parallèle sont présentés ci-dessous.

I. Analyse globale de la traduction des cellules déplétées en eRF3 ou Upf1

Je commencerai, dans ce chapitre, par présenter les expériences de mise au point de la technique du *ribosome profiling* utilisée pour la réalisation de l'étude du traductome et du transcriptome des cellules présentant un défaut de terminaison de la traduction par déplétion en facteur de terminaison de la traduction eRF3, ou des cellules présentant un défaut de dégradation des transcrits portant des codons stop prématurés NMD par déplétion en Upf1. Je décrirai ensuite les stratégies que nous avons utilisées pour l'analyse bio-informatique des résultats des expériences de *ribosome profiling* et de *RNA-seq* menées en parallèle. Je présenterai enfin les résultats obtenus grâce à ces expériences, d'une part, sous la forme d'un résumé plus ou moins détaillé puis ensuite sous la forme d'un article qui sera soumis prochainement pour publication.

1. Stratégie d'étude

Comme je l'ai décrit dans l'introduction, la méthode du *ribosome profiling* permet de suivre précisément la traduction de l'ensemble des ARNm cellulaires dans une condition donnée à un instant donné. Elle consiste à cartographier les sites de liaison des ribosomes sur les ARNm en cours de traduction. Cela est possible grâce au fait que le ribosome protège l'ARNm auquel il est lié sur une zone d'environ 28 à 30 nt, 12 nucléotides en 5' du codon situé au site P du ribosome et environ 15 nucléotides en 3' de codon. Le séquençage des fragments obtenus après digestion des polysomes à la RNase permet ainsi de déterminer au nucléotide près la position de chaque ribosome associé à un ARNm et, pour un ARNm donné, la quantité de ribosomes qui sont en train de le traduire. En couplant cette approche à des expériences de *RNA-seq* réalisées sur les mêmes

échantillons de cellules, il est possible de déterminer l'efficacité de traduction pour l'ensemble du transcriptome et de cartographier les régions codantes.

Cependant, l'efficacité de traduction est un paramètre qui ne dépend pas seulement du nombre de ribosomes associés à l'ARNm mais aussi de la vitesse d'élongation des ribosomes. En effet, lorsque l'élongation est bloquée, la mesure de la densité de ribosomes associés à une région codante donne une mesure indirecte de l'efficacité de traduction de chaque transcrit. Néanmoins, l'empreinte obtenue pour chaque ribosome, permet de retrouver la phase de lecture de ces ribosomes sur une région codante et de détecter les phénomènes de translecture des codons stop, particulièrement importants pour notre analyse. Pour cette raison, nous avons décidé de figer la traduction à l'étape de l'élongation par la cycloheximide.

2. Mise au point des expériences de *ribosome profiling* et de *RNA-seq*

La déplétion en eRF3a et Upf1 des cellules humaines HCT116 p53⁺ (cette lignée avait été utilisée dans l'étude précédente en raison de la stabilité de son cycle cellulaire) a été réalisée par la méthode d'interférence à ARN en utilisant des plasmides exprimant des sh-ARN dirigés contre les ARNm d'eRF3a (isoforme d'eRF3 essentielle à la terminaison de la traduction dans les cellules humaines) ou d'Upf1. La déplétion en eRF3 ayant déjà été mise au point dans le laboratoire, j'ai commencé par mettre au point les conditions permettant d'obtenir une déplétion efficace en Upf1 dans les cellules HCT 116. J'ai utilisé pour cela un mélange de différents plasmides exprimant des sh-ARN dirigés contre l'ARNm d'Upf1. Après électroporation des plasmides dans des cellules HCT116, j'ai analysé l'efficacité de la déplétion en Upf1 par Western blot à différents temps après électroporation. J'ai pu définir par Western blot qu'au temps de 72 h après électroporation, on avait le meilleur taux d'extinction d'Upf1. Sur la **Figure 23**, on observe que la déplétion d'eRF3 et d'Upf1 sont toutes les deux efficaces, le sh-eRF3a induit une diminution d'un facteur 5 de l'expression de la protéine eRF3a et le sh-Upf1 induit une diminution d'un facteur 3,5 de l'expression de la protéine Upf1.

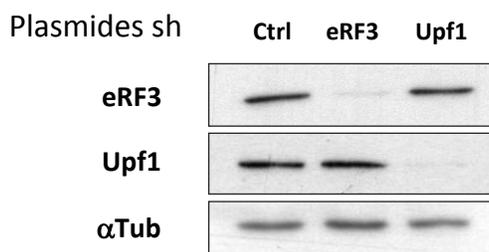


Figure 23: Analyse par Western blot de la déplétion en eRF3a et Upf1 dans les cellules HCT116. L'expression des protéines eRF3 et d'Upf1 72h après électrolocation des cellules avec les plasmides sh.

La première étape des expériences de *Ribosome profiling* a consisté à bloquer la traduction dans les cellules contrôles, déplétées en eRF3 ou en Upf1, par ajout de cycloheximide dans le but de figer les ribosomes en cours d'élongation sur les ARNm. Les cellules ont été ensuite lysées et la fraction cytoplasmique récupérée. Les lysats obtenus ont été ensuite séparés en deux fractions. Avant d'aller plus loin dans les manipulations une aliquote des cellules déplétées m'a permis de vérifier l'efficacité de déplétion en eRF3 et Upf1 de chaque expérience par Western blot.

La première a été utilisée pour extraire les ARN totaux et donc pour mesurer le niveau d'expression de chaque transcrite (transcriptome). Pour cela après extraction les ARN ont été fragmentés de façon aléatoire puis utilisés pour préparer des banques d'ADNc qui ont été soumises à un séquençage à haut débit (**Figure 24, RNA-Seq**). La deuxième fraction a été utilisée pour préparer les empreintes de ribosomes qui permettront de suivre la traduction de chaque ARNm. Pour cela, les lysats ont été incubés avec la RNase I qui hydrolyse les régions d'ARN simple brin accessibles à l'exception des régions qui sont associées aux protéines liant l'ARN et celles associées aux ribosomes. Les ribosomes ont été ensuite purifiés par centrifugation sur coussin de saccharose. Une fois les fragments d'ARNm protégés isolés, ils ont été séparés sur gel de polyacrylamide et la bande du gel qui correspond à 28-30 nt a été découpée. Les empreintes de ribosomes ont été ensuite purifiées et déplétées en fragments correspondant aux ARN ribosomiques (ARNr) contaminants grâce à un kit RiboZero (Illumina). Ces empreintes ont été ensuite utilisées pour préparer les banques d'ADNc qui ont été soumises à un séquençage à haut débit (**Figure 24, Ribo-seq**).

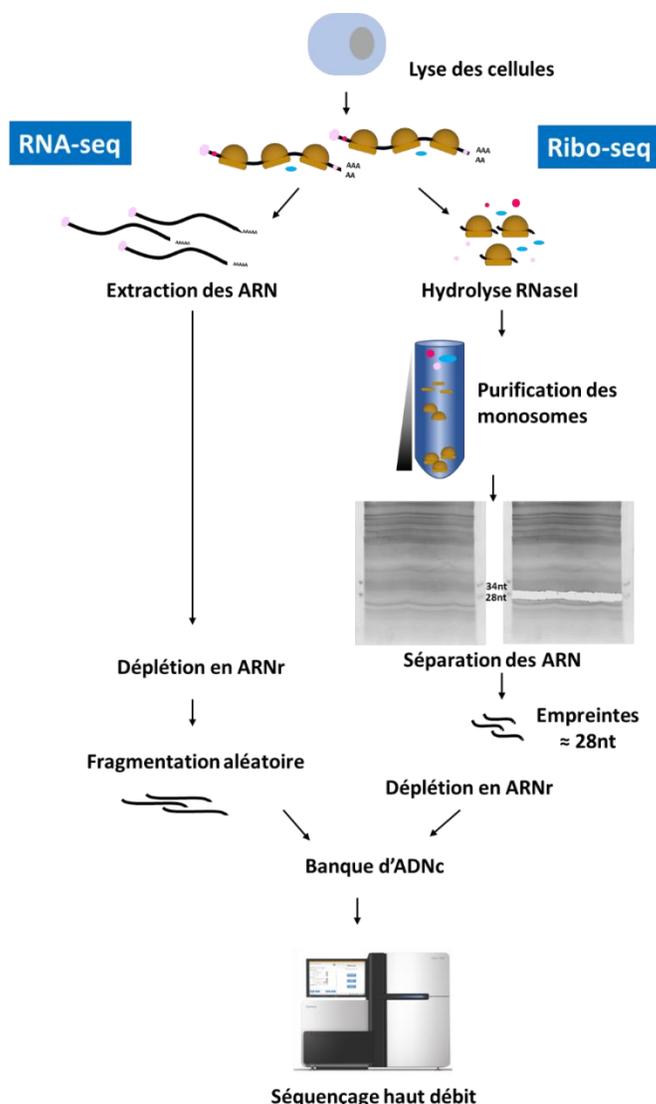


Figure 24 : Schéma général de la préparation des ARN pour le Ribo-seq et le RNA-seq. Après lyse des cellules traitées par la cycloheximide, les ARN totaux ont été extraits, fragmentés, transformés en banques d'ADNc qui ont été amplifiées et séquencées (RNA-seq, à gauche). Les mêmes extraits cellulaires ont été hydrolysés par la RNase I, puis les monosomes ont été purifiés par ultracentrifugation sur coussin de saccharose et les fragments d'ARN empreintes ont été extraits pour être séparés sur un gel de polyacrylamide. La bande d'ARN qui correspond à 28 nt a été découpée et les empreintes ribosomiques ont été transformées en banques d'ADNc qui ont été déplétées en fragments correspondant aux ARN ribosomiques contaminants, amplifiées et séquencées pour le Ribo-seq (à droite).

Des vérifications de la qualité de l'expérience ont été réalisées à chaque étape : analyse de la qualité et de la quantité des ARN obtenus pour le *RNA-seq* et le *Ribo-seq*, analyse de la longueur moyenne des empreintes après purification, analyse de la qualité des banques d'ADNc.

Les banques d'ADNc ont été réalisées par la plate-forme IMAGIF de l'Université Paris-Sud, le séquençage à haut débit a été fait par la plate-forme de séquençage de l'Institut de Biologie de l'ENS en utilisant la technologie Illumina.

3. Analyse bioinformatique du *RNA-seq* et du *Ribo-seq*

Les résultats du séquençage, *Ribo-seq* ainsi que les séquences correspondant au *RNA-Seq*, ont été soumis à une analyse bioinformatique dans le but d'identifier les ARNm et les quantifier. L'analyse bioinformatique des séquences brutes a été faite avec l'aide de Pierre Bertin, bioinformaticien dans l'équipe d'Olivier Namy à l'I2BC à Orsay.

Dans un premier temps, nous avons établi un pipeline (*workflow*) pour le *RNA-seq*. Il a permis d'éliminer la séquence de l'adaptateur en 3' utilisé lors des constructions des banques d'ADNc, d'éliminer les contaminations par les ARN ribosomiques restant et par les ARN non codants en alignant les séquences issues du premier filtre contre les séquences d'ARNr et d'ARN non codant disponibles dans les bases de données. Les séquences restantes ont été alignées sur le génome humain (version Hg38, *Genome Reference Consortium Human Build 38*).

Nous avons également établis un *workflow* pour le *Ribo-seq* (**Figure 25**), qui permettait d'éliminer la séquence de l'adaptateur en 3', de sélectionner les fragments de taille entre 24 et 35 nt, d'éliminer les contaminations par les ARNr et les ARN non codants et enfin d'aligner les séquences qui ont passés les différents filtres contre les séquences du génome humain Hg38, avec des paramètres d'alignement unique (un ribosome en cours de traduction ne peut se trouver que sur une seule position et ne peut donc protéger qu'un seul fragment d'ARN) et seulement deux mésappariements tolérés.

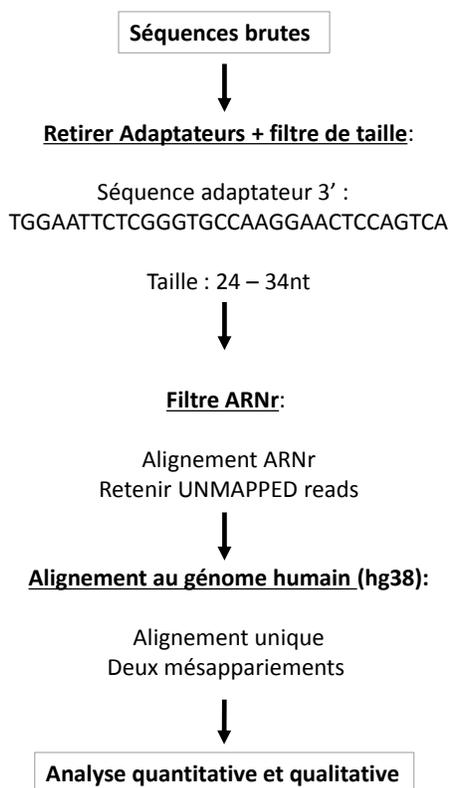


Figure 25 : Workflow d’analyse des séquences du *ribosome profiling*. Ce workflow permet d’éliminer les séquences des adaptateurs en 3’, de sélectionner les fragments de taille 24-34 nt, d’éliminer les ARNr et les ARN non codants et les ARN qui s’alignent à plusieurs endroits sur le génome humain.

Les séquences des fragments alignés ainsi obtenues à la fin de ce *workflow* ont été utilisées pour l’analyse qualitative et quantitative. L’analyse du *RNA-seq* a été utilisée comme référence pour l’interprétation des résultats du *Ribo-seq*. Elle nous a permis en particulier d’éliminer les variations dues à des modifications du taux de transcrits. On a pu ensuite mesurer la densité de ribosomes par molécule d’ARNm en calculant le ratio entre le nombre de séquences protégées par les ribosomes pour un ARNm donné et le nombre de séquences obtenues pour le même ARNm à partir du *RNA-seq*.

4. Analyse d'une première série d'expériences de *Ribosome profiling*

Pour être valides, les expériences de *ribosome profiling* doivent être réalisées sur plusieurs réplicats biologiques (2 au minimum). J'ai donc commencé par préparer les ARN pour l'analyse du traductome et du transcriptome pour deux expériences indépendantes sur des cellules contrôles, déplétées en eRF3 ou en Upf1 cultivées en parallèle. J'ai préparé les ARN protégés par les ribosomes pour le *Ribo-seq* et les ARN totaux pour le RNA-seq en collaboration avec l'équipe du Dr Olivier Namy à l'I2BC (Orsay) qui avait déjà mis au point ces techniques chez la levure.

Les 12 banques d'ADNc (*RNA-seq* et *Ribo-seq* des 2 réplicats dans trois conditions) ont été séquencées sur deux lignes de séquençage en multiplexant tous les échantillons d'un même réplicat biologique. Chaque ligne a été séquencée en quatre réplicats techniques et les séquences ont été "poolées" à la fin pour être analysées.

Il faut savoir que la technique de séquençage utilisée (Nextseq 500 Illumina) permet de générer de l'ordre de 4×10^8 séquences par ligne. Cela signifie qu'en multiplexant les 6 échantillons de *Ribo-seq* et de *RNA-seq* d'un même réplicat, on pouvait obtenir de l'ordre de 6×10^7 séquences brutes pour chaque échantillon, ce qui est largement suffisant pour avoir une bonne profondeur pour le transcriptome humain (environ 22 000 gènes).

Pour le *Ribo-seq*, les séquences brutes ont été ensuite filtrées selon le *workflow* décrit **Figure 25**. Les résultats de chaque étape sont représentés dans le **Tableau V**.

Pour le premier réplicat de *Ribo-seq* (noté -1), on a obtenu entre 7 et 10×10^7 séquences brutes pour chaque échantillon. Environ 40 % de ces séquences avait la bonne taille qui correspond à la taille de protection d'un ribosome (entre 25 et 34 nucléotides). La contamination par les ARNr et non codants représentait environ 20 % des séquences ayant la bonne taille. A la fin, les séquences qui se sont alignées à un seul endroit sur le génome humain ne représentaient plus que 4 % des séquences brutes pour les cellules déplétées en eRF3 ou en Upf1 et 11 % pour les cellules contrôles. En revanche, pour le deuxième réplicat (noté -2), la contamination par les ARNr et les ARN non codant a été plus élevée et les séquences uniques après alignement sur le génome humain étaient toutes du même ordre, entre 3 et 4%.

Tableau V : Récapitulatif des résultats du *Ribo-seq* des deux premières expériences.

	25<read<34			Alignement ARNr		Alignement hg38	
	Séquences brutes	Filtre taille	%	Séquences ARNr	% contamination	Séquences uniques	%
eRF3_1	73 000 000	24 000 000	32,8	14 000 000	19	3 000 000	4
eRF3_2	91 000 000	52 000 000	57	42 000 000	46	3 000 000	3,3
Ctrl_1	102 000 000	39 000 000	38,2	12 000 000	11,7	10 000 000	11
Ctrl_2	59 000 000	19 000 000	32,2	11 000 000	18,6	2 000 000	3,4
Upf1_1	73 000 000	25 000 000	34,2	16 000 000	22	3 000 000	4,1
Upf1_2	91 000 000	48 000 000	53,7	36 000 000	39,5	4 000 000	4,4

La contamination en ARNr représentait une proportion importante dans ces deux expériences. Elle est probablement due au fait que lors de la préparation des échantillons j’ai découpé une bande relativement large aux alentours du marqueur de taille 28 nt et que malgré la déplétion des ARNr en utilisant le kit RiboZéro, cela n’a pas suffi à éliminer une grande partie de ces contaminants.

Il faut se rappeler que la profondeur de séquençage est l’étape limitante dans l’analyse des résultats de *ribosome profiling*. Il est donc important d’adapter le nombre de séquences au niveau d’expression des transcrits d’intérêt et aux questions biologiques posées. Contrairement aux techniques d’analyses par *microarray*, dans le cas du séquençage à haut débit, tous les fragments sont en compétition pour être séquencés. Cela signifie qu’un fragment peu représenté va être moins séquencé qu’un fragment plus abondant. La contamination des échantillons en ARNr va donc diminuer considérablement la profondeur de séquençage des empreintes ribosomiques. C’est typiquement ce que l’on a observé pour ces deux expériences de *Ribo-seq*. La couverture des transcrits par les séquences obtenues n’était pas suffisante pour entreprendre une analyse

qualitative ou quantitative de ces expériences. Il m'a fallu donc refaire deux autres expériences de *Ribo-seq* en augmentant la profondeur de séquençage.

Pour le *RNA-seq*, après séquençage et analyse bioinformatique des séquences, nous avons obtenu le même ordre de grandeur de séquences brutes que pour le *Ribo-seq*, entre 7 et 10 x 10⁷ (**Tableau VI**). En comparaison avec le *Ribo-seq*, la contamination par les ARNr et l'ARN non-codant était relativement faible dans toutes les conditions et les deux réplicats. Ce résultat était tout à fait attendu parce que l'on a moins manipulé ces ARN lors de la préparation des ARN totaux. Au final, la proportion de séquences uniques qui s'alignent sur le génome humain est plus importante par rapport au *Ribo-seq* et représente environ 20 % des séquences brutes (**Tableau VI**).

Tableau VI : Récapitulatif des résultats du *RNA-seq* des deux premières expériences.

	Alignement ARNr			Alignement hg38	
	Séquences brutes	Séquences ARNr	% contamination	Séquences uniques	%
eRF3_1	81 000 000	3 000 000	3,7	17 000 000	21
eRF3_2	75 000 000	3 000 000	4	16 000 000	21,3
Ctrl_1	83 000 000	3 000 000	3,6	19 000 000	23
Ctrl_2	81 000 000	16 000 000	19,7	13 000 000	16
Upf1_1	91 000 000	2 000 000	2,2	17 000 000	18,7
Upf1_2	67 000 000	1 000 000	1,5	11 000 000	16,4

Une fois les séquences obtenues (dites *reads*), ces dernières sont comptabilisées et normalisées dans le but de pouvoir les comparer dans l'analyse différentielle. En effet, des biais comme le nombre total de *reads* alignés pour un échantillon et les artéfacts de PCR lors de la préparation des banques d'ADNc peuvent fausser les résultats. Pour y remédier, il faut diviser chaque échantillon par le nombre total de *reads* obtenus.

De plus, comme on séquence des petits fragments, il faut également diviser le nombre de séquences alignées sur un transcrit donné par la taille du transcrit. Ces ajustements nous ont permis d'obtenir une estimation normalisée du nombre de séquence ou RPKM, pour *Reads Per Kilobase and per Million (of reads)* et de tracer les graphiques de dispersion en comparant les comptages du premier réplikat biologique avec le deuxième. On observe que la dispersion des échantillons entre les deux expériences est linéaire avec des coefficients de corrélation de Pearson (R) supérieur à 0,995 pour le *RNA-seq* et moins bon, de 0,716 à 0,837 pour le *Ribo-seq* (**Figure 26**).

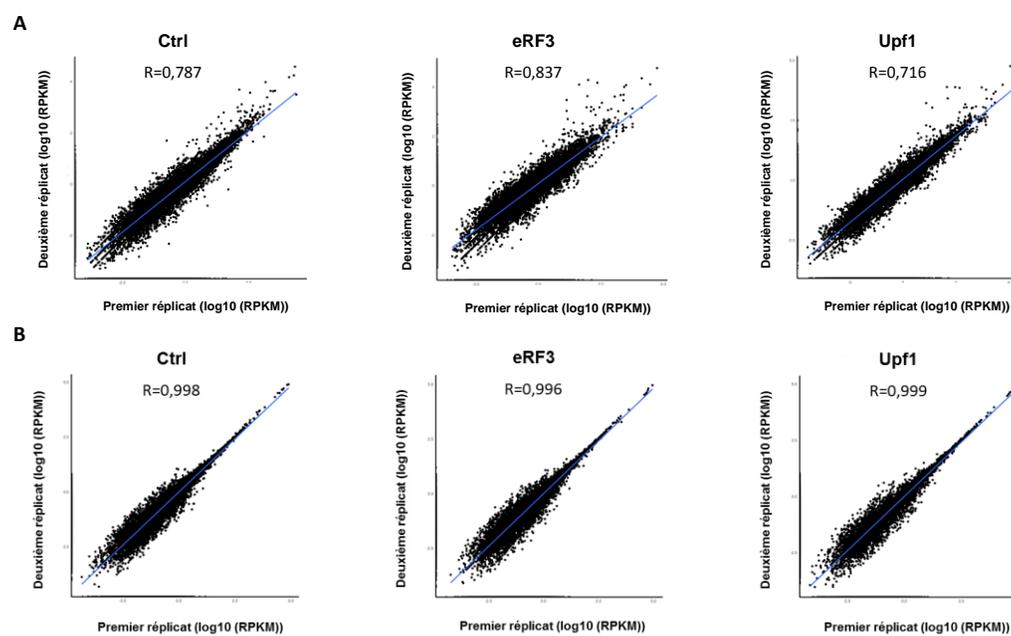


Figure 26 : Comparaison de la reproductibilité entre deux réplikat biologiques. A. Graphique de dispersion entre les deux échantillons *Ribo-seq*. B. Graphique de dispersion entre les deux échantillons de *RNA-seq*.

Ainsi, pour le *RNA-seq*, la reproductibilité entre les deux échantillons était très bonne et la profondeur de séquençage des deux premiers réplikat biologiques suffisait pour notre étude.

En revanche, la profondeur de séquençage des deux premières expériences de *Ribo-seq* était faible, du fait de leur contamination en ARNr. La solution de recommencer le séquençage des mêmes banques d'ADNc en augmentant la profondeur, c'est à dire en diminuant le multiplexage,

n'allait pas significativement arranger les résultats (on augmenterait aussi les séquences contaminantes).

5. Analyse d'une deuxième série d'expériences de *Ribosome profiling*

Nous avons décidé de faire deux autres expériences de *Ribosome profiling* dans les trois conditions, en diminuant d'une part la largeur de la bande de fragments d'ARN empreintes découpée sur le gel de polyacrylamide, et d'autre part, en augmentant la profondeur de séquençage de 6 fois par rapport aux deux premières expériences, c'est à dire en séquençant chaque échantillon sur une seule ligne. Les résultats du séquençage et du traitement bioinformatique des séquences sont représentés dans le **Tableau VII**.

Tableau VII : Récapitulatif des résultats du *RNA-seq* des deux deuxièmes expériences

Séquences brutes	425,000,000
Filtre de taille	295 000 000
%	69,5 %
Séquences ARNr	40 000 000
% Contamination	9,5 %
Séquences non alignées ARNr	255 000 000
Séquences multiples hg38	10 000 000
Séquences uniques hg38	120 000 000
%	28 %

Nous avons obtenu plus de 4×10^8 séquences brutes pour chaque échantillon et environ 70 % des séquences ont la taille de fragments de protection de ribosomes. La contamination par les ARNr et les ARN non codants représente moins de 10 %. Enfin, les séquences qui se sont alignées à un seul

endroit du génome humain représentent 28 % des séquences brutes soit $1,2 \times 10^8$ séquences. Cette couverture était largement suffisante pour entreprendre une analyse qualitative et quantitative de ces deux dernières expériences de *Ribosome profiling*.

6. Validation des résultats du *Ribosome profiling*

Pour avoir confiance dans les résultats du *Ribosome profiling*, il faut tout d'abord prouver que les séquences obtenues sont bien issues des empreintes ribosomiques et non pas de n'importe quel autre fragment d'ARN issu de la protection des ARN par un autre complexe ribonucléoprotéique d'une taille proche de celle du ribosome. Pour cela, deux contrôles sont indispensables : le premier consiste à vérifier la distribution de la taille des fragments d'empreintes (**Figure 27**), alors que, le deuxième consiste à vérifier leur périodicité (**Figure 28**).

En regardant la distribution des séquences. On observe que dans toutes les conditions, la distribution des fragments de tailles entre 25 et 34 nt présente un pic aux alentours de 28 nt. Ce qui signifie que la majorité des fragments obtenus ont une taille qui correspond à la taille de protection d'un ribosome.

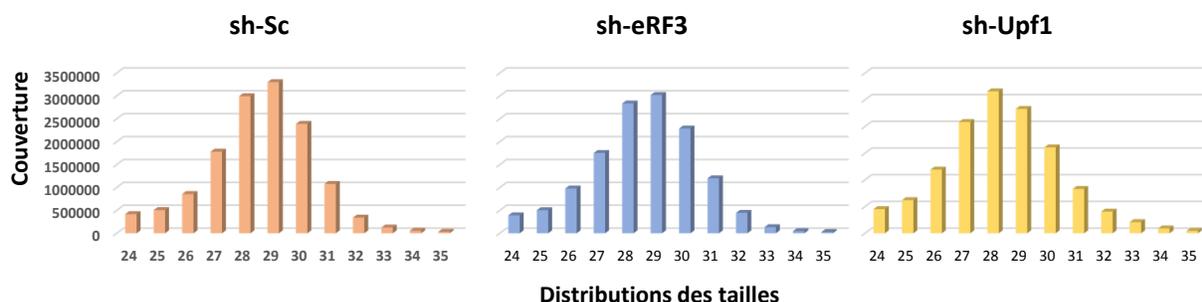


Figure 27 : Distribution des séquences de *Ribo-seq*. La distribution de taille des empreintes de ribosome présente un pic entre 28 et 29 nt.

De plus le ribosome, contrairement aux autres complexes ribonucléoprotéiques qui protègent l'ARN, est caractérisé par un mouvement périodique de 3 nucléotides qui correspond au déplacement codon par codon le long de l'ARNm. Cette signature du mouvement du ribosome

devrait être détectée dans nos séquences de *Ribo-seq*. Pour déterminer cette périodicité de 3 nt, on a fait une analyse métagénomique de ces séquences sur les ORF. Pour cela, les extrémités 5' des séquences d'empreinte ont été empilées et la couverture de chaque position pour chaque gène a été additionnée, en utilisant le codon d'initiation ou le codon stop comme point de référence (position 0).

De cette façon, les extrémités 5' des séquences devraient se retrouver à -12 nt en amont du site P du ribosome si le codon d'initiation est notre point de référence (**Figure 28A**) et à -15 nt en amont du site A du ribosome dans le cas où c'est le codon stop qui est utilisé comme point de référence (**Figure 28B**).

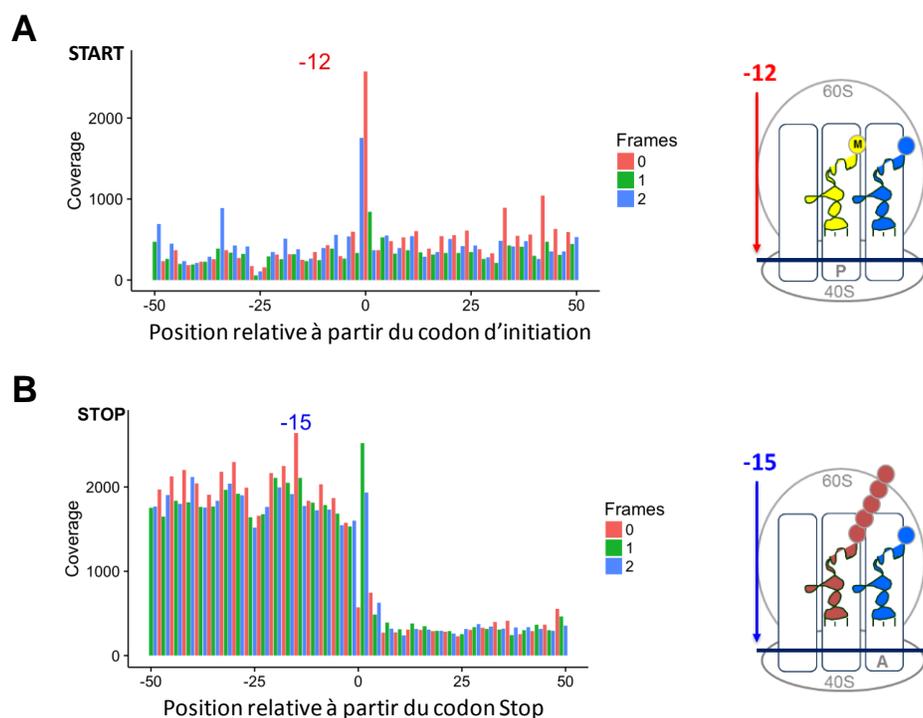


Figure 28 : Mise en évidence de la périodicité des séquences de *Ribo-seq*. L'extrémité 5' des reads se trouve à -12 et -15 nt des sites P et A du ribosome respectivement.

Dans ces analyses de métagène, les gènes avec moins de 10 séquences alignées dans la région codante ont été exclus. On observe aussi une variation de distribution de ces séquences le long de l'ORF, avec une augmentation au niveau du codon d'initiation et une décroissance du signal après

le codon stop. Ces caractéristiques sont observées pour nos expériences : une périodicité émerge bien que faiblement tous les 3 nt sur la même phase (phase rouge) en aval du point de référence pour l'initiation et en amont pour le codon de terminaison (**Figure 28**). La faiblesse de la périodicité du phasage des empreintes de *Ribosome profiling* est fréquemment observée dans le cas de génomes complexes comme le génome humain (Olivier Namy, communication personnelle). Cette périodicité est bien plus marquée pour des génomes plus petits comme celui des levures. On sait maintenant que cette faiblesse de périodicité est due à une digestion insuffisante des ARNm par la RNase I. Une digestion plus efficace par la RNase I améliore beaucoup le signal de phasage mais diminue considérablement la quantité des empreintes analysables. On remarque aussi dans nos expériences, l'augmentation du signal au site d'initiation et sa décroissance après le codon stop.

Il nous est donc possible de retrouver le cadre de lecture à l'intérieur des phases codantes en prenant comme information la position du premier nucléotide de chaque séquence. Les résultats obtenus montrent que nos séquences de *Ribo-seq* sont bien issues de la couverture de ribosome et nous permettent de conforter la fiabilité de nos résultats.

7. Détection des transcrits exprimés dans les cellules HCT116

En procédant à l'analyse des données, je me suis heurtée à un premier problème qui est d'identifier le "bon" transcrit pour chaque gène. En effet, plusieurs transcrits, qui proviennent de l'utilisation de différents promoteurs ou d'épissages alternatifs, sont décrits pour chaque gène dans les banques de données du génome humain. Cependant, tous les transcrits d'un gène ne sont pas présents dans le même type cellulaire. De plus, le nombre ainsi que la structure des uORF peut varier entre les différents transcrits d'un même gène. Il est donc primordial d'identifier les isoformes des transcrits effectivement exprimés dans les cellules HCT116 avant de procéder à l'analyse différentielle. Pour cela, nous avons dans un premier temps utilisé l'annotation APPRIS (*Annotation of principal and Alternative Splice Isoforms*) (Rodriguez *et al.*, 2012), qui est un algorithme basé sur des critères de conservation phylogénique, de structure et de fonctions, permettant d'identifier un seul transcrit comme isoforme principale pour chaque gène. Pour les gènes avec plus d'un transcrit principal, nous avons combiné cette algorithme avec une deuxième approche qui nous permet de sélectionner le transcrit le plus exprimé dans les cellules HCT116 en utilisant les comptages normalisés de *RNA-seq* dans les cellules contrôles. Par exemple, pour la sélection du transcrit du gène ASNS (**Figure**

29), en utilisant la couverture de *RNA-seq*, les séquences de *RNA-seq* s'alignent uniquement sur les exons. On remarque que seul le 4^{ème} transcrit sur la **Figure 29** porte tous les exons correspondant aux signaux de *RNA-seq*. C'est donc cette isoforme qui est exprimée dans les cellules HCT116 pour le gène ASNS.

