



HAL
open science

Composition et dynamique des complexes protéiques impliqués dans le "nonsense-mediated mRNA decay" chez la levure *Saccharomyces cerevisiae*

Marine Dehecq

► **To cite this version:**

Marine Dehecq. Composition et dynamique des complexes protéiques impliqués dans le "nonsense-mediated mRNA decay" chez la levure *Saccharomyces cerevisiae*. Biochimie, Biologie Moléculaire. Sorbonne Université, 2018. Français. NNT : 2018SORUS538 . tel-02489767

HAL Id: tel-02489767

<https://theses.hal.science/tel-02489767>

Submitted on 24 Feb 2020

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.

Sorbonne Université

Ecole doctorale Complexité du Vivant (ED515)

Unité de Génétique des interactions macromoléculaires – Institut Pasteur

Composition et dynamique des complexes protéiques impliqués dans le "*nonsense-mediated mRNA decay*" chez la levure *Saccharomyces cerevisiae*

Par Marine DEHECQ

Thèse de doctorat de Biochimie et Biologie moléculaire

Dirigée par Dr. Cosmin SAVEANU

Présentée et soutenue publiquement le 21 septembre 2018

Devant un jury composé de :

Mme. BEHM-ANSMANT Isabelle, Chargée de recherche

Rapporteur

M. SERAPHIN Bertrand, Directeur de recherche

Rapporteur

Mme. EPHRUSSI Anne, Senior scientist

Examineur

M. LE CROM Stéphane, Professeur

Examineur

M. SAVEANU Cosmin, Directeur de recherche

Directeur de thèse

Remerciements

Les quatre dernières années (presque 5), ont été très riches tant sur le plan professionnel que personnel. J'ai appris beaucoup de chose, j'ai travaillé avec des gens sympathique et qui m'ont tous encouragée, bousculée, soutenue.

Merci Cosmin pour ta confiance dès le commencement, pour ton encadrement toujours en douceur, pour l'enthousiasme face au projet, pour la liberté que tu m'as accordée. J'ai beaucoup aimé travailler avec toi ! Merci et bravo de m'avoir fait comprendre que l'analyse bio-info c'est des manips et qu'en plus ça peut être vraiment sympa (après un certain temps de pratique...).

Alain, je n'aurais pas pu rêver meilleur passage chez les « Jacquier ». Merci de m'avoir intégrée dans ton laboratoire. Et surtout merci d'être un challenge à chaque réunion de labo, tes sourcils froncés m'ont poussée à être toujours plus convaincante.

Chère Laurence, c'était un plaisir d'apprendre à tes cotés. Quel exemple de rigueur, de patience et d'organisation ! Je t'embrasse bien fort.

Abdelkader, merci pour toutes les bonnes nouvelles que tu m'as apportées à chaque nouvelle aquisition, sans toi, cette thèse aurait été bien différente. MS c'est pas Mass Spectrometry... c'est Marine Sourit !

Micheline, tu es une super collègue ! Présente aussi bien pour blaguer, pour parler musique (« ma fille Lucie... », haha !), que pour me prêter une oreille attentive et bienveillante. Fais attention à toi 😊

Gwenaël, merci pour l'initiation au CRAC, pour les discussions (nombreuses) de paillasse et bonne continuation dans tes projets futurs.

Franck, merci pour ton soutien quand je finissais tard et pour ton aide sous toutes ses formes (souches, manips, papier, discussions...) sur le projet. Bonne continuation !

Antonia, le soleil du labo ! À bientôt à Carmona !

Maité, Estelle, Léna, contente de vous avoir rencontrées. À une prochaine !

Alicia, Élodie, Valentine, Quentin, Célia, Adeline, merci de m'avoir montré la voie et pour tous les moments de complicité dus à l'expérience partagée.

Christophe, Thomas, Simon, Varun, merci pour vos petits conseils dans ma bataille pour apprivoiser R et le terminal !

Lucia, Claudine, Birame, Monique, Jennyfer, Salomé, merci pour les litres et les litres de YPD préparés, aucune contamination en 4 ans pour les erlens de 2L... ça c'est du travail de pro !

Merci à tous les membres du 4eme étage Fernbach qui ont tous contribué, de près ou de loin, à la réussite de cette thèse.

Merci à l'ARC pour mon financement de 4eme année.

Merci à mon comité de thèse, Frédéric Devaux, Marc Graille et Guilhem Janbon pour m'avoir suivie et encouragée tout au long de ce parcours. Merci Guilhem pour la relecture de ma thèse.

Merci à Isabelle Behm-Ansmant, Bertrand Séraphin, Anne Ephrussi et Stéphane Le Crom d'avoir accepté d'évaluer mon travail de thèse. C'est un honneur pour moi de pouvoir échanger avec vous.

Anne, tu as toujours été présente pour m'encourager dans ma carrière scientifique, de mon premier stage à l'EMBL jusqu'à mon prochain post-doc. Je te remercie particulièrement pour ta bienveillance.

Merci à Hervé Le Hir et à Joanne Kanaan pour le travail et les discussions que l'on a mené ensemble. Cela m'a permis de mieux comprendre le fonctionnement de l'hélicase U_{pf}i. Je me réjouis que la collaboration porte ses fruits et surtout qu'elle continue !

Enfin, merci à mes amis, à la collog du bonheur, à tous les bébés et les filles que j'ai entraîné, à la gym, pour tous les moments partagés et qui m'ont permis de garder un équilibre dans ma vie.

Merci à ma famille, mes parents, mes sœurs, mon frère.

Merci Romain.

Sommaire

Remerciements	2
Sommaire	4
Introduction	7
I. Importance des contrôles pour l'expression des gènes	8
II. La dégradation cytoplasmique des ARNs.	11
1. Dé-adénylation	12
2. Dégradation 5' vers 3'	15
3. Dégradation 3' vers 5'	16
4. Régulation des voies de dégradation des ARNs.....	17
III. Dégradation des ARNs aberrants	18
1. Mécanisme de la traduction.....	18
2. Contrôles dépendant de la traduction	21
3. Activation du NMD au niveau de ses substrats.....	24
4. Facteurs associés au NMD	26
5. Upf1, facteur majeur du NMD	27
6. Upf2, plateforme pour des interactions protéiques	30
7. Upf3.....	31
8. Le cycle de phosphorylation d'Upf1	32
9. Le recrutement de Smg5, Smg6 et Smg7 déclenche la dégradation des ARNs.....	36
10. Modèle EJC-dépendant du NMD : SURF-DECID	38
11. Modèle EJC-indépendant du NMD	39
IV. Fonctions physiologiques et patho-physiologiques du NMD et de ses facteurs.....	40
Objectifs de la thèse	44
Développement technologique	45
Analyse des données de spectrométrie de masse	46
Résultats	56
I. Recrutement d'Upf1 au niveau des ARNs cibles	57
1. Upf1 interagit avec les 3'-UTR de centaines d'ARNs.....	57
2. Les mutations du domaine hélicase d'Upf1 n'affectent pas ses interactions avec les autres partenaires.....	65
II. Complexes protéiques impliqués dans le NMD et leur dynamique	69
1. Stratégie utilisée et introduction aux résultats.....	69
2. Article de thèse.....	75
Abstract	75

Introduction	76
Results	79
Quantitative mass spectrometry reveals the global composition of Upf1 associated complexes	79
Upf1-associated proteins are grouped in two mutually exclusive complexes.	81
The Upf1 cysteine-histidine rich N-terminal domain is an interaction « hub » <i>in vivo</i> .	84
Nmd4 and Ebs1 are tightly associated with Upf1-HD-Cter.....	86
Potential Smg5-6-7 homologs can be essential for NMD efficiency under limiting conditions	88
Detector forms by binding of a Upf2/Upf3 heterodimer to Upf1 <i>in vivo</i>	91
Binding of Effector to NMD substrates depends on Upf2.....	92
Discussion	95
Affinity purification strategies for NMD	95
Detector formation is both similar and distinct from SURF/DECID.....	96
The importance of the C-terminal Upf1 domain in yeast and mammalian NMD.....	97
A molecular switch around Upf1 for the Detector to Effector transition.....	98
Conclusion.....	99
Materials and Methods	100
Yeast and bacterial strains.....	100
Plasmids construction for yeast experiments	100
Plasmids construction for <i>in vitro</i> interaction studies	100
TAP affinity purifications	101
Mass spectrometry, sample preparation, acquisition and data analysis	101
Reverse transcription and quantitative PCR.....	102
NMD efficiency calculation	103
Polysome fractionation.....	103
Acknowledgments.....	103
Author Contributions.....	104
Conflict of interest.....	104
References	105
Expanded view Figure.....	111
Appendix Material and Methods.....	115
Appendix Figures	123
Appendix Tables	130
Discussion générale et perspectives	136
I. Résumé des résultats	137
II. Discussion et perspectives	138
1. Ubr1 est associé spécifiquement à Upf2	138

2. Le complexe CK2, responsable de la phosphorylation de Upf2 ?	140
3. Hrr25 fait partie de l'effecteur	141
4. Plusieurs hypothèses existent pour le switch	142
4.1. Échange des deux complexes ou échanges des facteurs autour de Upf1 ?	142
4.2. Dimérisation de Upf1	143
4.3. Rôle de la phosphorylation d'Upf2	144
4.4. Ebs1 comme intermédiaire entre le Détecteur et l'Effecteur	144
4.5. Perspectives pour une meilleure compréhension du switch.....	146
5. Quelle est la fonction de l'Effecteur non fixé aux ARNs ?.....	146
6. Lsm1-7 et Pat1 sont présents sur les mêmes ARNs que Upf1	147
7. La question récurrente du lien entre les p-bodies et le NMD.....	147
8. Nmd4, régulateur de l'association d'Upf1 à l'ARN	148
Conclusion.....	150
Références	152
Annexes	185
Liste des abréviations	186
Table des Figures	189
Table des tableaux.....	191
Résumé.....	192
Abstract	192

Introduction

Introduction

Il y a 60 ans, Francis Crick, s'appuyant sur les travaux de plusieurs chercheurs, proposait un des principes fondamentaux de la biologie moléculaire. Il définit la directionnalité du mécanisme permettant l'expression des gènes dans une cellule vivante. L'acide désoxyribonucléique (ADN) contient l'information génétique de chaque cellule organisée sous forme de gènes disposés le long des chromosomes. La transcription de cet ADN en acide ribonucléique messenger (ARNm) sert de matrice pour la traduction en protéines, qui ont ensuite un rôle dans le fonctionnement de la cellule (Crick, 1958, 1970; Nirenberg and Matthaei, 1961). Chez les organismes eucaryotes, ce processus suit un mécanisme par étapes consécutives liées les unes aux autres (Moore and Proudfoot, 2009, **Figure 1**). L'ARN polymérase II interagit avec les promoteurs des gènes et permet l'ouverture de l'ADN et l'initiation de la transcription de celui-ci en ARN immature. De manière concomitante, l'ARN naissant subit plusieurs étapes de maturation dans le noyau, ajout d'une coiffe de 7-méthylguanosine triphosphate (m^7 -Gppp) en 5', épissage des introns et poly-adénylation de l'extrémité 3'. Les différents facteurs impliqués dans ces étapes de maturations sont recrutés par l'ARN polymérase II elle-même en cours de transcription *via* son domaine C-terminal (CTD). Ces étapes de maturations sont cruciales pour l'export des ARNs dans le cytoplasme, pour la traduction des ARNm en protéines et pour protéger les ARNs de la dégradation. Le temps de vie des ARNs et des protéines est très variable avec une influence considérable sur l'expression des gènes.

I. Importance des contrôles pour l'expression des gènes

La production des protéines en quantité et de qualité nécessaire pour la vie d'une cellule dépend du bon déroulement des étapes indiquées précédemment. Des erreurs sont possibles à chaque niveau. C'est pourquoi, des systèmes de contrôle détectent ces erreurs, les corrigent ou les éliminent. La qualité des ARNs par exemple est très contrôlée (**Figure 1**) afin d'éviter la traduction d'ARN dit « aberrant ». Par le terme aberrant, on considère ici les ARNs qui, étant apparentés à des ARNm qui codent pour des protéines cellulaires, ne servent pas à cette fonction. Des exemples d'ARNs aberrants sont présentés dans les paragraphes suivants.

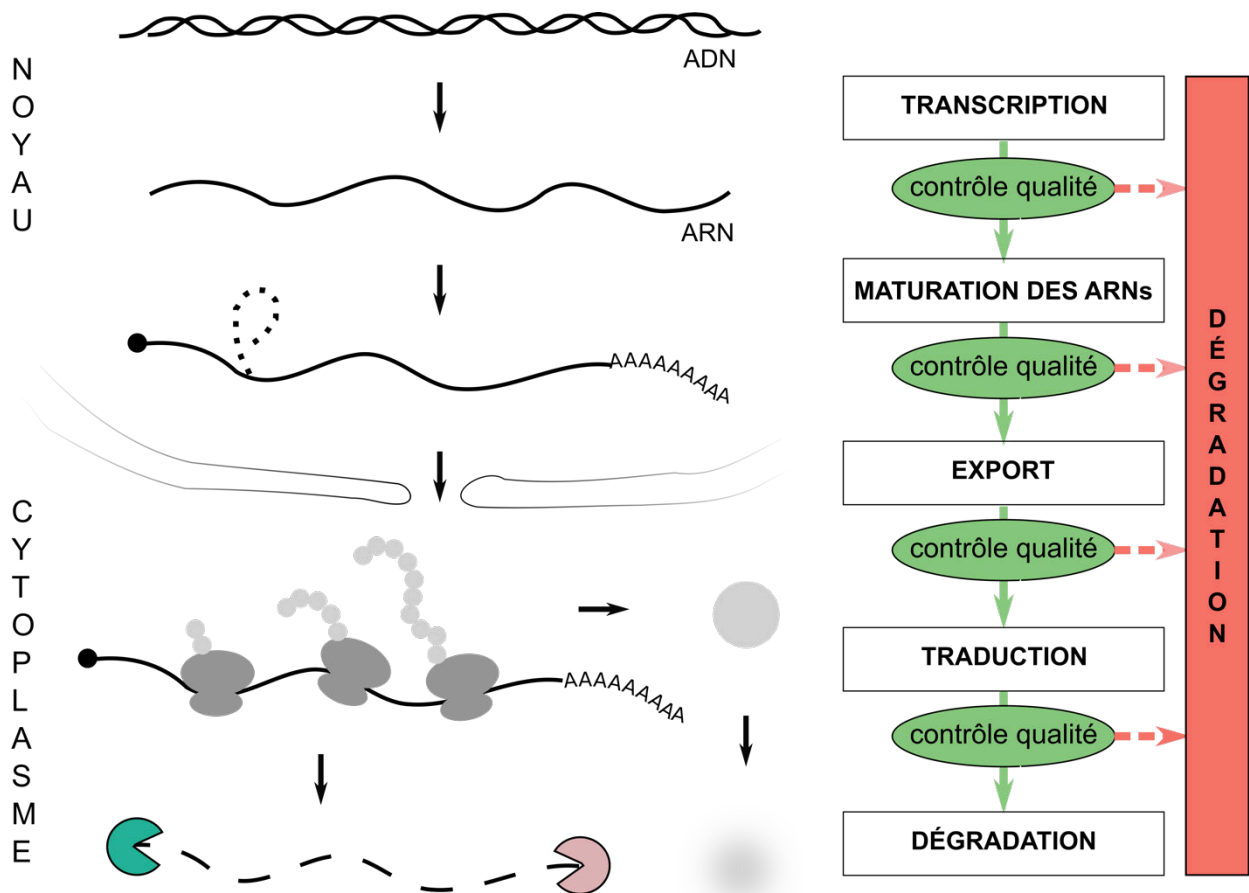


Figure 1 : Étapes de la vie d'un ARNm dans la cellule et mécanismes de contrôle qualité associés

À chaque étape, les mécanismes de surveillances permettent ou pas à l'ARN de passer à l'étape suivante. Dans le cytoplasme, tous les ARNs sont finalement dégradés par leurs deux extrémités par des exonucléases non spécifiques.

La transcription est la plus grande source d'erreur pour la synthèse des ARNs aberrants avec quatre erreurs pour un million de bases transcrites (Gout et al., 2017) c'est-à-dire cent fois plus que les erreurs de répliation de l'ADN (Lynch et al., 2016). Les erreurs peuvent être des insertions ou des délétions de base qui provoquent ensuite des changements du cadre de lecture pour la traduction et le codage de codon Stop précoce (PTC : *Premature Termination Codon*). Ces erreurs sont très délétères car elles peuvent aboutir à la production de protéines tronquées (Hall and Thein, 1994). L'effet de ces mutations dépend du gène dans lequel elle a lieu et si cela aboutit à la production d'une version dominant-négative de la protéine ou non.

Dans la plupart des cas, les protéines tronquées n'ont pas d'effet sur le phénotype, mais elles peuvent aussi être toxiques, en cas d'agrégation par exemple (Leeuwen et al., 1998) ou si elles entraînent une perte de fonction par effet de dose (Fan et al., 2001). Chez la levure *Saccharomyces cerevisiae* (*S. cerevisiae*), les erreurs de transcription sont associées à un stress protéotoxique ainsi qu'à une augmentation du vieillissement (Vermulst et al., 2015).

En plus des erreurs de fidélité de la transcription, le développement du séquençage des ARNs à haut débit a permis d'identifier des catégories d'ARN non traduits provenant de la forte activité de l'ARN polymérase II. Celle-ci s'associe aux régions d'ADN dépourvues de nucléosomes et peut initier des événements de transcription sur le brin sens comme sur le brin anti-sens (Neil et al., 2009; Xu et al., 2009), en amont ou en aval des sites de démarrage de la transcription canonique (Zhang and Dietrich, 2005; Pelechano et al., 2013; Malabat et al., 2015). C'est la transcription « pervasive » (résumé dans Jensen et al., 2013). Ces transcrits plus ou moins longs peuvent avoir une terminaison de la transcription anormale et, dans ce cas, sont reconnus par un complexe de surveillance des arrêts précoces de transcription et dégradés. Ces transcrits sont appelés les transcrits cryptiques non stables (CUT : *Cryptic Unstable Transcripts*), ils ne sont pas visibles dans des conditions sauvages car dégradés trop rapidement. Pour les observer, il faut inactiver le contrôle nucléaire (Wyers et al., 2005; Davis and Ares, 2006; Houalla et al., 2006). Parfois, les transcrits pervasifs échappent à ce contrôle précoce et sont exportés vers le cytoplasme. Dans d'autres cas, les transcrits ont une terminaison normale et sont de tout point de vue normaux par rapport aux systèmes de contrôle qualité nucléaire. Une fois arrivés dans le cytoplasme, ces divers types de transcrits aberrants entrent en traduction, ce qui déclenche leur dégradation rapide. Cette dégradation commence par l'enlèvement de la coiffe et est poursuivie par la dégradation par Xrn1, une exonucléase 5' vers 3' non spécifique. Si on observe une stabilisation d'un transcrit dans des cellules qui n'ont pas de Xrn1, on parle alors de XUT (*Xrn1-dependent Unannotated Transcripts*; van Dijk et al., 2011). Certains transcrits pervasifs sont plus résistants à la dégradation et restent suffisamment stables dans le cytoplasme pour être détectés, on les appelle les SUTs (*Stable Unannotated Transcripts*; Neil et al., 2009; Xu et al., 2009). Les XUTs et les SUTs sont des catégories non-exclusives, de nombreux transcrits appartiennent aux deux classes.

Les différentes étapes de maturation de l'ARN peuvent aussi être altérées. L'ajout de la coiffe est un mécanisme en deux étapes, dans un premier temps l'association d'une enzyme triphosphatase et d'une guanylyl transférase (Cet1/Ceg1 chez la levure *S. cerevisiae*) ajoute le

guanosyl-triphosphate puis Abd1 (*S. cerevisiae*) transfère un groupement méthyle formant la coiffe m⁷-GpppN (résumé dans Ramanathan et al., 2016). En cas de dysfonctionnement, par exemple si la méthylation n'a pas lieu, un mécanisme de contrôle nucléaire permet d'enlever la coiffe de guanine en clivant l'ARN rétablissant ainsi une extrémité 5'-monophosphate, qui sera dégradée par une exonucléase nucléaire. Chez les mammifères la protéine DXO assure seule ce contrôle (Jiao et al., 2013), chez la levure de boulanger, il existe deux systèmes redondants, la voie Rai1/Rat1 (Jiao et al., 2010) et la voie Dxo (Chang et al., 2012). Si l'épissage n'est pas correctement réalisé, des contrôles de qualité nucléaires évitent l'export de ces ARNs vers le cytoplasme. La poly-adénylation peut aussi être perturbée ou arriver prématurément à des sites cryptiques (Ozsolak et al., 2010). Un ARN non poly-adénylé n'est pas protégé et il est aussitôt dégradé par l'exosome nucléaire (Hilleren et al., 2001). Un ARN non exporté voit sa queue poly(A) allongée ce qui a pour conséquence le recrutement de l'exosome nucléaire qui dégrade les ARNs (Jensen et al., 2001). Par ailleurs les liens existant entre les différentes étapes sont aussi des contrôles, à l'image du complexe de jonction des exons (EJC) chez la plupart des eucaryotes qui est déposé pendant l'épissage sur les ARNs et participe à l'export efficace des ARNs dans le cytoplasme (Le Hir et al., 2001).

Enfin la traduction comporte aussi son lot d'erreur, une erreur par millier de bases (Zaher and Green, 2009). Le problème peut provenir des ARNs de transfert (ARNt) si l'anticodon et l'acide-aminé fixé ne correspondent pas. Le ribosome seul peut aussi engendrer des erreurs et incorporer un acide-aminé à la place d'un codon Stop ou décaler le cadre de lecture (résumé dans Dunkle and Dunham, 2015).

Les ARNs portants des erreurs ou résultants du manque de précision de la transcription et des étapes de maturations sont considérés comme aberrants et seront dégradés. La dégradation de ces ARNs utilise les voies classiques de dégradation des ARNs de la cellule dans le noyau et dans le cytoplasme.