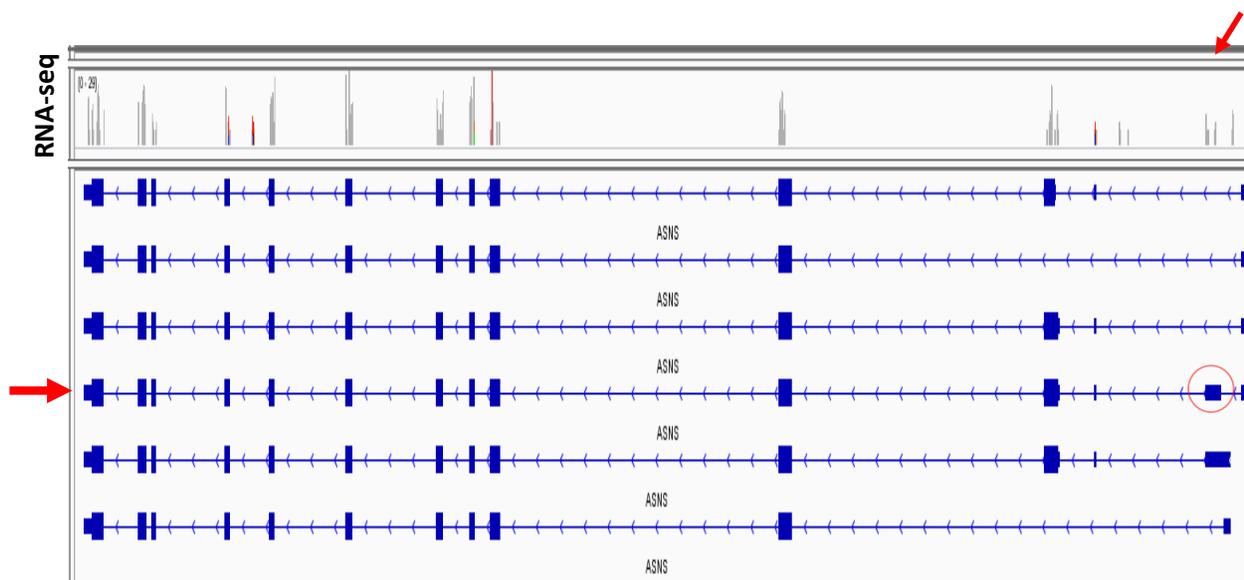


Figure 29 : Sélection des transcrits exprimés dans les cellules HCT116. *Le signal de RNA-seq qui s'aligne uniquement sur les exons nous permet d'identifier le transcrit exprimé dans les cellules HCT116 à partir des données sur les transcrits d'ASNS recueillies dans la base de données ENSEMBL.*

En combinant ces deux approches, non seulement, nous avons pu déterminer qu'il y a environ 13 000 transcrits qui sont exprimés dans les cellules HCT116 sur les 22 000 gènes du génome humain, mais aussi, nous avons pu identifier ces transcrits dans les banques de données.

8. Identification des uORF fonctionnelles dans les cellules HCT116

L'étape suivante a été l'identification des transcrits portant des uORF traduites, donc fonctionnelles, dans les cellules HCT116. Pour cela, nous avons réalisé un pipeline d'analyse qui permet dans un premier temps de détecter toutes les uORF potentielles présentes dans les régions

5'UTR des 13 000 transcrits détectés dans les cellules HCT116. Ensuite, en prenant les résultats du *Ribo-seq*, nous avons cherché tous les signaux de traduction dans ces régions 5'UTR.

Avec l'aide de Pierre Bertin, nous avons donc commencé par établir une liste d'uORF potentielles *in silico*, en prenant comme critère de base qu'une uORF est plus petite que l'ORF principale avec une taille minimale de 2 codons, et qu'elle est comprise entre un codon d'initiation et un codon stop en phase. Nous avons fixé certaines caractéristiques pour notre recherche des codons d'initiation : ce peut être soit un codon AUG classique entouré d'un contexte non défini, soit un codon AUG entouré du contexte de Kozak (RCCaugG), soit encore un codon d'initiation non-AUG dans un contexte de Kozak favorable à l'initiation. Dans ce dernier cas tous les codons sont possibles, seul le contexte de Kozak permet de définir le codon d'initiation. Nous avons fait ensuite un contrôle visuel de toutes les ORF détectées par le pipeline informatique et supprimé toutes les uORF dont le codon d'initiation non-AUG ne contenait pas au moins deux nucléotides du codon AUG. Il est à noter que nous en avons trouvé très peu. Cette analyse *in silico* nous a permis de trouver une liste de 8827 transcrits portant des uORF ce qui représentent 40 % des 22 000 gènes annotés pour le génome humain. Cette proportion est tout à fait cohérente avec ce qui avait été proposé dans la littérature (Iacono et al., 2005; Timothy G. Johnstone et al., 2016; Raj et al., 2016).

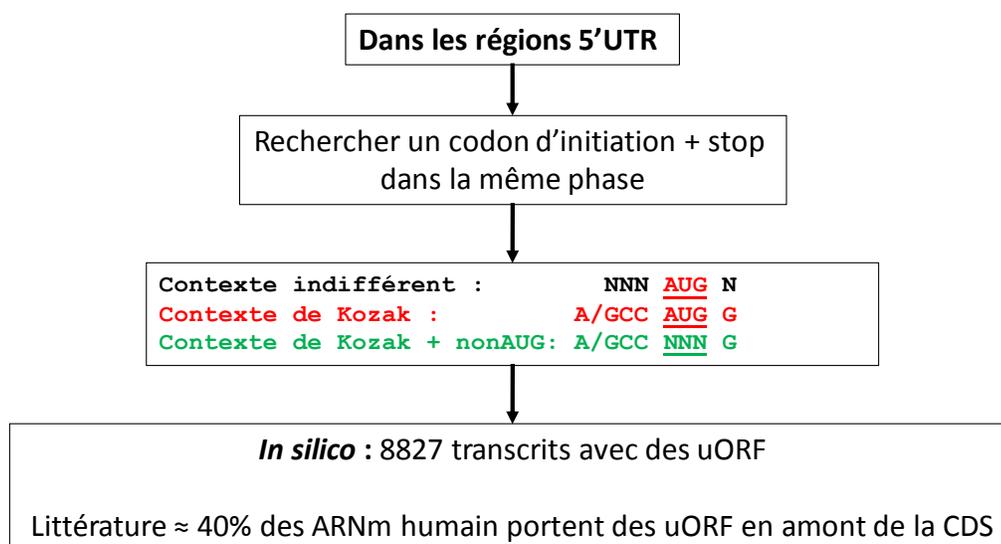


Figure 30 : Pipeline de détection des uORF *in silico*. Il permet de détecter les uORF présentes dans la région 5'UTR des transcrits exprimés dans les cellules HCT116.

Pour identifier les transcrits portant des uORF **réellement** traduites dans les cellules HCT116, nous avons mis en place un autre pipeline d'analyse permettant de rechercher dans mes données de *Ribo-seq* les signaux de traduction dans la région 5'UTR des transcrits et de faire l'intersection de ces signaux avec les uORF retrouvées *in silico* (**Figure 31**).

J'ai ainsi identifié 2726 transcrits portant des uORF qui sont effectivement traduites dans les cellules HCT116, ce qui représente 21% des 13 000 gènes exprimés dans ces cellules. Cette analyse met en lumière la difficulté d'extraire les bonnes informations non seulement à partir des banques de données mais aussi à partir des différentes données bioinformatiques publiées dans la littérature.

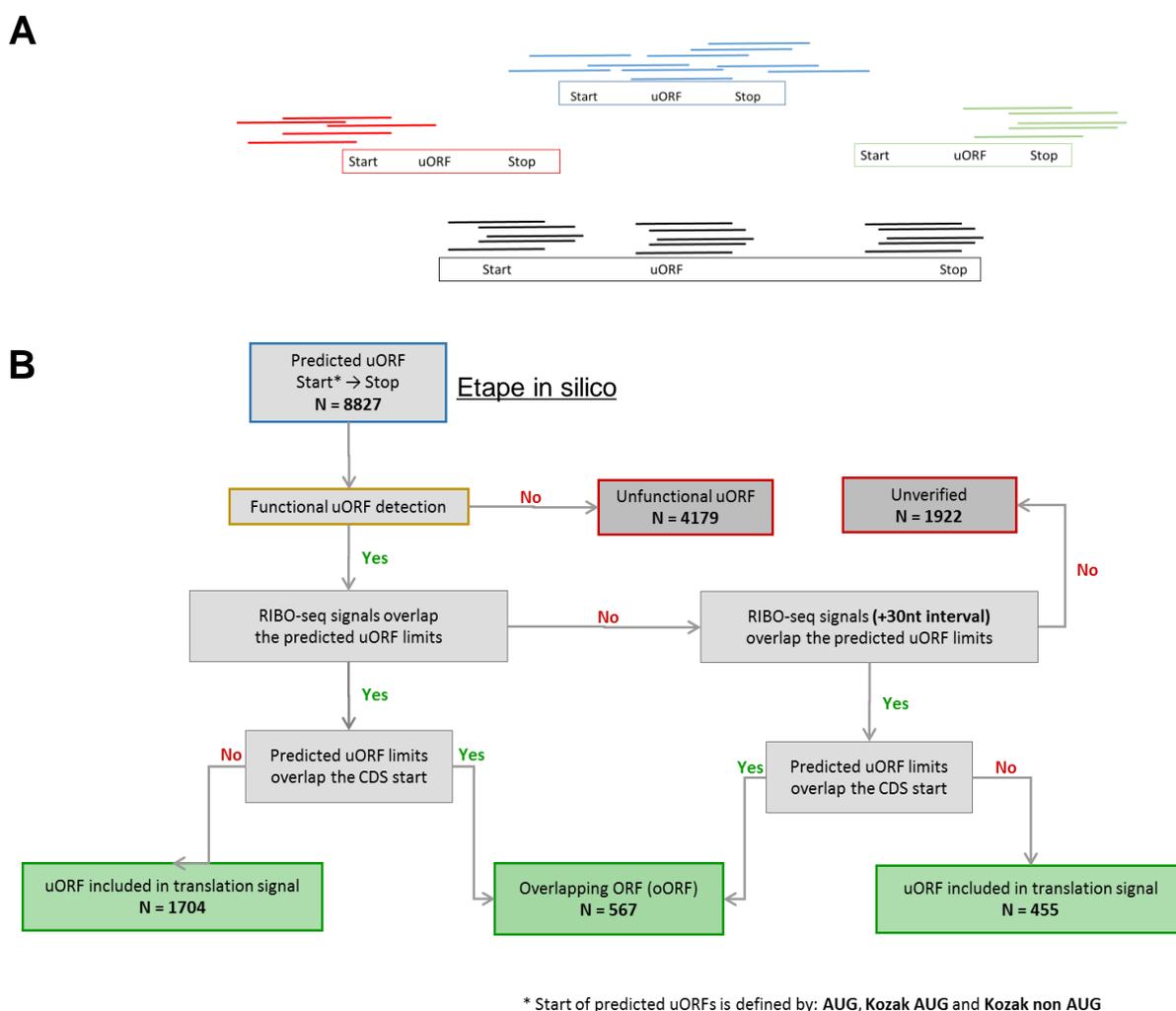


Figure 31 : Pipeline de détection des uORF fonctionnelles. *A. Différentes classes de reads sur les uORF traduites. B. Pipeline de détection des signaux de traduction présents dans la région 5'UTR. Ce pipeline permet de faire la corrélation avec les uORF trouvées in silico.*

Ce pipeline nous a permis de classer ces uORF en différentes classes : celles qui sont entièrement en amont de l'ORF principale, c'est le cas pour 2159 transcrits, et celles qui chevauchent l'ORF principale, c'est le cas de 567 transcrits. De plus, ce pipeline nous a permis d'identifier le statut des codons d'initiation des uORF traduites. Ces résultats montrent qu'une très faible proportion (3,6 %) de ces uORF ont un codon AUG dans un contexte favorable, dit contexte de Kozak. Il faut remarquer ici que les codons d'initiation AUG des deux uORF présentes dans l'ARNm d'ATF4, et qui servent souvent de modèle de régulation traductionnelle par les uORF, sont dans un contexte de Kozak qui est conservé chez les mammifères. Cependant, le résultat marquant de cette analyse des transcrits portant des uORF est que 9,4% des uORF traduites sont initiées par des codons non AUG (**Figure 32**), or d'après la littérature, l'initiation à des codons alternatifs est un événement rare chez les eucaryotes (Ivanov et al., 2011; Tikole and Sankararamakrishnan, 2006). Dans notre étude, le codon CUG constitue la plus grande proportion, soit 48 %, des uORF initiés à des codons non-AUG. Quant au codon AUU, il est le moins fréquent et constitue que 12 % des codons d'initiation non-AUG. Ces proportions sont proches de ce qui est décrits dans les autres articles de la littérature (Peabody, 1989; Raj et al., 2016).

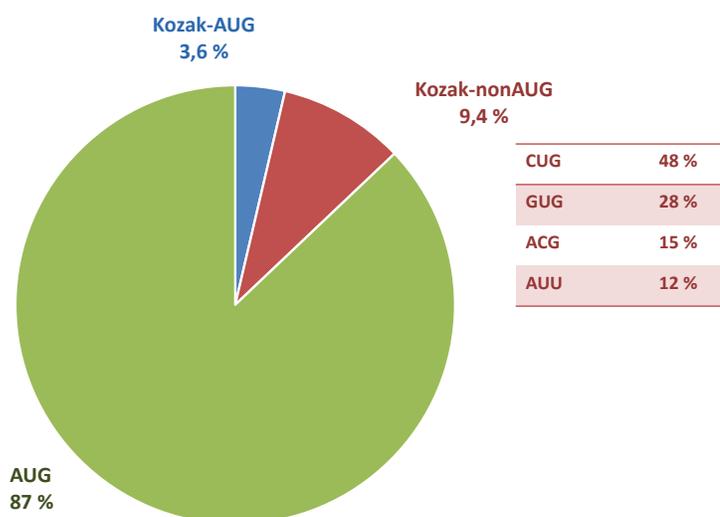


Figure 32 : Statut du codon d'initiation des uORF fonctionnelles. 90,6 % des uORF sont initiés par un codon AUG dont 3,6 % sont dans un contexte de Kozak et 9,4 % sont initiés par des codons non-AUG avec un contexte Kozak.

On s'est ensuite intéressé au contexte nucléotidique qui entoure les codons d'initiation de ces 2726 transcrits portant des uORF pour voir s'il y a un contexte particulier qui entoure les codons d'initiations de ces uORF. Pour cela on a récupéré les séquences qui entourent les codons d'initiation des uORF en prenant 13 nucléotides en amont et 11 nucléotides en aval des codons d'initiation des uORF. Les résultats de cette analyse montrent que pour les codons d'initiation AUG qui ne sont pas entouré du contexte de Kozak, aucun contexte particulier n'est retrouvé si ce n'est une particulière richesse en nucléotides G et C dans la séquence qui entoure ces codons (**Figure 33A**).

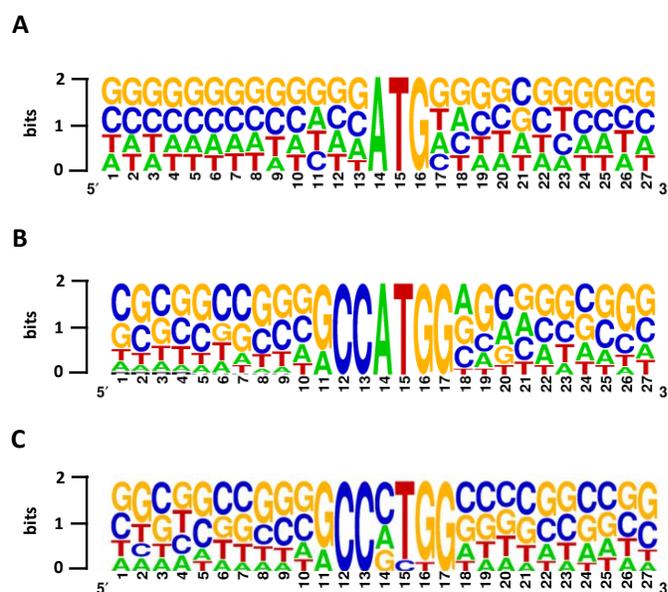


Figure 33 : Analyse des séquences entourant les codons d'initiation des uORF fonctionnelles.
A. Séquence entourant les codons d'initiation canoniques AUG. **B.** Séquence entourant les codons d'initiation AUG dans un contexte Kozak. **C.** Séquence entourant les codons non-AUG dans un contexte Kozak.

De même les codons AUG ou non-AUG dans un contexte de Kozak (**Figure 33 B et C**) présentent une certaine richesse en G et C dans la région qui les entoure sans avoir pour autant un motif consensus détectable. On peut rapprocher de cette observation, celle faite par M. Kozak sur

la particulière richesse en G et C des ARNm mammifères présentant de longues régions 5'UTR (Kozak, 1990).

9. Analyse de l'effet des facteurs eRF3 et Upf1 sur l'expression des gènes

Les analyses de l'expression différentielle des gènes dans les cellules HCT116 ont été réalisées à partir du transcriptome que nous avons établi et des *reads* du *RNA-seq* et du *Ribo-seq*. Les résultats de la comparaison entre les cellules déplétées en eRF3a par rapport au contrôle, pour le *RNA-seq* et pour le *Ribo-seq*, montrent que 2689 gènes ont une variation significative du niveau de l'ARNm (1574 sont augmentés et 1115 diminués) et 2649 gènes présentent une modification significative de la traduction (1450 sont augmentés et 1199 diminués). La plupart de ces gènes (2141) sont modifiés dans les deux analyses.

Pour les cellules déplétées en Upf1, nous avons identifiés 784 gènes qui ont une variation significative du niveau de l'ARNm (533 sont augmentés et 251 diminués) et 815 gènes qui présentent une modification significative de la traduction (502 sont augmentés et 313 diminués). Il est intéressant de noter,

- que les gènes cibles de eRF3 sont trois fois plus nombreuses que celles d'Upf1 que ce soit pour le transcriptome ou pour le traductome,

- que seulement environ 5% des gènes sont en même temps des cibles d'eRF3 et d'Upf1.

Ces résultats montrent que bien qu'ils fassent partie du même complexe SURF, les cibles d'eRF3 et d'Upf1 sont très différentes. Ce résultat est renforcé par l'analyse ontologique des cibles des deux facteurs qui montre que les cibles d'eRF3 sont impliquées dans le cycle cellulaire, les voies de signalisation; la réponse aux stress et la mort programmée des cellules alors qu'il nous a été impossible de mettre en évidence un enrichissement significatif des cibles d'Upf1 dans un processus cellulaire particulier. Ce dernier résultat suggère qu'Upf1 agit sur des gènes participant à une grande variété de processus cellulaire. Pour eRF3, sont lien avec le cycle cellulaire, les voies de signalisation comme la voie mTOR et la réponse aux stress avaient déjà été montré par notre équipe ainsi que par d'autres groupes (Ait Ghezala et al., 2012; Chauvin et al., 2007; Kikuchi et al., 1988).

Ainsi, eRF3 et Upf1 semblent avoir des fonctions très différentes aussi bien dans la régulation de la traduction que dans celles de la stabilité des ARNm. Les résultats décrits dans l'article qui suit

montrent ces différences, aussi bien au niveau de l'analyse différentielle de l'expression globale des gènes, qu'au niveau des gènes dont la traduction est régulée par des uORF.

Article 1 : En préparation

Involvement of translation termination factor eRF3 and nonsense-mediated mRNA decay factor Upf1 in the translational control of uORFs carrying mRNAs: a genome wide analysis

Affaf Aliouat^{1,2}, Pierre Bertin³, Isabelle Hatin³, Mohamed Doulazmi^{1,2}, Samia Salhi^{1,2}, Olivier Namy³ and Olivier Jean-Jean^{1,2*}

¹Sorbonne Universités, UPMC Univ Paris 06, UMR 8256, Biological Adaptation and Aging, Paris, France.

²CNRS, UMR 8256, Biological Adaptation and Aging, Paris, France.

³Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, Gif sur Yvette cedex , France.

Running Title: Translational control of uORFs carrying mRNAs

Keywords: eRF3, GSPT1, translation termination, Upf1, Nonsense-mediated mRNA decay, uORF, ribosome profiling

*Corresponding author: Tel: 33144272299; Fax: 33144278381 ; Email: olivier.jean-jean@upmc.fr

Involvement of translation termination factor eRF3 and nonsense-mediated mRNA decay factor Upf1 in the translational control of uORF carrying mRNAs: a genome wide analysis

Affaf Aliouat^{1,2}, Pierre Bertin³, Isabelle Hatin³, Mohamed Doulazmi^{1,2}, Samia Salhi^{1,2}, Olivier Namy³ and Olivier Jean-Jean^{1,2*}

¹Sorbonne Universités, UPMC Univ Paris 06, UMR 8256, Biological Adaptation and Aging, Paris, France.

²CNRS, UMR 8256, Biological Adaptation and Aging, Paris, France.

³Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, Gif sur Yvette cedex , France.

*Corresponding author: Tel: 33144272299; Fax: 33144278381 ; Email: olivier.jean-jean@upmc.fr

Running Title: Translational control of uORF carrying mRNAs

Keywords: eRF3, GSPT1, translation termination, Upf1, Nonsense-mediated mRNA decay, uORF, ribosome profiling

Abstract

In addition to its role in translation termination, eRF3 has been implicated in the nonsense-mediated mRNA decay (NMD) pathway through its interaction with Upf1. NMD is a RNA quality control mechanism, which detects and degrades aberrant mRNAs as well as some normal transcripts including those containing upstream open reading frames in their 5' leader sequence and those that harbor long 3' UTR. In this study, we evaluated translation by ribosome profiling and mRNA level by RNA-seq in human cells subjected to either eRF3a- or Upf1 depletion. Our bioinformatics analyses allow to delineate the features of the transcripts controlled by eRF3a and Upf1 and to compare the impact of each of these factors on gene expression. We also drew a transcriptome-wide map of uORFs and obtained a list of 2726 functional uORFs in our reference HCT116 transcriptome. Among those, only a small fraction contained common targets for both eRF3a and Upf1. We also detected ~10% of uORFs with non-AUG start codons. Our results provide strong support for the notion that different classes of transcripts bearing uORFs are regulated either by translational processes involving translation termination or by NMD.

Introduction

Termination is the final step of translation and occurs when a stop codon enters the A site of the elongating ribosome. Two release factors, eRF1 and eRF3, associated in a complex are required for this process. In a first step, eRF3 GTPase activity enhances stop codon recognition by eRF1 (Inge-Vechtsov *et al*, 2003; Salas-Marco & Bedwell, 2004). Then, eRF1 activates the peptidyl transferase center of the ribosome which triggers the release of the polypeptide. After GTP hydrolysis and dissociation of eRF3, the ATP binding cassette protein ABCE1 recruited to the post-termination complex, facilitates peptide release and ribosome recycling by splitting the 80S ribosomes into 60S and 40S subunits (Heuer *et al*, 2017; Kiosze-Becker *et al*, 2016; Preis *et al*, 2014). In mammals, two distinct proteins, eRF3a and eRF3b, are able to function as release factors. However, it has been shown that eRF3a is the main factor acting in translation termination and that the expression level of eRF3a controls the availability of an active termination complex by modulating eRF1 protein stability (Chauvin *et al*, 2005). The GSPT1 gene encoding eRF3a is expressed in a proliferation-dependent manner (Hoshino *et al*,

1989a), and eRF3a is subjected to proteasomal degradation when not associated with eRF1 (Chauvin & Jean-Jean, 2008). All these data suggest that eRF3a plays a key role in the modulation of translation termination efficiency.

In connection with its role in translation termination, eRF3 has also been implicated in mRNA decay through its interaction with the poly(A)-binding protein PABPC1 during canonical translation termination (Cosson *et al*, 2002), and in the nonsense-mediated mRNA decay (NMD) pathway through its interaction with Upf1 when termination takes place at premature stop codons (Ivanov *et al*, 2008). Nonsense-mediated mRNA decay is a RNA quality control mechanism, which detects and degrades aberrant mRNAs as well as some normal transcripts including those containing upstream open reading frames (uORFs) in their 5' leader sequence and those that harbor long 3' UTR (Mendell *et al*, 2000; Lejeune, 2017; Rahmani *et al*, 2009).

The NMD core complex consists of Upf1, Upf2, and Upf3 proteins interacting with each other, with translation termination factors and, in multicellular organisms, with the exon-junction complex (EJC) and SMG proteins. At premature stop codons, Upf1, an RNA helicase whose ATPase activity is essential for NMD, is recruited to the translation termination complex in an inactive form that must then be activated by the Upf2-Upf3 complex (Chamieh *et al*, 2008). Upf3 bound to EJC during splicing, associates with Upf2 in the cytoplasm. The presence of an EJC, 24 nt 5' of an exon-exon junction, allows Upf2 to bridge the EJC to Upf1 associated to the translation termination complex and stimulate its RNA helicase activity (Kervestin & Jacobson, 2012). These interactions promote Upf1 phosphorylation by SMG proteins and the recruitment of SMG6, an endonuclease that cleaves the mRNA in the vicinity of the premature termination codon. The resulting mRNA fragments are then targeted by mRNA decay enzymes.

It has been shown that Upf1 competes with PABPC1 for interacting with eRF3 (Singh *et al*, 2008) and that the distance between the terminating ribosome and the poly(A) tail-associated PABPC1 is a key determinant to triggers NMD. The expansion of this distance in mRNA with abnormal premature stop codons, uORFs, or long 3'UTRs, favors Upf1-eRF3 association and hence NMD (Mühlemann & Jensen, 2012). In addition to its role in NMD activation, recent results provide evidence that ATP hydrolysis by Upf1 is required for efficient translation termination and ribosome release at a premature termination codon (Serdar *et al*, 2016).

Using microarray screens, we have shown that, in human cells, translation termination disruption caused by eRF3a depletion induces upregulation of the transcriptional activator ATF4 and of its targeted genes (Ait Ghezala *et al*, 2012). We provided evidence that the effects on ATF4 expression promoted by eRF3a depletion was partly due to the modification of ATF4

mRNA translational control at uORFs and partly to a defect in NMD activation increasing ATF4 mRNA stability (Ait Ghezala *et al*, 2012). These microarray studies also revealed that in addition to ATF4, the expression of more than 10% of the genes with a detectable mRNA level, were significantly affected at the translational level by eRF3a depletion. The list of these transcripts includes a number of mRNAs carrying uORFs.

Through their physical association and their involvement in translation termination and NMD, eRF3 and Upf1 are regulating the protein and mRNA levels of a significant number of genes and contribute to the fine-tuning of their expression. Although several recent studies, including our microarray analysis in eRF3-deficient cells, drew up a comprehensive catalog of eRF3- and NMD-regulated transcripts in human cells (Ait Ghezala *et al*, 2012; Celik *et al*, 2017; Colombo *et al*, 2017; Malabat *et al*, 2015; Tani *et al*, 2012), it is not known yet at the genome scale in what extent both of these factors affect translational control and what is the subset of genes that are regulated either by only one of these factors or by both. To address these issues, we evaluated translation by ribosome profiling and mRNA level by RNA-seq in human cells subjected to either eRF3a- or Upf1 depletion. Our bioinformatics analyses allow to delineate the features of the transcripts controlled by eRF3a and Upf1 and to compare the impact of each of these factors on gene expression. We also performed a transcriptome-wide map of uORFs that yield new insights into the mechanism by which uORFs control translation. Indeed, our results provide strong support for the notion that different class of transcripts bearing uORFs are regulated either by translational processes involving translation termination or by NMD.

Results

Genome-wide analysis of the transcriptome and translome of eRF3a- and Upf1-depleted human cells

To explore how eRF3 and Upf1 functions in translational regulation, we performed a transcriptome-wide map of ribosome-protected mRNA sequences in the human HCT116 cultured cell line. For this purpose, we performed shRNA-mediated knockdowns for eRF3a and Upf1 mRNAs. A reference and control data set (Ctrl) was generated using a non-silencing shRNA sequence (Ait Ghezala *et al*, 2012). Three days after electroporation with the shRNA expressing plasmids, cells were blocked in elongation through cycloheximide treatment. In all knockdowns, Western blot analysis of electroporated cells showed efficient depletion of the

targeted eRF3a and Upf1 proteins (Figure 1A). Cell lysates were used to prepare ribosomal footprint libraries and mRNA libraries. In addition, we performed sucrose gradient fractionation of the lysates to ascertain the quality of the polysomes after cycloheximide treatment. The polysome profiles revealed an identical number of polysomal peaks in all three conditions. However, a reduced number of ribosomes per polysomal peak was observed for eRF3a- and Upf1-depleted cells, in comparison to control cells, providing direct evidence for the negative impact of eRF3a and Upf1 depletion on global mRNA translation (Figure 1B). Two biological replicates of ribosomal footprint and mRNA libraries were generated and subjected to deep sequencing to obtain Ribo-seq and RNA-seq data sets, respectively. Ribo-seq and RNA-seq reads were aligned against the human genome assembly hg38. After discarding all reads mapping to rRNA and all reads with multiple alignments, we obtained $\sim 1.8 \cdot 10^7$ reads per sample for the RNA-seq and $\sim 1.2 \cdot 10^8$ reads per sample for the Ribo-seq (see supplementary Table 1). Our Ribo-seq and RNA-seq data sets were highly reproducible across deep-sequencing replicates, with Pearson correlation coefficient (r) values > 0.9 (Supplementary Figure 1). Read data were therefore pooled over replicates for further analysis. For Ribo-seq data, ribosome-protected fragment-length assays generated a peak of 29-30 nucleotides for all samples as expected for mammalian ribosomal footprints (Figure 1C).

For most of the genes, annotations of human genome hg38 (Genome Reference Consortium Human Build 38) provide more than one transcript (199,234 transcripts for 20,310 coding genes in hg38 assembly) derived from multiple transcription start sites, alternative splicing and alternative polyadenylation sites. In addition, some of the annotated transcripts are poorly supported by primary data. However, for a given cell type, only a sub-set of these transcripts are expressed and constitute its transcriptome. Thus, to obtain a transcriptome-wide translational landscape of ribosome-protected mRNA fragments, we have first established a transcriptome map of HCT116 cells using the RNA-seq data sets. We first used the APPRIS annotation (Rodriguez *et al*, 2013) keeping only transcripts containing the "appris_principal" tags. This allowed us to eliminate the poorly-supported transcripts and select the most representative ones. Then, for genes with more than one transcript remaining, we selected the most expressed transcript using the normalized counts of transcripts in control HCT116 cells, divided by their length to obtain comparable numbers. Comparison of the resulting transcriptome of control cells with those of eRF3a- and UpF1-depleted cells showed only a very small number of differences in the set of selected transcripts. Using Ribo-seq data and visual inspection, these differences, mainly due to alternative splicing differences in the coding sequence, were standardized by choosing the transcript selected in control cell transcriptome.

In fine, we selected a HCT116 transcriptome comprising 12,996 transcripts, all with annotated protein-coding sequences.

Differential expression analysis of eRF3a and Upf1-depleted cells

Once the transcriptome established, the datasets of ribosome footprints were mapped to the transcriptomes of HCT116 cells. A first differential expression analysis of mRNA level (RNA-seq) and translational (Ribo-seq) changes was conducted using DESeq2 (Love et al. 2014) to compute log₂ fold changes (log₂FC) between two conditions, either eRF3-depleted (KD-eRF3a) versus Ctrl or Upf1-depleted (KD-Upf1) versus Ctrl. The adjusted p-value (p adj) threshold for significant changes was set to 0.05. For eRF3a-depleted cells, we identified 2,688 genes that had significant changes in mRNA level (1,574 up-regulated and 1,114 down-regulated) and 2,645 genes that had significant changes in translation (1,450 up-regulated and 1,195 down-regulated) (See supplementary data file S1). The vast majority of these genes (2,144) exhibits significant changes, at both, mRNA and translational levels (Figure 2A). For Upf1-depleted cells, we identified 784 genes with significant changes in mRNA level (533 up-regulated and 251 down-regulated) and 815 genes that had significant changes in translation (502 up-regulated and 313 down-regulated), 563 were significantly changed, at both, mRNA and translational levels (Figure 2B). Interestingly, at the mRNA level as well as at the translational level, we found a low overlap of only ~250 targets between both conditions, eRF3a or Upf1 depletion (Figure 2C). The fact that these two factors have non-overlapping sets of targets shows that they are involved in different cellular processes and may have radically separate functions in the regulation of translation and mRNA levels.

Thus, we next explored the gene-regulatory pathways that are affected by eRF3 or Upf1 depletion. Gene ontology and KEGG Pathway analyses of eRF3a targets showed particular enrichment for genes associated with cell cycle, signaling pathways, cellular response to stress, programmed cell death, and protein metabolic processes (Figure 3, and supplementary data file S2). This is agreement with the reported literature showing that eRF3 is involved in cell cycle progression (Hoshino *et al*, 1989b; Kikuchi *et al*, 1988), in mTOR signaling pathway (Chauvin *et al*, 2007a) and was recently shown to interfere with the cellular response to stress through its effect on ATF4 translational control (Ait Ghezala *et al*, 2012). Conversely, the pathway analysis of Upf1 targets did not show any highly significant enrichment for particular cellular processes. On the contrary, Upf1 targets are found in highly diverse biological pathways including regulation of transcription and translation, cell growth, cell proliferation, cell motility,

transmembrane transport, and cell survival. Several Upf1 targets are associated with diseases such as diabetes, neoplasms, hypertension, and neurodegenerative disorders.

To further explore how eRF3 and Upf1 function in translational regulation, we next compared genome-wide transcriptional (RNA-seq) and translational differences (Ribo-seq). This comparison was performed independently for mRNA coding sequences (CDS) and 5'UTRs in order to explore with more detail the effect of eRF3a and Upf1 depletion on uORF translation (Figure 4 A-D). The scatter plots (RNA-seq vs Ribo-seq) of eRF3a/Ctrl show that some genes displayed a high ($\log_2FC > 2$) increase in translation in their 5'UTR and in their CDS as well (Figure 4 A and B). In contrast, almost none of Upf1 differentially expressed targets exhibit such a high increase in translation (Figure 4 C and D) suggesting that the impact of Upf1 is much lower than that of eRF3a.