II. La dégradation cytoplasmique des ARNs.

La dégradation des ARNs dans le cytoplasme suit deux voies distinctes (**Figure 2**). Chez la levure *S. cerevisiae*, ces deux voies impliquent une première étape commune de dé-adénylation (Decker and Parker, 1993; Muhrad and Parker, 1992) par les complexes Pan2/Pan3 et Ccr4/Pop2/Not (Brown and Sachs, 1998; Tucker et al., 2001). Pour la voie majeure de dégradation chez les eucaryotes, la dé-adénylation des ARNs est suivie de l'hydrolyse de la coiffe (décapping) en 5' par l'hétérodimère Dcp1/Dcp2 et le grignotage de

l'ARN de l'extrémité 5' vers le 3' par l'exonucléase Xrn1 (Beelman et al., 1996; Couttet et al., 1997; van Dijk et al., 2002; Dunckley and Parker, 1999; Hsu and Stevens, 1993; Muhlrud et al., 1994, 1995; Steiger et al., 2003). Ce mécanisme de dégradation est secondé par la voie de l'exosome cytoplasmique qui permet après dé-adénylation de dégrader les ARNs de l'extrémité 3' vers le 5' (Anderson and Parker, 1998) aidée du complexe SKI (Ski2, 3, 8) et de la protéine Ski7. Ces deux voies existent chez tous les eucaryotes et leur importance varie en fonction des espèces. La dégradation des ARNs dans le cytoplasme est le dernier levier de régulation de l'expression des gènes. Le temps de vie d'un ARN varie en fonction des espèces d'ARN, de trois minutes à plus de cent minutes chez la levure *S. cerevisiae* (Wang et al., 2002).

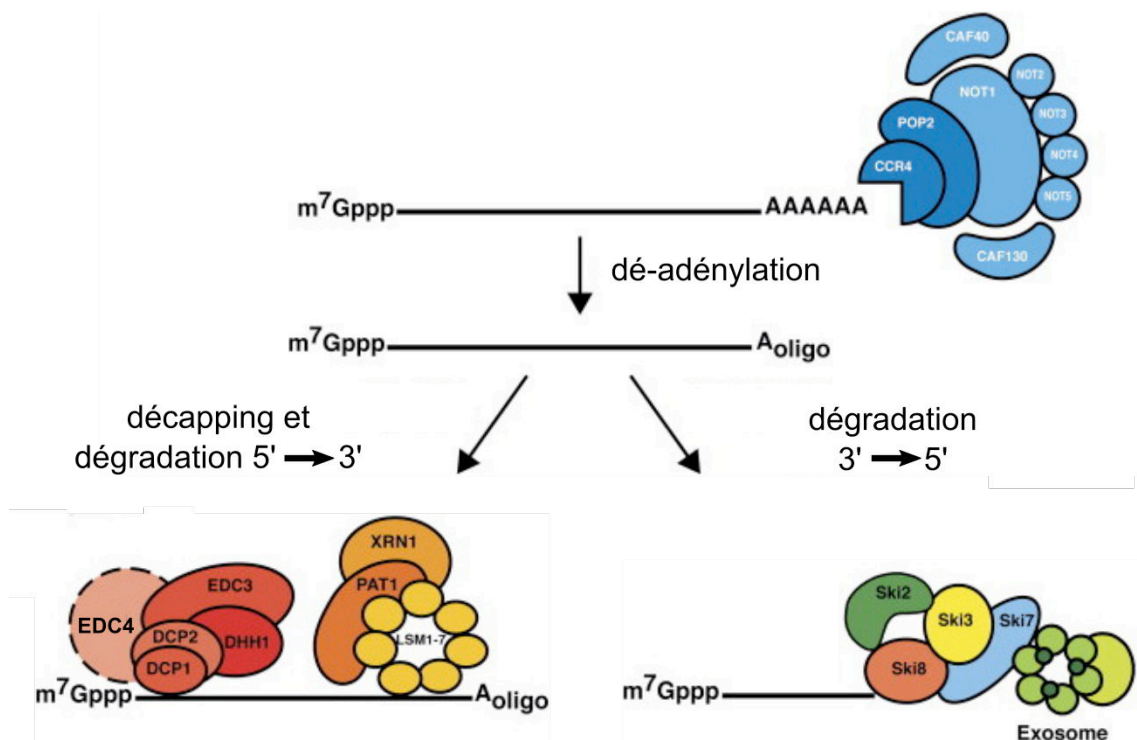


Figure 2 : Voies générales de dégradation des ARNs cytoplasmiques

(Adapté de Parker and Sheth, 2007)

1. Dé-adénylation

La dé-adénylation est le processus permettant la dégradation de la queue poly(A) des ARNs. La queue poly(A) est ajoutée co-transcriptionnellement dans le noyau par une ARN polymérase particulière. Sa taille est homogène au sein d'une espèce mais peut varier entre les espèces, environ 70 adénines pour la levure *S. cerevisiae* et jusqu'à 250 chez d'autres eucaryotes. Elles sont liées à des protéines de liaison des queues poly(A), les PABPs

(*Poly(A)-Binding Protein*), Pab1 chez *S. cerevisiae* (Sachs et al., 1986) et PABPC1 chez les métazoaires (Grange et al., 1987). La queue poly(A) et les protéines qui interagissent sur cette partie ont un double rôle dans le métabolisme des ARNs dans le cytoplasme, d'une part elles stabilisent les ARNs et activent l'initiation de la traduction *via* la formation du complexe de pré-initiation (**Figure 4A**), d'autre part, elles participent au recrutement des protéines initiant la dé-adénylation (Sachs and Davis, 1989) et donc la dégradation des ARNs.

Chez la levure *S. cerevisiae*, plusieurs complexes protéiques sont impliqués dans l'hydrolyse de la queue poly(A), Pan2/Pan3 (Boeck et al., 1996; Brown et al., 1996) et Ccr4/Pop2/Not (Tucker et al., 2001).

Pan2/Pan3 est recruté au niveau de la queue poly(A) *via* l'interaction de Pan3 avec Pab1 (Siddiqui et al., 2007). L'hydrolyse de la queue poly(A) est catalysée par le domaine exonucléase de Pan2 qui appartient à la famille des RNAses D. Pan3 est la sous-unité régulatrice du complexe. L'activité de Pan2/Pan3 est promue par Pab1 (Boeck et al., 1996), et en condition normale le complexe permet la dégradation de la première partie de la queue poly(A) non liée aux protéines Pab1 (**Figure 3B**, Yi et al., 2018).

Le complexe Ccr4/Pop2/Not est composé de deux sous-unités catalytiques, Ccr4 et Pop2 (CAF1 chez les mammifères, Bianchin et al., 2005; Daugeron et al., 2001; Thore et al., 2003) qui interagissent (Basquin et al., 2012). Les deux protéines ont une activité de dé-adénylase similaire *in vitro* mais elles semblent ne pas avoir la même spécificité de substrats au moins chez la levure. En effet, Ccr4 est active pour la dé-adénylation des régions poly(A) riches en PABPs alors que Pop2 permet de dé-adényler les régions pauvres en PABPs (**Figure 3A**, Webster et al., 2018). Leur action est complémentaire, cependant, chez la levure *S. cerevisiae* la délétion de *CCR4* a plus d'impact que celle de *POP2* sur la croissance (Tucker et al., 2002) et indiquerait un rôle moins important pour cette dernière. Le complexe Ccr4/Pop2/Not est recruté au niveau des ARNs par des interactions avec des protéines liant le 3'-UTR, par exemple les protéines de la famille des Puf (*Pumilio Family*) chez la levure *S. cerevisiae* (Olivas and Parker, 2000). Chez les mammifères, le complexe Ccr4/Caf1/Not est aussi recruté au niveau d'ARNs précis (Goldstrohm et al., 2006; Stowell et al., 2016; Wahle and Winkler, 2013) pour initier la dé-adénylation rapide dans divers processus biologiques, dont l'embryogenèse, la réponse immunitaire et la prolifération cellulaire (Belloc et al., 2008; Carballo et al., 1998; Subtelny et al., 2014). Les protéines Not1-5, Caf130 et Caf40 sont des cofacteurs pour la dé-adénylation conservées chez les eucaryotes. Not1 est une protéine plateforme (Bai et al., 1999), elle interagit avec Pop2 (Caf1) et d'autres sous-unités Not du

complexe (Basquin et al., 2012; Petit et al., 2012). Les sous-unités Not2, Not3 et Not5 de levure activent l'hydrolyse de la coiffe et donc la dégradation des ARNs *via* une interaction avec le facteur Pat1 (Alhusaini and Coller, 2016). Les autres sous-unités n'ont pas de fonctions définies.

La délétion isolée de *CCR4* ou de *PAN2* entraîne un allongement des queues poly(A) (Boeck et al., 1996; Tucker et al., 2001), ce qui semble montrer une redondance fonctionnelle entre les deux complexes. Cependant, la dé-adénylation complète des ARNs ne peut être assurée que par la combinaison des deux enzymes. Des expériences de pulse-chase pour l'inactivation des deux complexes montrent finalement un mécanisme coopératif, mineur chez la levure (Tucker et al., 2001) et la drosophile et plus important chez l'Homme. Le complexe Pan2/3 agit le premier pour une première phase de dé-adénylation rapide de 90 nucléotides à environ 65 chez la levure (Brown and Sachs, 1998; Yi et al., 2018) puis Ccr4/Pop2/Not est recruté dans une deuxième phase (Tucker et al., 2002; Yi et al., 2018). On observe ensuite une pause lorsque la queue poly(A) est réduite à 10-12 bases (Decker and Parker, 1993).

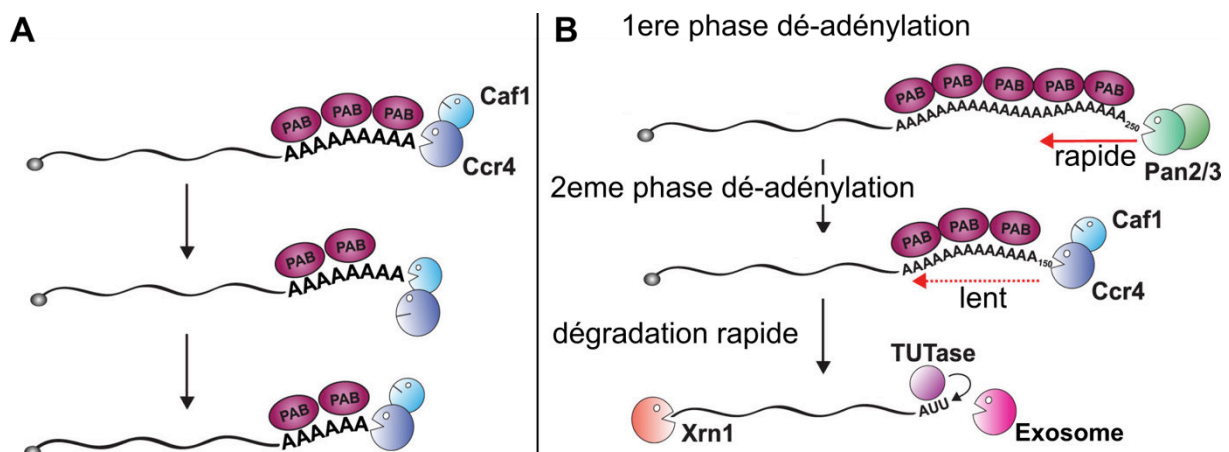


Figure 3 : Schéma des modèles de dé-adénylation des ARNs

(Adapté de Bresson and Tollervey, 2018). **A.** Coopération de Pop2 (Caf1) et Ccr4 pour la dé-adénylation de la queue poly(A) liée aux PABPs. Ccr4 permet d'enlever les PABPs pour dé-adényler et Caf1 permet la dé-adénylation des régions non protégées suite à l'action de Ccr4. **B.** Rôle des deux complexes de dé-adénylation. Pan2/3 est responsable de la dé-adénylation précoce et rapide des régions initialement non protégées par les PABPs puis Ccr4/Pop2(Caf1)/Not prend le relais comme indiqué en A. Les TUTases (*Terminal Uridyl Transferases*) sont des protéines présentes chez l'humain permettant d'ajouter des U à la fin des transcrits après dé-adénylation pour accélérer la dégradation par l'exosome. La levure *S. cerevisiae* n'a pas de TUTase connue.

2. Dégradation 5' vers 3'

Après la pause de dé-adénylation à 10-12 adénines, le complexe Pat1/Lsm1-7 est recruté à la place de Pab1. Cet échange est probablement un signal permettant l'activation de l'hydrolyse de la coiffe (Chowdhury et al., 2007; Tharun and Parker, 2001) et l'initiation de la dégradation du bout 5' vers l'extrémité 3'.

L'hydrolyse de la coiffe peut être aussi indépendante de la dé-adénylation préalable. Par exemple le transcrit RPS28B est dégradé rapidement par activation du décapping dans une voie dépendante de la protéine Rps28 et du cofacteur de décapping Edc3 (Badis et al., 2004; He et al., 2014).

Le complexe d'hydrolyse de la coiffe est composé de deux protéines principales, Dcp1 et Dcp2. Dcp2 est la sous-unité catalytique (van Dijk et al., 2002) permettant le clivage de la coiffe libérant le m⁷-Gpp de l'extrémité 5'-monophosphate du transcrit (She et al., 2006, 2008). Dcp1 interagit avec Dcp2 et promeut son activité (Deshmukh et al., 2008; She et al., 2004, 2008). Les 300 premiers nucléotides de Dcp2 sont suffisants pour initier l'hydrolyse de la coiffe (Dunckley and Parker, 1999), c'est la région conservée entre les espèces. Dcp1 et Dcp2 forment une holoenzyme régulée par de multiples cofacteurs : Dhh1, Edc3, Pat1, Lsm1-7, Scd6, Sbp1, Mrt1-3 (Hatfield et al., 1996 pour Mrt1-3; He and Jacobson, 2015), Edc4 (Chang et al., 2014), etc. Ces cofacteurs sont impliqués dans les différentes étapes permettant l'hydrolyse de la coiffe, le raccourcissement de la queue poly(A) évoqué précédemment, l'inhibition de la traduction et le décrochage des facteurs d'initiation de la traduction protégeant la coiffe. Dhh1, par exemple, est un activateur de l'hydrolyse de la coiffe (Coller et al., 2001), il a un rôle d'inhibition précoce de l'initiation de la traduction qui permettrait d'affaiblir les interactions entre la coiffe et les facteurs d'initiation la rendant plus accessible à Dcp1/Dcp2. Scd6 et Sbp1, en liant eIf4g (**Figure 4A**) auraient aussi un effet d'inhibition de la traduction favorisant l'accès à la coiffe (Nissan et al., 2010). Pat1, comme expliqué précédemment permet le recrutement de Dcp1/Dcp2 et Xrn1 au niveau des ARNs (Charenton et al., 2017 pour Dcp2/Pat1; Fromont-Racine et al., 2000; Ho et al., 2002; Krogan et al., 2006 pour Pat1/Xrn1; Tarassov et al., 2008; Tong et al., 2004) où la queue poly(A) est raccourcie. Il contribue aussi à l'inhibition de la traduction (Nissan et al., 2010) et il active Dcp2. Lsm1-7 promeut l'activation de Dcp2 par Pat1 et permet de protéger l'extrémité 3' de l'ARN pendant la dégradation par Xrn1. Edc3 interagit avec Dcp2, permet de lever son auto-inhibition (Harigaya et al., 2010; He and Jacobson, 2015) et il est présent dans la forme active de l'holoenzyme (Charenton et al., 2016). Chez la levure *S. cerevisiae*, Dcp2 a une longue

extension C-terminale non conservée chez les autres eucaryotes et non requise pour l'hydrolyse de la coiffe. Cette région possède des domaines supplémentaires d'interactions pour plusieurs cofacteurs comme Pat1, Edc3 ou Upf1 (voir **partie III.5.** « Upf1, facteur majeur du NMD » de l'introduction) qui pourraient avoir un rôle dans le choix de certains substrats particuliers. Le domaine C-terminal pourrait aussi avoir un rôle dans des fonctions accessoires de Dcp2, comme son lien possible avec la transcription (Grousl et al., 2009; Shalem et al., 2011) chez la levure.

Plusieurs équipes ont tenté de figer le complexe Dcp1/Dcp2 dans une forme active (Charenton et al., 2016; Mugridge et al., 2016; Valkov et al., 2017; Wurm et al., 2017), cela a abouti à un mécanisme d'activation de Dcp1/Dcp2 par les cofacteurs Edc1 et Edc3 *via* la levée de l'auto-inhibition et la stabilisation du complexe Dcp1/Dcp2 sur les ARNs (Mugridge et al., 2018).

Une fois la coiffe dégradée, le complexe d'hydrolyse de la coiffe laisse la place à l'exonucléase Xrn1 (Stevens, 1978). Xrn1 est l'exonucléase cytoplasmique majeure, elle est processive, directionnelle et requiert une extrémité 5'-monophosphate libre (Stevens, 1980). Chez les mammifères, la coordination de l'hydrolyse de la coiffe et de la dégradation pourrait être organisée au niveau de Edc4 (*alias* Hedls), protéine plateforme permettant l'interaction de Dcp1/Dcp2 et de Xrn1 (Chang et al., 2014). La dégradation de l'ARN par Xrn1 est co-translationnelle (Hu et al., 2009; Pelechano et al., 2015), elle suit les ribosomes.

3. Dégradation 3' vers 5'

La deuxième voie de dégradation des ARNs est la voie de l'exosome. L'exosome cytoplasmique de levure est composé de dix sous-unités dont Rrp44 (Dis3) qui possède une activité exonucléolytique et endonucléolytique (Bousquet-Antonelli et al., 2000; Lebreton et al., 2008; Schaeffer et al., 2009; Schneider et al., 2009), et neuf autres protéines formant un tunnel permettant l'adressage de l'ARN au niveau du site catalytique de Dis3 (Bonneau et al., 2009). L'activité endonucléolytique pourrait ne pas être requise pour l'activité cytoplasmique de Dis3, d'ailleurs, l'équivalent mammifère Dis3L, uniquement cytoplasmique, n'a pas conservé cette capacité (résumé dans Zinder and Lima, 2017). Pour la dégradation des ARNs cytoplasmiques, l'exosome requiert l'aide du complexe SKI. Ce complexe est composé de Ski2, Ski3 et Ski8 et interagit avec l'exosome *via* Ski7, une protéine faisant le pont entre les deux complexes. Les protéines du complexe SKI forment aussi un canal permettant de diriger l'ARN à dégrader vers l'exosome (Halbach et al., 2013). La protéine Ski2 est probablement une hélicase ARN et elle pourrait permettre de dérouler les structures de l'ARN pour faciliter

ensuite la dégradation. Le recrutement de l'exosome a lieu au niveau des extrémités 3' non protégées ainsi que sur certains ARNs *via* la liaison de protéines sur des motifs spécifiques (Chen et al., 2001), cependant les mécanismes sont encore peu clairs.

4. Régulation des voies de dégradation des ARNs

Les caractéristiques permettant d'orienter un ARN vers la voie 5' vers 3' ou la voie de l'exosome ne sont pas connues. Il semble que certains ARNs soient plus sensibles à l'une ou l'autre des voies. L'importance des deux voies est aussi différente en fonction des espèces. Ainsi la levure utilise principalement la voie de décapping dépendant de la dé-adénylation tandis que chez les plantes c'est la voie de l'exosome qui semble majoritaire. Pourquoi existe-t-il de telles différences ? Quels sont les éléments impliqués dans la régulation de ces mécanismes ?

La disponibilité des facteurs peut être une explication pour le choix différentiel des voies selon les espèces ou lorsque des modifications de l'environnement affectent leur abondance. Par exemple, chez la levure, la dégradation des ARNs a lieu principalement par la voie 5' vers 3' sauf si celle-ci est perturbée (par exemple mutation d'une enzyme du décapping), dans ce cas, l'exosome cytoplasmique prend le relais. La dégradation est aussi régulée *in cis* *via* des éléments de l'ARN directement (Cheng et al., 2017). Un exemple déjà évoqué est le recrutement des facteurs de la dé-adénylation *via* une interaction avec des protéines liant les 3'-UTR. Par exemple, une étude récente montre l'effet activateur de la protéine Mmi1 sur la dégradation d'un ARN de *Schizosaccharomyces pombe* (*S. pombe*). Mmi1 interagit avec une épingle d'ARN située dans le 3'-UTR et accélère le recrutement du complexe Ccr4/Pop2/Not (Stowell et al., 2016, 2018). Dans d'autres cas de régulation, le court-circuitage de la dé-adénylation par le décapping permet de dégrader rapidement des ARNs spécifiques (Badis et al., 2004; He et al., 2014). Le taux de dégradation des ARNs est aussi influencé par l'efficacité de la traduction en lien avec l'optimisation des codons le long de la séquence codante (Presnyak et al., 2015). De plus en plus d'articles décrivent des liens entre l'usage des codons, l'efficacité de la traduction et la dégradation des ARNs (Bazzini et al., 2016; Mishima and Tomari, 2016; Presnyak et al., 2015; Radhakrishnan and Green, 2016). Par exemple, une étude propose un mécanisme différentiel des deux dé-adénylases Ccr4 et Pop2 dans le cas d'une demi-vie courte ou longue pour les ARNs (Webster et al., 2018). Cette étude montre que la dé-adénylation des ARNs optimisés pour l'usage de leur codon, traduit efficacement et possédant plus de PABPs sur leur queue poly(A) (Beilharz and Preiss, 2007) utilise

préférentiellement la dé-adénylase Ccr4 alors que les ARNs non-optimisés sont dé-adénylés par Pop2. La différence est que Ccr4 doit enlever au fur et à mesure les PABPs ce qui ralentit sa progression et donc ralentit la dégradation alors que Pop2, recruté sur des queues poly(A) pauvres en PABPs, permet une dé-adénylation rapide suivie de la dégradation des ARNs (**Figure 3**).

III. Dégradation des ARNs aberrants

Dans le cas des ARNs aberrants, il existe des voies de surveillances permettant de recruter spécifiquement la voie de l'hydrolyse de la coiffe ou la voie de l'exosome cytoplasmique. Les ARNs aberrants sont des ARNs contenant des erreurs de séquence ou issues des manquements aux étapes de transcription, maturation ou export. Des contrôles nucléaires existent et permettent l'élimination d'une partie des ARNs aberrants, les autres sont exportés dans le cytoplasme. Leur traduction est alors le dernier rempart pour détecter des erreurs. Trois mécanismes de contrôle majeurs des ARNs existent dans le cytoplasme et dépendent de la traduction.