Detection of translated uORFs

We then decided to draw up a comprehensive map of uORFs in HCT116 cells with the aim to determine those that are regulated either by translation termination or by NMD. The translation initiation mechanism stringently select AUG codons for initiation, whereas non-AUG initiation sites are known to be rare and gene-specific. The ability of ribosomes to initiate at non-AUG codons was documented a long time ago (Anderson & Buzash-Pollert, 1985; Hann *et al*, 1988) and functional non-AUG initiation codons have been identified using phylogenetic analysis (Ivanov *et al*, 2011). However, the efficiency of non-AUG initiation in general was shown to be very low and measurements of translation initiation in mammalian cell indicates that, even in optimal nucleotide context, non-AUG initiation efficiency is at least 20 times lower than that of AUG (Kozak, 1989). Thus, the non-AUG codons encountered in the mRNA will be used for translation initiation, provided that it is surrounded by a suitable consensus sequence called the "Kozak context", i. e., RCCNNNG with R being a purine nucleotide and NNN the initiation codon (Kozak, 1986). Hence, different codons may interact with the translational machinery so as to lead to initiation: the canonical AUG initiation codon, in a Kozak context or not, and the non-AUG codons preferably in a Kozak context.

To compute the detection of uORFs with translation signals in HCT116 cells, we used a two steps strategy. We first predicted uORFs as a sequence in mRNA 5'UTRs starting either by a canonical AUG start or by the consensus Kozak context RCCNNNG and ending with an in frame stop codon. This led to a list of 8,827 transcripts with predicted uORFs. In parallel, for each transcript, we processed the Ribo-seq data to detect translation signals in the 5'UTRs, regardless of the sequence information. We then compared the Ribo-seq detected signals with

the "predicted uORFs" to determine truly translated uORFs, i. e., a "predicted uORF" containing Ribo-seq signals. This task was achieved using the algorithm presented in Figure 5 and homemade Python 3 scripts. However, assigning a codon as an initiation site is also dependent on the ribosome footprint coverage threshold that is used to distinguish starts from non-starts. Thus, we set the threshold of read coverage per nucleotide to 10. The overall result mapped 2,726 functional uORFs, among which 567 overlapping the CDS start signal were indentified. One can note that a high number (4,179) of transcripts with predicted uORFs were not selected through our algorithm. This is due either to the fact that these transcripts were not expressed in HCT116 cells, for the majority, or to the absence of ribosome footprint coverage in region of the predicted uORF. Moreover, among the signals detected in 5'UTRs by Ribo-seq, 1922 signals (noted "unverified" in Figure 5) could not be attributed to uORFs either through our algorithm or by visual inspection. Because we were unable to decide whether these signals corresponded to actual translation, they were excluded of our analysis.

Our analysis of functional uORFs allows us to evaluate the use of non-AUG start codons for uORFs translation. As shown in Figure 6, non-AUG codons in Kozak context that were actually used for translation initiation represented 9.4% of the initiation sites of functional uORFs whereas only a few translation initiation events have been detected at non-AUG codons for CDS (Touriol *et al*, 2003; Menschaert *et al*, 2013). Analysis of the non-AUG codon usage at identified uORFs revealed that only four codons, near-cognate to AUG (CUG, GUG, ACG, and AUU), are used, with an enrichment of the CUG codon (Figure 6A).

Comparison of the non-AUG starts with canonical AUG codons surrounded by a perfect Kozak context, showed that the latter are surprisingly low, representing only 3.6% of all starts. This low rate of strong initiator AUGs questions the regulatory role of the functional uORFs. Indeed, the mRNA of the transcriptional activator ATF4, which is the best studied example of translational regulation by uORFs (Harding *et al*, 2000; Lu *et al*, 2004; Vatterm & Wek, 2004), carries two uORFs with initiator AUGs in strong Kozak contexts that are conserved throughout mammalian evolution (Ait Ghezala *et al*, 2012). Whatever the start codon used, the analysis of the sequence extending 13 nucleotides upstream and 10 nucleotides downstream of the initiation codon failed to reveal clear consensus motifs but point out a particular G-C enrichment (Figure 6B).

When examining the uORF carrying mRNAs that are differentially expressed ($p \text{ adj.} > 0.05$) in the Ribo-seq data sets of eRF3a- and Upf1-depleted cells, we found that, among the 596 transcripts for eRF3a depletion and 186 for Upf1 depletion, only a small fraction (63) was common to both conditions (Figure 7 and supplementary data file S3) including ATF4 and

IFRD1 mRNAs. This result confirms that different transcripts are regulated at the translational level by either eRF3a or Upf1 and suggested that three classes of uORFs co-exist, the one regulating eRF3a targets, the second one acting on mRNAs targeted by Upf1 and the third one involved in the controls of both eRF3a- and Upf1-driven processes. It has been previously described that the expression of ATF4 and IFRD1 mRNAs is controlled by regulatory uORFs involving translation efficiency and by NMD (Ait Ghezala *et al*, 2012; Zhao *et al*, 2010). Thus, to validate our differential expression analysis of Ribo-seq data, we performed Western blot analysis of ATF4 and IFRD1 proteins in eRF3a- and Upf1-depleted cells. We also examined the expression of the short inhibitory form (LIP) of the transcriptional regulator C/EBP β whose translation is controlled by an uORF contained in the long activating forms (LAP and LAP*) coding sequence (Figure 8A). Note that this regulatory uORF escapes our bioinformatic analysis because our algorithm did not include uORF search downstream of the CDS start codon. The results presented in Figure 8B confirmed our differential analysis, showing an increase in ATF4 and IFRD1 translation for both knockdown conditions. Moreover, for C/EBP β translation, we observed an increase in the LIP form in eRF3a-depleted cells only (Figure 8B). This result was expected as the ribosomes translating the LAP and LAP* forms likely displaced the EJC that could have triggered NMD.

Discussion

In the present study, we established the first transcriptome-wide map of translation in human cells depleted either in the translation termination factor eRF3a or in Upf1, a key actor of the NMD pathway, using the of method Ribosome profiling (Guo *et al*, 2010; Ingolia *et al*, 2009). The RNA-seq and Ribo-seq data sets allow us to compare for both, eRF3a and Upf1 knockdown, the modifications of mRNA and translation levels for the 12,996 transcripts expressed in HCT116 cells.

We have previously shown using microarrays that translation termination has a regulatory impact on gene expression and that this effect is directly linked to its role in translation (Ait Ghezala *et al*, 2012). In this study, the differential analysis of eRF3a-depleted versus control cells allowed us to increase the accuracy of these results and reveals that ~80% of the genes that are regulated at the mRNA level, are also affected at the translational level. Interestingly, among the 60% of ~the up-regulated genes, we found a number of positive regulators of transcription including ATF4, ATF6, IFRD1, BTG1, EGR2, C/EBP β , C/EBP γ , and

transcription factors with a Forkhead box (Figure 4 and supplementary file S1). The translational changes of these transcription factors might cause important changes in the rate of transcription of their target genes, which could then indirectly affect their translational level. This combined transcriptional regulation and translational control of transcriptional regulators provides for versatility in regulating the gene expression. An example of this versatility is given by ATF4 for which this combination of transcriptional regulation and translational control allow to selectively repress or activate key regulatory genes that are essential for maintaining the balance between stress remediation and apoptosis (Dey *et al*, 2010). In addition, these transcriptional regulators are downstream targets of various signaling pathways, which integrate a number of kinases. This may explain the particular enrichment of the differentially expressed genes from eRF3a-depleted sample in genes involved in stress response, signaling pathways, and kinase binding (Figure 3 and Supplementary data file S2).

The results of the differential analysis of Upf1-depleted versus control cells give an overview of its impact on gene expression that is contrasted with that of eRF3a. First, a much lower number of genes (~800 for Upf1 versus ~2600 for eRF3a) exhibits significant changes in mRNA and translation levels, 55% only being modified for both. This result is close to that of Upf1 targets previously found in human cells using BRIC-seq, a method for directly measuring RNA stability (Tani *et al*, 2012). However, a very recent study, describe a higher number of human Upf1 targets (1717 with $\log_2FC > 0.5$) using a combination of knockdown and rescue of several NMD factors, including Upf1, SMG6 and SMG7 (Colombo *et al*, 2017). The second difference with eRF3a is that Upf1 targets are involved in very diverse biological pathways with no clear and strong enrichment for some particular cellular processes as also noticed by Tani *et al*. (2012).

The third important discrepancy between eRF3a and Upf1 depletion effect is that these two factors have non-overlapping sets of targets (only ~250 targets are common to both factors in the differential analyses). Of course, Upf1 and eRF3 are key actors of different mechanisms, i. e., NMD and translation termination, respectively. Nevertheless, physically, they are associated in the same translation termination SURF complex (Kashima *et al*, 2006), and functionally, NMD requires termination of translation to proceed (Lejeune, 2017). Moreover, a recent study has pointed out the role of Upf1 in the efficiency of translation termination and ribosome release at a premature termination codon (Serdar *et al*, 2016). One can argue that eRF3 is acting on genes that are sensitive to translational control whereas Upf1 is acting on genes whose expression is mostly controlled by mRNA decay. Indeed, these two processes are separated at the cellular level: the latter takes place at the first round of translation, just after the mRNA exit

from the nucleus, at the early stage of its cytoplasmic period which in contrast mainly concerns the former. The low impact of Upf1 on translational control could be reflected in the very low number of differentially expressed targets that display a high increase in translation ($\log_2FC > 1$) (Figure 4 C and D).

The goal of the study was also to establish a comprehensive map of the functional uORFs of human HCT116 cells and to bring new insights on how translation termination and NMD may influence uORF functions. Systematic searches for uORFs primarily relied upon *in silico* approaches, including protein prediction from transcript sequences and evolutionary conservation analysis. Indeed, these sequence-based analyses have identified uORFs in the transcriptomes of a variety of organisms and have shown that genes with a uORF represent 40 to 60% of the human genome (Iacono *et al*, 2005; Brar *et al*, 2012). But, as noticed by Wethmar *et al.*, compared to the number of predicted uORFs, remarkably few have been functionally studied (Wethmar, 2014). The recent ribosome footprinting studies have shown that a fraction of ribosome footprints lie outside of canonical protein-coding regions and appear to be derived from ribosomes translating the 5'UTR sequences (Bazzini *et al*, 2012; Brar *et al*, 2012; Ingolia *et al*, 2011; Lee *et al*, 2012). These footprints correspond to either uORFs or to alternative main ORFs translation initiation 5' of the conventional start sites (Sendoel *et al*, 2017). However, it is still unclear what fraction of uORFs undergoes translation, and, besides mRNA sequence elements, what are the factors that influence uORF function.

Using a computational analysis based on Ribo-seq data, we obtained a list of 2726 functional uORFs in our reference HCT116 transcriptome. Because non-AUG codons are poor initiator codons, we decided to retain only those surrounded by a perfect Kozak consensus sequence. Using this strategy, we found almost 10% of uORFs with non-AUG start codons. A high proportion of non-AUG initiation sites, representing 2/3 of all identified starts, was also described recently using puromycin treatment of cells to release elongating ribosomes and to enrich the ribosomal footprint library in translation initiation sites (Fritsch *et al*, 2012). However, in this study, this surprising high level of non canonical initiation starts correspond only to the newly identified initiation sites and not to the initiation sites of the entire transcriptome. In addition, a wide variety of non-AUG codons, near-cognate to AUG, were found including AGG, AUA, AAG and AUC codons whereas we did not find any of these codons in our analysis. The rules that govern translation initiation have been established experimentally more than two decades ago and were confirmed by numerous studies since then. These experiments clearly showed that the CUG codon is the best initiator codon and is also the most frequently encountered as non-AUG start for main coding sequences of genes (Touriol *et al*, 2003).

However, translation initiation at CUG codon requires a highly favorable initiation context (Kozak, 1989). This was recently confirmed in an genome-wide analysis of translation initiation sites showing that only 3% of CUG initiation sites identified in 5'UTRs produced detectable translation initiation signal (Michel *et al*, 2014). This suggests that a number of the putative initiation sites described in Fritsch *et al*. report (Fritsch *et al*, 2012) are not actually used for translation. Nevertheless, the mechanisms by which ribosomes select non-AUG codons for translation initiation, are largely unknown and it is possible that the translation apparatus uses non-AUG starts with higher frequency in some particular growth conditions, as recently reported for tumor initiation (Sendoel *et al*, 2017).

Another central finding of our differential expression analyses is that uORF carrying mRNAs are relatively rare, representing less than 25% of the transcripts that are differentially expressed in eRF3a- and Upf1-depleted cells. Among those, only a small fraction was a common target of both eRF3a and Upf1. This result reinforces the idea that the two different subsets of transcripts are regulated at the translational level by separate mechanisms. It is possible that regulation of eRF3a targets occurs via changes in the balance of initiation factors. It is well known that the efficiency of post-termination reinitiation relies on the abundance in ternary complex eIF2-GTP-tRNA_i^{Met} that is controlled by the phosphorylation status of eIF2 (Vattem & Wek, 2004). Another factor, eIF3, also plays a role in determining reinitiation efficiency after uORF translation (Roy *et al*, 2010). On the other hand, Upf1 target regulation could be sensitive to other types of factors such as SMG factors acting on Upf1 phosphorylation and triggering mRNA degradation (Lejeune, 2017).

Yet, the extent and function of translation within the 5'UTR remain largely unexplored. Thus, we believe that our comparative study on the translational effect of eRF3a and Upf1 depletion in human cells will provide important new informations to understand the processes regulating gene expression at the translational level.

Materials and methods

Plasmids and short interfering RNA

Plasmid expressing small interfering RNAs sh-3a1 targeting eRF3a mRNA have been previously described (Chauvin *et al*, 2005). Plasmids pSUPERpuro-hUpf1/I and pSUPERpuro-hUpf1/II targeting two different sequences in human Upf1 (Paillusson, 2005) and pSUPERcontrol with the non-silencing shRNA sequence (5'-ATTCTCCGAACGTGTCACG-3') used as negative control shRNA (sh-Ctrl) were a generous gift of O. Mühlemann.

Cell culture, electroporation and transfection

The HCT116 cell line (ATCC number : CCL-247) was maintained in McCoy medium (Invitrogen) supplemented with 10% fetal calf serum, 100 mg/ml streptomycin and 100 U/ml penicillin at 37°C under 5% CO₂ atmosphere. Electroporation of cells was performed with a Gene pulser II electroporation system (Bio-Rad) using 4.8×10^6 cells and 20 µg of plasmid DNA. eRF3 depletion was performed with sh-3a1 (Chauvin *et al*, 2007). Depletion of hUpf1 was induced by co-transfection of an equimolar mix of pSUPERpuro-hUpf1/I and pSUPERpuro-hUpf1/II plasmids (Paillusson, 2005). The electroporated cells were collected 72 h after electroporation.

Western blot analysis

Cell pellets were resuspended in 100 µl of a lysis buffer (50 mM TrisHCl pH 8, 120 mM NaCl, 1% NP40, 1 mM EDTA) containing, complete EDTA-free cocktail of protease inhibitors and phosphatase inhibitor cocktail (Roche) and 1 mg/ml pepstatine. Cells were lysed on ice, centrifuged for 20 min at 16.000g and supernatant retained as cell extracts. Protein concentrations of extracts were determined using MicroBCA Reagent (Sigma), bovine serum albumin being used as standard. For each sample, 30 µg of total protein were loaded on polyacrylamide gel and subjected to electrophoresis. Western blotting was then performed as described (Chauvin *et al*, 2007). Antibodies directed against human eRF3 were previously described (Chauvin *et al*, 2007). Mouse monoclonal antibodies against α -Tubulin were purchased from GE Healthcare. Antibodies directed against human IFRD1 and ATF4 were purchased from Proteintech. Antibodies directed against human Upf1 were purchased from Cell Signaling. Proteins were detected by chemiluminescence and exposure to X-ray film.

Polysomes preparation

Polysome fractionation was performed according to Verrier and Jean-Jean (2000) with minor modifications. Three days after electroporation, two 100-mm plates of HCT116 cells at 70% cell confluence were used for sucrose gradient centrifugation experiments. After 90 min incubation in fresh medium, cycloheximide (100 µg/ml final concentration) was added for 10 min. Cells were then washed with PBS and collected by trypsinization in the presence of cycloheximide, and pelleted. An aliquot fraction of the dry cell pellet was resuspended in 300 µl of lysis buffer (50 mM Tris-HCl, pH 7.4, 50 mM KCl, 10 mM Mg-acetate, 1 mM DTT, 0,1% Nonidet P40) containing 60 units of RNase inhibitor (Promega) and 100 µg/ml of

cycloheximide and lysed by vortexing on ice twice for 30 s. Nuclei and cell debris were removed by centrifugation at 3,000 x g for 10 min., and a 150 μ L aliquot fraction of supernatant was layered onto a 12-ml 15-50% (w/v) sucrose gradient. The gradient was centrifuged at 39,000 rpm in a SW41 Beckman rotor for 2.75 h at 4°C. Optical density at 254 nm was monitored by pumping the gradient through a Retriever 500 (Teledyne Isco) fraction collector.

Purification of ribosome-protected RNA fragments

Ribosome Profiling was performed as described previously (Baudin-Baillieu *et al*, 2016). Briefly, polysomes were prepared from two biological replicates of control, eRF3 and Upf1 depleted cells, and aliquoted at 60 OD260 units per tube and rapidly frozen and stored at -80°C. Monosomes were prepared by digesting polysome extracts for 1 hour at room temperature with 15 units RNaseI (Ambion) per OD unit. Digested polysomes were purified on a 24% sucrose cushion in sucrose polysomes extraction buffer (50 mM Tris-Acétate pH 7.6, 50 mM NH₄CL, 12 mM MgCl₂) and centrifuge for 2 h 15 min at 4°C, 100.000 rpm in a TLa110 rotor using OptimaMAX-XP Beckman centrifuge, to pellet the 80S monosomes. Each pellet was washed and resuspended in 750 μ l of sucrose polysomes extraction buffer by pipetting up and down. Ribosome protected fragments (RPFs) of mRNA were purified using acid phenol chloroform, and ethanol precipitation, the RNA was recovered in 500 μ l of TE (10 mM Tris-HCl pH 7) containing SUPERase-IN 0.1 U/ μ l (Ambion) and stored at -20°C. RNA fragments were loaded on 17% polyacrylamide gel containing 7 M urea and subjected to electrophoresis for about 6h at 150V with heating at 65°C. A 28 nt RNA oligonucleotide (oNTI 199 5'-AUGUACACGGAGUCGACCCGCAACGCGA-3') was used as a size marker to demarcate the ribosome footprint region. After migration, the gel was stained in 1/10.000 dilution of SYBER Gold (Invitrogen) for 30 min, RPFs were visualized with a UV lamp at 300 nm, and the 28 nt RNA fragments were excised. The excised gel pieces were loaded into a pierced 0.5 mL tubes inside a 2 mL tube, and centrifuged at full speed. Then, the RNA was eluted by soaking gel debris overnight at 4°C in an elution buffer (300 mM Sodium Acetate pH 5.5, 1 mM EDTA) and RNA was recovered by filtrating the eluate on a spin-X column (Costar), at full speed. RNA was then precipitated with glycogen ethanol overnight at -20°C.

Finally, RPFs were depleted of major rRNA contamination present within the 28 nt gel fragments by subtractive hybridization using biotinylated oligonucleotides (Table 1) and were recovered by reacting with MageneShere Paramagnetic streptavidin particles (Ribo-Zero Gold rRNA Removal Kit Illumina). The supernatants containing the ribosome footprints were

recovered and the RNA was precipitated with glycogen (20 mg/ml) in ethanol overnight at -20°C.

Purification of total RNA

Total RNA was purified from HCT116 cell extracts by hot acid phenol extraction. An equal volume of acid phenol was added to the extract that was vortexed for 1 h at 65°C, then centrifuged 10 min at full speed. Then, the RNA was extracted by adding an equal volume of chloroform to the first supernatant, vortexed for 5 min and centrifuged at full speed. The RNA was then precipitated overnight at -20°C with 0.3 M Sodium Acetate pH 5.2 and 3 volumes ethanol 100%. Centrifuged at full speed for 20 min at 4°C, RNA were solubilized in 500 µl TE (10 mM Tris-HCl pH 7, 1 mM EDTA) containing SUPERase-IN 0.1 U/µl (Ambion) and subjected to rRNA depletion, as described above.

Library construction and sequencing

cDNA library from 100 ng RNA was prepared by the Imagif platform (Institut de Biologie Intégrative de la Cellule, Gif sur Yvette, France), using the TruSeq Small RNA Sample Prep Kit with 3' sRNA Adapter (Illumina) according to the manufacturer's protocol. RNA integrity and quality was verified using Bioanalyzer Small RNA Analysis kit (Agilent). Sequencing was performed at the IBENS genomic platform (Institut de Biologie de l'École normale supérieure, Paris, France) on a Illumina Nextseq 500 and with a single reads 75 cycle run. A minimum of $7 \cdot 10^8$ reads were achieved for bioinformatics analysis.

Data analysis

Four sequencing libraries were prepared from purified RPFs and total RNA, for each biological replicates of eRF3- depleted, Upf1-depleted and control cells. The libraries were sequenced using illumina next-seq technology, reads from transcriptome (RNA-seq) and ribosome profiling (RFPs) libraries were trimmed using Cutadapt (Martin *et al*, 2011), to remove the 3' adapter of the reads «TGGAATTCTCGGGTGCCAAGGAACTCCAGTCA». Reads derived from rRNA were removed using Bowtie (Langmead *et al*, 2009), aligning to the rRNA reference downloaded from the Ensembl BIOMART database (Kinsella *et al*, 2011). The remaining non-ribosomal reads were then aligned to the human reference genome hg38 using TopHat2 (Kim *et al*, 2013), allowing 2 mismatches and keeping only uniquely mapped reads. At the end, we obtained $1.2 \cdot 10^8$ RPFs mapping principally to the CDS. These reads were used for subsequent analysis. In parallel, we obtained transcriptomic data, generating $8 \cdot 10^7$ reads

from total RNA. As for ribosome profiling, we retained only uniquely mapped reads excluding rRNA sequences. After discarding all reads mapping to rRNA and all reads with multiple alignments, we obtained $1.8 \cdot 10^7$ reads.

Differential gene expression

Raw counts were obtained using FeatureCounts (Liao *et al*, 2014) based on hg38 human annotations. The counts were realized on the exons (feature) and regrouped by transcripts (metafeature) keeping only stranded counts. Then, the counts were normalized using DESeq2 (Love *et al*, 2014) through the SARTools package using R 3.3.1 (Varet *et al*, 2016). RNA-seq and RIBO-seq counts and their normalizations were computed separately. P-value threshold for significant changes was set to $5 \cdot 10^{-2}$. Samples correlations were computed using the SERE statistics (Schulze *et al*, 2012).

Detection of principal transcript isoform

Since human genome annotations provide, for most of genes, more than one transcript, we attempted to select the most representative transcripts to compute this analysis. We first used the APPRIS annotation (Rodriguez *et al*, 2013) keeping only transcripts containing the “appris_principal” tags. Then, for genes with more than one transcript remaining, we selected the most expressed transcript. This selection was done using the normalized count of transcripts in control cell (Ctrl), dividing by their length to obtain a comparable number. Transcripts of a same gene were then compare between each others and the most covered (RIBO-seq counts) was selected.

Acknowledgements

A. A. was supported by a Contrat Doctoral from the Université Pierre et Marie Curie.

References

- Ait Ghezala H, Jolles B, Salhi S, Castrillo K, Carpentier W, Cagnard N, Bruhat A, Fournoux P & Jean-Jean O (2012) Translation termination efficiency modulates ATF4 response by regulating ATF4 mRNA translation at 5' short ORFs. *Nucleic Acids Res.* **40**: 9557–9570
- Anderson CW & Buzash-Pollert E (1985) Can ACG serve as an initiation codon for protein synthesis in eucaryotic cells? *Mol. Cell. Biol.* **5**: 3621–3624

- Baudin-Baillieu A, Hatin I, Legendre R & Namy O (2016) Translation Analysis at the Genome Scale by Ribosome Profiling. *Methods Mol. Biol. Clifton NJ* **1361**: 105–124
- Bazzini AA, Lee MT & Giraldez AJ (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**: 233–237
- Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT & Weissman JS (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**: 552–557
- Celik A, Baker R, He F & Jacobson A (2017) High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA N. Y. N* **23**: 735–748
- Chamieh H, Ballut L, Bonneau F & Le Hir H (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**: 85–93
- Chauvin C & Jean-Jean O (2008) Proteasomal degradation of human release factor eRF3a regulates translation termination complex formation. *RNA N. Y. N* **14**: 240–245
- Chauvin C, Salhi S, Le Goff C, Viranaicken W, Diop D & Jean-Jean O (2005) Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. *Mol. Cell. Biol.* **25**: 5801–5811
- Chauvin C, Salhi S & Jean-Jean O (2007) Human eukaryotic release factor 3a depletion causes cell cycle arrest at G1 phase through inhibition of the mTOR pathway. *Mol. Cell. Biol.* **27**: 5619–5629
- Colombo M, Karousis ED, Bourquin J, Bruggmann R & Mühlemann O (2017) Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA N. Y. N* **23**: 189–201
- Cosson B, Berkova N, Couturier A, Chabelskaya S, Philippe M & Zhouravleva G (2002) Poly(A)-binding protein and eRF3 are associated in vivo in human and *Xenopus* cells. *Biol. Cell* **94**: 205–216
- Dey S, Baird TD, Zhou D, Palam LR, Spandau DF & Wek RC (2010) Both transcriptional regulation and translational control of ATF4 are central to the integrated stress response. *J. Biol. Chem.* **285**: 33165–33174
- Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J & Brosch M (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* **22**: 2208–2218
- Guo H, Ingolia NT, Weissman JS & Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835–840
- Hann SR, King MW, Bentley DL, Anderson CW & Eisenman RN (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**: 185–195

- Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M & Ron D (2000) Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells. *Mol. Cell* **6**: 1099–1108
- Heuer A, Gerovac M, Schmidt C, Trowitzsch S, Preis A, Kötter P, Berninghausen O, Becker T, Beckmann R & Tampé R (2017) Structure of the 40S-ABCE1 post-splitting complex in ribosome recycling and translation initiation. *Nat. Struct. Mol. Biol.* **24**: 453–460
- Hoshino S, Miyazawa H, Enomoto T, Hanaoka F, Kikuchi Y, Kikuchi A & Ui M (1989a) A human homologue of the yeast GST1 gene codes for a GTP-binding protein and is expressed in a proliferation-dependent manner in mammalian cells. *EMBO J.* **8**: 3807–3814
- Hoshino S, Miyazawa H, Enomoto T, Hanaoka F, Kikuchi Y, Kikuchi A & Ui M (1989b) A human homologue of the yeast GST1 gene codes for a GTP-binding protein and is expressed in a proliferation-dependent manner in mammalian cells. *EMBO J.* **8**: 3807–3814
- Iacono M, Mignone F & Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* **349**: 97–105
- Inge-Vechtormov S, Zhouravleva G & Philippe M (2003) Eukaryotic release factors (eRFs) history. *Biol. Cell* **95**: 195–209
- Ingolia NT, Ghaemmaghami S, Newman JRS & Weissman JS (2009) Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**: 218–223
- Ingolia NT, Lareau LF & Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**: 789–802
- Ivanov IP, Firth AE, Michel AM, Atkins JF & Baranov PV (2011) Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* **39**: 4220–4234
- Ivanov PV, Gehring NH, Kunz JB, Hentze MW & Kulozik AE (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* **27**: 736–747
- Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G & Ohno S (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* **20**: 355–367
- Kervestin S & Jacobson A (2012) NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* **13**: 700–712
- Kikuchi Y, Shimatake H & Kikuchi A (1988) A yeast gene required for the G1-to-S transition encodes a protein containing an A-kinase target site and GTPase domain. *EMBO J.* **7**: 1175–1182

- Kim D-Y, Kim W, Lee K-H, Kim S-H, Lee H-R, Kim H-J, Jung Y, Choi J-H & Kim K-T (2013) hnRNP Q regulates translation of p53 in normal and stress conditions. *Cell Death Differ.* **20**: 226–234
- Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P & Flicek P (2011) Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database J. Biol. Databases Curation* **2011**: bar030
- Kiosze-Becker K, Ori A, Gerovac M, Heuer A, Nürenberg-Goloub E, Rashid UJ, Becker T, Beckmann R, Beck M & Tampé R (2016) Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. *Nat. Commun.* **7**: 13248
- 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 (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.* **9**: 5073–5080
- Langmead B, Trapnell C, Pop M & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25
- Lee S, Liu B, Lee S, Huang S-X, Shen B & Qian S-B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* **109**: E2424–E2432
- Lejeune F (2017) Nonsense-mediated mRNA decay at the crossroads of many cellular pathways. *BMB Rep.* **50**: 175–185
- Liao Y, Smyth GK & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**: 923–930
- Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550
- Lu PD, Harding HP & Ron D (2004) Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J. Cell Biol.* **167**: 27–33
- Malabat C, Feuerbach F, Ma L, Saveanu C & Jacquier A (2015) Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* **4**:
- Martin F, Barends S, Jaeger S, Schaeffer L, Prongidi-Fix L & Eriani G (2011) Cap-assisted internal initiation of translation of histone H4. *Mol. Cell* **41**: 197–209
- Mendell JT, Medghalchi SM, Lake RG, Noensie EN & Dietz HC (2000) Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol. Cell. Biol.* **20**: 8944–8957
- Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappe J, Gevaert K & Van Damme P (2013) Deep Proteome Coverage Based on Ribosome Profiling Aids Mass Spectrometry-based Protein and Peptide Discovery and Provides Evidence of

Alternative Translation Products and Near-cognate Translation Initiation Events. *Mol. Cell. Proteomics* **12**: 1780–1790

- Michel AM, Andreev DE & Baranov PV (2014) Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics* **15**: 380
- Mühlemann O & Jensen TH (2012) mRNP quality control goes regulatory. *Trends Genet. TIG* **28**: 70–77
- Paillusson A (2005) A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res.* **33**: e54–e54
- Preis A, Heuer A, Barrio-Garcia C, Hauser A, Eyler DE, Berninghausen O, Green R, Becker T & Beckmann R (2014) Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* **8**: 59–65
- Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S & Hanson J (2009) Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol.* **150**: 1356–1367
- Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink J-J, Lopez G, Valencia A & Tress ML (2013) APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* **41**: D110–117
- Roy B, Vaughn JN, Kim B-H, Zhou F, Gilchrist MA & Arnim AG Von (2010) The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. *RNA* **16**: 748–761
- Salas-Marco J & Bedwell DM (2004) GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol. Cell. Biol.* **24**: 7769–7778
- Schulze SK, Kanwar R, Gölzenleuchter M, Therneau TM & Beutler AS (2012) SERE: single-parameter quality control and sample comparison for RNA-Seq. *BMC Genomics* **13**: 524
- Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, Levorse J, Dill BD, Schramek D, Molina H, Weissman JS & Fuchs E (2017) Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**: 494–499
- Serdar LD, Whiteside DL & Baker KE (2016) ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nat. Commun.* **7**: 14021
- Singh G, Rebbapragada I & Lykke-Andersen J (2008) A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* **6**: e111
- Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y & Akimitsu N (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol.* **9**: 1370–1379

- Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats A-C & Vagner S (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell* **95**: 169–178
- Varet H, Brillet-Guéguen L, Coppée J-Y & Dillies M-A (2016) SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One* **11**: e0157022
- Vattem KM & Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11269–11274
- Wethmar K (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* **5**: 765–778
- Zhao C, Datta S, Mandal P, Xu S & Hamilton T (2010) Stress-sensitive Regulation of IFRD1 mRNA Decay Is Mediated by an Upstream Open Reading Frame. *J. Biol. Chem.* **285**: 8552–8562

Table 1 : Biotinylated oligos targeted against abundant human rRNA

>Oligo 1	5' CCG GCT ATC CGA GGC CAA CCG /3Bio/ 3'
>Oligo 2	5' CGA TCA GAA GGA CTT GGG CCC /3Bio/ 3'
>Oligo 3	5' GCG CTT GGC GCC AGA AGC GAG /3Bio/ 3'
>Oligo 4	5' GGC GCC GAG AGG CAA GGG GCG /3Bio/ 3'
>Oligo 5	5' CTA CAG CAC CCG GTA TTC CCA /3Bio/ 3'
>Oligo 6	5' TCA GAC AGG CGT AGC CCC GGG /3Bio/ 3'
>Oligo 7	5' AGT AGT GGT ATT TCA CCG GCG /3Bio/ 3'

Legend to figures

Figure 1. **A.** Western blot analysis of eRF3 and Upf1 depletion in control, eRF3 and Upf1 depleted cells, α -Tub serves as loading control. **B.** Polysome profiling of control, eRF3a and Upf1 depleted cells. **C.** Length distribution of ribosome footprint sequences after excluding nuclease-generated rRNA contaminations in control, eRF3 and Upf1 depleted cells.

Figure 2. Venn diagram of differentially expressed genes in eRF3 and Upf1 depleted cells.

Figure 3. Pathway analysis of eRF3a targets (padj. > 0.05). **A.** KEGG pathway and **B.** Gene ontology of biological processes.

Figure 4. Fold changes in gene expression for eRF3a-depleted vs control cells (KD-eRF3a vs Ctrl) or Upf1-depleted vs control cells (KD-Upf1 vs Ctrl). Biplots show log₂FC in Ribo-seq (y axis) and RNA-seq (x axis) levels, both in the 5'UTR and the CDS.