1. Mécanisme de la traduction

Après l'export dans le cytoplasme, les ARNm sont traduits par les ribosomes. Tout d'abord la petite sous-unité du ribosome (40S pour *S. cerevisiae*) chargée avec l'ARN de transfert (ARNt) initiateur (anticodon de la méthionine AUG) s'assemble sur l'ARN au niveau de la coiffe. Elle est accompagnée de nombreux facteurs de pré-initiation dont eIf4e qui lie directement le m⁷-Gppp et eIf4g stabilisant le complexe formé grâce à une interaction avec les PABPs (Tarun and Sachs, 1997; Tarun et al., 1997). L'ARN serait circularisé (Wells et al., 1998), et cela permettrait une bonne stabilité du complexe de pré-initiation importante pour scanner le 5'-UTR jusqu'à la méthionine initiatrice (codon Start). Une fois arrivée au codon Start, la grande sous-unité du ribosome est recrutée pour former le ribosome (80S pour *S. cerevisiae*) qui commence ensuite la traduction (**Figure 4A**). Le ribosome entre ensuite dans une phase d'élongation (**Figure 4B**), à chaque cycle un ARNt correspondant au codon suivant de l'ARN entre dans le site accepteur (site A) du ribosome chargé de son acide-aminé et escorté par le facteur d'élongation eEfla. La liaison peptidique entre l'acide-aminé n-1 et n est réalisée au sein du ribosome au niveau du site peptidyl-transférase (site P). La catalyse de cette réaction ainsi que l'action de facteur d'élongation apporte l'énergie nécessaire au ribosome pour transloquer de trois nucléotides, permettant à un nouvel ARNt n+1 de se

positionner au niveau du site A vide. Le ribosome poursuit la traduction jusqu'au codon Stop des ARNs marquant la fin de la séquence codant la protéine. Dans le code génétique classique, trois codons (UAA, UAG, UGA) indiquent la fin de la traduction et n'ont pas d'ARNt correspondant. À la place, les facteurs de terminaison de la traduction, eRF1 et eRF3 (*eukaryotic Release Factor 1 & 3*), sont incorporés au site A du ribosome, permettent le relâchement du polypeptide (**Figure 4C**) et initient les premières étapes du recyclage du ribosome. Le complexe eRF1/eRF3/GTP est recruté au niveau du site A du ribosome (Mitkevich et al., 2006) ; eRF1 reconnaît le codon Stop (Bertram et al., 2000; Blanchet et al., 2015) car sa structure similaire à celle d'un ARNt (Cheng et al., 2009), lui permet de rentrer dans la poche du ribosome ; eRF3 *via* sa liaison au GTP apporte l'énergie nécessaire pour la réaction. L'hydrolyse du GTP en GDP et Pi (Phosphate inorganique) modifie la conformation de eRF1 au sein du complexe (Alkalaeva et al., 2006). La nouvelle organisation de eRF1/eRF3 permet alors au domaine GGQ (Glycine-Glycine-Glutamine(Q)) de eRF1 de s'immiscer au niveau du site de liaison peptidique afin de catalyser la réaction d'hydrolyse de la liaison de l'ARNt avec son acide-aminé (Frolova et al., 1999; Salas-Marco and Bedwell, 2004) permettant au polypeptide de se détacher du ribosome (Jin et al., 2010). Une fois le polypeptide relâché, le ribosome reste accroché à l'ARN et d'autres mécanismes permettent son recyclage (résumé dans Hellen, 2018 ; **Figure 4D**). Après libération du peptide naissant, eRF3 se décroche de eRF1 et le facteur ABCE1 (Rli1 chez la levure *S. cerevisiae*) prend le relais. ABCE1 est une ATPase, elle interagit avec eRF1 et la sous-unité 40S du ribosome (Brown et al., 2015; Preis et al., 2014; Shao et al., 2016). ABCE1 initie la séparation entre la grande sous-unité du ribosome et la petite. Grâce à la liaison de l'ATP puis à son hydrolyse, ABCE1 se lie au ribosome en terminaison puis change de conformation pour adopter un état dit ouvert (Rees et al., 2009), qui conduit à des modifications des facteurs de la terminaison comme eRF1. Ces modifications pourraient affaiblir les liaisons entre les sous-unités du ribosome et entraîner le départ de la sous-unité 60S. L'hydrolyse d'ATP par ABCE1 permet ensuite son propre recyclage (Barthelme et al., 2011). La sous-unité 40S du ribosome reste accrochée à l'ARNm ainsi qu'à l'ARNt situé en position P. Les facteurs d'initiation de la traduction eIf1, eIf1a et eIf3 sont alors recrutés, leurs interactions avec la sous-unité 40S affaiblissent les liaisons avec l'ARNm et l'ARNt qui sont alors dissociés (Pisarev et al., 2007, 2010).

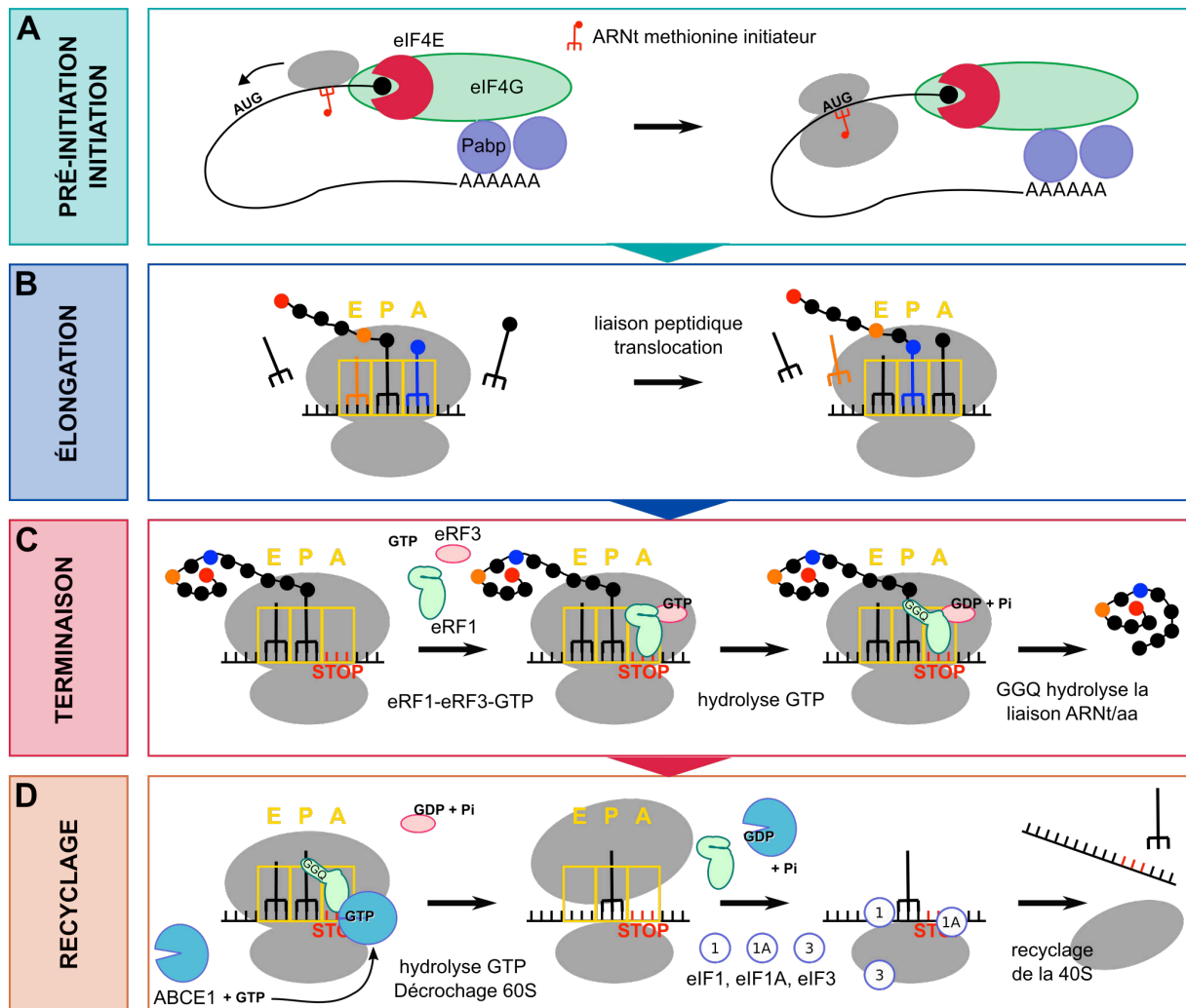


Figure 4 : Processus de traduction des ARNm eucaryotes

A. Pré-initiation et initiation. La coiffe des ARNs exportés vers le cytoplasme interagit avec le complexe de liaison de la coiffe, CBC (*Cap Binding Complex*) composé de Cbp20 et Cbp80, dans les premières étapes de l'initiation de la traduction ces facteurs sont remplacés par eIF4e, eIF4g et d'autres pour stabiliser le transcrit dans le cytoplasme et permettre sa traduction efficace. La petite sous-unité du ribosome scanne le 5'-UTR jusqu'au codon Start où la grande sous-unité la rejoint et débute la traduction. **B.** Pendant l'élongation, un ARNt apporte à chaque cycle l'acide-aminé suivant correspondant au codon sur l'ARNm aidé par des facteurs d'élongation qui permettent aussi la translocation du ribosome le long de l'ARN. **C.** La terminaison a lieu au niveau du codon Stop. Les facteurs eRF1 et eRF3 reconnaissent le ribosome au niveau du Stop et catalysent le relâchement du polypeptide produit correspondant à la protéine codée par l'ARN. Elle est assistée par les PABPs, notamment *via* une interaction entre eRF3 et Pab1. **D.** Le recyclage du ribosome permet de recommencer un cycle de traduction sur un autre ARN. Il est initié par ABCE1 (Rli1).

2. Contrôles dépendant de la traduction

Les trois mécanismes de contrôle associés à la traduction sont le *No-Go mRNA Decay* (NGD), le *Non-Stop mRNA Decay* (NSD) et le *Nonsense-Mediated mRNA Decay* (NMD). Ces contrôles qualité ont, d'une part, un effet sur l'ARN aberrant qui sera dégradé et d'autre part, sur la protéine nouvellement produite qui sera aussi éliminée par des contrôles qualité des protéines associés pour le NGD et le NSD. Pour le NMD, il n'y a pas de résultats clairs montrant une dégradation spécifique des protéines tronquées néo-synthétisées.

Le NGD est un mécanisme s'occupant des arrêts de la traduction pendant l'élongation. L'arrêt peut être dû à une structure sur l'ARN, tige-boucle, pseudoknot, qui bloque l'avancée du ribosome (**Figure 5B**, Doma and Parker, 2006). Ce blocage va entraîner une collision de ribosome en amont du blocage qui est importante pour déclencher le NGD (Simms et al., 2017). Le ribosome bloqué n'ayant pas de codon Stop au site accepteur, la terminaison selon le mécanisme classique ne peut pas avoir lieu. La pause prolongée du ribosome bloqué déclenche un clivage endonucléolytique de l'ARNm à proximité du site de blocage (Doma and Parker, 2006). La coupure génère un fragment 3' rapidement dégradé par Xrn1 et un fragment 5' avec le ribosome bloqué. Le recyclage du ribosome bloqué est initié par le couple de protéines Dom34 (Pelota pour les mammifères)/Hbs1 respectivement paralogues de eRF1 et eRF3 (Graille et al., 2008 pour Dom34; Lee et al., 2007 pour Hbs1). L'interaction de Dom34 et Hbs1 au niveau du site A du ribosome permet de recruter ABCE1 (Rli1) pour la dissociation du ribosome comme décrit dans la **Figure 4D**. Le fragment 5' libre de l'ARN est ensuite dégradé par l'exosome (Doma and Parker, 2006; Schmidt et al., 2016). Dom34 ne possède pas le domaine GGQ d'eRF1 permettant l'hydrolyse de la liaison ARNt/acide-aminé au site P du ribosome. Cela aboutit à la persistance du peptide naissant et du dernier ARNt incorporé, accroché à la sous-unité 60S du ribosome (Shoemaker et al., 2010). Ce ribosome particulier est alors reconnu par un complexe de protéines appelé RQC (*Ribosome-based Quality Control*) permettant l'extraction du peptide naissant et le recyclage de la sous-unité 60S (résumé dans Defenouillère and Fromont-Racine, 2017).