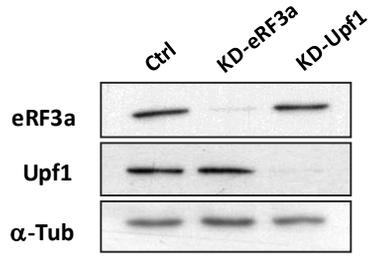
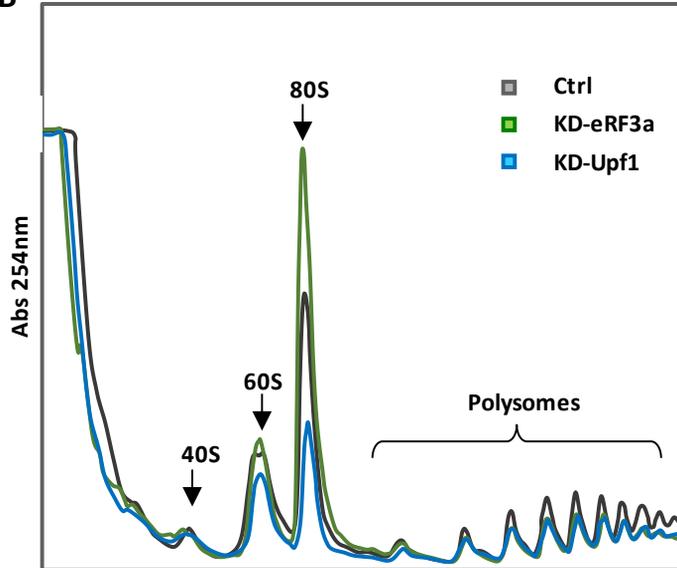
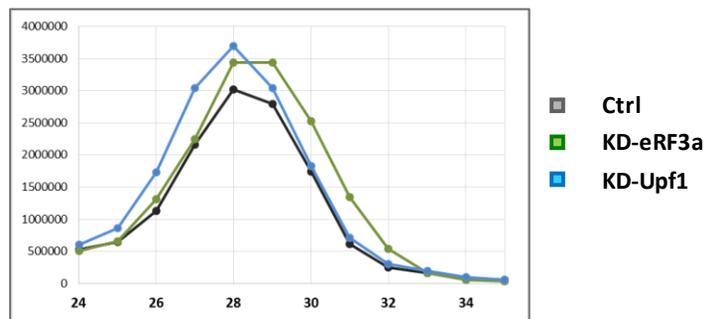
Figure 5. Algorithm used for the functional detection of either upstream ORF (uORF), or CDS-overlapping ORF (oORF). The respective annotations are as follows: AUG (without surrounding Kozak consensus sequence), Kozak AUG (AUG codon in a Kozak context), or Kozak non-AUG (non-AUG codon in a Kozak context). These codons were searched for in the 5' leader sequence of transcripts expressed in HCT116 cells.

Figure 6. Translation of regulatory uORFs. **A.** Codon distribution of upstream translation initiation sites (left). Relative frequency of non-AUG start codons of translated uORF (right). **B.** Weblogos depicting the sequence context of initiation codons (13 bases upstream and 11 bases downstream) of uORFs.

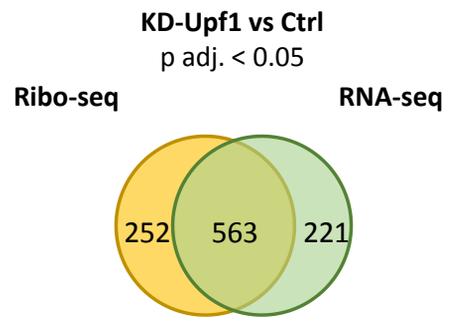
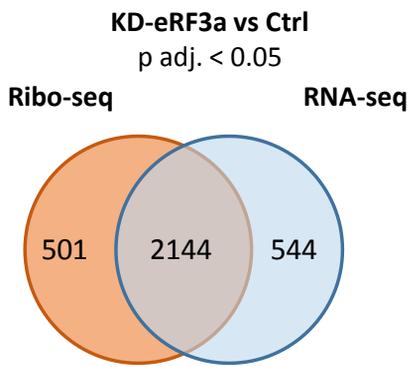
Figure 7. Venn diagram of differentially expressed transcripts carrying functional uORFs in eRF3 and Upf1 depleted cells.

Figure 8. Western blot analysis of ATF4, IFRD1 and C/EBP β in eRF3a- and Upf1-depleted cells. **A.** Schematic representation of the organisation of uORFs in ATF4, IFRD1 and C/EBP β mRNAs. **B.** Western blot analysis of ATF4 (left panels), IFRD1 (central panel), and C/EBP β (right panel) in control cells (Ctrl), eRF3a-depleted cells (KD-eRF3a) and Upf1-depleted cells (KD-Upf1). α -Tub served as a loading control.

Supplementary Figure 1. Statistical analysis of duplicates. Ribosome or RNA densities of two independent replicates were plotted for comparison. Biplot comparing RPKM (reads per million per kilobase of transcript) on a logarithmic scale between two biological replicates of Ribo-seq (up) and RNA-seq (down) experiments carried out on independent libraries sequenced fourth.

A**B****C****Figure 1**

A



B

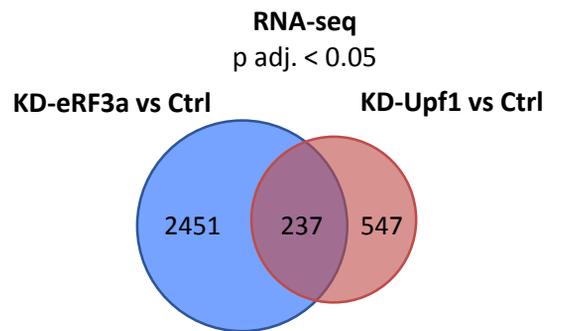
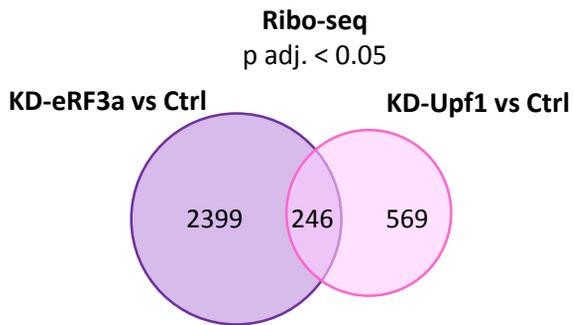
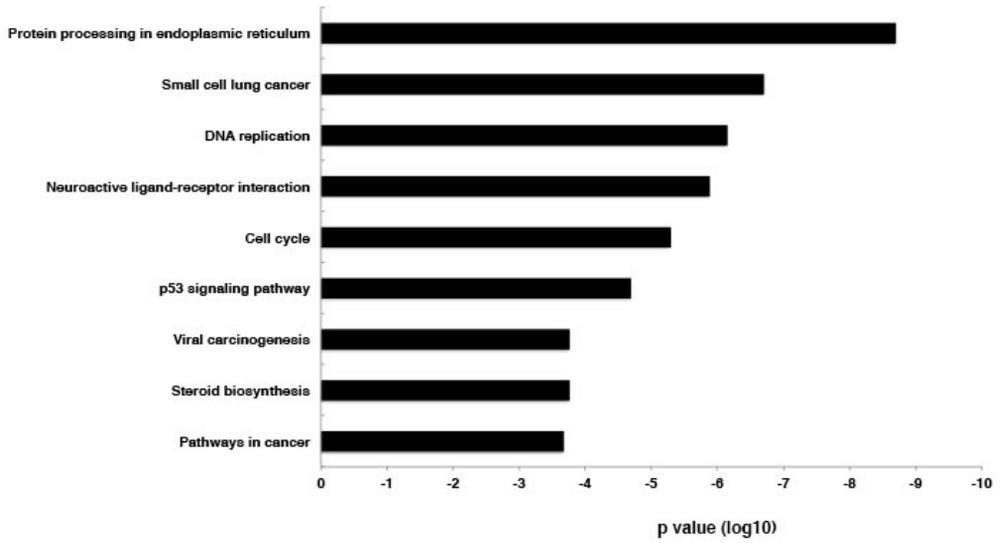


Figure 2

A

Significantly Enriched in KEGG pathways



B

Significantly Enriched Gene Ontology Biological Processes

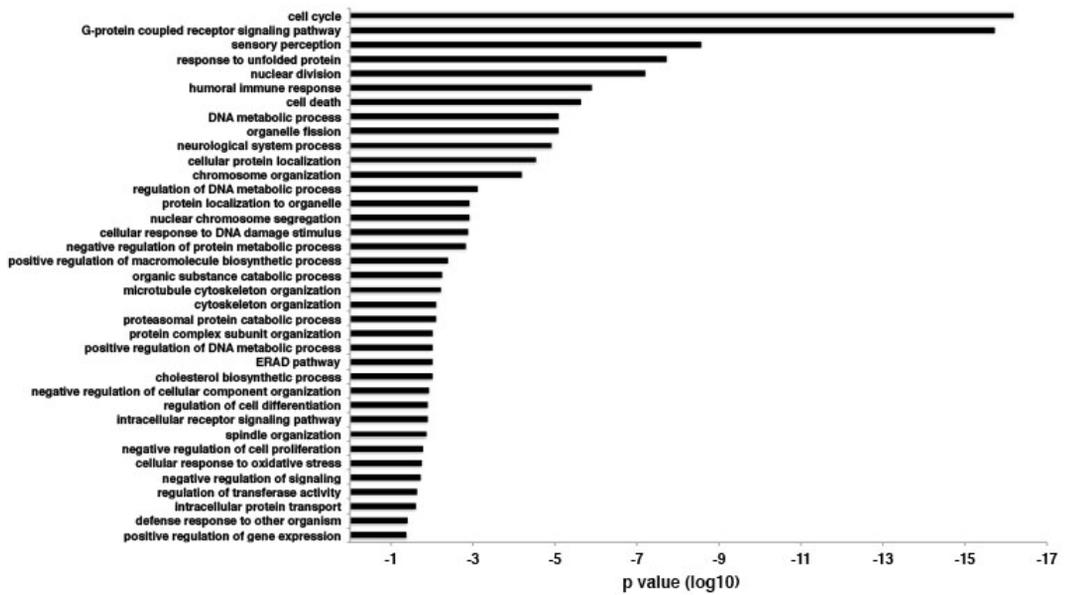
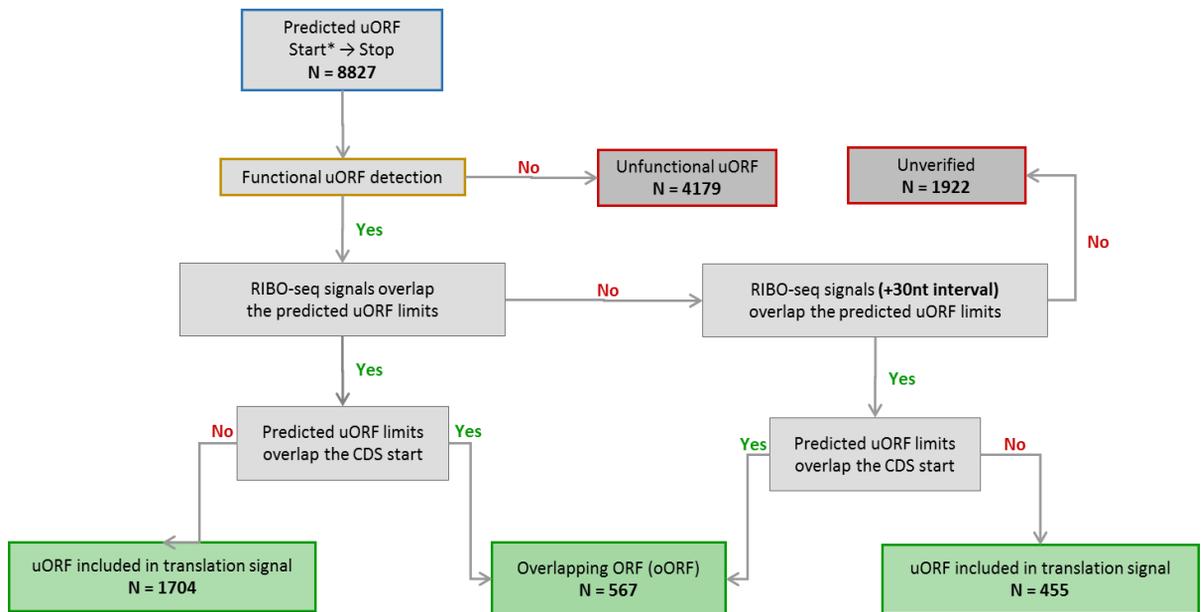


Figure 3



* Start of predicted uORFs is defined by: AUG, Kozak AUG and Kozak non AUG

Figure 5

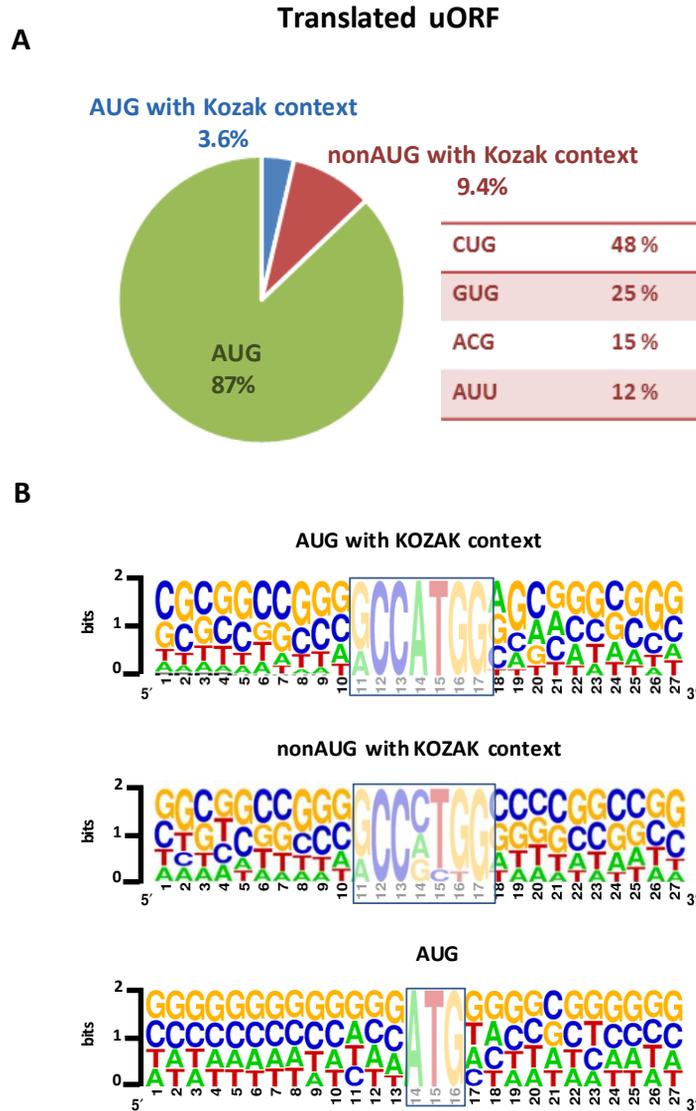


Figure 6

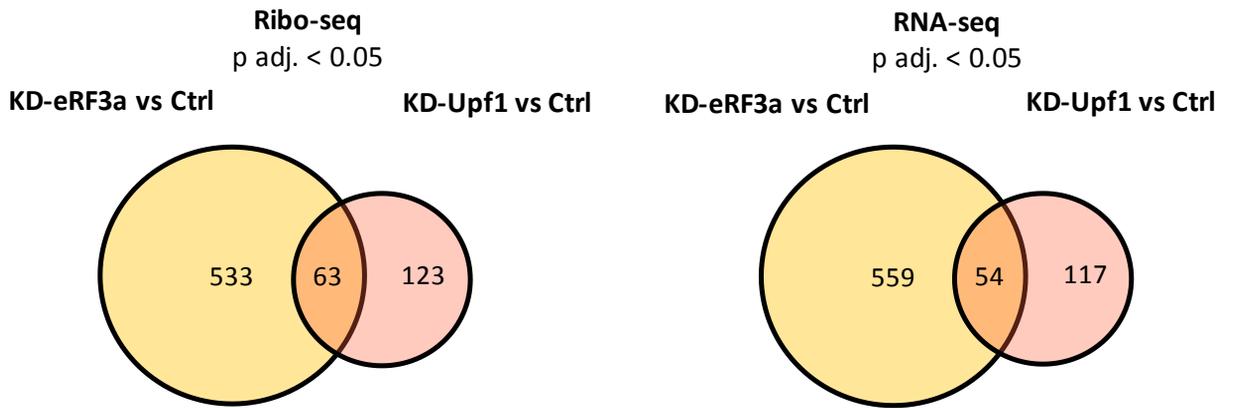


Figure 7

A

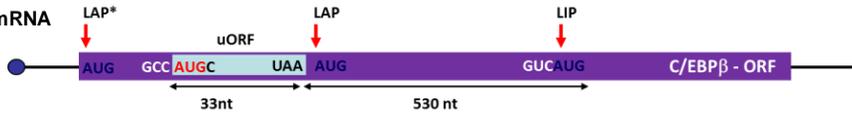
ATF4 mRNA



IFRD1 mRNA



C/EBPβ mRNA



B

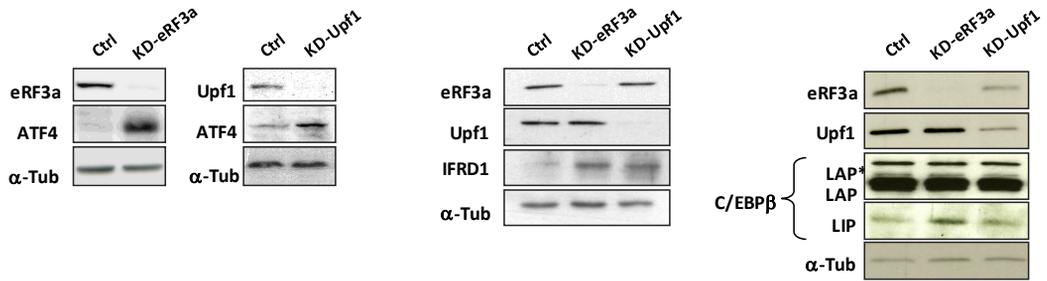
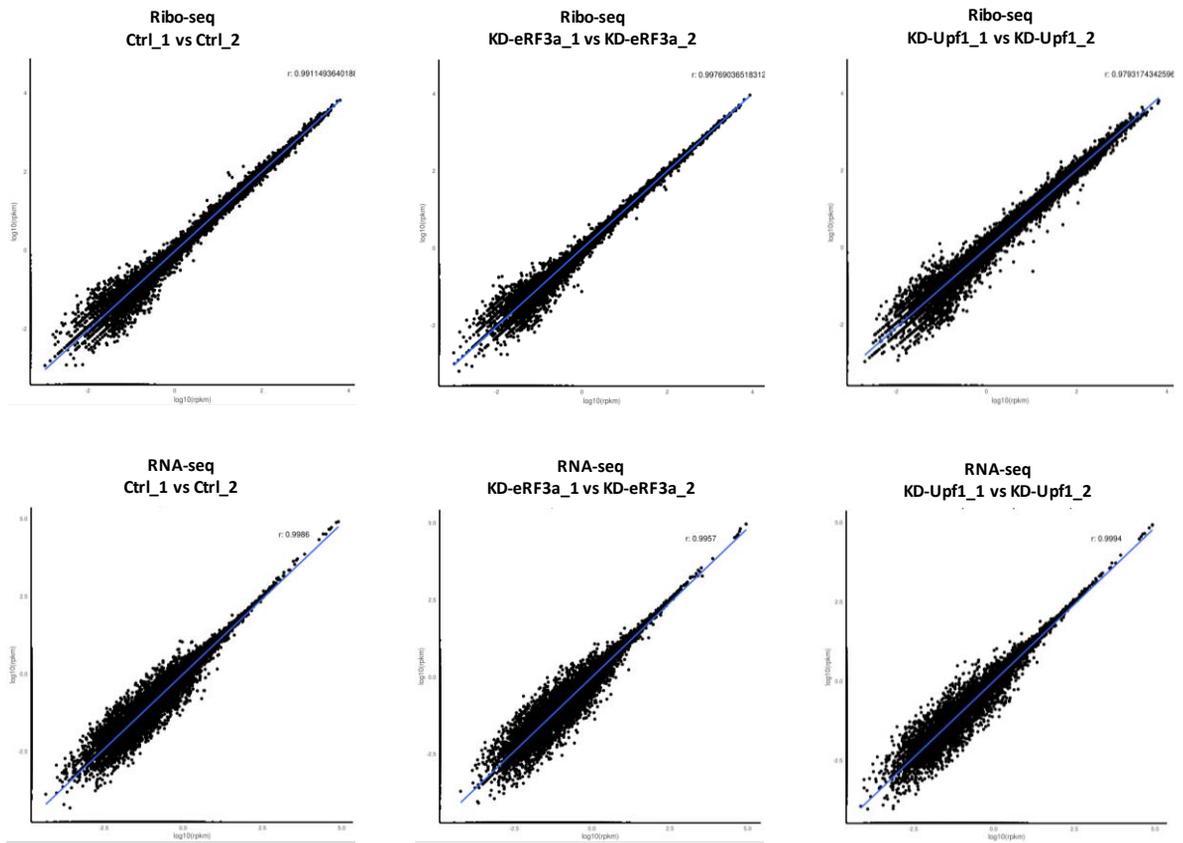


Figure 8



Supplementary Figure 1

	Samples	Total Read	Trim Adapt (25-34)				Mapping rRNA		Mapping genome	
			Adapter present + size filter ok	%	Too short	Too long	Mapped read	% contamination	Mapped (Uniquely)	% (of raw data)
Ribo-seq	sh-Sc_1	496037564	366124433	73.8	5229054	10913389	45870113	12.5	133733098	27.0
	sh-Sc_2	411820853	269795611	65.5	10183184	20115833	25493141	9.4	114401028	27.8
	sh-Upf1_1	386563970	221204420	57.2	6897235	13589959	49671167	22.5	78854057	20.4
	sh-Upf1_2	444079333	309649847	69.7	8880901	34570906	33528160	10.8	139575763	31.4
	sh-eRF3_1	406326972	309337455	76.1	9869052	14126353	40761681	13.2	137032369	33.7
	sh-eRF3_2	406058053	296980600	73.1	9043793	14467973	48034296	16.2	136422829	33.6
RNA-seq	sh-Sc_1	81 517 731					3 219 405	3.95	17 130 126	21.88
	sh-Sc_2	75 204 811					3 316 119	4.41	15 911 405	22.13
	sh-Upf1_1	82 630 892					2 898 637	3.51	18 716 052	23.47
	sh-Upf1_2	81 568 719					15 628 651	19.16	13 431 650	20.37
	sh-eRF3_1	91 595 387					2 339 477	2.55	17 120 904	19.18
	sh-eRF3_2	67 317 174					1 502 035	2.23	11 389 315	17.31

Supplementary Table 1

II. Effet de la déplétion en Upf1 sur la formation du ribosome 80S

Afin de vérifier l'intégrité des polysomes au cours de mes expériences de *ribosome profiling*, j'ai été amenée à analyser le profil des polysomes (*polysome profiling*) des cellules contrôles, déplétées en eRF3 et déplétées en Upf1. L'expérience de *polysome profiling* donne des informations qualitatives sur l'efficacité de la traduction et aussi sur les mécanismes qui peuvent modifier cette efficacité. Pour l'analyse du profil des polysomes, les extraits de cellules dont la traduction a été figée par la cycloheximide sont fractionnés par ultracentrifugation sur un gradient de saccharose 15-50%. Les gradients sont développés par passage à travers un système de lecture et d'enregistrement de l'absorbance à 254 nm. Une quinzaine de fractions sont collectées à intervalles réguliers (voir l'article en Annexe pour les détails expérimentaux). Le profil d'absorbance permet d'observer les différents pics correspondant aux sous-unités ribosomiques 40S et 60S libres, aux ribosomes 80S libres ainsi qu'aux polysomes portant un nombre croissant de ribosomes en cours de traduction (**Figure 34A**).

La comparaison des profils polysomiques des cellules normales ou déplétées en eRF3 ou d'Upf1 montre que la déplétion en eRF3 comme la déplétion en Upf1 induit dans les cellules HCT116 une diminution globale de la quantité de polysomes par rapport aux cellules contrôles (**Figure 34A**). Cette diminution est le signe d'une diminution de l'efficacité de traduction. Cependant, ce qui a attiré notre attention, c'est la différence observée dans le profil des sous unités libres du ribosome. Il avait déjà été observé par notre équipe que la déplétion en eRF3 des cellules humaines entraîne une augmentation des ribosomes libres concomitante avec la diminution des polysomes, ce qui traduit un défaut d'initiation de la traduction dans ces cellules. Il avait été montré que ce défaut d'initiation de la traduction est dû à l'inhibition de la voie de signalisation mTOR (Chauvin et al., 2007).

Dans le cas de la déplétion en Upf1, nous observons une diminution significative de la quantité de ribosome 80S libres par rapport aux cellules contrôles (**Figure 34A**). Cette diminution a été observée de façon reproductible et nous avons pu quantifier chaque fraction libre et calculer à la fois pour chaque sous-unité ribosomique et pour le ribosome 80S le rapport de surface fraction libre/total des fractions libres (**Figure 34B**). On observe dans les cellules déplétées en Upf1 une diminution de 30% des ribosomes 80S libres par rapport aux cellules contrôles, et dans le même

temps, une augmentation proportionnelle des sous unités 40S et 60S. Ce résultat suggère une altération dans le processus de traduction due à une altération de la formation du ribosome 80S.

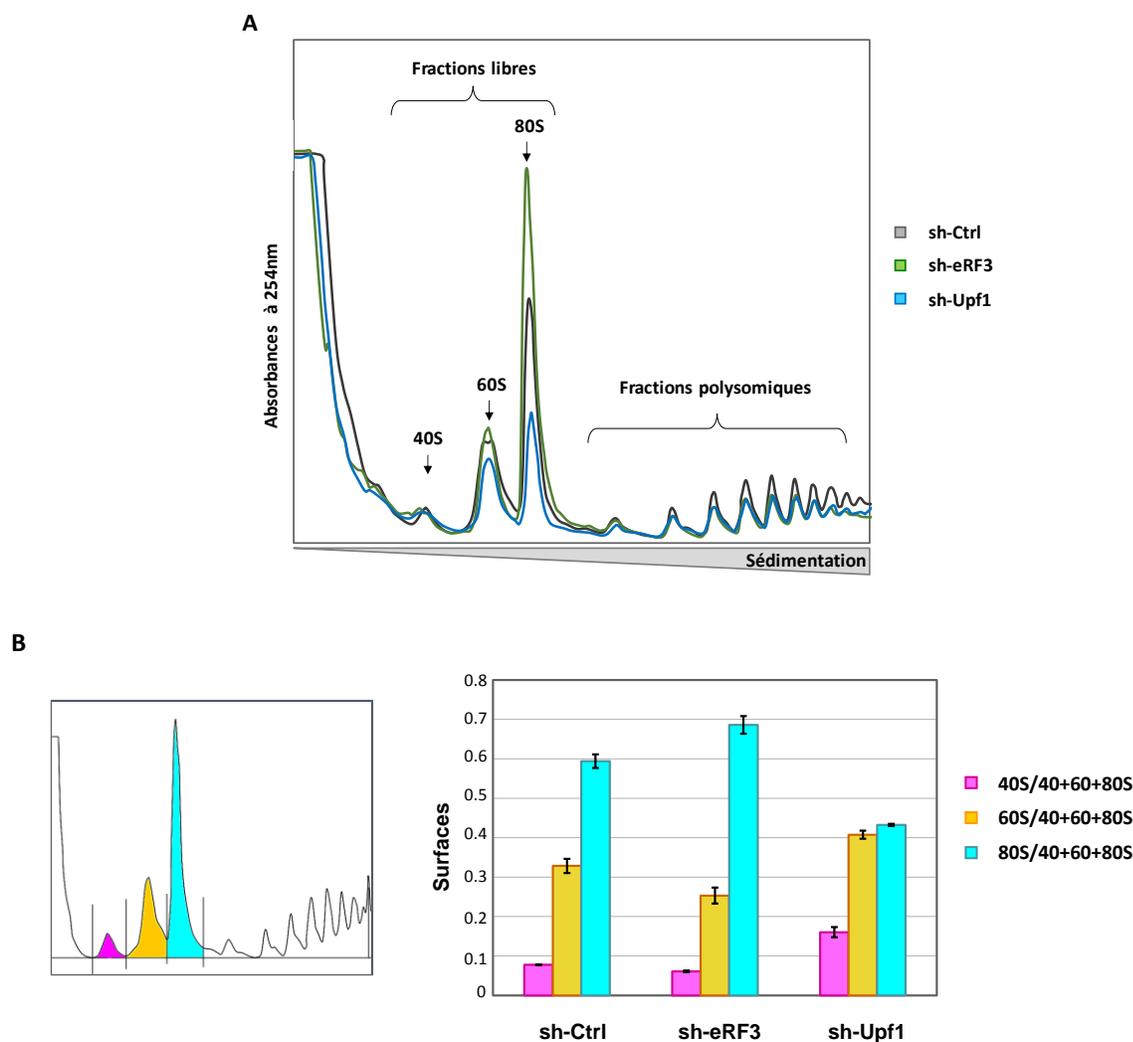


Figure 34 : Analyse de la déplétion en eRF3 et en Upf1. *A. Comparaison des profils polysomiques des cellules contrôles ou déplétées en eRF3 et Upf1. B. Quantification des fractions libres 40, 60 et 80S des cellules contrôles ou déplétées en eRF3 et Upf1. A gauche, la quantification de chaque fraction libre et de l'ensemble des polysomes ont été réalisées en déterminant la surface sur quatre profils de polysomes. A droite, rapport de chaque fraction libre sur l'ensemble des fractions libres.*

Chez les mammifères, le facteur d'initiation de la traduction eIF6 régule l'assemblage du ribosome entier sur l'ARNm au cours de la dernière étape de l'initiation de la traduction. Grâce à son activité anti-associative, eIF6 lie la grande sous unité ribosomique 60S pour empêcher son association prématurée avec la petite sous unité ribosomique 40S et ainsi prévient la formation d'un ribosome 80S inactif. Le relargage d'eIF6 de la 60S est assuré par le facteur RACK1 (*Receptor for Activated C Kinase 1*), après interaction entre RACK1 et le complexe PKC (*Protéine kinase C*) qui phosphoryle eIF6 lié à la 60S et ainsi permet sa dissociation. PKC activée subit une translocation de l'endomembrane à la sous unité 40S portant RACK1, et active la sous unité 60S en phosphorylant eIF6, étape nécessaire à son relargage (Miluzio et al., 2009) (**Figure 35**).

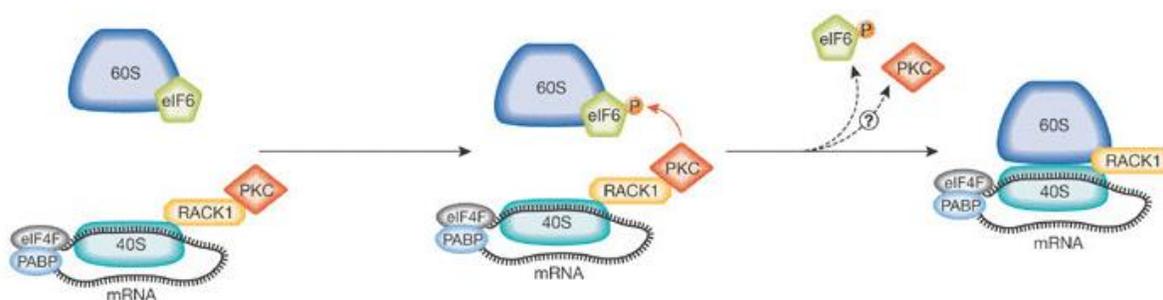


Figure 35 : Le facteur eIF6 prévient l'assemblage du ribosome. Lorsque PKC est activée elle se lie à son récepteur RACK fixé sur la 40S, ce qui lui permet de phosphoryler eIF6, une fois phosphorylé eIF6 se détache de la 60S ce qui permet au 80S de se former et favoriser la traduction (Miluzio et al., 2009).

Il avait été observé par étude du profil des polysomes (**Figure 36**) qu'une modification de l'abondance d'eIF6 entraînait des variations des sous unités libres et des ribosomes 80S dans les cellules (Ceci et al., 2003).

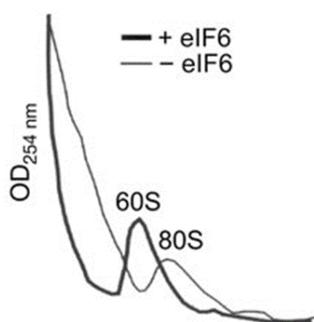


Figure 36 : La déplétion en eIF6 favorise la formation des ribosomes 80S, en revanche, sa surexpression augmente les sous-unités 60S libres (Ceci et al., 2003).

Dans nos expériences, l'augmentation des sous-unités ribosomiques libres et la diminution des ribosomes 80S pouvait être due à une augmentation de l'expression du facteur eIF6 dans les cellules déplétées en Upf1. Pour vérifier cette hypothèse, j'ai comparé le taux d'expression du facteur eIF6 entre les cellules déplétées ou non en Upf1. Cette analyse a été réalisée à la fois au niveau des ARNm par PCR quantitative en temps réel (qPCR) et au niveau protéique par Western blot. Mes résultats montrent qu'il n'y a pas de différence dans le taux d'expression d'eIF6, ni au niveau des ARNm, ni au niveau des protéines, j'ai donc pu exclure l'hypothèse de la surexpression de eIF6 dans les cellules déplétées en Upf1 (Résultats non montrés).

J'ai ensuite analysé l'hypothèse d'une plus grande rétention du facteur eIF6 au niveau de la sous-unité 60S dans les cellules déplétées en Upf1. Pour cela j'ai séparé les différentes fractions ribosomiques et polysomiques par centrifugation sur gradient de sucrose 15-50 %, puis j'ai extrait les protéines de ces différentes fractions et analysé la distribution d'eIF6 entre les deux profils. J'ai également vérifié l'expression de la protéine RACK1 qui se lie à la sous-unité 40S ainsi que de différentes protéines constitutives de la petite ou de la grande sous-unité ribosomique, ces dernières me servent de contrôle pour délimiter dans quelle fraction ribosomique ma protéine se trouve (**Figure 37**).

Les résultats de ces expériences montrent qu'effectivement il n'y a aucune différence de distribution du facteur eIF6 entre les profils ribosomiques des cellules contrôles et les cellules déplétées en Upf1. On remarque qu'eIF6 est présente dans la fraction 60S qui correspond aux fractions 4 et 5 dans les deux conditions. D'autre part si on regarde le taux d'expression des protéines ribosomiques dans la fraction des ARNs lié au polysomes, on remarque qu'il y a moins de protéines ribosomiques dans les cellules déplétées en Upf1 par rapport aux cellules contrôle ce qui confirme que la traduction est diminuée dans les cellules déplétées en Upf1.

La quantité totale en eIF6 ne semble pas varier dans les cellules déplétées en Upf1 par rapport aux cellules contrôles. De même, l'association d'eIF6 aux sous-unités 60S ne semble pas modifiée. Cependant, on ne peut pas exclure une modification de l'état de phosphorylation d'eIF6 dans les cellules déplétées en Upf1.

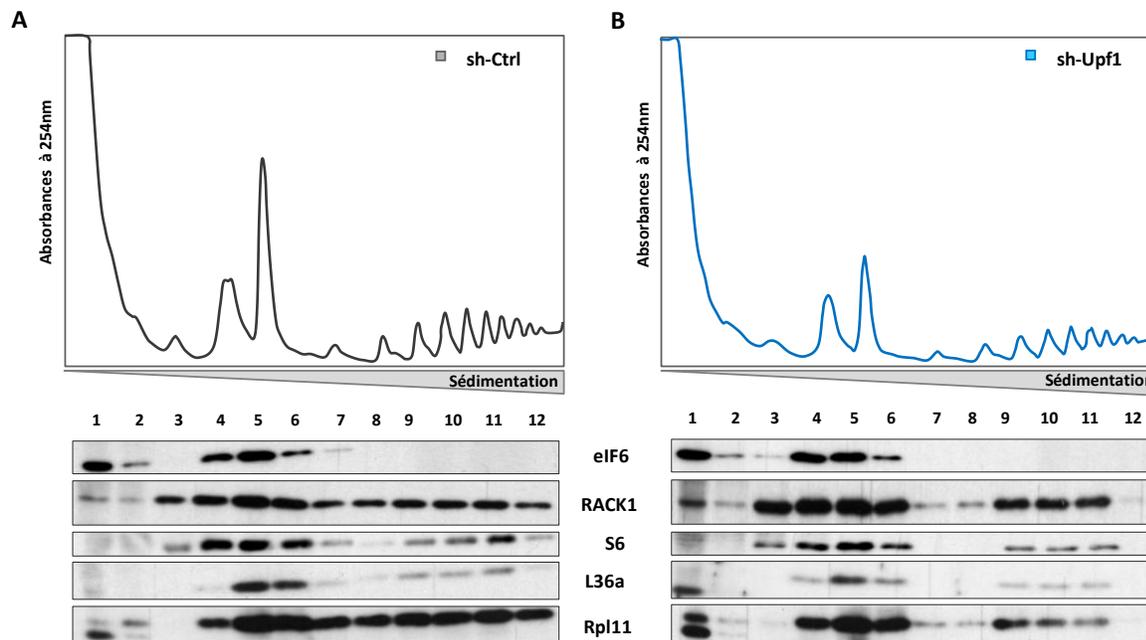


Figure 37 : Localisation de la protéine eIF6 sur un profil de polysomes. *A. Profil polysomique des cellules contrôle. B. Profil polysomique des cellules déplétées en Upf1. Panneaux du bas : analyse de l'expression des protéines eIF6, Rack1 et des protéines ribosomiques S6, L36a et Rpl11 par Western blot.*

Une autre explication pourrait être que l'assemblage des sous-unités du ribosome et en particulier de la sous unité 60S est altéré lors de la déplétion en Upf1. Cette altération de la sous-unité 60S empêcherait un assemblage correct des ribosomes entraînant une diminution des formes libres des ribosomes 80S inactifs comme des formes associées aux polysomes. De nombreuses protéines entrent en jeu dans le processus d'assemblage des ribosomes et une ou plusieurs de ces protéines pourrait être la cible de la NMD.

Discussion

La terminaison de la traduction est la dernière étape de la synthèse protéique mais c'est une étape essentielle parce qu'elle permet la libération de la protéine nouvellement synthétisée, ainsi que le recyclage des composants du ribosome. La terminaison de la traduction a aussi une fonction importante dans la régulation traductionnelle, en favorisant la translecture à certains codons stop et en intervenant dans la fonction régulatrice des uORF par son effet sur la réinitiation post-terminaison.

Chez les eucaryotes, la terminaison est assurée par les facteurs eRF1 et eRF3 qui se fixent sous forme d'un complexe au site A du ribosome. Le facteur eRF3, par son activité GTPase est indispensable au processus de terminaison et semble être le facteur limitant de l'efficacité de ce processus. En plus de son rôle dans la terminaison de la traduction, eRF3 est impliqué dans d'autres mécanismes liés à la traduction tels que la stabilité des ARNm par son interaction avec la protéine PABP, et le contrôle de la qualité des ARNm par son interaction avec Upf1.

La protéine Upf1 est aussi une protéine essentielle. C'est non seulement un des acteurs clé de la NMD mais il a été montré que son activité ATPasique est nécessaire pour une terminaison de la traduction efficace et la libération des ribosomes aux codons de terminaison prématurés (Serdar et al., 2016). La protéine Upf1 est physiquement et fonctionnellement associée à eRF3 : d'une part, elle interagit avec eRF3 dans le complexe SURF, et d'autre part, les deux protéines sont impliquées dans la régulation post-transcriptionnelle de l'expression des gènes, notamment, ceux qui portent des uORF. Cependant les mécanismes moléculaires de ces régulations restent encore largement inconnus et le rôle de l'association physique de ces deux facteurs dans ces mécanismes n'a jamais été évalué.

L'objectif de ma thèse était de comprendre, par des analyses de la traduction à l'échelle du génome, le rôle de ces deux facteurs dans le contrôle traductionnel médié par les uORF.

Le *ribosome profiling* a permis récemment de faire des avancées considérables dans l'étude de la traduction, notamment dans la détection des régions codantes pour des transcrits qui étaient supposés être non-codants, dans la détections des mécanismes de régulations post-transcriptionnelles, et dans la détection de séquences traduites dans les régions 5'UTR et 3'UTR d'un nombre non négligeable de transcrits (Bazzini et al., 2014, 2012; Brar et al., 2012; Guo et al., 2010; Guydosh and Green, 2014; Ingolia et al., 2014, 2011; Lee et al., 2012). Cependant, malgré l'abondante littérature récente sur les uORF, aucune analyse à grande échelle n'a été réalisée dans

des conditions qui affectaient leur expression, en dehors des analyses dans les conditions de stress (Andreev et al., 2015; Gerashchenko et al., 2012). J'ai donc utilisé cette méthode pour analyser le transcriptome de cellules dont la terminaison de la traduction et la NMD étaient affectées.

1. Identification des uORF dans le transcriptome des cellules HCT116

Bien que plusieurs études récentes aient réalisé des études à grande échelle de la traduction des uORF, la plus part de ces études se limitent à un catalogue d'uORF prédites qui montre souvent la conservation de ces uORF à travers les espèces mais rarement, le lien entre la traduction de ces uORF et le transcriptome réel des cellules. Une conséquence est que la proportion d'uORF traduites est encore peu connue à ce jour, en particulier pour les génomes de mammifères.

Pour mieux comprendre le rôle de ces uORF dans la traduction, nous avons d'abord définis le transcriptome exact des cellules HCT116, puis développé un pipeline d'analyse qui nous permettait de déterminer la liste des uORF traduites dans ce transcriptome.

Le transcriptome des cellules humaines contient 8827 transcrits portant au moins une uORF prédite, en amont ou chevauchant l'ORF principale, ce qui est tout à fait cohérent avec ce qui a été trouvé par d'autres prédictions (Calvo et al., 2009; Ye et al., 2015). Ces uORF représentent une proportion de 40 % des transcrits exprimés dans un type cellulaire donné, et ce pourcentage important souligne la nécessité de faire une étude fonctionnelle de ces uORF.

Il est clairement établi que l'initiation de la traduction chez l'eucaryote requière un contexte de traduction favorable (Wold et al., 1986) celui-ci est d'autant plus important lorsque le codon d'initiation est un codon non canonique (non AUG). Il est également établi qu'au moins deux appariement du codon initiateur avec l'anticodon de l'ARNt^{Met} initiateur. La première position semble ne pas être importante et n'affecte pas la capacité de servir de codon initiateur, bien que, lors de mésappariement, la Cytosine soit clairement favorisé par rapport à U ou G. La deuxième position est plus sensible à la substitution et ne peut pas tolérer la présence d'un résidu de purine, la Cytosine représente aussi le plus efficace 2^{ème} nucléotide de tous les codons d'initiation non-AUG, enfin, la troisième position n'affecte pas ou peu la capacité de servir d'initiateur (Peabody, 1989). Dans cette analyse nous avons donc identifié toutes les uORF, qui s'initient soit à un codon AUG classique entouré d'un contexte non défini, soit à un codon AUG entouré du contexte de

Kozak (RCCaugG), ou encore à un codon d'initiation non-AUG dans un contexte de Kozak favorable à l'initiation.

Le pipeline que nous avons réalisé nous permettait d'identifier les uORF traduites à partir des 8827 transcrits portant des uORF prédites *in silico*, et des signaux de traduction, issus du *ribosome profiling*, présents dans les régions 5' non traduites des ARNm. Cette approche nous a permis de définir une liste de 2726 transcrits portant des uORF traduites, dans les cellules HCT116. Ce résultat diffère des résultats de l'analyse réalisée chez les vertébrés, dans laquelle les auteurs trouvent plus de 8000 transcrits portant des uORF traduites malgré le fait qu'ils n'aient pris que les uORF initiées à un codon AUG dans un contexte indifférent (Timothy G Johnstone et al., 2016). Ce résultat peut-être expliqué par le fait, que contrairement à nous, cette étude est globale et porte sur la totalité du transcriptome humain, les auteurs ne se sont ni restreints aux transcrits exprimés dans les cellules étudiées, ni à la cohérence de ces transcrits.

En effet, en analysant finement leur liste d'uORF considérées comme étant fonctionnelles nous nous sommes rendu compte que le tiers des transcrits étaient classés TSL5 selon la classification donnée par la base de données génomique ENSEMBL. Cette notion de score de cohérence des transcrits a été introduite afin de remédier à la présence de plusieurs transcrits pour un même gène dans les banques de données. Lors d'une analyse à grande échelle, il est, important d'évaluer les transcrits. Ainsi, le Transcript Support Level (TSL) est une méthode qui permet d'évaluer le niveau de pertinence d'une annotation d'un transcrit humain. Cette méthode est basée sur un score d'alignement des transcrits et des EST (Expressed Sequence Tag) sur le transcrit référence de GENECODE. Le score est établi en fonction du degré d'alignement, ce qui permet de classer les transcrits par ordre de cohérence décroissant de TSL1 à TSL5. TSL1, pour un variant dont la séquence couvre entièrement le transcrit référence, alors que TSL5, est attribué à un variant qui ne couvre pas ou que partiellement le transcrit référence. Dans l'étude de Johnstone *et al.*, certains transcrits classés TSL5 portent un nombre impressionnant d'uORF, jusqu'à 55 (Timothy G Johnstone et al., 2016). Ceci est dû au fait que ces transcrits ne contiennent pas la région 5' de la séquence codante et que les uORFs trouvées correspondent à des codons d'initiation et des codons de terminaison statistiquement répartis dans les trois phases de la séquence codante. Cette remarque montre l'importance qu'il y a à réaliser une identification du transcriptome le plus précisément possible des cellules que l'on utilise.

Notre détection d'uORF traduites a donc été réalisée uniquement sur les ~13 000 transcrits exprimés dans les cellules HCT116. Ces transcrits ont été sélectionnés sur la base de l'algorithme APPRIS qui permet d'écarter de la sélection les transcrits non soutenus par des données fiables de séquençage, et par la confrontation de cette sélection avec nos données du *RNA-seq*.

2. Usage des codons d'initiation non canoniques par les uORF fonctionnelles

En analysant les prévalences des codons d'initiation de ces uORF, on a constaté qu'environ 10 % des uORF traduites sont initiées à des codons non-AUG. Cette proportion est très importante par rapport à celle des ORF principales initiées à des codons non-AUG. En effet, il a été montré que seulement 0,1% des séquences codantes des ARNm démarrent leur traduction à des codons non-AUG chez les eucaryotes (Tikole and Sankararamakrishnan, 2006). Les protéines traduites à partir de codons non canoniques semblent être impliquées dans la régulation de l'expression des gènes, c'est le cas des transcrits codant les facteurs de transcription Myc ou Egr1, ou encore de celui du facteur de croissance FGF2 (Touriol et al., 2003, pour une revue). Notre résultat était inattendu car on pensait que la machinerie de traduction sélectionnait rigoureusement les codons AUG pour l'initiation, alors que les codons non-AUG étaient rarement traduits. Bien que la capacité des ribosomes à initier à des codons non AUG ait déjà été montrée (Anderson and Buzash-Pollert, 1985) et que des codons d'initiation non-AUG fonctionnels aient déjà été identifiés (Ivanov et al., 2011), l'efficacité de l'initiation non-AUG était tout de même supposée être très faible. Il a été montré que, même dans un contexte nucléotidique optimal, c'est à dire le contexte de Kozak, l'efficacité d'initiation aux codons non-AUG est au moins 20 fois inférieure à celle des codon AUG (Kozak, 1989). Cette notion a été confirmée par une analyse bioinformatique récente compilant les données existante de *Ribosome profiling* sur les site d'initiation de la traduction (Michel et al., 2014). Cette étude montre pour le codon non canonique CUG une faible efficacité d'initiation, même lorsqu'il est dans un contexte favorable.

Notre analyse des séquences qui entourent les codons d'initiation des uORF montre que mis à part le contexte Kozak, on ne retrouve aucun autre motif consensus. La même observation a été rapportée dans une autre étude récente des sites d'initiation de la traduction dans des cellules humaines (Fritsch et al., 2012).

En revanche, dans cette analyse réalisée avec des cellules traitées par la puromycine, qui permet d'inhiber l'élongation et de ne conserver que les ribosomes au site d'initiation, Fritsch et collaborateurs trouvent une proportion étonnamment élevée de nouvelles uORF (uORF non encore prédites) dont le site d'initiation est un codon non-AUG, le codon AUG représente à peine plus d'un tiers de tous les codons d'initiation (Fritsch et al., 2012). Ces statistiques sont cependant à prendre avec prudence parce qu'ils n'englobent pas les uORF déjà connues, qui sont le plus souvent prédites sur la base de la présence d'un codon d'initiation AUG exclusivement. De plus, en l'absence de signal d'élongation de la traduction, on peut aussi se poser la question sur la réalité de l'initiation à tous les sites putatifs décrits dans cet article. Nous n'avons trouvé dans notre étude aucune uORF débutant par les codons AGG, AUA, AAG ou AUC pourtant décrits par Fritsch et collaborateurs.

Enfin, dans notre analyse, un certain nombre d'empreintes ribosomiques ne correspondant à aucun codon d'initiation d'une uORF potentielle au niveau de l'empreinte ou dans son voisinage immédiat (+/- 30 nucléotides) représentaient 31 % des signaux trouvés dans les 5'UTR. Nous avons recherché qu'elle pouvait être l'origine de ces signaux, translecture du codon stop d'une uORF en amont ou décalage de phase de lecture lors de la traduction d'une uORF. Cependant, aucune de ces hypothèses n'a résisté à l'analyse fine de ces signaux. On peut alors imaginer que ces signaux pourraient correspondre à des empreintes de ribosomes «fantômes», c'est à dire non impliqués dans un processus de traduction. Il semble que de tels signaux, déjà décrits par certains transcrits (Olivier Namy, communication personnelle), peuvent être supprimés en réalisant une expérience de *ribosome profiling* contrôle avec des cellules dont la traduction est inhibée.

3. Rôles respectifs d'eRF3 et Upf1 dans le contrôle traductionnel par les uORF

Un des résultats marquants de notre analyse comparative des modifications du transcriptome et du traductome lors de la déplétion en eRF3 et Upf1 est le peu de cibles communes de ces deux facteurs qui pourtant sont associés physiquement dans le complexe SURF. De même, on retrouve un petit nombre de cibles communes dans les transcrits portant des uORF suggérant qu'il existe deux classes d'uORF, l'une impliquant la terminaison de la traduction et l'autre la NMD. Cette différence notable pourrait s'expliquer par les différents mécanismes dans lesquels ces facteurs interviennent. D'une part, pour eRF3, son rôle dans le contrôle traductionnel par les uORF fait intervenir le mécanisme de réinitiation post-terminaison. Ce mécanisme est, comme on l'a vu dans l'introduction, dépendant

non seulement d'un certain nombre de paramètres portés par l'uORF (en particulier sa taille et le contexte entourant le codon stop), la séquence en aval (en particulier, la distance la séparant du codon d'initiation suivant), de la répartition des uORF dans le cas d'uORF multiples, mais aussi d'un certain nombre de facteurs, agissant en trans, nécessaires à l'initiation, comme les facteurs eIF2 et eIF3. La régulation traductionnelle des cibles d'eRF3 pourraient être particulièrement sensible à l'un ou l'autre ou à une combinaison de ces paramètres et de ces facteurs.

D'autre part, pour Upf1, les uORF apportent un codon stop prématuré déclenchant la dégradation de l'ARNm par la NMD. Ce mécanisme est peu sensible aux caractéristiques des uORF comme leur taille et la distance du codon d'initiation suivant, et n'est sans doute pas non plus influencé par les facteurs mis en jeu lors du processus de reinitiation. Les paramètres liés à l'uORF qui influent sur le déclenchement de la NMD sont la présence d'une jonction exon-exon en aval de l'uORF, et la présence d'un EJC à proximité de cette jonction, D'autres paramètres indépendants de l'uORF interviennent également, qui sont les facteurs qui favorisent l'activité d'Upf1 comme les protéines SMG qui régulent son état de phosphorylation, enfin les protéines impliquées dans le processus de dégradation de l'ARNm. Les cibles d'Upf1 réuniraient alors les caractéristiques nécessaires au déclenchement de la NMD.

4. Déplétion en eRF3 et translecture des codons stop

Il avait été suggéré par le laboratoire lors d'une précédente étude sur la régulation traductionnelle de l'ARNm d'ATF4 que la déplétion en eRF3 entraînait la translecture du codon stop de l'uORF1 et favorisait ainsi la traduction de l'ORF d'ATF4 (Ait Ghezala et al., 2012). Un des enjeux de notre étude du transcriptome des cellules HCT116 déplétées en eRF3 était d'évaluer l'importance de la translecture au niveau de l'expression du génome entier. Il aurait en particulier été intéressant de connaître l'impact de ce processus sur la traduction des uORF régulatrices. De plus, une étude récente du laboratoire d'Olivier Namy appliquant la méthode de *Ribosome profiling* à des levures portant une mutation du facteur eRF3 montrait clairement des signaux de translecture pour une centaine de gènes (Baudin-Baillieu et al., 2014). Notre stratégie pour les expériences de *Ribosome profiling* a donc été d'immobiliser la traduction dans la phase d'élongation pour pouvoir collecter les empreintes ribosomiques des ribosomes situés sur les codons stop. Or, l'analyse fine de nos résultats de *Ribo-seq* ne nous a pas permis de mettre en évidence de modification significative des

empreintes ribosomiques dans les régions situées immédiatement en aval des codons stop des uORF, dans les cellules déplétées en eRF3. Plusieurs facteurs peuvent expliquer cet échec.

D'une part, la quantité des empreintes obtenues dans nos expériences dont une partie ne correspondait pas de façon évidente à de réels signaux de traduction mais probablement à des artefacts que l'on ne pouvait pas éliminer sur la base de l'absence de phasage traductionnel de 3 en 3. Cette difficulté à obtenir un phasage clair des empreintes de ribosomes en cours de traduction a été observée pour les génomes complexes comme le génome humain ou comme celui de la drosophile. Les génomes plus simples comme celui de la levure ne présentent pas un tel inconvénient et permettent une analyse beaucoup plus fine de la traduction par la méthode de *Ribosome profiling*.

D'autre part, la translecture d'un codon stop est un évènement relativement rare, même dans les cellules déplétées en eRF3, ne représentant qu'un faible pourcentage des évènements de terminaison. Si il est relativement facile de le détecter grâce à des gènes rapporteurs portant une activité enzymatique facilement mesurable, il est beaucoup plus difficile à mettre en évidence en comptant un nombre d'empreintes dont une partie correspond à du bruit de fond. Le choix du seuil de détection des signaux significatifs, exprimé en nombre de *reads* par nucléotide, pourrait masquer les empreintes faibles correspondant aux signaux de translecture. La solution de choisir un seuil trop bas a l'inconvénient de diminuer de beaucoup la qualité de l'analyse. La solution décrite plus haut pour les empreintes "fantômes" de réaliser une expérience contrôle en inhibant la traduction de façon non spécifique, avec de l'EDTA par exemple, qui permet de bloquer complètement l'association des ribosomes aux ARNm, pourrait permettre de supprimer artificiellement le bruit de fond mais augmente la complexité de l'analyse bioinformatique.

Enfin, il serait peut être utile de compléter notre analyse des empreintes des ribosomes en cours d'élongation par une analyse des empreintes des ribosomes en cours d'initiation, en utilisant des inhibiteurs de l'élongation comme la harringtonine ou la puromycine. Cette méthode alternative de *Ribosome profiling* permettrait de mieux fixer les sites d'initiation de la traduction, particulièrement pour les uORF, et d'évaluer avec une meilleure précision leur niveau de traduction.

La méthode *Ribosome profiling* est une méthode encore relativement récente, qui met en jeu des techniques de biologie moléculaire fines dont tous les détails expérimentaux ne sont pas encore fixés (choix de la RNase pour les empreintes, choix de l'inhibiteur de la traduction, protocole pour

une suppression des fragments d'ARN ribosomiques contaminants, protocole pour la réalisation des banques de cDNA et pour leur amplification, choix des contrôles, etc.) et des analyses bioinformatiques complexes dont, là aussi, tous les outils ne sont pas établis. Même s'il existe maintenant un certain nombre de programmes informatiques mis à la disposition des utilisateurs de cette technique, et des bases de données génomiques de plus en plus fiables, les particularités de chaque projet tant au niveau des génomes et des cellules étudiés qu'au niveau de l'objet de l'étude, uORF, translecture, conditions de stress, infection virale, etc., font que chaque utilisateur de cette méthode est amené à construire lui-même les outils informatiques adaptés à son étude.

L'importance du *Ribosome profiling* dans notre connaissance approfondie de la traduction fait que cette méthode est actuellement utilisée par de nombreux laboratoires qui y apporteront sans doute de nombreuses améliorations.

Bibliographie

Bibliographie

- Ait Ghezala, H., Jolles, B., Salhi, S., Castrillo, K., Carpentier, W., Cagnard, N., Bruhat, A., Fafournoux, P., Jean-Jean, O., 2012. Translation termination efficiency modulates ATF4 response by regulating ATF4 mRNA translation at 5' short ORFs. *Nucleic Acids Res.* 40, 9557–9570. doi:10.1093/nar/gks762
- Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., Pestova, T.V., 2006. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell* 125, 1125–1136. doi:10.1016/j.cell.2006.04.035
- Ameri, K., Harris, A.L., 2008. Activating transcription factor 4. *Int. J. Biochem. Cell Biol.* 40, 14–21. doi:10.1016/j.biocel.2007.01.020
- Amrani, N., Ghosh, S., Mangus, D.A., Jacobson, A., 2008. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* 453, 1276–1280. doi:10.1038/nature06974
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Le Hir, H., Andersen, G.R., 2006. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313, 1968–1972. doi:10.1126/science.1131981
- Anderson, C.W., Buzash-Pollert, E., 1985. Can ACG serve as an initiation codon for protein synthesis in eucaryotic cells? *Mol. Cell. Biol.* 5, 3621–3624.
- Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E., Cormican, P., Morris, D.W., Shatsky, I.N., Baranov, P.V., 2015a. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* 4. doi:10.7554/eLife.03971
- Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E., Cormican, P., Morris, D.W., Shatsky, I.N., Baranov, P.V., 2015b. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* 4. doi:10.7554/eLife.03971
- Asano, K., 2014. Why is start codon selection so precise in eukaryotes? *Transl. Austin Tex* 2, e28387. doi:10.4161/trla.28387
- Azzalin, C.M., Reichenbach, P., Khoraiuli, L., Giulotto, E., Lingner, J., 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798–801. doi:10.1126/science.1147182
- Baek, I.C., Kim, J.K., Cho, K.-H., Cha, D.-S., Cho, J.-W., Park, J.K., Song, C.-W., Yoon, S.K., 2009. A novel mutation in Hr causes abnormal hair follicle morphogenesis in hairpoor mouse, an animal model for Marie Unna Hereditary Hypotrichosis. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 20, 350–358. doi:10.1007/s00335-009-9191-8
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., Le Hir, H., 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12, 861–869. doi:10.1038/nsmb990
- Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., Tampe, R., 2011a. Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proc. Natl. Acad. Sci.* 108, 3228–3233. doi:10.1073/pnas.1015953108
- Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., Tampe, R., 2011b. Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proc. Natl. Acad. Sci.* 108, 3228–3233. doi:10.1073/pnas.1015953108
- Baserga, S.J., Benz, E.J., 1992. Beta-globin nonsense mutation: deficient accumulation of mRNA occurs despite normal cytoplasmic stability. *Proc. Natl. Acad. Sci. U. S. A.* 89, 2935–2939.
- Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret, D., Namy, O., 2014. Genome-wide translational changes induced by the prion [PSI⁺]. *Cell Rep.* 8, 439–448. doi:10.1016/j.celrep.2014.06.036
- Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., Giraldez, A.J., 2014. Identification of small ORFs in

- vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993. doi:10.1002/embj.201488411
- Bazzini, A.A., Lee, M.T., Giraldez, A.J., 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336, 233–237. doi:10.1126/science.1215704
- Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.-P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., Karcher, A., Thomm, M., Hopfner, K.-P., Green, R., Beckmann, R., 2012a. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 482, 501–506. doi:10.1038/nature10829
- Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.-P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., Karcher, A., Thomm, M., Hopfner, K.-P., Green, R., Beckmann, R., 2012b. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 482, 501–506. doi:10.1038/nature10829
- Bersano, A., Ballabio, E., Bresolin, N., Candelise, L., 2008. Genetic polymorphisms for the study of multifactorial stroke. *Hum. Mutat.* 29, 776–795. doi:10.1002/humu.20666
- Beznošková, P., Cuchalová, L., Wagner, S., Shoemaker, C.J., Gunišová, S., von der Haar, T., Valášek, L.S., 2013. Translation Initiation Factors eIF3 and HCR1 Control Translation Termination and Stop Codon Read-Through in Yeast Cells. *PLoS Genet.* 9, e1003962. doi:10.1371/journal.pgen.1003962
- Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., Peltz, S.W., 2000. Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA N. Y.* N 6, 1226–1235.
- Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W., Kulozik, A.E., 2010. NMD: RNA biology meets human genetic medicine. *Biochem. J.* 430, 365–377. doi:10.1042/BJ20100699
- Bianchin, C., Mauxion, F., Sentis, S., Séraphin, B., Corbo, L., 2005. Conservation of the deadenylase activity of proteins of the Caf1 family in human. *RNA N. Y.* N 11, 487–494. doi:10.1261/rna.7135305
- Blais, J.D., Filipenko, V., Bi, M., Harding, H.P., Ron, D., Koumenis, C., Wouters, B.G., Bell, J.C., 2004. Activating Transcription Factor 4 Is Translationally Regulated by Hypoxic Stress. *Mol. Cell. Biol.* 24, 7469–7482. doi:10.1128/MCB.24.17.7469-7482.2004
- Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M.J., Namy, O., 2015. New insights into stop codon recognition by eRF1. *Nucleic Acids Res.* 43, 3298–3308. doi:10.1093/nar/gkv154
- Bonetti, B., Fu, L., Moon, J., Bedwell, D.M., 1995. The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 251, 334–345. doi:10.1006/jmbi.1995.0438
- Brar, G.A., Weissman, J.S., 2015a. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664. doi:10.1038/nrm4069
- Brar, G.A., Weissman, J.S., 2015b. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664. doi:10.1038/nrm4069
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., Weissman, J.S., 2012a. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557. doi:10.1126/science.1215110
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., Weissman, J.S., 2012b. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557. doi:10.1126/science.1215110
- Brito, M., Malta-Vacas, J., Carmona, B., Aires, C., Costa, P., Martins, A.P., Ramos, S., Conde, A.R., Monteiro, C., 2005. Polyglycine expansions in eRF3/GSPT1 are associated with gastric cancer susceptibility. *Carcinogenesis* 26, 2046–2049. doi:10.1093/carcin/bgi168
- Brown, A., Fernández, I.S., Gordiyenko, Y., Ramakrishnan, V., 2016. Ribosome-dependent activation of stringent control. *Nature* 534, 277–280. doi:10.1038/nature17675
- Brown, C.E., Sachs, A.B., 1998. Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. *Mol. Cell. Biol.* 18, 6548–6559.

- Brown, C.M., Stockwell, P.A., Trotman, C.N., Tate, W.P., 1990. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res.* 18, 6339–6345.
- Brown, P.O., Botstein, D., 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21, 33–37. doi:10.1038/4462
- Bulygin, K.N., Khairulina, Y.S., Kolosov, P.M., Ven'yaminova, A.G., Graifer, D.M., Vorobjev, Y.N., Frolova, L.Y., Kisselev, L.L., Karpova, G.G., 2010. Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome. *RNA N. Y. N* 16, 1902–1914. doi:10.1261/rna.2066910
- Büttner, K., Wenig, K., Hopfner, K.-P., 2006. The exosome: a macromolecular cage for controlled RNA degradation. *Mol. Microbiol.* 61, 1372–1379. doi:10.1111/j.1365-2958.2006.05331.x
- Calvo, S.E., Pagliarini, D.J., Mootha, V.K., 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci.* 106, 7507–7512. doi:10.1073/pnas.0810916106
- Camier, S., Séraphin, B., 2007. [Destroy this (RNA) message after reading it!]. *Med. Sci. MS* 23, 850–856. doi:10.1051/medsci/20072310850
- Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L.A., Offenhäuser, N., Marchisio, P.C., Biffo, S., 2003. Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* 426, 579–584. doi:10.1038/nature02160
- Cenik, C., Cenik, E.S., Byeon, G.W., Grubert, F., Candille, S.I., Spacek, D., Alsallakh, B., Tilgner, H., Araya, C.L., Tang, H., Ricci, E., Snyder, M.P., 2015. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* 25, 1610–1621. doi:10.1101/gr.193342.115
- Chauvin, C., Jean-Jean, O., 2008. Proteasomal degradation of human release factor eRF3a regulates translation termination complex formation. *RNA N. Y. N* 14, 240–245. doi:10.1261/rna.728608
- Chauvin, C., Salhi, S., Jean-Jean, O., 2007. Human eukaryotic release factor 3a depletion causes cell cycle arrest at G1 phase through inhibition of the mTOR pathway. *Mol. Cell. Biol.* 27, 5619–5629. doi:10.1128/MCB.00035-07
- Chauvin, C., Salhi, S., Le Goff, C., Viranaicken, W., Diop, D., Jean-Jean, O., 2005. Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. *Mol. Cell. Biol.* 25, 5801–5811. doi:10.1128/MCB.25.14.5801-5811.2005
- Chavatte, L., Seit-Nebi, A., Dubovaya, V., Favre, A., 2002. The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J.* 21, 5302–5311.
- Chen, Z., Dong, J., Ishimura, A., Daar, I., Hinnebusch, A.G., Dean, M., 2006. The Essential Vertebrate ABCE1 Protein Interacts with Eukaryotic Initiation Factors. *J. Biol. Chem.* 281, 7452–7457. doi:10.1074/jbc.M510603200
- Cheng, Z., Saito, K., Pisarev, A.V., Wada, M., Pisareva, V.P., Pestova, T.V., Gajda, M., Round, A., Kong, C., Lim, M., Nakamura, Y., Svergun, D.I., Ito, K., Song, H., 2009. Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.* 23, 1106–1118. doi:10.1101/gad.1770109
- Chiu, W.-L., Wagner, S., Herrmannová, A., Burela, L., Zhang, F., Saini, A.K., Valásek, L., Hinnebusch, A.G., 2010. The C-terminal region of eukaryotic translation initiation factor 3a (eIF3a) promotes mRNA recruitment, scanning, and, together with eIF3j and the eIF3b RNA recognition motif, selection of AUG start codons. *Mol. Cell. Biol.* 30, 4415–4434. doi:10.1128/MCB.00280-10
- Corcos, L., Solier, S., 2005. **Épissage alternatif, pathologie et thérapeutique moléculaire.** *médecine/sciences* 21, 253–260. doi:10.1051/medsci/2005213253
- Cougot, N., van Dijk, E., Babajko, S., Séraphin, B., 2004. “Cap-tabolism.” *Trends Biochem. Sci.* 29, 436–444. doi:10.1016/j.tibs.2004.06.008
- Culbertson, M.R., Leeds, P.F., 2003. Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.* 13, 207–214.