D'autres situations aboutissent aussi à un blocage de l'élongation et donc à un événement de NGD. Il a été montré qu'une suite de codons rares pouvait ralentir la traduction jusqu'à l'arrêt (Doma and Parker, 2006; Herrick et al., 1990). De même, une succession d'adénosines (codon AAA) sur l'ARN entraîne la traduction d'une suite d'acides-aminés chargés (Lysine) pouvant rester coincée dans le tunnel de sortie du polypeptide du ribosome *via* des interactions

électrostatiques et aboutir à un arrêt de l'élongation. On retrouve cette dernière configuration dans le cas d'ARNs sans codon Stop mais poly-adiénylés (Guydosh and Green, 2017).

Le NSD reconnaît les ARNs n'ayant pas de codon Stop (**Figure 5C**, van Hoof et al., 2002). Ces transcrits pourraient être assez nombreux dans la cellule et provenir de la reconnaissance de sites cryptiques de poly-adiénylation qui peuvent arriver au niveau de la séquence codante (Frischmeyer et al., 2002; Graber et al., 1999; Mahadevan et al., 1997; Ozsolak et al., 2010; Sparks and Dieckmann, 1998), d'ARNs où le codon Stop a été décodé par erreur par un acide-aminé (*readthrough* : Baudin-Baillieu et al., 2014), ou d'ARNs en voie de dégradation du côté 3'. Les ribosomes en fin de transcrit sans codon Stop s'arrêtent avec leur site accepteur vide. Les étapes permettant la reconnaissance de ce ribosome bloqué sont encore peu définies. Le complexe Dom34/Hbs1 serait recruté au niveau du site A (Saito et al., 2013) permettant d'initier le recyclage du ribosome de la même manière que pour le NGD et la dégradation du peptide naissant. Il n'y a pas de coupure endonucléolytique, le complexe SKI est recruté au niveau de la petite sous-unité du ribosome et permettrait ensuite de recruter l'exosome *via* l'interaction avec Ski7. La protéine Ski7 chez la levure a une structure proche du facteur d'élongation eEfla et du facteur de terminaison eRF3. Une hypothèse est que Ski7 serait recruté au niveau du site A vide pour permettre le recrutement du complexe SKI et de l'exosome au niveau du 3' de l'ARN aberrant (van Hoof et al., 2002).

Le NGD et le NSD sont très liés. Ils font intervenir le même groupe de protéines et il est parfois difficile de discriminer entre leurs cibles car elles peuvent se mélanger. Par exemple, après un événement de NGD, la coupure de l'ARN tronque celui-ci au niveau du blocage qui se retrouve être alors un parfait substrat du NSD pour les ribosomes suivants. Le cas du blocage du ribosome dans la queue poly(A) peut aboutir à un événement de NGD comme décrit précédemment (Guydosh and Green, 2017) mais aussi à un événement de NSD si le ribosome traduit jusqu'au bout de la queue poly(A).

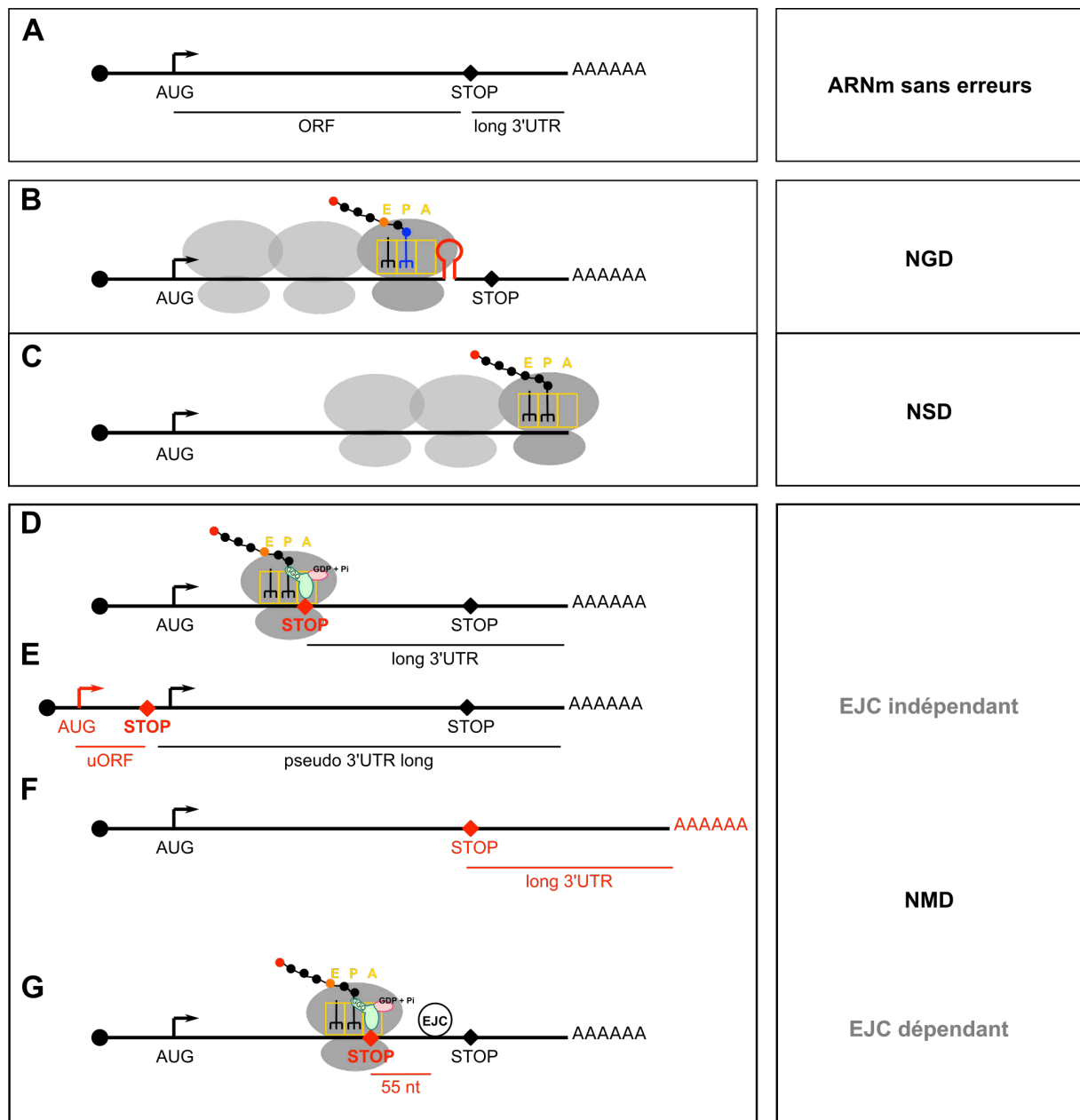


Figure 5 : Voies cytoplasmiques de surveillance des ARNs

A. Schéma d'un ARNm mature traduit. **B.** Activation du NGD par le blocage du ribosome au niveau d'une séquence bloquante (structure, suite de codon rare, etc.) dessinée en rouge. **C.** Activation du NSD par l'absence de codon Stop et le blocage du ribosome à l'extrémité du transcrit. **D.** à **G.** Activation du NMD par l'existence d'une terminaison de la traduction non efficace au niveau du PTC (Stop rouge) causée par un allongement du 3'-UTR ou pseudo 3'-UTR (**D.**, **E.** et **F.**) ou à la présence du complexe EJC à 55 nucléotides en aval du PTC (**G.**). eRF1 et eRF3 sont représentés comme en **Figure 4.**

3. Activation du NMD au niveau de ses substrats

Le NMD est un mécanisme de contrôle de la qualité des ARNs en cours de traduction (Ruiz-Echevarria et al., 1996; Zhang et al., 1997) dans le cytoplasme. Ce mécanisme a été initialement identifié comme reconnaissant les ARNs ayant des codons Stop précoces (PTC : *Premature Termination Codon* ; Losson and Lacroute, 1979). Ces PTCs peuvent provenir de mutations, de l'inclusions d'un intron (He et al., 1993; Sayani et al., 2008) ou de changements de phase du ribosome sur l'ARNm (*frameshift*, Belew et al., 2011). C'est aussi un mécanisme permettant de nettoyer la cellule des transcrits aberrants générés par le manque de précision de l'initiation et de la terminaison de la transcription. Ces transcrits incluent ; des ARNs codant des petites ORFs en amont (*uORF*, Gaba et al., 2005; Guan et al., 2006; Oliveira and McCarthy, 1995; Ruiz-Echevarría and Peltz, 2000; Welch and Jacobson, 1999) ou en aval de l'ORF annoté (iTSS: *internal Transcription Start Site*, Malabat et al., 2015) hors phase rapidement interrompues par un codon Stop ; les ARNs pervasifs ayant échappés à la surveillance nucléaire (XUTs, SUTs) et les ARNs dont le 3'-UTR a été allongé suite à l'oubli du site de poly-adénylation (Muhlrad and Parker, 1999a). Chez la levure, 47% des gènes sont susceptibles de générer des ARNs substrats de NMD (Malabat et al., 2015) ce qui, en quantité, correspond à environ 5 à 10 % des transcrits. Le déclenchement du NMD conduit à la dégradation de l'ARN très rapidement par plusieurs voies de dégradation.

Les facteurs du NMD sont universellement conservés chez les eucaryotes. On peut trouver des homologues anciens des protéines cœurs chez les archées ce qui tend à montrer l'existence du NMD chez la majeure partie des espèces. Cependant, plusieurs modèles ont été proposés pour le mécanisme moléculaire chez les levures et chez les métazoaires. Le point commun entre ces modèles est la discrimination entre un événement de terminaison de la traduction précoce au niveau des ARNs substrats et une terminaison « normale ». Les chercheurs ont longtemps pensé que des facteurs de dégradation se liaient sur la séquence codante au niveau des DSE (*Downstream Sequence Element*), et permettaient la dégradation des ARNs (Hagan et al., 1995; Hennigan and Jacobson, 1996; Zhang et al., 1995). Dans le cas d'un ARN normal, la traduction enlève ces facteurs et l'ARN est stable alors que dans le cas d'un PTC, des facteurs de dégradation restent accrochés et initient la dégradation rapidement. Ce modèle ne concerne très probablement qu'une minorité de transcrits.