- Czaplinski, K., Majlesi, N., Banerjee, T., Peltz, S.W., 2000. Mtt1 is a Upf1-like helicase that interacts with the translation termination factors and whose overexpression can modulate termination efficiency. *RNA N. Y. N* 6, 730–743.
- Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., Peltz, S.W., 1998. The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* 12, 1665–1677.
- Das, S., Maitra, U., 2001. Functional significance and mechanism of eIF5-promoted GTP hydrolysis in eukaryotic translation initiation. *Prog. Nucleic Acid Res. Mol. Biol.* 70, 207–231.
- Daugeron, M.C., Mauxion, F., Séraphin, B., 2001. The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* 29, 2448–2455.
- De Benedetti, A., Harris, A.L., 1999. eIF4E expression in tumors: its possible role in progression of malignancies. *Int. J. Biochem. Cell Biol.* 31, 59–72.
- des Georges, A., Hashem, Y., Unbehaun, A., Grassucci, R.A., Taylor, D., Hellen, C.U.T., Pestova, T.V., Frank, J., 2014. Structure of the mammalian ribosomal pre-termination complex associated with eRF1.eRF3.GDPNP. *Nucleic Acids Res.* 42, 3409–3418. doi:10.1093/nar/gkt1279
- Dever, T.E., Green, R., 2012. The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* 4, a013706–a013706. doi:10.1101/cshperspect.a013706
- Dong, J., Qiu, H., Garcia-Barrio, M., Anderson, J., Hinnebusch, A.G., 2000. Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. *Mol. Cell* 6, 269–279.
- Dong, Z., Zhang, J.-T., 2006. Initiation factor eIF3 and regulation of mRNA translation, cell growth, and cancer. *Crit. Rev. Oncol. Hematol.* 59, 169–180. doi:10.1016/j.critrevonc.2006.03.005
- Dugeon, G., Jean-Jean, O., Frolova, L., Le Goff, X., Philippe, M., Kisselev, L., Haenni, A.L., 1997. Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA. *Nucleic Acids Res.* 25, 2254–2258.
- Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., Weissman, J.S., 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* 2. doi:10.7554/eLife.01179
- Durand, S., Lykke-Andersen, J., 2013. Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat. Struct. Mol. Biol.* 20, 702–709. doi:10.1038/nsmb.2575
- Dziembowski, A., Lorentzen, E., Conti, E., Séraphin, B., 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat. Struct. Mol. Biol.* 14, 15–22. doi:10.1038/nsmb1184
- Eberle, A.B., Lykke-Andersen, S., Mühlemann, O., Jensen, T.H., 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* 16, 49–55. doi:10.1038/nsmb.1530
- Eulalio, A., Behm-Ansmant, I., Izaurralde, E., 2007. P bodies: at the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.* 8, 9–22. doi:10.1038/nrm2080
- Fatscher, T., Boehm, V., Gehring, N.H., 2015. Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. *Cell. Mol. Life Sci.* 72, 4523–4544. doi:10.1007/s00018-015-2017-9
- Feng, T., Yamamoto, A., Wilkins, S.E., Sokolova, E., Yates, L.A., Münzel, M., Singh, P., Hopkinson, R.J., Fischer, R., Cockman, M.E., Shelley, J., Trudgian, D.C., Schödel, J., McCullagh, J.S.O., Ge, W., Kessler, B.M., Gilbert, R.J., Frolova, L.Y., Alkalaeva, E., Ratcliffe, P.J., Schofield, C.J., Coleman, M.L., 2014. Optimal translational termination requires C4 lysyl hydroxylation of eRF1. *Mol. Cell* 53, 645–654. doi:10.1016/j.molcel.2013.12.028
- Finoux, A.-L., Séraphin, B., 2006. In vivo targeting of the yeast Pop2 deadenylase subunit to reporter transcripts induces their rapid degradation and generates new decay intermediates. *J. Biol. Chem.* 281, 25940–25947. doi:10.1074/jbc.M600132200

- Franks, T.M., Singh, G., Lykke-Andersen, J., 2010. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell* 143, 938–950. doi:10.1016/j.cell.2010.11.043
- Freistroffer, D.V., Pavlov, M.Y., MacDougall, J., Buckingham, R.H., Ehrenberg, M., 1997. Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J.* 16, 4126–4133. doi:10.1093/emboj/16.13.4126
- Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., Brosch, M., 2012a. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 22, 2208–2218. doi:10.1101/gr.139568.112
- Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., Brosch, M., 2012b. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 22, 2208–2218. doi:10.1101/gr.139568.112
- Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., Conti, E., 2005. SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell* 17, 537–547. doi:10.1016/j.molcel.2005.01.010
- Furic, L., Rong, L., Larsson, O., Koumakpayi, I.H., Yoshida, K., Brueschke, A., Petroulakis, E., Robichaud, N., Pollak, M., Gaboury, L.A., Pandolfi, P.P., Saad, F., Sonenberg, N., 2010. eIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14134–14139. doi:10.1073/pnas.1005320107
- Gardner, L.B., 2008. Hypoxic Inhibition of Nonsense-Mediated RNA Decay Regulates Gene Expression and the Integrated Stress Response. *Mol. Cell Biol.* 28, 3729–3741. doi:10.1128/MCB.02284-07
- Gatfield, D., Izaurralde, E., 2004. Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature* 429, 575–578. doi:10.1038/nature02559
- Geballe, A.P., Morris, D.R., 1994. Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* 19, 159–164. doi:10.1016/0968-0004(94)90277-1
- Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., Kulozik, A.E., 2005. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* 20, 65–75. doi:10.1016/j.molcel.2005.08.012
- Gehring, N.H., Lamprinaki, S., Hentze, M.W., Kulozik, A.E., 2009. The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.* 7, e1000120. doi:10.1371/journal.pbio.1000120
- Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., Kulozik, A.E., 2003. Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* 11, 939–949.
- Gerashchenko, M.V., Gladyshev, V.N., 2014. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res.* 42, e134. doi:10.1093/nar/gku671
- Gerashchenko, M.V., Lobanov, A.V., Gladyshev, V.N., 2012a. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17394–17399. doi:10.1073/pnas.1120799109
- Gerashchenko, M.V., Lobanov, A.V., Gladyshev, V.N., 2012b. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17394–17399. doi:10.1073/pnas.1120799109
- Giorgi, C., Moore, M.J., 2007. The nuclear nurture and cytoplasmic nature of localized mRNPs. *Semin. Cell Dev. Biol.* 18, 186–193. doi:10.1016/j.semcdb.2007.01.002
- Gosselin, P., Oulhen, N., Jam, M., Ronzca, J., Cormier, P., Czjzek, M., Cosson, B., 2011. The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS. *Nucleic Acids Res.* 39, 3496–3503. doi:10.1093/nar/gkq1306
- Gunišová, S., Beznosková, P., Mohammad, M.P., Vlčková, V., Valášek, L.S., 2016. In-depth analysis of *cis*-determinants that either promote or inhibit reinitiation on *GCN4* mRNA after translation of its four short uORFs. *RNA* 22, 542–558. doi:10.1261/rna.055046.115

- Guo, H., Ingolia, N.T., Weissman, J.S., Bartel, D.P., 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840. doi:10.1038/nature09267
- Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., Lander, E.S., 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251. doi:10.1016/j.cell.2013.06.009
- Guydosh, N.R., Green, R., 2014a. Dom34 Rescues Ribosomes in 3' Untranslated Regions. *Cell* 156, 950–962. doi:10.1016/j.cell.2014.02.006
- Guydosh, N.R., Green, R., 2014b. Dom34 Rescues Ribosomes in 3' Untranslated Regions. *Cell* 156, 950–962. doi:10.1016/j.cell.2014.02.006
- Hannan, K.M., Brandenburger, Y., Jenkins, A., Sharkey, K., Cavanaugh, A., Rothblum, L., Moss, T., Poortinga, G., McArthur, G.A., Pearson, R.B., Hannan, R.D., 2003. mTOR-dependent regulation of ribosomal gene transcription requires S6K1 and is mediated by phosphorylation of the carboxy-terminal activation domain of the nucleolar transcription factor UBF. *Mol. Cell. Biol.* 23, 8862–8877.
- Hayden, C.A., Jorgensen, R.A., 2007. Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biol.* 5, 32. doi:10.1186/1741-7007-5-32
- Hershey, J.W., Asano, K., Naranda, T., Vornlocher, H.P., Hanachi, P., Merrick, W.C., 1996. Conservation and diversity in the structure of translation initiation factor EIF3 from humans and yeast. *Biochimie* 78, 903–907.
- Hinnebusch, A.G., 2014. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* 83, 779–812. doi:10.1146/annurev-biochem-060713-035802
- Hinnebusch, A.G., 2011. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* 75, 434–467. doi:10.1128/MMBR.00008-11
- Hinnebusch, A.G., 2006. eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem. Sci.* 31, 553–562. doi:10.1016/j.tibs.2006.08.005
- Hinnebusch, A.G., Dever, T.E., Asano, K., 2007. 9 Mechanism of Translation Initiation in the Yeast *Saccharomyces cerevisiae*. *Cold Spring Harb. Monogr. Arch.* 48, 225–268.
- Holcik, M., Sonenberg, N., 2005. Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol.* 6, 318–327. doi:10.1038/nrm1618
- Hoshino, S., Imai, M., Kobayashi, T., Uchida, N., Katada, T., 1999. The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J. Biol. Chem.* 274, 16677–16680.
- Hoskins, A.A., Moore, M.J., 2012. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.* 37, 179–188. doi:10.1016/j.tibs.2012.02.009
- Huber, R., Pietsch, D., Panterodt, T., Brand, K., 2012. Regulation of C/EBP β and resulting functions in cells of the monocytic lineage. *Cell. Signal.* 24, 1287–1296. doi:10.1016/j.cellsig.2012.02.007
- Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., Izaurralde, E., 2008. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA N. Y. N* 14, 2609–2617. doi:10.1261/rna.1386208
- Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S., Press, W.H., 2015. Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLoS Genet.* 11, e1005732. doi:10.1371/journal.pgen.1005732
- Iacono, M., Mignone, F., Pesole, G., 2005. uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* 349, 97–105. doi:10.1016/j.gene.2004.11.041
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., Weissman, J.S., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.

- Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., Weissman, J.S., 2014a. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379. doi:10.1016/j.celrep.2014.07.045
- Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., Weissman, J.S., 2014b. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379. doi:10.1016/j.celrep.2014.07.045
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 324, 218–223. doi:10.1126/science.1168978
- Ingolia, N.T., Lareau, L.F., Weissman, J.S., 2011a. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi:10.1016/j.cell.2011.10.002
- Ingolia, N.T., Lareau, L.F., Weissman, J.S., 2011b. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi:10.1016/j.cell.2011.10.002
- Irigoyen, N., Firth, A.E., Jones, J.D., Chung, B.Y.-W., Siddell, S.G., Brierley, I., 2016. High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS Pathog.* 12, e1005473. doi:10.1371/journal.ppat.1005473
- Ishigaki, Y., Li, X., Serin, G., Maquat, L.E., 2001. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106, 607–617.
- Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., Maquat, L.E., 2008. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* 133, 314–327. doi:10.1016/j.cell.2008.02.030
- Isken, O., Maquat, L.E., 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* 9, 699–712. doi:10.1038/nrg2402
- Isken, O., Maquat, L.E., 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* 21, 1833–1856. doi:10.1101/gad.1566807
- Ivanov, I.P., Firth, A.E., Michel, A.M., Atkins, J.F., Baranov, P.V., 2011a. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* 39, 4220–4234. doi:10.1093/nar/gkr007
- Ivanov, I.P., Firth, A.E., Michel, A.M., Atkins, J.F., Baranov, P.V., 2011b. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* 39, 4220–4234. doi:10.1093/nar/gkr007
- Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., Kulozik, A.E., 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* 27, 736–747. doi:10.1038/emboj.2008.17
- Izaurralde, E., Mattaj, I.W., 1995. RNA export. *Cell* 81, 153–159.
- Jackson, R.J., Hellen, C.U.T., Pestova, T.V., 2012. Termination and post-termination events in eukaryotic translation. *Adv. Protein Chem. Struct. Biol.* 86, 45–93. doi:10.1016/B978-0-12-386497-0.00002-5
- Jackson, R.J., Hellen, C.U.T., Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127. doi:10.1038/nrm2838
- Janosi, L., Hara, H., Zhang, S., Kaji, A., 1996. Ribosome recycling by ribosome recycling factor (RRF)—an important but overlooked step of protein biosynthesis. *Adv. Biophys.* 32, 121–201.
- Janzen, D.M., Frolova, L., Geballe, A.P., 2002. Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Mol. Cell. Biol.* 22, 8562–8570.
- Jerbi, S., Jolles, B., Bouceba, T., Jean-Jean, O., 2016. Studies on human eRF3-PABP interaction reveal the influence of eRF3a N-terminal glycine repeat on eRF3-PABP binding affinity and the lower affinity

- of eRF3a 12-GGC allele involved in cancer susceptibility. *RNA Biol.* 13, 306–315. doi:10.1080/15476286.2015.1137421
- Jia, Y., Polunovsky, V., Bitterman, P.B., Wagner, C.R., 2012. Cap-dependent translation initiation factor eIF4E: an emerging anticancer drug target. *Med. Res. Rev.* 32, 786–814. doi:10.1002/med.21260
- Jin, H., Kelley, A.C., Loakes, D., Ramakrishnan, V., 2010. Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8593–8598. doi:10.1073/pnas.1003995107
- Johnstone, T.G., Bazzini, A.A., Giraldez, A.J., 2016. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* 35, 706–723. doi:10.15252/embj.201592759
- Johnstone, T.G., Bazzini, A.A., Giraldez, A.J., 2016. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* 35, 706–723. doi:10.15252/embj.201592759
- Joshi, B., Cameron, A., Jagus, R., 2004. Characterization of mammalian eIF4E-family members. *Eur. J. Biochem.* 271, 2189–2203. doi:10.1111/j.1432-1033.2004.04149.x
- Jurica, M.S., Moore, M.J., 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14.
- Jurica, M.S., Moore, M.J., 2002. Capturing splicing complexes to study structure and mechanism. *Methods San Diego Calif* 28, 336–345.
- Kahvejian, A., Roy, G., Sonenberg, N., 2001. The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb. Symp. Quant. Biol.* 66, 293–300.
- Kapp, L.D., Lorsch, J.R., 2004. GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J. Mol. Biol.* 335, 923–936.
- Karagoyozov, L., Godfrey, R., Bohmer, S.-A., Petermann, A., Holters, S., Ostman, A., Bohmer, F.-D., 2008. The structure of the 5'-end of the protein-tyrosine phosphatase PTPRJ mRNA reveals a novel mechanism for translation attenuation. *Nucleic Acids Res.* 36, 4443–4453. doi:10.1093/nar/gkn391
- Karcher, A., Schele, A., Hopfner, K.-P., 2008. X-ray structure of the complete ABC enzyme ABCE1 from *Pyrococcus abyssi*. *J. Biol. Chem.* 283, 7962–7971. doi:10.1074/jbc.M707347200
- Kaul, G., Pattan, G., Rafeequi, T., 2011. Eukaryotic elongation factor-2 (eEF2): its regulation and peptide chain elongation. *Cell Biochem. Funct.* 29, 227–234. doi:10.1002/cbf.1740
- Keeling, K.M., Lanier, J., Du, M., Salas-Marco, J., Gao, L., Kaenjak-Angeletti, A., Bedwell, D.M., 2004. Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in *S. cerevisiae*. *RNA N. Y. N* 10, 691–703.
- Khoshnevis, S., Gross, T., Rotte, C., Baierlein, C., Ficner, R., Krebber, H., 2010. The iron-sulphur protein RNase L inhibitor functions in translation termination. *EMBO Rep.* 11, 214–219. doi:10.1038/embor.2009.272
- Kikuchi, Y., Shimatake, H., Kikuchi, A., 1988. A yeast gene required for the G1-to-S transition encodes a protein containing an A-kinase target site and GTPase domain. *EMBO J.* 7, 1175–1182.
- Kilberg, M.S., Shan, J., Su, N., 2009. ATF4-dependent transcription mediates signaling of amino acid limitation. *Trends Endocrinol. Metab. TEM* 20, 436–443. doi:10.1016/j.tem.2009.05.008
- Kim, K.M., Cho, H., Choi, K., Kim, J., Kim, B.-W., Ko, Y.-G., Jang, S.K., Kim, Y.K., 2009. A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation. *Genes Dev.* 23, 2033–2045. doi:10.1101/gad.1823409
- Kimball, S.R., 1999. Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.* 31, 25–29.
- Kolosov, P., Frolova, L., Seit-Nebi, A., Dubovaya, V., Kononenko, A., Oparina, N., Justesen, J., Efimov, A., Kisselev, L., 2005. Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res.* 33, 6418–6425. doi:10.1093/nar/gki927
- Komar, A.A., Hatzoglou, M., 2015. Exploring Internal Ribosome Entry Sites as Therapeutic Targets. *Front. Oncol.* 5. doi:10.3389/fonc.2015.00233
- Kong, C., Ito, K., Walsh, M.A., Wada, M., Liu, Y., Kumar, S., Barford, D., Nakamura, Y., Song, H., 2004. Crystal structure and functional analysis of the eukaryotic class II release factor eRF3 from *S. pombe*. *Mol. Cell* 14, 233–245.

- Kononenko, A.V., Mitkevich, V.A., Dubovaya, V.I., Kolosov, P.M., Makarov, A.A., Kisselev, L.L., 2008. Role of the individual domains of translation termination factor eRF1 in GTP binding to eRF3. *Proteins* 70, 388–393. doi:10.1002/prot.21544
- Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299, 1–34.
- Kozak, M., 2001. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res.* 29, 5226–5232.
- Kozak, M., 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8301–8305.
- Kozak, M., 1989. The scanning model for translation: an update. *J. Cell Biol.* 108, 229–241.
- Kozak, M., 1987a. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15, 8125–8148.
- Kozak, M., 1987b. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol. Cell. Biol.* 7, 3438–3445.
- Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292.
- Kryuchkova, P., Grishin, A., Eliseev, B., Karyagina, A., Frolova, L., Alkalaeva, E., 2013. Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic Acids Res.* 41, 4573–4586. doi:10.1093/nar/gkt113
- Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., Gehring, N.H., 2006. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *RNA N. Y. N* 12, 1015–1022. doi:10.1261/rna.12506
- Lareau, L.F., Hite, D.H., Hogan, G.J., Brown, P.O., 2014. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* 3, e01257. doi:10.7554/eLife.01257
- Lazaris-Karatzas, A., Montine, K.S., Sonenberg, N., 1990. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* 345, 544–547. doi:10.1038/345544a0
- Le Bacquer, O., Martineau, Y., Mamane, Y., 2006. [When translation arises from its TORpor]. *Med. Sci. MS* 22, 514–518. doi:10.1051/medsci/2006225514
- Le Goff, X., Philippe, M., Jean-Jean, O., 1997. Overexpression of human release factor 1 alone has an antisuppressor effect in human cells. *Mol. Cell. Biol.* 17, 3164–3172.
- Le Hir, H., 2000. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* 19, 6860–6869. doi:10.1093/emboj/19.24.6860
- Le Hir, H., Moore, M.J., Maquat, L.E., 2000. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* 14, 1098–1108.
- Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., Qian, S.-B., 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* 109, E2424–E2432. doi:10.1073/pnas.1207846109
- Lewis, J.D., Izaurralde, E., 1997. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* 247, 461–469.
- Liang, H., Wong, J.Y., Bao, Q., Cavalcanti, A.R.O., Landweber, L.F., 2005. Decoding the Decoding Region: Analysis of Eukaryotic Release Factor (eRF1) Stop Codon-Binding Residues. *J. Mol. Evol.* 60, 337–344. doi:10.1007/s00239-004-0211-8
- Liljas, A., 2013. Zooming in on eukaryotic translation initiation. *Nat. Struct. Mol. Biol.* 20, 1141–1142. doi:10.1038/nsmb.2684
- Linde, L., Kerem, B., 2011. Nonsense-mediated mRNA decay and cystic fibrosis. *Methods Mol. Biol. Clifton NJ* 741, 137–154. doi:10.1007/978-1-61779-117-8_10
- Liu, T., Li, R., Pan, T., Liu, D., Petersen, R.B., Wong, B.-S., Gambetti, P., Sy, M.S., 2002. Intercellular transfer of the cellular prion protein. *J. Biol. Chem.* 277, 47671–47678. doi:10.1074/jbc.M207458200

- Loughran, G., Sachs, M.S., Atkins, J.F., Ivanov, I.P., 2012. Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res.* 40, 2898–2906. doi:10.1093/nar/gkr1192
- Lovett, P.S., Rogers, E.J., 1996. Ribosome regulation by the nascent peptide. *Microbiol. Rev.* 60, 366–385.
- Lu, P.D., Harding, H.P., Ron, D., 2004. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J. Cell Biol.* 167, 27–33. doi:10.1083/jcb.200408003
- Lykke-Andersen, J., Shu, M.D., Steitz, J.A., 2000. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* 103, 1121–1131.
- Lykke-Andersen, S., Jensen, T.H., 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* 16, 665–677. doi:10.1038/nrm4063
- Majumdar, R., Bandyopadhyay, A., Maitra, U., 2003. Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J. Biol. Chem.* 278, 6580–6587. doi:10.1074/jbc.M210357200
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., Lührmann, R., 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205–2208. doi:10.1126/science.1077783
- Malta-Vacas, J., Aires, C., Costa, P., Conde, A.R., Ramos, S., Martins, A.P., Monteiro, C., Brito, M., 2005. Differential expression of the eukaryotic release factor 3 (eRF3/GSPT1) according to gastric cancer histological types. *J. Clin. Pathol.* 58, 621–625. doi:10.1136/jcp.2004.021774
- Mamane, Y., Petroulakis, E., Martineau, Y., Sato, T.-A., Larsson, O., Rajasekhar, V.K., Sonenberg, N., 2007. Epigenetic activation of a subset of mRNAs by eIF4E explains its effects on cell proliferation. *PloS One* 2, e242. doi:10.1371/journal.pone.0000242
- Maquat, L.E., Hwang, J., Sato, H., Tang, Y., 2010. CBP80-promoted mRNP rearrangements during the pioneer round of translation, nonsense-mediated mRNA decay, and thereafter. *Cold Spring Harb. Symp. Quant. Biol.* 75, 127–134. doi:10.1101/sqb.2010.75.028
- Maquat, L.E., Li, X., 2001. Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA N. Y. N* 7, 445–456.
- Marchildon, F., Lala, N., Li, G., St-Louis, C., Lamothe, D., Keller, C., Wiper-Bergeron, N., 2012. CCAAT/enhancer binding protein beta is expressed in satellite cells and controls myogenesis. *Stem Cells Dayt. Ohio* 30, 2619–2630. doi:10.1002/stem.1248
- Martin, F., Barends, S., Jaeger, S., Schaeffer, L., Prongidi-Fix, L., Eriani, G., 2011. Cap-assisted internal initiation of translation of histone H4. *Mol. Cell* 41, 197–209. doi:10.1016/j.molcel.2010.12.019
- Matheisl, S., Berninghausen, O., Becker, T., Beckmann, R., 2015. Structure of a human translation termination complex. *Nucleic Acids Res.* 43, 8615–8626. doi:10.1093/nar/gkv909
- Maurin, A.-C., Jousse, C., Balage, M., Averous, J., Parry, L., Bruhat, A., Cherasse, Y., Zeng, H., Zhang, Y., Harding, H., Ron, D., Fafournoux, P., 2005. [GCN2 regulates feeding behavior to maintain amino acid homeostasis in omnivores]. *Med. Sci. MS* 21, 799–801. doi:10.1051/medsci/20052110799
- McGlinicy, N.J., Smith, C.W.J., 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* 33, 385–393. doi:10.1016/j.tibs.2008.06.001
- McGlinicy, N.J., Tan, L.-Y., Paul, N., Zavolan, M., Lilley, K.S., Smith, C.W., 2010. Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics* 11, 565. doi:10.1186/1471-2164-11-565
- Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., Dietz, H.C., 2001. Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* 10, 99–105.
- Mendell, J.T., ap Rhys, C.M.J., Dietz, H.C., 2002. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* 298, 419–422. doi:10.1126/science.1074428

- Mendell, J.T., Medghalchi, S.M., Lake, R.G., Noensie, E.N., Dietz, H.C., 2000. Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol. Cell. Biol.* 20, 8944–8957.
- Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., Dietz, H.C., 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* 36, 1073–1078. doi:10.1038/ng1429
- Michel, A.M., Andreev, D.E., Baranov, P.V., 2014. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics* 15. doi:10.1186/s12859-014-0380-4
- Michel, A.M., Baranov, P.V., 2013. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale: Ribosome profiling. *Wiley Interdiscip. Rev. RNA* 4, 473–490. doi:10.1002/wrna.1172
- Miettinen, T.P., Björklund, M., 2015. Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res.* 43, 1019–1034. doi:10.1093/nar/gku1310
- Millevoi, S., Vagner, S., 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.* 38, 2757–2774. doi:10.1093/nar/gkp1176
- Miluzio, A., Beugnet, A., Volta, V., Biffo, S., 2009. Eukaryotic initiation factor 6 mediates a continuum between 60S ribosome biogenesis and translation. *EMBO Rep.* 10, 459–465. doi:10.1038/embor.2009.70
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., Tollervy, D., 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* 91, 457–466.
- Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.-H., Hinnebusch, A.G., Lorsch, J.R., 2010. The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol. Cell* 39, 950–962. doi:10.1016/j.molcel.2010.08.021
- Mitkevich, V.A., Kononenko, A.V., Petrushanko, I.Y., Yanvarev, D.V., Makarov, A.A., Kisselev, L.L., 2006. Termination of translation in eukaryotes is mediated by the quaternary eRF1*eRF3*GTP*Mg²⁺ complex. The biological roles of eRF3 and prokaryotic RF3 are profoundly distinct. *Nucleic Acids Res.* 34, 3947–3954. doi:10.1093/nar/gkl549
- Moore, M.J., Proudfoot, N.J., 2009. Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell* 136, 688–700. doi:10.1016/j.cell.2009.02.001
- Mottagui-Tabar, S., Tuite, M.F., Isaksson, L.A., 1998. The influence of 5' codon context on translation termination in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 257, 249–254.
- Munzarová, V., Pánek, J., Gunišová, S., Dányi, I., Szamecz, B., Valášek, L.S., 2011. Translation Reinitiation Relies on the Interaction between eIF3a/TIF32 and Progressively Folded cis-Acting mRNA Elements Preceding Short uORFs. *PLoS Genet.* 7, e1002137. doi:10.1371/journal.pgen.1002137
- Nakielný, S., Dreyfuss, G., 1997. Nuclear export of proteins and RNAs. *Curr. Opin. Cell Biol.* 9, 420–429.
- Namy, O., Hatin, I., Rousset, J.-P., 2001. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep.* 2, 787–793. doi:10.1093/embo-reports/kve176
- Nanda, J.S., Saini, A.K., Muñoz, A.M., Hinnebusch, A.G., Lorsch, J.R., 2013. Coordinated movements of eukaryotic translation initiation factors eIF1, eIF1A, and eIF5 trigger phosphate release from eIF2 in response to start codon recognition by the ribosomal preinitiation complex. *J. Biol. Chem.* 288, 5316–5329. doi:10.1074/jbc.M112.440693
- Nilsen, T.W., Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi:10.1038/nature08909
- Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., Silhavy, D., 2009. Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol. Biol.* 71, 367–378. doi:10.1007/s11103-009-9528-4

- Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., Weissman, J.S., Bukau, B., 2011. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* 147, 1295–1308. doi:10.1016/j.cell.2011.10.044
- Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., Ohno, S., 2003. Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol. Cell* 12, 1187–1200.
- Pain, V.M., 1996. Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* 236, 747–771.
- Palam, L.R., Baird, T.D., Wek, R.C., 2011. Phosphorylation of eIF2 Facilitates Ribosomal Bypass of an Inhibitory Upstream ORF to Enhance CHOP Translation. *J. Biol. Chem.* 286, 10939–10949. doi:10.1074/jbc.M110.216093
- Palii, S.S., Kays, C.E., Deval, C., Bruhat, A., Fafournoux, P., Kilberg, M.S., 2009. Specificity of amino acid regulated gene expression: analysis of genes subjected to either complete or single amino acid deprivation. *Amino Acids* 37, 79–88. doi:10.1007/s00726-008-0199-2
- Pantopoulos, K., 2004. Iron metabolism and the IRE/IRP regulatory system: an update. *Ann. N. Y. Acad. Sci.* 1012, 1–13.
- Parent, A., Bisailon, M., 2006. [Synergy between transcription and mRNA processing events]. *Med. Sci. MS* 22, 626–632. doi:10.1051/medsci/20062267626
- Park, J.-E., Yi, H., Kim, Y., Chang, H., Kim, V.N., 2016. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol. Cell* 62, 462–471. doi:10.1016/j.molcel.2016.04.007
- Parker, R., Sheth, U., 2007. P bodies and the control of mRNA translation and degradation. *Mol. Cell* 25, 635–646. doi:10.1016/j.molcel.2007.02.011
- Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., Ramakrishnan, V., 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell* 26, 41–50. doi:10.1016/j.molcel.2007.03.018
- Peabody, D.S., 1989. Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* 264, 5031–5035.
- Peixeiro, I., Inacio, A., Barbosa, C., Silva, A.L., Liebhaber, S.A., Romao, L., 2012. Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. *Nucleic Acids Res.* 40, 1160–1173. doi:10.1093/nar/gkr820
- Pelechano, V., Wei, W., Steinmetz, L.M., 2015. Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. *Cell* 161, 1400–1412. doi:10.1016/j.cell.2015.05.008
- Pestova, T.V., Kolupaeva, V.G., 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* 16, 2906–2922. doi:10.1101/gad.1020902
- Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E., Hellen, C.U., 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403, 332–335. doi:10.1038/35002118
- Pestova, T.V., Lorsch, J.R., Hellen, C.U., 2007. 4 The Mechanism of Translation Initiation in Eukaryotes. *Cold Spring Harb. Monogr. Arch.* 48, 87–128.
- Peter, D., Igreja, C., Weber, R., Wohlbold, L., Weiler, C., Ebertsch, L., Weichenrieder, O., Izaurralde, E., 2015. Molecular architecture of 4E-BP translational inhibitors bound to eIF4E. *Mol. Cell* 57, 1074–1087. doi:10.1016/j.molcel.2015.01.017
- Pisarev, A.V., Hellen, C.U.T., Pestova, T.V., 2007. Recycling of eukaryotic posttermination ribosomal complexes. *Cell* 131, 286–299. doi:10.1016/j.cell.2007.08.041
- Pisarev, A.V., Kolupaeva, V.G., Pisareva, V.P., Merrick, W.C., Hellen, C.U.T., Pestova, T.V., 2006. Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* 20, 624–636. doi:10.1101/gad.1397906
- Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U.T., Pestova, T.V., 2010. The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. *Mol. Cell* 37, 196–210. doi:10.1016/j.molcel.2009.12.034

- Pisareva, V.P., Pisarev, A.V., Hellen, C.U.T., Rodnina, M.V., Pestova, T.V., 2006. Kinetic analysis of interaction of eukaryotic release factor 3 with guanine nucleotides. *J. Biol. Chem.* 281, 40224–40235. doi:10.1074/jbc.M607461200
- Poyry, T.A.A., 2004. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev.* 18, 62–75. doi:10.1101/gad.276504
- Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., Beckmann, R., 2014. Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* 8, 59–65. doi:10.1016/j.celrep.2014.04.058
- Pyronnet, S., Imataka, H., Gingras, A.C., Fukunaga, R., Hunter, T., Sonenberg, N., 1999. Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* 18, 270–279. doi:10.1093/emboj/18.1.270
- Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S., Hanson, J., 2009. Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol.* 150, 1356–1367. doi:10.1104/pp.109.136036
- Raj, A., Wang, S.H., Shim, H., Harpak, A., Li, Y.I., Engelmann, B., Stephens, M., Gilad, Y., Pritchard, J.K., 2016. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *eLife* 5. doi:10.7554/eLife.13328
- Rajkowsch, L., Vilela, C., Berthelot, K., Ramirez, C.V., McCarthy, J.E.G., 2004. Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J. Mol. Biol.* 335, 71–85.
- Raney, A., Law, G.L., Mize, G.J., Morris, D.R., 2002. Regulated Translation Termination at the Upstream Open Reading Frame in S-Adenosylmethionine Decarboxylase mRNA. *J. Biol. Chem.* 277, 5988–5994. doi:10.1074/jbc.M108375200
- Rebbapragada, I., Lykke-Andersen, J., 2009. Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell Biol.* 21, 394–402. doi:10.1016/j.ceb.2009.02.007
- Rehwinkel, J., Raes, J., Izaurralde, E., 2006. Nonsense-mediated mRNA decay: target genes and functional diversification of effectors. *Trends Biochem. Sci.* 31, 639–646. doi:10.1016/j.tibs.2006.09.005
- Richardson, J.P., Stansfield, I., Innes, S., Minella, O., Bertram, G., 2001. Endless possibilities: translation termination and stop codon recognition. *Microbiology* 147, 255–269. doi:10.1099/00221287-147-2-255
- Rogers, G.W., Richter, N.J., Lima, W.F., Merrick, W.C., 2001. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* 276, 30914–30922. doi:10.1074/jbc.M100157200
- Rooijers, K., Loayza-Puch, F., Nijtmans, L.G., Agami, R., 2013. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat. Commun.* 4, 2886. doi:10.1038/ncomms3886
- Rousseau, D., Gingras, A.C., Pause, A., Sonenberg, N., 1996. The eIF4E-binding proteins 1 and 2 are negative regulators of cell growth. *Oncogene* 13, 2415–2420.
- Roy, B., Vaughn, J.N., Kim, B.-H., Zhou, F., Gilchrist, M.A., 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. doi:10.1261/rna.2056010
- Roy, B., von Arnim, A.G., 2013. Translational Regulation of Cytoplasmic mRNAs. *Arab. Book* 11, e0165. doi:10.1199/tab.0165
- Rozovsky, N., Butterworth, A.C., Moore, M.J., 2008. Interactions between eIF4AI and its accessory factors eIF4B and eIF4H. *RNA N. Y. N* 14, 2136–2148. doi:10.1261/rna.1049608
- Rufener, S.C., Mühlemann, O., 2013. eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* 20, 710–717. doi:10.1038/nsmb.2576
- Ruiz-Echevarria, M.J., Czaplinski, K., Peltz, S.W., 1996. Making sense of nonsense in yeast. *Trends Biochem. Sci.* 21, 433–438.
- Sachs, M.S., 2006. Downstream control of upstream open reading frames. *Genes Dev.* 20, 915–921. doi:10.1101/gad.1427006

- Salas-Marco, J., Bedwell, D.M., 2004. GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol. Cell. Biol.* 24, 7769–7778. doi:10.1128/MCB.24.17.7769-7778.2004
- Sanvito, F., Piatti, S., Villa, A., Bossi, M., Lucchini, G., Marchisio, P.C., Biffo, S., 1999. The beta4 integrin interactor p27(BBP/eIF6) is an essential nuclear matrix protein involved in 60S ribosomal subunit assembly. *J. Cell Biol.* 144, 823–837.
- Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., Amieux, P.S., 2009. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13939–13944. doi:10.1073/pnas.0907143106
- Sato, H., Maquat, L.E., 2009. Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes Dev.* 23, 2537–2550. doi:10.1101/gad.1817109
- Schafer, S., Adami, E., Heinig, M., Rodrigues, K.E.C., Kreuchwig, F., Silhavy, J., van Heesch, S., Simaite, D., Rajewsky, N., Cuppen, E., Pravenec, M., Vingron, M., Cook, S.A., Hubner, N., 2015. Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat. Commun.* 6, 7200. doi:10.1038/ncomms8200
- Scheper, G.C., Proud, C.G., 2002. Does phosphorylation of the cap-binding protein eIF4E play a role in translation initiation? *Eur. J. Biochem.* 269, 5350–5359.
- Scheper, G.C., van Kollenburg, B., Hu, J., Luo, Y., Goss, D.J., Proud, C.G., 2002. Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA. *J. Biol. Chem.* 277, 3303–3309. doi:10.1074/jbc.M103607200
- Schoenberg, D.R., Maquat, L.E., 2012. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* 13, 246–259. doi:10.1038/nrg3160
- Seit-Nebi, A., 2002. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep.* 3, 881–886. doi:10.1093/embo-reports/kvf178
- Seit-Nebi, A., Frolova, L., Justesen, J., Kisselev, L., 2001. Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucleic Acids Res.* 29, 3982–3987.
- Serdar, L.D., Whiteside, D.L., Baker, K.E., 2016. ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nat. Commun.* 7, 14021. doi:10.1038/ncomms14021
- Serin, G., Gersappe, A., Black, J.D., Aronoff, R., Maquat, L.E., 2001. Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol. Cell. Biol.* 21, 209–223. doi:10.1128/MCB.21.1.209-223.2001
- Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., Burge, C.B., 2013. Widespread regulation of translation by elongation pausing in heat shock. *Mol. Cell* 49, 439–452. doi:10.1016/j.molcel.2012.11.028
- Shatkin, A.J., 1976. Capping of eucaryotic mRNAs. *Cell* 9, 645–653.
- Shatkin, A.J., Manley, J.L., 2000. The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842. doi:10.1038/79583
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311.
- Shirokikh, N.E., Alkalaeva, E.Z., Vassilenko, K.S., Afonina, Z.A., Alekhina, O.M., Kisselev, L.L., Spirin, A.S., 2010. Quantitative analysis of ribosome-mRNA complexes at different translation stages. *Nucleic Acids Res.* 38, e15. doi:10.1093/nar/gkp1025
- Shoemaker, C.J., Green, R., 2011. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci.* 108, E1392–E1398. doi:10.1073/pnas.1113956108
- Silva, A.L., Ribeiro, P., Inacio, A., Liebhaber, S.A., Romao, L., 2008. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA* 14, 563–576. doi:10.1261/rna.815108

- Silva, A.L., 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. doi:10.1016/j.febslet.2008.12.058
- Simon, E., Camier, S., Séraphin, B., 2006. New insights into the control of mRNA decapping. *Trends Biochem. Sci.* 31, 241–243. doi:10.1016/j.tibs.2006.03.001
- Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., Tamarkin-Ben-Harush, A., Viollet, B., Dikstein, R., 2015. Translational Tolerance of Mitochondrial Genes to Metabolic Energy Stress Involves TISU and eIF1-eIF4GI Cooperation in Start Codon Selection. *Cell Metab.* 21, 479–492. doi:10.1016/j.cmet.2015.02.010
- Sivagnanasundaram, S., Morris, A.G., Gaitonde, E.J., McKenna, P.J., Mollon, J.D., Hunt, D.M., 2000. A cluster of single nucleotide polymorphisms in the 5' leader of the human dopamine D3 receptor gene (DRD3) and its relationship to schizophrenia. *Neurosci. Lett.* 279, 13–16.
- Sonenberg, N., Hinnebusch, A.G., 2009. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* 136, 731–745. doi:10.1016/j.cell.2009.01.042
- Sonenberg, N., Rupprecht, K.M., Hecht, S.M., Shatkin, A.J., 1979. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4345–4349.
- Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., Barford, D., 2000. The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100, 311–321.
- Soto-Rifo, R., Rubilar, P.S., Limousin, T., de Breyne, S., Décimo, D., Ohlmann, T., 2012. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J.* 31, 3745–3756. doi:10.1038/emboj.2012.220
- St Johnston, D., 1995. The intracellular localization of messenger RNAs. *Cell* 81, 161–170.
- Steitz, J.A., 1969. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* 224, 957–964.
- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T.K., Hein, M.Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., Mann, M., Ingolia, N.T., Weissman, J.S., 2012. Decoding human cytomegalovirus. *Science* 338, 1088–1093. doi:10.1126/science.1227919
- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T.K., Hein, M.Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., Mann, M., Ingolia, N.T., Weissman, J.S., 2012. Decoding Human Cytomegalovirus. *Science* 338, 1088–1093. doi:10.1126/science.1227919
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., Bartel, D.P., 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71. doi:10.1038/nature13007
- Sun, X., Perlick, H.A., Dietz, H.C., Maquat, L.E., 1998. A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10009–10014.
- Tange, T.Ø., Shibuya, T., Jurica, M.S., Moore, M.J., 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA N. Y. N* 11, 1869–1883. doi:10.1261/rna.2155905
- Tian, B., Graber, J.H., 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3, 385–396. doi:10.1002/wrna.116
- Tikole, S., Sankaramakrishnan, R., 2006. A survey of mRNA sequences with a non-AUG start codon in RefSeq database. *J. Biomol. Struct. Dyn.* 24, 33–42. doi:10.1080/07391102.2006.10507096
- Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhrad, D., Parker, R., 2002. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *EMBO J.* 21, 1427–1436. doi:10.1093/emboj/21.6.1427
- Unbehaun, A., Borukhov, S.I., Hellen, C.U.T., 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. doi:10.1101/gad.1255704

- Urlaub, G., Mitchell, P.J., Ciudad, C.J., Chasin, L.A., 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* 9, 2868–2880.
- Valásek, L.S., 2012. “Ribozomin”-translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr. Protein Pept. Sci.* 13, 305–330.
- van Dijk, E., Le Hir, H., Séraphin, B., 2003. DcpS can act in the 5′-3′ mRNA decay pathway in addition to the 3′-5′ pathway. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12081–12086. doi:10.1073/pnas.1635192100
- van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. *Trends Genet. TIG* 30, 418–426. doi:10.1016/j.tig.2014.07.001
- Vattem, K.M., Wek, R.C., 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11269–11274. doi:10.1073/pnas.0400541101
- Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi:10.1038/nrg3185
- von Arnim, A.G., Jia, Q., Vaughn, J.N., 2014. Regulation of plant translation by upstream open reading frames. *Plant Sci. Int. J. Exp. Plant Biol.* 214, 1–12. doi:10.1016/j.plantsci.2013.09.006
- von der Haar, T., Gross, J.D., Wagner, G., McCarthy, J.E.G., 2004. The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nat. Struct. Mol. Biol.* 11, 503–511. doi:10.1038/nsmb779
- von der Haar, T., McCarthy, J.E.G., 2002. Intracellular translation initiation factor levels in *Saccharomyces cerevisiae* and their role in cap-complex function. *Mol. Microbiol.* 46, 531–544.
- von Manteuffel, S.R., Gingras, A.C., Ming, X.F., Sonenberg, N., Thomas, G., 1996. 4E-BP1 phosphorylation is mediated by the FRAP-p70s6k pathway and is independent of mitogen-activated protein kinase. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4076–4080.
- Wada, M., Ito, K., 2014. A genetic approach for analyzing the co-operative function of the tRNA mimicry complex, eRF1/eRF3, in translation termination on the ribosome. *Nucleic Acids Res.* 42, 7851–7866. doi:10.1093/nar/gku493
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509
- Wang, X., Flynn, A., Waskiewicz, A.J., Webb, B.L., Vries, R.G., Baines, I.A., Cooper, J.A., Proud, C.G., 1998. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. *J. Biol. Chem.* 273, 9373–9377.
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484
- Watatani, Y., Ichikawa, K., Nakanishi, N., Fujimoto, M., Takeda, H., Kimura, N., Hirose, H., Takahashi, S., Takahashi, Y., 2008. Stress-induced Translation of ATF5 mRNA Is Regulated by the 5′-Untranslated Region. *J. Biol. Chem.* 283, 2543–2553. doi:10.1074/jbc.M707781200
- Weil, D., 2007. [GW bodies and stress granules, two cytoplasmic structures for mRNA degradation and storage in mammalian cells]. *J. Soc. Biol.* 201, 349–358. doi:10.1051/jbio:2007905
- Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., Bartel, D.P., 2016. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep.* 14, 1787–1799. doi:10.1016/j.celrep.2016.01.043
- Wells, S.E., Hillner, P.E., Vale, R.D., Sachs, A.B., 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* 2, 135–140.
- Wen, Y., Liu, Y., Xu, Y., Zhao, Y., Hua, R., Wang, K., Sun, M., Li, Y., Yang, S., Zhang, X.-J., Kruse, R., Cichon, S., Betz, R.C., Nöthen, M.M., van Steensel, M.A.M., van Geel, M., Steijlen, P.M., Hohl, D., Huber, M., Dunnill, G.S., Kennedy, C., Messenger, A., Munro, C.S., Terrinoni, A., Hovnanian, A., Bodemer, C., de Prost, Y., Paller, A.S., Irvine, A.D., Sinclair, R., Green, J., Shang, D., Liu, Q., Luo, Y., Jiang, L., Chen, H.-D., Lo, W.H.-Y., McLean, W.H.I., He, C.-D., Zhang, X., 2009. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat. Genet.* 41, 228–233. doi:10.1038/ng.276

- Wethmar, K., 2014. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* 5, 765–778. doi:10.1002/wrna.1245
- Wethmar, K., Bégay, V., Smink, J.J., Zaragoza, K., Wiesenthal, V., Dörken, B., Calkhoven, C.F., Leutz, A., 2010a. C/EBPbetaDeltaORF mice--a genetic model for uORF-mediated translational control in mammals. *Genes Dev.* 24, 15–20. doi:10.1101/gad.557910
- Wethmar, K., Smink, J.J., Leutz, A., 2010b. Upstream open reading frames: molecular switches in (patho)physiology. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 32, 885–893. doi:10.1002/bies.201000037
- Will, C.L., Luhrmann, R., 2011. Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.* 3, a003707–a003707. doi:10.1101/cshperspect.a003707
- Wilson, D.N., Nierhaus, K.H., 2006. The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis. *Cell. Mol. Life Sci. CMLS* 63, 2725–2737. doi:10.1007/s00018-006-6125-4
- Wittmann, J., Hol, E.M., Jäck, H.-M., 2006. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell. Biol.* 26, 1272–1287. doi:10.1128/MCB.26.4.1272-1287.2006
- Wold, W.S., Deutscher, S.L., Takemori, N., Bhat, B.M., Magie, S.C., 1986. Evidence that AGUAUAUGA and CCAAGAUGA initiate translation in the same mRNA region E3 of adenovirus. *Virology* 148, 168–180.
- Wolin, S.L., Walter, P., 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* 7, 3559–3569.
- Wong, L.E., Li, Y., Pillay, S., Frolova, L., Pervushin, K., 2012. Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. *Nucleic Acids Res.* 40, 5751–5765. doi:10.1093/nar/gks192
- Yamashita, A., Chang, T.-C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.-Y.A., Shyu, A.-B., 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.* 12, 1054–1063. doi:10.1038/nsmb1016
- Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T., Kurata, R., Hirano, H., Anderson, P., Ohno, S., 2009. SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev.* 23, 1091–1105. doi:10.1101/gad.1767209
- Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., Ohno, S., 2001. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* 15, 2215–2228. doi:10.1101/gad.913001
- Ye, Y., Liang, Y., Yu, Q., Hu, L., Li, H., Zhang, Z., Xu, X., 2015. Analysis of human upstream open reading frames and impact on gene expression. *Hum. Genet.* 134, 605–612. doi:10.1007/s00439-015-1544-7
- Yepiskoposyan, H., Aeschmann, F., Nilsson, D., Okoniewski, M., Mühlemann, O., 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA N. Y. N* 17, 2108–2118. doi:10.1261/rna.030247.111
- Zhao, C., Datta, S., Mandal, P., Xu, S., Hamilton, T., 2010. Stress-sensitive Regulation of IFRD1 mRNA Decay Is Mediated by an Upstream Open Reading Frame. *J. Biol. Chem.* 285, 8552–8562. doi:10.1074/jbc.M109.070920
- Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtsov, S., Kisselev, L., Philippe, M., 1995. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.* 14, 4065–4072.
- Zinzalla, V., Stracka, D., Oppliger, W., Hall, M.N., 2011. Activation of mTORC2 by association with the ribosome. *Cell* 144, 757–768. doi:10.1016/j.cell.2011.02.014

- Zoncu, R., Efeyan, A., Sabatini, D.M., 2011. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat. Rev. Mol. Cell Biol.* 12, 21–35. doi:10.1038/nrm3025
- Zoschke, R., Watkins, K.P., Barkan, A., 2013. A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 25, 2265–2275. doi:10.1105/tpc.113.111567

Annexe

Article 2 : soumis à RNA Biology

**Co-translational deadenylation of mRNAs in human cells: the
example of MYC mRNA**

Béatrice Jolles, Affaf Aliouat, Samia Salhi and Olivier Jean-Jean*

Sorbonne Universités, UPMC Univ Paris 06, CNRS-UMR8256, Biological Adaptation and
Ageing - Institut de Biologie Paris Seine (B2A-IBPS), 7 quai Saint Bernard, 75005 Paris, France

Co-translational deadenylation of mRNAs in human cells: the example of MYC mRNA

Béatrice Jolles, Affaf Aliouat, Samia Salhi and Olivier Jean-Jean*

Sorbonne Universités, UPMC Univ Paris 06, CNRS-UMR8256, Biological Adaptation and Ageing - Institut de Biologie Paris Seine (B2A-IBPS), 7 quai Saint Bernard, 75005 Paris, France

Béatrice Jolles: beatrice.jolles@upmc.fr

Affaf Aliouat: affaf.aliouat@upmc.fr

Samia Salhi: samia.salhi@upmc.fr

*Corresponding author: Tel: 33144272299; Fax: 33144278381; Email: olivier.jean-jean@upmc.fr

Running title: co-translational deadenylation of MYC mRNA

Abstract

The earliest step in the mRNA degradation process is deadenylation, a progressive shortening of the mRNA poly(A) tail by deadenylases. The question of when deadenylation takes place remains open. Deadenylation can occur either independently of translation, following disassembly of translating polysomes, or during the active translation process. In this study, we attempted to answer this question by analyzing the poly(A) tail length distribution of various mRNAs isolated from the polysomal fractions of polysome profiling experiments performed with human cultured cells. Our analysis shows that, for all the genes tested with the only exception of MYC, the poly(A) tail length distribution does not change with the number of ribosomes carried by the mRNA suggesting that the poly(A) tail length does not correlate with translation efficiency. Conversely, for MYC mRNA, we observed a poly(A) tail length decrease in the fractions containing the largest polysomes. Because the fractions with the highest number of ribosomes are also those for which translation termination takes place, we analyzed the poly(A) tail length distribution in polysomal fractions of cells depleted in translation termination factor eRF3. Our results show that the shortening of MYC mRNA poly(A) tail is reduced in eRF3-depleted cells. These findings suggest that MYC mRNA is co-translationally deadenylated and that the deadenylation process requires translation termination to proceed.

Keywords: mRNA degradation, poly(A) tail, deadenylation, translation, translation termination, MYC, eRF3

INTRODUCTION

Cytoplasmic mRNA degradation is one of the post-transcriptional processes that modulate gene expression. Once exported to the cytoplasm, the mammalian messengers are either translated or degraded or stored. These processes are tightly intertwined and our knowledge of the sequence of events that determine the post-transcriptional fate of mRNA is largely incomplete. At the 3'-end of most eukaryotic mRNAs, the poly(A) tail is a key actor in these processes. By promoting the closed loop configuration of mRNAs through its binding to the initiation factor eIF4G, the cytoplasmic poly(A) binding protein PABP facilitates the recycling of terminating ribosomes towards initiation and enhances translation efficiency ¹. Conversely, by recruiting deadenylation complexes, PABP initiates the shortening of the poly(A) tail ². Deadenylation is the initial and rate-limiting step for major mRNA degradation pathways ³. Two cytoplasmic deadenylation complexes degrade the poly(A) tail: PAN2–PAN3 complex shortens the 3' poly(A) tail to approximately 100 nt that are then trimmed by the TOB-CAF1-CCR4-NOT complex ⁴. In this two-step deadenylation process, both complexes are recruited by cytoplasmic PABP through either PAN3-PABP or TOB-PABP interactions. However, alternative mechanisms of deadenylation involve the targeted recruitment of deadenylation complexes either by sequence-specific RNA-binding proteins (RBPs) or by the miRNA pathway ⁵. In human cells, an example of regulated recruitment of the CAF1-CCR4-NOT complex is provided by the AU-rich element-mediated mRNA decay. The AU-rich elements (AREs) found in the 3' untranslated region of short-lived mRNAs, such as FOS and MYC mRNAs, are the most common destabilizing elements and more than 20 different ARE-RBPs have been identified so far.

The common view of mRNA degradation process was that mRNAs are protected from degradation throughout the translation process and that translational inhibition and disassembly

of translating polysomes precedes mRNA decay. This simplistic model was recently challenged by the observations that, in normal growth conditions, decapped intermediates of mRNA degradation are associated with translating ribosomes in *Drosophila* and yeast cells^{6,7}.

Additional experimental evidences seem to strongly argue that translational repression is not a prerequisite for mRNA decay. Using a reporter transcript, Funakoshi *et al.*⁸ have shown a competitive interaction between translation termination complex and deadenylase complexes for PABP binding. From these results, the authors proposed a model that couples translation termination to deadenylation: as translation termination proceeds, a PABP-bound translation termination complex is released from PABP and recruited to the translating ribosome thus allowing a deadenylase complex to bind to the PABP molecule, and to initiate poly(A) tail degradation⁸. This coupling model leads to the speculative view that the poly(A) tail shortening is incremented at each translation termination event, which potentially eliminates one PABP molecule and therefore uncovers ~20 adenosine residues susceptible to deadenylase attack^{8,9}. However, as noted by Brook and Gray¹⁰, most mRNAs are likely translated by many more ribosomes than required to evict all the poly(A) tail-bounded PABP, suggesting that such eviction does not occur at every termination event. This remark is particularly judicious for mRNAs with short poly(A) tail and/or long half-life.

To date, only few examples of a direct connection between translation and deadenylation have been identified. For instance, during the S and M phases of synchronized human cells, poly(A) tail length dynamics influences translational control for a subset of genes with cell-cycle functions¹¹. In addition, it has long been known that deadenylation is a major means to silence translation in oocytes and early embryos^{12,13}. This was recently confirmed by the use of high-throughput methods for poly(A) tail length profiling such as TAIL-seq and PALseq^{14,15}. Subtelny *et al.*¹⁵ have shown that poly(A) tail length strongly correlates with translational efficiency in early stages of zebrafish and frog embryos but that this coupling disappears in

gastrulating embryos and is absent in the non-embryonic samples studied, including mouse liver and a sampling of mammalian cell lines. This latter result was confirmed by the global poly(A) tail distribution analysis of Chang *et al.*¹⁴ showing that, in HeLa and NIH 3T3 cells, the poly(A) tail length is not correlated to translational rates, hence suggesting that deadenylation and translation are not coupled .

In the present work, we investigated deadenylation during active translation in human cells. The poly(A) tail length distribution of various mRNAs was analyzed using the Ligase-Mediated Poly(A) Test (LM-PAT) on fractions of polysome profiling experiments. In agreement with the global analyses described above, we found that the poly(A) tail length of mRNAs was largely not affected by the ribosome density, except for MYC mRNA. In this particular case, the poly(A) tail length decreases strongly in the fractions containing the largest polyribosomes. What is more, this decrease is partially alleviated by the silencing of translation termination factor eRF3, suggesting that co-translational deadenylation takes place only on the translated MYC mRNAs that have experienced termination.

RESULTS

Co-translational deadenylation of MYC mRNA

In order to precisely evaluate the relationship between translation and deadenylation during normal cell growth, we have studied the poly(A) tail length of mRNAs engaged in an active translation process. The ribosomes of HCT116 human cells in the exponential growing phase were freed on mRNAs with cycloheximide and cell extracts were loaded onto 15-50% sucrose gradients that were subjected to centrifugation to obtain the polysome profiles (Fig. 1). RNAs were then extracted from the polysomal fractions of the gradient. The fraction carrying a single ribosome per mRNA molecule and co-sedimentating with the peak of empty ribosomes

(80S peak on the profile shown in Fig. 1) was omitted as it contained very low amounts of mRNAs (data not shown). The cDNA of RNAs from polysomal fractions was obtained using the Ligase-Mediated Poly(A) test (LM-PAT) that is briefly described in Figure 1. Then, after amplification with gene specific forward primers of the resulting cDNA, the distribution of poly(A) tail lengths was analyzed by agarose gel electrophoresis. The choice of the genes to be analyzed was dictated by several technical constraints. Our first criterion was the gene expression level in HCT116 cells. The abundance of a particular mRNA was obtained from the results of our previous microarray experiments on total and polysome-associated mRNAs in HCT116 cells¹⁶. Poorly expressed genes, i. e., with a very low level of mRNA, were discarded due to the elevated number of cycles required during PCR amplification of the cDNA. Secondly, the choice of the forward primer for PCR led us to eliminate mRNAs with highly repeated sequence motifs in their 3' untranslated region and particularly those with long stretches of A residues. The third criterion was the absence of multiple polyadenylation sites. In fact, multiple polyadenylation sites gave complex poly(A) tail profiles. For JUN mRNA, which harbors several polyadenylation sites, we chose the forward primer so that to amplify the poly(A) tail of the longest mRNA isoform which is also the most abundant isoform.

With these criteria, we analyzed a dozen of mRNAs including those encoding c-Myc (MYC), β -actin (ACTB), eukaryotic elongation factor 1G (EEF1G), ribosomal proteins S6 and L35 (RPS6 and RPL35 respectively) and c-Jun (JUN). The poly(A) tail distribution for these mRNAs are shown in Figure 2. For a given mRNA species, a similar distribution of mRNAs between light and heavy polysome fractions, reflecting translation efficiency, was reproducibly obtained in the different experiments performed. In addition, the size and distribution of the poly(A) tail lengths were also similar throughout the experiments and a typical experiment is presented in Figure 2. We observed that the poly(A) tail lengths varied between 100 and 250 A residues, the longest being for JUN mRNA and the shortest for RPS6 mRNA. As clearly

illustrated by the 3D surface plot analysis of the agarose gel images (Fig. 2), ACTB and JUN mRNAs, which concentrated in the heavy polysome fractions, were likely very efficiently translated whereas RPL35 mRNA which was distributed almost equally in all polysomal fractions appeared to be translated with a much lower efficiency.

Interestingly, with the exception of MYC mRNA, for the other mRNA species (ACTB, EEF1G, RPS6, RPL35 and JUN), the distribution of poly(A) tail lengths did not considerably change in the different polysomal fractions. This is shown by the analysis of the density profiles of the agarose gel lanes (Figure 2, only shown for the last four fractions corresponding to heavy polysomes). The presented density profiles of the gel lanes were performed without O. D. calibration. This allowed us to normalize the profiles on the most intense band corresponding to the mRNAs devoid of poly(A) tail. Note that, in the experimental conditions we used for LM-PAT cDNA synthesis (see Materials and Methods section), the intensity of the band of mRNA without poly(A) tail was proportional to the amount of mRNA present in the analyzed fraction.

For MYC mRNA, the superimposition of the density profiles showed that the poly(A) tail length distribution clearly changed in the last two lanes corresponding to the fractions containing mRNAs translated by more than 7 ribosomes (Fig. 2). In these two fractions which contained the most actively translated mRNAs, the length of the poly(A) tails was globally decreased. The accumulation of MYC mRNAs with short poly(A) tail in these last two lanes suggested that some MYC mRNA molecules were deadenylated concomitantly to their translation. These results also suggested that the other tested mRNA species did not undergo the same deadenylation process. Because the fractions with the highest number of ribosomes were also those for which translation termination took place, we supposed that MYC mRNAs were co-translationally deadenylated once they had experienced translation termination.

Influence of translation termination on MYC mRNA deadenylation

To evaluate the influence of the translation termination process on MYC mRNA deadenylation, we performed the same poly(A) tail analysis on mRNAs extracted from cells depleted in translation termination factor eRF3a which is the active isoform of translation termination factor eRF3 in mammalian cells¹⁷. Human HCT116 cells were electroporated with plasmid expressing sh-RNA 3a1 targeting eRF3a mRNA¹⁷. Three days after electroporation, cell lysates were subjected to fractionation by centrifugation on sucrose gradient and LM-PAT analysis was performed on RNA extracted from the gradient fractions. eRF3a depletion was ascertained by Western blot analysis (see supplementary Fig. S1). Representative experiments of poly(A) tail length distribution analysis for eRF3a depleted samples are presented in Figure 3 for the same genes as in Figure 2. The depletion of eRF3a which leads to the decrease in translation termination efficiency, did not profoundly alter the distribution of ribosomes on the mRNAs (compare the 3D surface plots of Figures 2 and 3). In addition, with the exception of MYC mRNA, eRF3a depletion did not modify the distribution of poly(A) tail lengths which are similar to those displayed in Figure 2. This shows that, for ACTB, EEF1G, RPS6, RPL35 and JUN mRNAs, translation termination does not interfere neither with poly(A) tail length and, likely, nor with mRNA deadenylation. The only difference in poly(A) tail length distribution is observed for MYC mRNA when the 3D surface plots and density profiles are compared between Figures 2 and 3. In the last two lanes, corresponding to the fractions containing mRNAs translated by more than 7 ribosomes, the shortening of the poly(A) tails was reduced when eRF3a was depleted (Fig. 3).

The comparison of the gel lane density profiles of MYC mRNAs as well as ACTB mRNAs for three independent experiments performed in both conditions, eRF3a depletion or control, is presented in Figure 4. Only the profiles of the last four fractions of the polysome profile, corresponding to heavy polysomes, are shown in Fig. 4. The agarose gel and 3D surface plot

images of experiment 1 are shown in Figure 2 and 3 for control and eRF3a-depleted cells, respectively, whereas agarose gel and 3D surface plot images of experiments 2 and 3 are shown in supplementary Figure S1. The density profiles in Fig. 4A clearly show that, for cells electroporated with the plasmid expressing the control sh-RNA (sh-Ctrl), the poly(A) tail length of MYC mRNAs is decreased in the last two fractions containing the mRNAs that terminate their translation whereas the poly(A) tail shortening of MYC mRNA is much less pronounced in eRF3a-depleted (sh-3a) cells. Conversely, the density profiles of ACTB mRNA (Fig. 4B) reveal that poly(A) tail length distribution is similar for all fractions and is not affected by translation termination. Altogether, these results suggest that MYC mRNA deadenylation is weakened when translation termination efficiency is impaired and tends to confirm that the co-translational deadenylation of MYC mRNA requires translation termination to proceed.

DISCUSSION

For most of mRNA decay pathways, the earliest step in cytoplasmic mRNA degradation is deadenylation, a progressive shortening of the mRNA poly(A) tail by deadenylases^{5,18}. However, for the vast majority of mRNA species very little is presently known on the specific mechanism of their deadenylation and when this process takes place, independently of translation, following disassembly of translating polysomes, or during the active translation process. In this study, we attempted to answer these questions by analyzing the poly(A) tail length distribution of various mRNAs isolated from the polysomal fractions of polysome profiling experiments. Except in the case of the moderately translated RPL35 mRNA, all the presented mRNAs are highly translated, which is evidenced by the fact that the largest polysomes (more than 7 ribosomes) are the most populated. For all mRNA species studied but MYC and ACTB mRNAs, the poly(A) tail appeared as a smear, whatever the polysome fraction

considered, showing that poly(A) tails are highly heterogeneous in length as previously reported^{14,15}. Our experiments also confirmed the observation that ribosomal protein genes such as RPS6 and RPL35 have short-tailed mRNAs¹⁵.

The global investigations allowing to measure poly(A) tail length at the genomic scale have shown that poly(A) tail length and translation efficiency are decoupled in non-embryonic cells^{14,15}. Considering that translation efficiency is proportional to the number of ribosomes per transcript, our analysis of polysome-associated mRNAs, i. e., actively translated mRNAs, shows that, for all of the genes tested with the only exception of MYC, the poly(A) tail length does not correlate with the number of ribosomes carried by the mRNA (Figure 2) which is in good agreement with the recent findings mentioned above. For MYC mRNA, the poly(A) tail length distribution clearly changed along the polysome profile. The poly(A) tails are relatively homogeneous in length, up to the fraction which contained the mRNAs carrying 6-7 ribosomes. In these fractions, poly(A) tails are clustered and form a distinct peak at ~250 nt in length. This peak progressively disappears in the last two fractions containing mRNAs translated by more than 7 ribosomes. Finally, the poly(A) tail length distribution appears as a regular smear in the last fraction (Figures 2 and 4A, sh-Ctrl). This result suggests that for MYC mRNA the length of the poly(A) tail is sensitive to translation efficiency and that the most actively translated mRNAs are subjected to poly(A) tail shortening by deadenylation. The shortening of MYC mRNA poly(A) tails during ongoing translation was previously proposed on the basis of Northern blot experiments performed with RNA from HeLa cells treated with protein synthesis inhibitors¹⁹. Our experiments confirm this view and furthermore allow to precisely associate MYC mRNA deadenylation with polysomes which are completing translation. Indeed, the only substantial difference between RNA molecules translated by 2 to 7 ribosomes and those translated by more than 7 ribosomes is that, on the latter, some ribosomes have completed the open reading frame translation and reached the stop codon. This raises the possibility that

deadenylation of MYC mRNA was coupled with the last step of translation, i. e., translation termination. Our results showing that a translation termination defect reduced the shortening of MYC mRNA poly(A) tail argue in this sense (Fig. 3 and Fig. 4A Myc – sh-3a1, the disappearance of the long poly(A) tail peak is delayed in eRF3a-depleted cells). Besides, it is noticeable that eRF3a depletion had no effect on the poly(A) tail length distribution of the other genes.

What could be the mechanism coupling translation termination and MYC mRNA deadenylation? It has been recently proposed that, upon translational termination, recruitment of translation termination complex to the ribosomes permits the binding of the deadenylase complexes to the 3' end of the poly(A) tail, thus coupling mRNA deadenylation with translation termination²⁰. However, this hypothesis concerns the general mechanism of deadenylation involving the binding of eRF3 on PABP and the deadenylation complex recruitment via their competitive binding to PABP. If true, we should have likely observed poly(A) tail shortening for other mRNAs than that of MYC which is not the case. MYC mRNA contains AU-rich elements in its 3'UTR and is targeted by sequence-specific RNA-binding proteins²¹. Recently, mammalian MYC mRNA has been shown to escape the two-step general deadenylation pathway and to be degraded through a specific decay process involving the TOB- Caf1-CCR4 deadenylation complex interacting with the sequence-specific RNA-binding proteins CPEB²². The binding of CPEB to CPEs (Cytoplasmic Polyadenylation Element) present in MYC mRNA 3' untranslated region allows the recruitment of TOB-Caf1-CCR4 complex. An hypothesis to explain the link between MYC mRNA deadenylation and translation termination is that translation termination would facilitate the recruitment of deadenylase complex on CPE-containing mRNAs. Nevertheless, the case of MYC mRNA seems to be different from that of JUN mRNA which also contains CPEs that can bind CPEB²³. Indeed, JUN mRNA poly(A) tail

length distribution is not changed along the polysome gradient profile and is not sensitive to eRF3a depletion (Figures 2 and 3).

In fine, we show that co-translational deadenylation of mRNAs is not a prevalent process but likely concerns some genes that use uncommon deadenylation pathways.

MATERIALS AND METHODS

Cell culture and cell electroporation

The HCT116 cell line (ATCC number: CCL-247) was maintained in McCoy medium (Invitrogen) supplemented with 10% fetal calf serum, 1 mM sodium pyruvate, 100 µg/ml streptomycin and 100 units/ml penicillin at 37°C under 5% CO₂ atmosphere. eRF3a silencing by plasmid expressing small interfering RNAs targeting eRF3a mRNA (sh-3a1) was previously described ²⁴. Plasmid pSUPERcontrol with the non-silencing shRNA sequence (5'-ATTCTCCGAACGTGTCACG-3') was used as negative control (sh-Ctrl). Electroporation of cells was performed with a Gene pulser II electroporation system (Bio-Rad) using 4.8 x 10⁶ cells and 20 µg of plasmid DNA. Electroporated cells were collected 72 h after electroporation.