Chez la levure *S. cerevisiae*, la distance entre le PTC et la fin du transcrit, la queue poly(A), est un facteur corrélé à la reconnaissance des cibles du NMD (Amrani et al., 2004; Muhlrad and Parker, 1999a). L'allongement du 3'-UTR pourrait perturber les interactions entre les

facteurs de terminaison et d'autres facteurs situés en aval ce qui empêcherait la terminaison efficace de la traduction. En particulier, il semble que la distance entre le PTC et la queue poly(A) pourrait empêcher l'interaction directe de Pab1 avec eRF3. Cette interaction permet d'activer l'hydrolyse du GTP lié à eRF3 (**Figure 4**), sa coupure aurait comme effet de ralentir la terminaison (Amrani et al., 2006; Hoshino et al., 1999; Kononenko et al., 2010). Cela permettrait alors aux facteurs du NMD de venir interagir (Czaplinski et al., 1998; Ivanov et al., 2008; Kashima et al., 2006; Singh et al., 2008; Wang et al., 2001). Ce modèle est appelé « Faux 3'-UTR » (Amrani et al., 2004).

Les données en faveur du modèle « Faux 3'-UTR » sont nombreuses. Par exemple, des expériences d'ancrage artificiel de la protéine Pab1 à proximité des PTCs sur un ARN rapporteur du NMD montrent une inhibition de la dégradation de l'ARN (Amrani et al., 2004; Behm-Ansmant et al., 2007a; Silva et al., 2008). De plus, le rapprochement physique de la queue poly(A) et du PTC par la formation de structure dans le 3'-UTR inhibe aussi le NMD (Eberle et al., 2008). Cependant, d'autres résultats remettent en question une partie des mécanismes proposés, notamment la compétition entre Pab1 et Upf1 pour la liaison avec eRF3 et son effet sur le NMD (Kervestin et al., 2012; Meaux et al., 2008; Roque et al., 2015). Des données plus récentes ont par ailleurs fait évoluer le modèle. En effet, il semblerait que l'interaction entre les PABPs et eRF3 ne soit pas requise pour inhiber le NMD au contraire de l'interaction PABP/eIf4g. C'est un autre mécanisme qui est proposé, dans lequel la circularisation de l'ARNm par les PABPs et eIf4g à proximité du codon Stop favorise une terminaison de la traduction efficace sans activer le NMD (Fatscher et al., 2014; Joncourt et al., 2014). Dans le cas d'un PTC, le site de circularisation et le codon Stop sont éloignés dans l'espace ce qui entraînerait un délai dans la terminaison de la traduction permettant le recrutement de la machinerie de NMD. Chez la levure *S. cerevisiae*, une autre règle s'ajoute encore au « faux 3'-UTR », l'importance de la longueur de l'ORF (Decourty et al., 2014). Cette étude teste l'effet de la longueur de la séquence codante sur la stabilité d'un ARN possédant un 3'-UTR de longueur fixe. Elle suggère que la stabilité de l'ARN est indirectement corrélée à la longueur de l'ORF et que cette différence dépend du NMD.

Au contraire de la levure, les premières analyses d'ARNs instables ont identifié un autre facteur déterminant pour reconnaître les PTCs dans les cellules de mammifères. Les PTCs étaient suivis d'un complexe de jonction des exons (EJC) (Sun and Maquat, 2000; Sun et al., 2000; Thermann et al., 1998; Zhang et al., 1998, 1998) situé à plus de 50-55 nucléotides en aval (Nagy and Maquat, 1998) et qui induisait la dégradation rapide. L'absence du site

d'épissage à 50-55 nucléotides du PTC conduisait à la stabilisation du transcrit. Les codons Stop normaux étaient localisés dans le dernier exon ou à moins de 50 nucléotides d'un intron en amont. La conséquence de cette règle de position est que les PTCs situés dans le dernier exon échappent à la détection par le NMD (Hall and Thein, 1994) alors que ceux localisés dans les autres exons initient le NMD à la condition qu'ils ne soient pas trop proches des sites d'épissage (Cheng et al., 1994). Le positionnement spécifique de l'EJC dans le cadre d'un PTC s'explique par le fait que les EJs sont nettoyés des régions codantes par les ribosomes en traduction (Dostie and Dreyfuss, 2002) mais conservés sur les régions en aval du codon Stop, non traduites. Une distance minimale est nécessaire entre le Stop et l'EJC pour induire le NMD car l'encombrement du ribosome est important et qu'il pourrait décrocher les EJs plus proches que 50 nucléotides par gêne stérique. Ce modèle de reconnaissance des PTCs implique ensuite le recrutement des facteurs du NMD par l'EJC pour initier le mécanisme.

D'autres études chez les eucaryotes multicellulaires ont montré l'existence d'un mécanisme de détection des PTCs indépendants de l'EJC, similaire au modèle de Faux 3'-UTR (Bühler et al., 2006; Eberle et al., 2008; Singh et al., 2008). Les ARNs substrats sont alors des ARNs sans EJC en aval des PTCs mais avec des 3'-UTR allongés. Des ARNs rapporteurs ayant insérés de manière synthétique une séquence longue (600-1700nt) dépourvue d'EJC après le Stop voient leur niveau diminué dû à l'accélération de leur dégradation par le NMD (Huang et al., 2011). Ce type de substrat est aussi présent de manière naturelle dans les cellules mammifères (Yepiskoposyan et al., 2011). Tous les types de substrats ARN décrits précédemment (**Figure 5D, E, F**) se retrouvent chez les mammifères, où on estime que 3 à 10% du transcriptome est affecté par le NMD (Mendell et al., 2004; Peccarelli and Kebaara, 2014; Ramani et al., 2009; Tani et al., 2012; Wittmann et al., 2006; Yepiskoposyan et al., 2011). Certains substrats de NMD combinent un long 3'-UTR et la présence d'EJC sur l'ARN, les deux mécanismes proposés ne sont pas exclusifs.

4. Facteurs associés au NMD

La reconnaissance des PTCs entraîne le recrutement au niveau des ARNs de facteur du NMD qui conduisent à l'inhibition de la traduction et à la dégradation des transcrits (Muhlrad and Parker, 1999b). Les premiers facteurs du NMD ont été identifiés chez la levure *S. cerevisiae* dans des cribles génétiques cherchant des conditions dans lesquels des codons Stop pouvaient être traversés par les ribosomes à une plus grande fréquence (Culbertson et al., 1980). Les mutations *upf* (*up-frameshift*) identifiées ont été caractérisées plus tard et les gènes

correspondant ont été nommés *UPF1*, *UPF2* et *UPF3* (Cui et al., 1995; Leeds et al., 1991, 1992). Six autres facteurs critiques pour le NMD ont été identifiés chez le ver *Caenorhabditis elegans* (*C. elegans*) dans des cribles de suppression de mutations variées (Hodgkin et al., 1989). Les mutations suppressives *smg* (*suppressors with morphogenic effect on genitalia*) entraînaient par ailleurs des malformations des appareils génitaux d'où leur nom. Ces mutations ont ensuite été caractérisées et les gènes correspondant ont été nommés *SMG1* à *7* (Cali et al., 1999; Hodgkin et al., 1989; Pulak and Anderson, 1993). *UPF1*, *UPF2* et *UPF3* sont les seuls facteurs conservés chez tous les eucaryotes y compris certains groupes très anciens ; ils sont les facteurs cœur du NMD (Behm-Ansmant et al., 2007b; Chen et al., 2008; Culbertson and Leeds, 2003; Kadlec et al., 2006). Parmi les gènes *SMGs*, *SMG2-3-4* sont respectivement les homologues d'*UPF1-2-3* (Aronoff et al., 2001; Page et al., 1999). Les autres *SMGs* n'ont pas d'équivalent évident chez la levure *S. cerevisiae*. Seul *SMG7* pourrait avoir un orthologue nommé *EBS1* (Luke et al., 2007). Chez les mammifères, les facteurs impliqués dans le NMD ont été retrouvés par analyse *in silico*, par homologie de séquence avec les facteurs *UPFs* et/ou *SMGs*. Tous les *SMGs* ont des orthologues chez les mammifères (Applequist et al., 1997; Chiu et al., 2003; Denning et al., 2001; Lykke-Andersen et al., 2000; Mendell et al., 2000; Perlick et al., 1996; Yamashita et al., 2001). En outre, beaucoup d'autres gènes ont été identifiés comme ayant un rôle dans le NMD, *SMG8*, *SMG9*, *PNRC2*, *MOV10*, *NMD4*, *DBP2*, *DHX34*, *RUVBL1/2*, *GNL2*, etc. (Casadio et al., 2015; Cho et al., 2009; Gregersen et al., 2014; He and Jacobson, 1995; Hug and Cáceres, 2014; Izumi et al., 2010; Longman et al., 2013; Yamashita et al., 2009). Les rôles de certains de ces facteurs dans le NMD seront détaillés dans les paragraphes suivants, pour les autres leurs fonctions dans le NMD sont encore très peu claires.

5. Upf1, facteur majeur du NMD

Upf1 est un facteur très conservé parmi les eucaryotes (**Figure 6**), l'homologie des séquences protéiques entre des espèces très différentes (*Homo sapiens*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *S. cerevisiae*, *C. elegans*) est de 40 à 62% contre 59 à 67% pour les protéines ribosomales (Culbertson and Leeds, 2003). Sa délétion entraîne chez tous les organismes une perte complète du NMD, par ailleurs elle est associée à une létalité embryonnaire chez la souris et le poisson zèbre *Danio rerio* (Medghalchi et al., 2001; Wittkopp et al., 2009). Chez les levures, en revanche, les souches mutante *upf1Δ* ont une croissance proche des souches sauvages (Cui et al., 1995; Leeds et al., 1991). La protéine

Upf1 est impliquée dans toutes les étapes du mécanisme, de la reconnaissance des cibles à la dégradation des transcrits. Pour ces raisons, il est considéré comme le facteur majeur du NMD.

Upf1 peut interagir avec de nombreux ARNs (Hwang et al., 2010b). Cependant, on observe une plus grande densité d'Upf1 lié sur les ARNs substrats de NMD, sa fixation est favorisée par l'existence d'un long 3'-UTR ou par la présence d'un complexe EJC en aval du PTC (Hogg and Goff, 2010; Kurosaki and Maquat, 2013). Par ailleurs, la traduction a un effet positif sur la liaison d'Upf1 uniquement sur les ARNs ayant un PTC (Kurosaki and Maquat, 2013). Plusieurs hypothèses existent pour le recrutement d'Upf1 sur les ARNs et son accumulation après un PTC. Des expériences d'immuno-précipitation d'ARN par Upf1 humain (Zünd et al., 2013) ont montré une redistribution d'Upf1 sur toute la longueur des ARNs après inhibition de la traduction. Ces données ont contribué à proposer un modèle dans lequel Upf1 se fixe sur tous les ARNs en première instance puis est enlevé des séquences codantes après la traduction pour ne rester associé qu'au niveau des 3'-UTR des ARNs. Malgré des arguments convaincants, d'autres données plaident pour un recrutement d'Upf1 directement au niveau des PTCs ou des 3'-UTR des ARNs traduits (Shigeoka et al., 2012) à l'image de MOV10, hélicase de la famille SF1, proche en séquence d'Upf1 qui est recruté au niveau des codons Stop précoces (Gregersen et al., 2014).

Le recrutement d'Upf1 au niveau des ARNs n'est pas suffisant pour initier un événement de NMD, un deuxième signal est nécessaire. Ce signal pourrait être apporté par des facteurs autres qui activeraient Upf1. En effet, Upf1 seul ne semble pas avoir d'effet sur la terminaison de la traduction *in vitro* (Neu-Yilik et al., 2017). De plus, la délétion des partenaires *UPF2* et *UPF3* a le même effet sur le NMD que la délétion de *UPF1*. En outre, il a été montré que l'activité ATPase d'Upf1 était importante pour différencier les ARNs avec PTC, des ARNs sans (Lee et al., 2015).