Polysome fractionation and RNA isolation

Polysome fractionation was performed according to Verrier and Jean-Jean ²⁵ with minor modifications. Three days after electroporation, two 100-mm plates of HCT116 cells at 70% cell confluence were used for sucrose gradient centrifugation experiments. After centrifugation, optical density (O. D.) at 254 nm was monitored by pumping the gradient through a Retriever 500 (Teledyne Isco) fraction collector. The flow rate was optimized in order to collect ~15 fractions of 0.8 mL. Each fraction was precipitated with 1.2 mL of isopropanol and kept at -80°C for further RNA analysis. For RNA isolation, the polysome fractions of the sucrose

gradients (Figure 1) were subjected to centrifugation at 18,000 x g for 1 hr, then the pellets were rinsed with 1 mL of 80% ethanol, dried and resuspended in 350 μ l of RAI buffer of NucleoSpin RNA II kit (Macherey-Nagel) containing 1% β -mercaptoethanol (v/v). RNA was purified according to the manufacturer's instructions and eluted in 40 μ l of RNase-free water.

LM-PAT cDNA synthesis

LM-PAT experiments were performed according to Sallés et al.²⁶. Firstly, 2 μ l of 5'-phosphorylated poly(dT)12-18 oligonucleotide (10 ng/ μ l in RNase-free water) was added to 5 μ l of each of the purified RNA fractions in order to saturate the entire poly(A) tails. The mixture was heat-denatured at 65°C for 10 min and immediately transferred to a 42°C water bath. Then, 13 μ l of a pre-warmed (42°C) mixture of 2 μ l of 10X RT buffer from the SuperScript First-Strand Synthesis System for RT-PCR (Life Technologies), 1 μ l of dNTP mix (10 mM each), 4 μ l of 25 mM MgCl₂, 2 μ l of 0.1 mM DTT, 1 μ l of RNase OUT, 1 μ l of 10 mM ATP, 1.7 μ l of T4 DNA ligase (New England Biochemicals) corresponding to 10 Weiss units, was added with gentle mixing. After incubation for 30 min at 42°C and while at 42°C, 1 μ l of an excess of oligo(dT) anchor oligonucleotide (5'-GCGAGCTCCGCGGCCGCG-T12-3') (200 ng/ μ l) was added and the reactions were further incubated at 12°C for 2 h. The oligo(dT) anchor mainly ligates specifically to the 5'-phosphorylated oligo(dT) hybridized at the 3' most end of the poly(A) tail. The reactions were then transferred to 42°C and pre-incubated for 2 min, before adding 1 μ l (200 units) of SuperScript II reverse transcriptase (Life Technologies). The reverse transcription reaction was then performed by incubation at 42°C for 1 h. Finally, the reverse transcriptase was heat inactivated at 70°C for 15 min.

The short incubation at 42°C before the addition of the reverse transcriptase minimizes priming of unligated oligo (dT) anchor at the very 5'-end of the poly(A) tail. Nevertheless, the small amount of oligo (dT) anchor priming at the 5'-end of the poly(A) tail allows to estimate

the length of the cDNA fragment in the absence of poly(A) tail and hence the length of the poly(A) tail ²⁶.

PAT-PCR analysis

PCR forward primers for specific genes were chosen according to the recommendations of Sallés and Strickland ²⁷. They were designed close to the 3' end of the mRNA in order to obtain 250 to 750 base pair long DNA fragments after amplification (Table I). Standard PCR was performed with GoTaq® G2 Green Master Mix (Promega), with 0.5 μ M forward gene specific primer, 0.5 μ M oligo(dT) anchor reverse primer and 1-2 μ l of template PAT cDNA in a 25 μ l reaction volume. Typically, the PCR conditions were 3 min at 95°C (initial denaturation), followed by 30 cycles of 30 s at 95°C, 45 s at 60°C, 1 min at 72°C, ending with a 7-min final extension at 72°C. PCR samples were then loaded onto a 2% agarose gel in Tris-Borate EDTA buffer and electrophoresis was performed for 3 h at 5 V/cm. After ethidium bromide staining and extensive washing of the gel, photographs were taken using a UV transilluminator Gel Doc XRT (BioRad). The image of the gel was further processed with ImageJ software. Surface plots were obtained using the Interactive 3D Surface Plot option of the ImageJ 1.49v program. Gel lane density profiles were obtained using the Gel Analyzer menu of ImageJ and identical rectangle areas were selected to analyze each gel lane. We also used the uncalibrated O. D. option in order to compare the relative intensities of poly(A) tail lengths. Images of lane profiles were then stacked to superimpose the profiles.

ACKNOWLEDGMENTS

A. A. was supported by a Contrat Doctoral from the Université Pierre et Marie Curie.

Funding: the Université Pierre et Marie Curie (UPMC) and the Centre National de la Recherche Scientifique (CNRS).

REFERENCES

1. Kahvejian A, Roy G, Sonenberg N. The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol* 2001; 66:293–300.
2. Chen C-YA, Shyu A-B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2011; 2:167–83.
3. Shyu AB, Belasco JG, Greenberg ME. Two distinct destabilizing elements in the c-fos message trigger deadenylation as a first step in rapid mRNA decay. *Genes Dev* 1991; 5:221–31.
4. Yamashita A, Chang T-C, Yamashita Y, Zhu W, Zhong Z, Chen C-YA, Shyu A-B. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol* 2005; 12:1054–63.
5. Wahle E, Winkler GS. RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. *Biochim Biophys Acta* 2013; 1829:561–70.
6. Antic S, Wolfinger MT, Skucha A, Hosiner S, Dorner S. General and MicroRNA-Mediated mRNA Degradation Occurs on Ribosome Complexes in *Drosophila* Cells. *Mol Cell Biol* 2015; 35:2309–20.
7. Hu W, Sweet TJ, Chamnongpol S, Baker KE, Collier J. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* 2009; 461:225–9.
8. Funakoshi Y, Doi Y, Hosoda N, Uchida N, Osawa M, Shimada I, Tsujimoto M, Suzuki T, Katada T, Hoshino S. Mechanism of mRNA deadenylation: evidence for a molecular

- interplay between translation termination factor eRF3 and mRNA deadenylases. *Genes Dev* 2007; 21:3135–48.
9. Hoshino S, Imai M, Kobayashi T, Uchida N, Katada T. The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J Biol Chem* 1999; 274:16677–80.
 10. Brook M, Gray NK. The role of mammalian poly(A)-binding proteins in co-ordinating mRNA turnover. *Biochem Soc Trans* 2012; 40:856–64.
 11. Park J-E, Yi H, Kim Y, Chang H, Kim VN. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol Cell* 2016; 62:462–71.
 12. Richter JD. CPEB: a life in translation. *Trends Biochem Sci* 2007; 32:279–85.
 13. Sachs A, Wahle E. Poly(A) tail metabolism and function in eucaryotes. *J Biol Chem* 1993; 268:22955–8.
 14. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* 2014; 53:1044–52.
 15. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 2014; 508:66–71.
 16. Ait Ghezala H, Jolles B, Salhi S, Castrillo K, Carpentier W, Cagnard N, Bruhat A, Fafournoux P, Jean-Jean O. Translation termination efficiency modulates ATF4 response by regulating ATF4 mRNA translation at 5' short ORFs. *Nucleic Acids Res* 2012; 40:9557–70.

17. Chauvin C, Salhi S, Le Goff C, Viranaicken W, Diop D, Jean-Jean O. Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. *Mol Cell Biol* 2005; 25:5801–11.
18. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007; 8:113–26.
19. Laird-Offringa IA, de Wit CL, Elfferich P, van der Eb AJ. Poly(A) tail shortening is the translation-dependent step in c-myc mRNA degradation. *Mol Cell Biol* 1990; 10:6132–40.
20. Osawa M, Hosoda N, Nakanishi T, Uchida N, Kimura T, Imai S, Machiyama A, Katada T, Hoshino S, Shimada I. Biological role of the two overlapping poly(A)-binding protein interacting motifs 2 (PAM2) of eukaryotic releasing factor eRF3 in mRNA decay. *RNA* 2012; 18:1957–67.
21. Barreau C, Paillard L, Osborne HB. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 2005; 33:7138–50.
22. Ogami K, Hosoda N, Funakoshi Y, Hoshino S. Antiproliferative protein Tob directly regulates c-myc proto-oncogene expression through cytoplasmic polyadenylation element-binding protein CPEB. *Oncogene* 2014; 33:55–64.
23. Zearfoss NR, Alarcon JM, Trifilieff P, Kandel E, Richter JD. A molecular circuit composed of CPEB-1 and c-Jun controls growth hormone-mediated synaptic plasticity in the mouse hippocampus. *J Neurosci* 2008; 28:8502–9.

24. Chauvin C, Salhi S, Jean-Jean O. Human eukaryotic release factor 3a depletion causes cell cycle arrest at G1 phase through inhibition of the mTOR pathway. *Mol Cell Biol* 2007; 27:5619–29.
25. Verrier SB, Jean-Jean O. Complementarity between the mRNA 5' untranslated region and 18S ribosomal RNA can inhibit translation. *RNA N Y N* 2000; 6:584–97.
26. Sallés FJ, Richards WG, Strickland S. Assaying the polyadenylation state of mRNAs. *Methods* 1999; 17:38–45.
27. Sallés FJ, Strickland S. Rapid and sensitive analysis of mRNA polyadenylation states by PCR. *PCR Methods Appl* 1995; 4:317–21.

FIGURE LEGENDS

Figure 1. Schematic description of experimental procedure. Polysomes from human HCT116 cell extracts were separated by centrifugation on a 15-50% sucrose gradient. RNA from polysomal fractions were isolated and subjected to Ligase-Mediated Poly(A) Test (LM-PAT). After amplification with gene specific forward primers of the resulting cDNA, the distribution of poly(A) tail lengths was visualized by agarose gel electrophoresis. The size of the DNA ladder is indicated in base pairs on the left, the number of ribosomes in each fraction is indicated below the gel. 3D surface plot analysis of the agarose gel image was obtained using ImageJ software with minimum set to 0%, maximum to 100% and z-scale to 1.0.

Figure 2. Distribution of poly(A) tail lengths of MYC, ACTB, EEF1G, RPS6, RPL35 and JUN mRNAs. RNA extracted from a polysome fractionation experiment was subjected to LM-PAT analysis and cDNAs were amplified with gene specific forward primers. For the six genes

presented, cDNAs were amplified from the same polysome fractionation experiment and LM-PAT reaction, only gene specific forward primers changed. For each gene indicated above the panels, the agarose gel with DNA ladder on the left (in base pairs), the 3D surface plot analysis of the agarose gel and the stacked image of the density profiles of the last four lanes corresponding to heavy polysomes are shown. In the density profile panel, the fraction containing 4-5 ribosomes is in black, the fraction 6-7 ribosomes in blue, the fraction 7-9 ribosomes in green and the fraction with more than 10 ribosomes (10+) in red. The density profiles of the gel lanes were performed without O. D. calibration. This allowed to normalize the profiles on the most intense band corresponding to mRNAs devoid of poly(A) tail (indicated by a dotted line).

Figure 3. Distribution of poly(A) tail lengths in eRF3a-depleted cells. Human HCT116 cells were electroporated with a plasmid expressing small interfering RNAs targeting eRF3a mRNA (sh-3a1). Three days after electroporation, a cell extract was fractionated on sucrose gradient and RNAs were extracted from the polysome fractions. RNAs were subjected to LM-PAT analysis and cDNAs were amplified with gene specific forward primers. For the six genes presented, cDNAs were amplified from the same polysome fractionation experiment and LM-PAT reaction, only gene specific forward primers changed. For each gene, MYC, ACTB, EEF1G, RPS6, RPL35 and JUN, indicated above the panels, the agarose gel with DNA ladder on the left (in base pairs), the 3D surface plot analysis of the agarose gel, and the stacked image of the density profiles of the last four lanes corresponding to heavy polysomes are shown. In the density profile panel, the fraction containing 4-5 ribosomes is in black, the fraction 6-7 ribosomes in blue, the fraction 7-9 ribosomes in green and the fraction with more than 10 ribosomes (10+) in red. The density profiles of the gel lanes were performed without O. D.

calibration allowing to normalize the profiles on the most intense band corresponding to mRNAs devoid of poly(A) tail (indicated by a dotted line).

Figure 4. Density profiles of MYC and ACTB poly(A) tails for three independent experiments in eRF3a-depleted and control cells. The montage of the density profiles of the last four fractions (4-5, 6-7, 7-9 and 10+ ribosomes) of 3 polysome fractionation experiments with either eRF3a-depleted (sh-3a1) or control (sh-Ctrl) cell extracts are shown for MYC mRNA (A) and ACTB mRNA (B). For MYC – sh-3a, exp. 3, the narrow peak in the middle of fraction 4-5 profile is due to a dust dot in the agarose gel (see the corresponding agarose gel in supplementary Figure S2). The density profiles of the gel lanes were performed without O. D. calibration allowing to normalize the profiles on the most intense band corresponding to mRNAs devoid of poly(A) tail (indicated by a dotted line). The agarose gel and 3D surface plot images of experiment 1 (Exp.1) are shown in Figure 2 and 3 for control and eRF3a-depleted cells, respectively. Agarose gel and 3D surface plot images of experiments 2 (Exp. 2) and 3 (Exp. 3) are shown in supplementary Figure S2.

Table I: PCR gene specific forward primers

Gene name	Accession Number	Primer (5' to 3')	PCR fragment length without poly (A) tail (nt)
MYC	NM_002467.4	AACCTCACAACCTTGGCTGAGTCTTG	408
ACTB	NM_001101.3	ACAGGGGAGGTGATAGCATTGCTT	268
RPS6	NM_001010.2	CGTATTGCTCTGAAGAAGCAGCGTA	234
EEF1G	NM_001404.4	TTGCCTTTCCGCTGAGTCCAGATTG	264
RPL35	NM_007209.3	GGAAATCCATTGCCCGTGTTCTCAC	278
JUN	NM_002228.3	GGACAGCCCACTGAGAAGTCAAACA	272

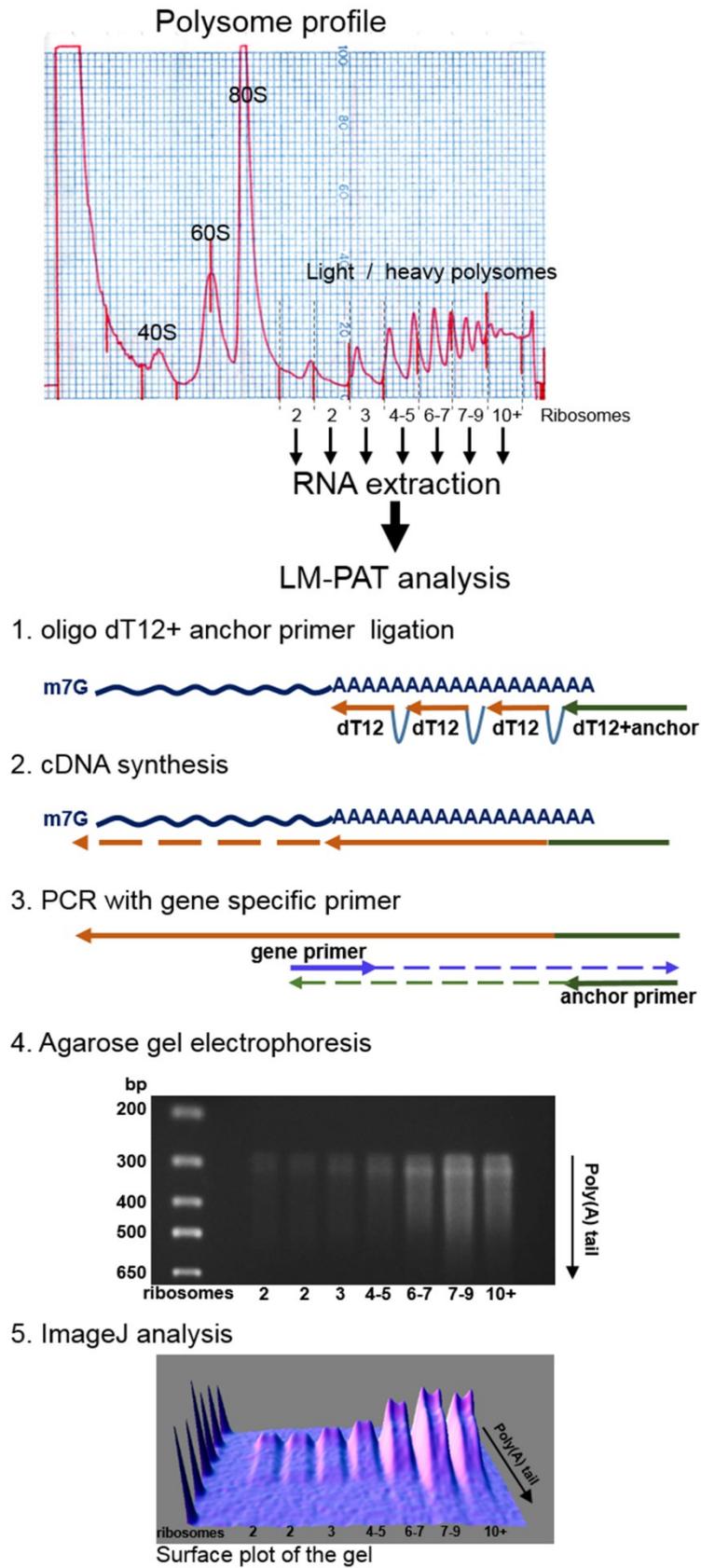


Figure 1

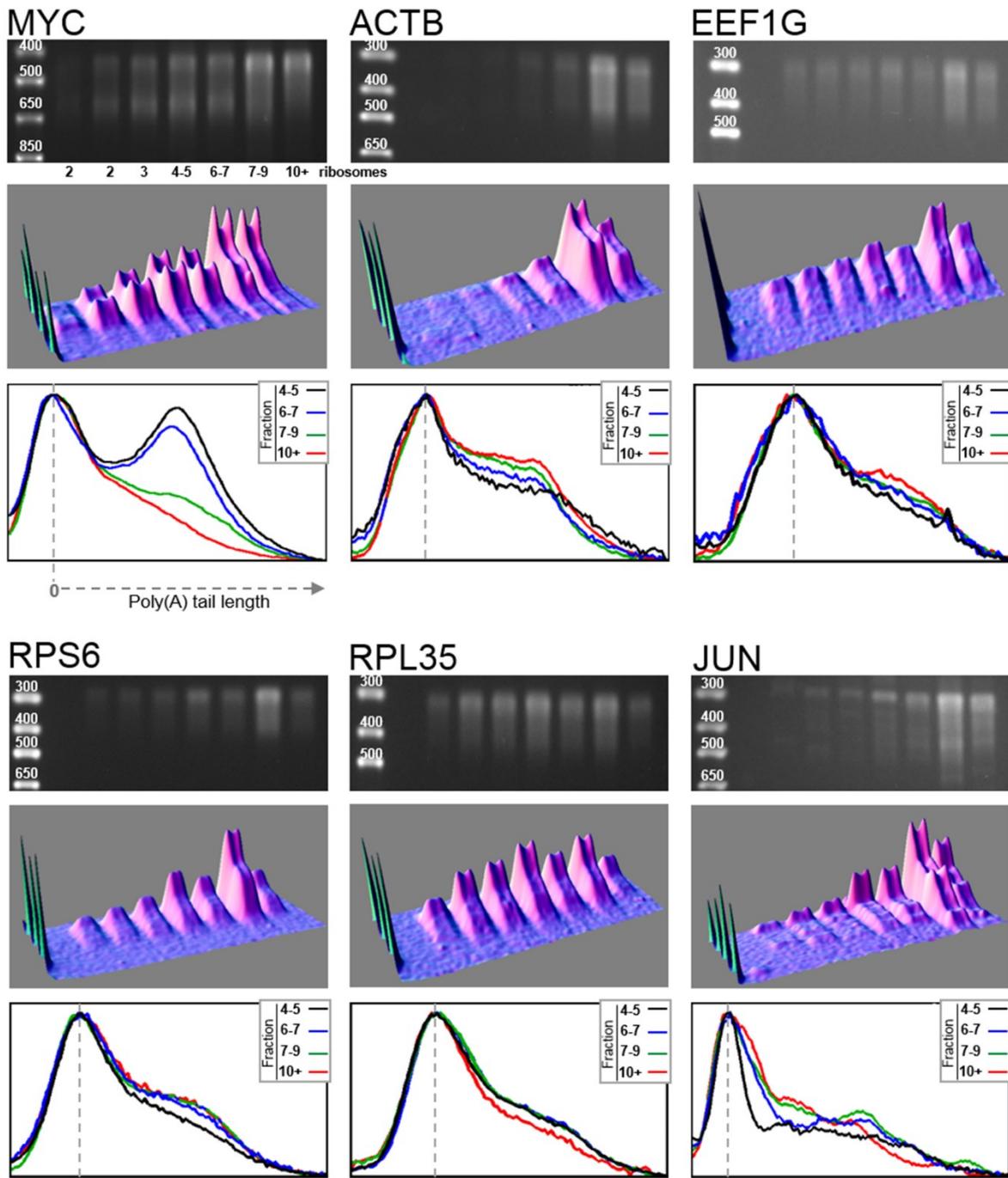


Figure 2

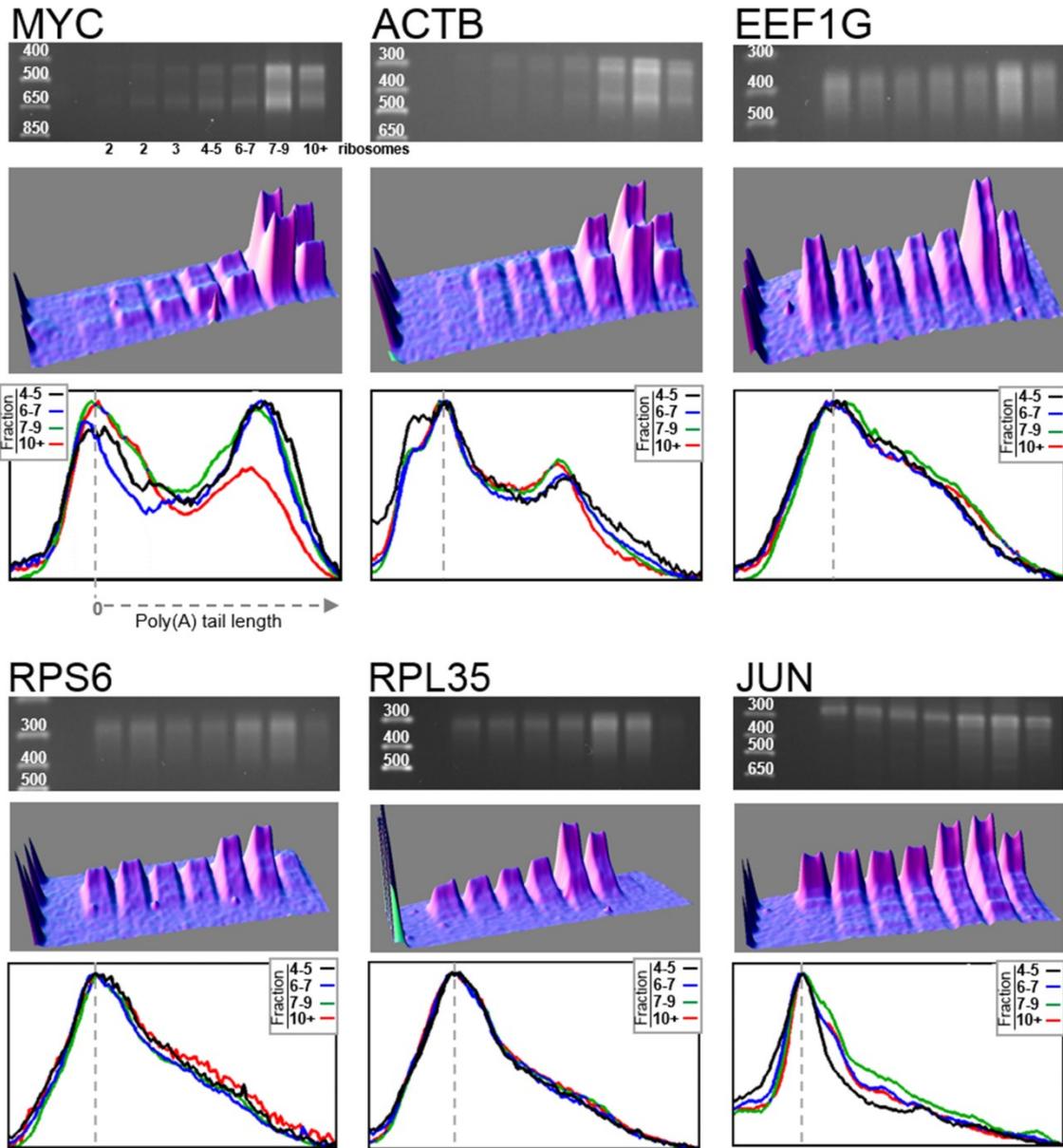
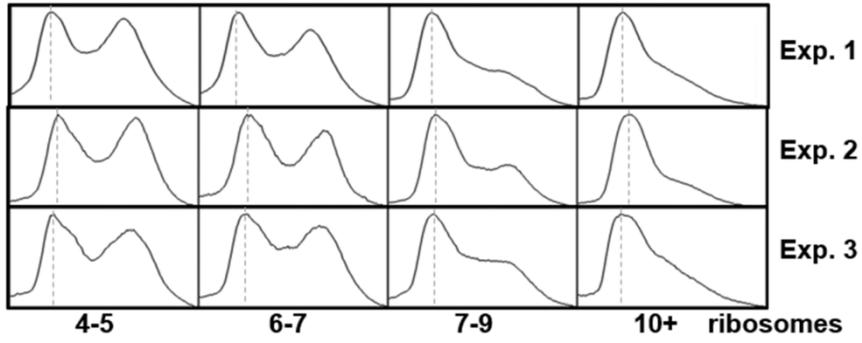
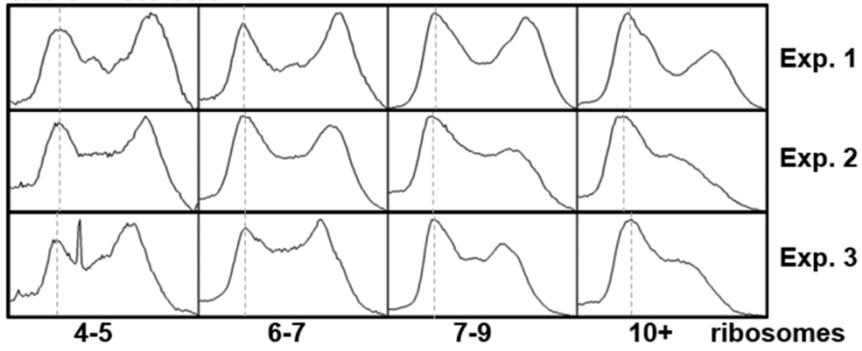
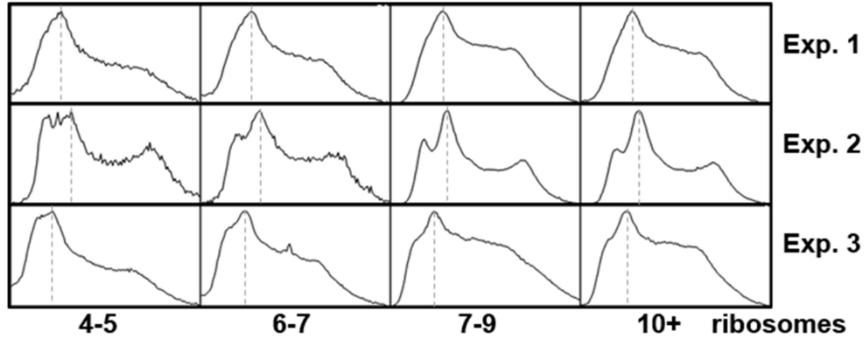
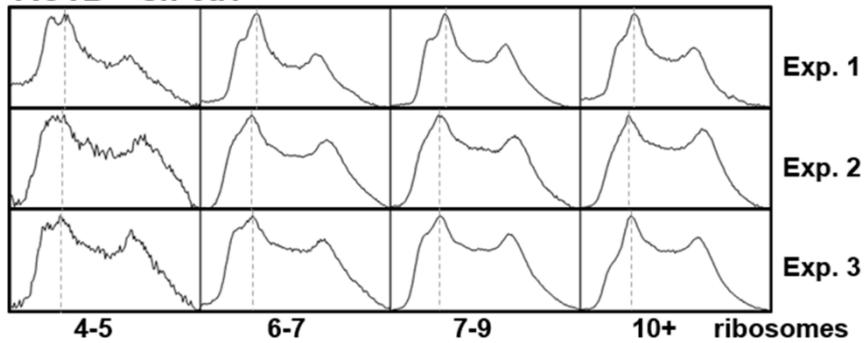
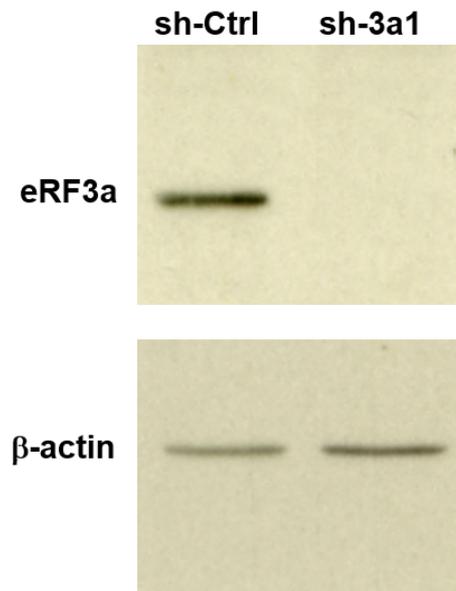


Figure 3

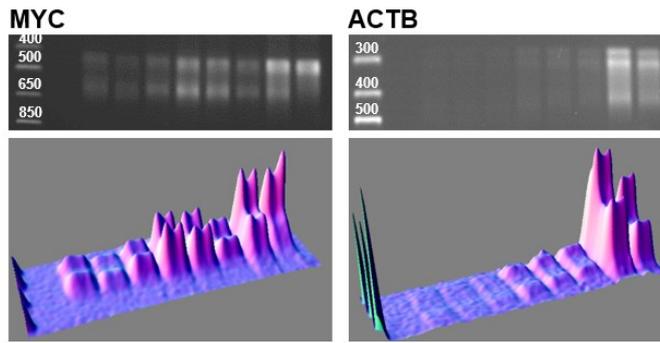
A**MYC – sh-Ctrl****MYC – sh-3a1****B****ACTB – sh-Ctrl****ACTB – sh-3a1****Figure 4**



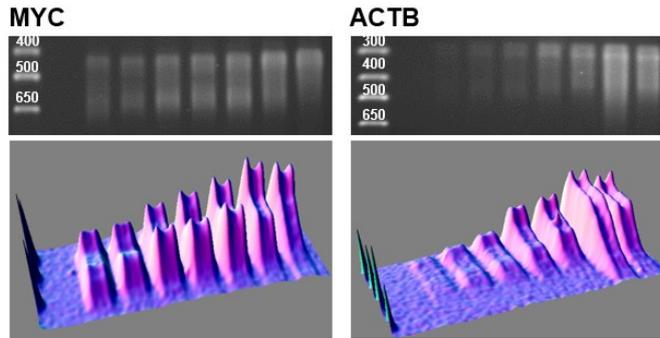
Supplementary Figure S1. Western blot analysis of eRF3a depletion in control (sh-Ctrl) and sh-3a1 electroporated cells. Cells were lysed by sonication on ice, centrifuged for 20 min at 16,000xg and supernatant retained as cell extracts. For each sample, 30 μ g of total protein were loaded on polyacrylamide gel and subjected to electrophoresis. Western blotting was then performed as described in reference 18 (main text). β -actin served as loading control.

A: sh-Ctrl

Exp. 2

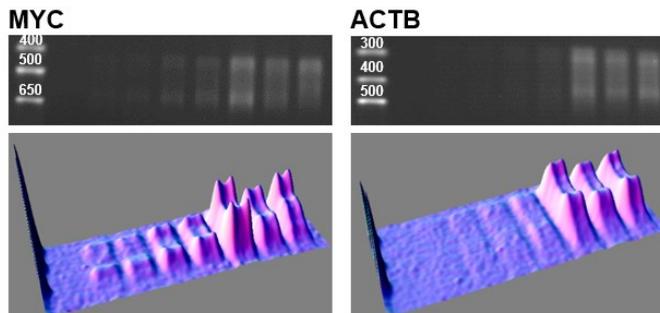


Exp. 3

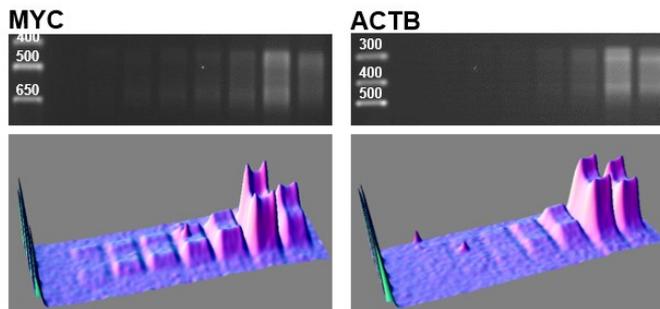


B: sh-3a1

Exp. 2



Exp. 3



Supplementary Figure S2. Agarose gel and 3D surface plot images for MYC and ACTB mRNAs, corresponding to the density profiles of experiments 2 and 3 (Exp. 2 and Exp. 3, respectively) shown in Figure 4. Cells were electroporated either with control plasmid (A: sh-Ctrl) or plasmid expressing sh-RNA directed against eRF3a mRNA (B: sh-3a1).