D'un point de vue structural, Upf1 contient deux principales régions conservées (**Figure 6A et B**). La première région conservée se situe du côté N-terminale de la protéine, c'est un domaine en doigt de zinc riche en cystéine et histidine (domaine CH). Des interactions avec Upf2 principalement, eRF1, eRF3 et Dcp2 de levure ont été décrites pour cette région (Czaplinski et al., 1998; He and Jacobson, 1995, 2015; He et al., 1996; Ivanov et al., 2008). La deuxième région conservée est un domaine ATPase/hélicase en position C-terminale de la protéine (domaine HD : *Helicase Domain*). Ce domaine hélicase appartient à la superfamille d'hélicase ARN et ADN SF1 et contient deux sous-parties RecA, RecA1 et RecA2 ainsi

qu'un site de liaison à l'ARN. *In vitro*, Upf1 peut se transloquer de 5' vers 3' sur des substrats ADNs et ARNs (Fiorini et al., 2012), elle appartient à la sous-famille des hélicase SF1B. Ces hélicases possèdent aussi deux protrusions caractéristiques formées par les régions 1B et 1C. Upf1 a une constante d'affinité pour son substrat très élevée, elle est assez lente mais très processive. Cette dernière caractéristique est particulière pour Upf1 et n'est pas partagée par toutes les hélicases de la famille (Kanaan et al., in revision). Les deux activités ATPase et hélicase sont requises pour le NMD (Weng et al., 1996a). Les structures en cristallographie d'Upf1 de levure et Upf1 humain suggèrent une organisation conservée entre les espèces (Chakrabarti et al., 2011). Le domaine CH est replié entre le domaine 1B et RecA2 de l'hélicase (**Figure 6C**). Les deux régions RecA forment deux modules se faisant face et liant la molécule d'ARN. Le domaine 1C est en surface de la protéine.

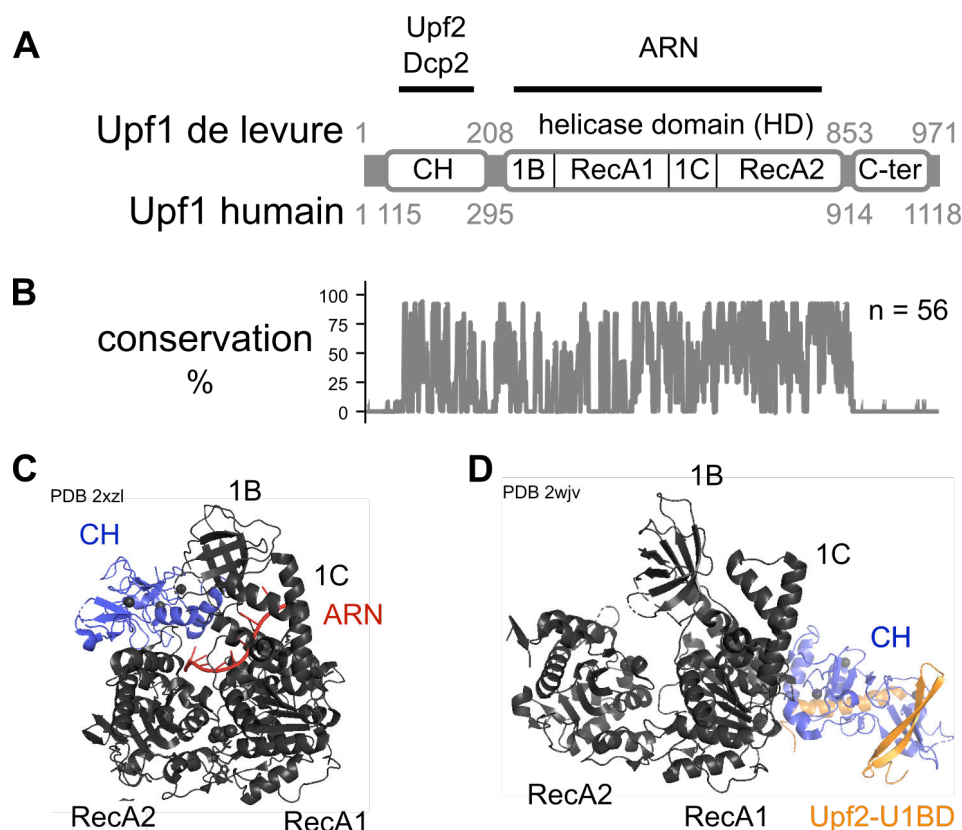


Figure 6 : Upf1 ; architecture, interactions, conservation et structures

A. Schéma de l'organisation en domaines d'Upf1 de la levure *S. cerevisiae*. Le domaine CH, riche en cystéine et histidine, interagit avec Upf2 et Dcp2. Le dessin n'est pas à l'échelle. **B.** Graphique du pourcentage de conservation des résidus d'Upf1. L'alignement a été réalisé à partir de 56 séquences différentes d'Upf1 appartenant à des phylums variés. Nous avons utilisé l'algorithme Mafft (via Jalview) pour calculer la similarité des résidus dans les séquences. **C.** Structure d'Upf1 de *S. cerevisiae* (54-851) couplée à un polymère de huit ribonucléotides U (PDB 2xz1, Chakrabarti et al.,

2011) Le domaine CH est coloré en bleu, l'ARN en rouge. Les billes noires localisées au niveau du CH sont des atomes de zinc. **D.** Structure du fragment d'Upf1 humain (115-914) couplé au fragment 1105-1198 (domaine U1BD) d'Upf2 humain (PDB 2wjv, Clerici et al., 2009). Le domaine CH d'Upf1 est représenté en bleu, le fragment d'Upf2 est en orange. La position des structures **C** et **D** a été alignée en fonctions des régions RecA1 et RecA2 en utilisant le logiciel PyMOL.

Les activités ATPase et hélicase d'Upf1 sont régulées par plusieurs facteurs. Des tests de l'activité en cellule unique ont notamment montré l'effet auto-inhibiteur des domaines CH et C-ter (Fiorini et al., 2013, 2015). L'inhibition par le domaine CH est levée en présence d'Upf2 (Fiorini et al., 2012, 2015). Des données de structure en présence d'Upf2 (**Figure 6C et D**) montrent d'ailleurs l'existence d'un changement de conformation d'Upf1 qui pourrait expliquer l'effet sur le domaine hélicase (Chakrabarti et al., 2011; Chamieh et al., 2008). Les facteurs de terminaison de la traduction eRF1 et eRF3 semblent aussi inhiber l'activité d'Upf1, au moins *in vitro* (Czaplinski et al., 1998). Upf1 possède aussi plusieurs sites et répétitions de motifs qui sont phosphorylés et qui participent à la régulation de l'activité hélicase d'Upf1 (Lasalde et al., 2013) ou à la liaison de différents cofacteurs. Deux autres régions du domaine HD, les protubérances 1B et 1C peuvent modifier la conformation d'Upf1 et influencer la liaison à l'ARN et les activités ATPase et hélicase (Chakrabarti et al., 2011; Cheng et al., 2007; Kanaan et al., in revision). Par exemple, une boucle non structurée au niveau du domaine 1B est localisée au centre de la structure et est impliquée dans la régulation de l'activité hélicase d'Upf1 (Gowravaram et al., 2018).

6. Upf2, plateforme pour des interactions protéiques

En tant que facteur cœur du NMD, Upf2 a d'autres rôles que la régulation de l'activité hélicase d'Upf1. La protéine contient trois domaines similaires au domaine MIF4G (*Middle domain of eukaryotic Initiation Factor 4G*) qui permet pour le facteur eIF4g de lier les autres facteurs d'initiation de la traduction et l'ARN. Le troisième motif MIF4G est le site de liaison d'Upf3. Une interaction avec la protéine Smg1 a aussi été décrite pour ce domaine chez l'Homme, les deux protéines pourraient interagir en même temps, sans compétition (Clerici et al., 2013). Upf2 possède aussi un domaine en position C-terminale permettant l'interaction avec le domaine CH d'Upf1, U1BD (**Figure 7A**). Upf2 relie Upf1 à Upf3 (Chamieh et al., 2008; He et al., 1997). Par ailleurs, des mutations dans le premier domaine MIF4G abolissent complètement le NMD (Fourati et al., 2014). Cela montre l'importance de ce domaine pour la

fonction d'Upf2. La région concernée est positionnée en surface de la protéine Upf2 et pourrait donc servir de plateforme pour l'interaction avec d'autres cofacteurs par exemple Dbp6 (Fourati et al., 2014), impliqué dans la biogenèse de la sous-unité 60S du ribosome chez *S. cerevisiae* (Kressler et al., 1998). Une équipe a aussi décrit une interaction directe d'Upf2 avec les facteurs de terminaison de la traduction (López-Perrote et al., 2016), comme Upf1, qui semble indiquer que Upf2 a un rôle dans les étapes précoces du NMD. Upf2 est une phospho-protéine (Chiu et al., 2003). Chez la levure *S. cerevisiae* la phosphorylation de plusieurs résidues de la région N-terminale semble importante pour le NMD (Wang et al., 2006). Cette région pourrait permettre l'interaction avec certaines protéines liant l'ARN comme Hrp1 (González et al., 2000; Wang et al., 2006) qui a un rôle dans la dégradation de certains ARNs rapporteurs (PGK1, PPR1) du NMD.

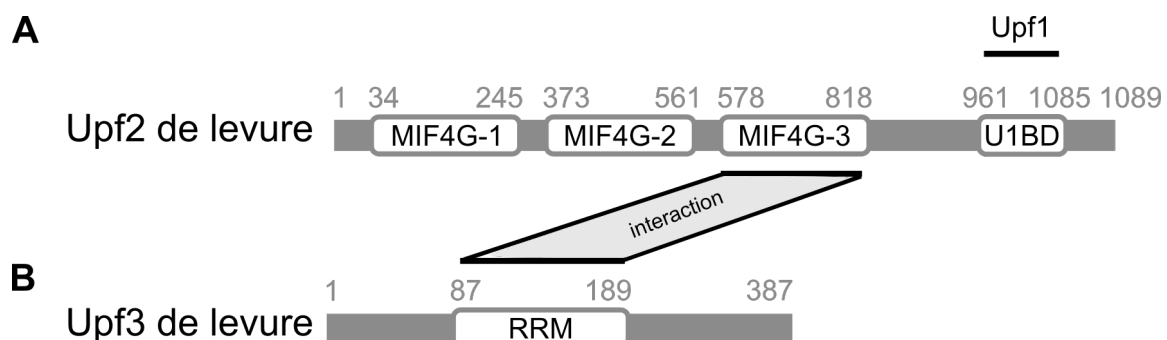


Figure 7 : Upf2 – Upf3

Organisation des domaines des protéines Upf2 (A.) et Upf3 (B.) de levure et leurs interactions.

7. Upf3

Upf3 est une protéine moins conservée entre les espèces par rapport à Upf1 et Upf2. Elle contient un domaine similaire à un motif de reconnaissance de l'ARN (RRM : *RNA Recognition Motif*) qui est le site d'interaction avec Upf2. Chez l'Homme, il existe deux paralogues, Upf3a et Upf3b, provenant d'un événement de duplication du génome. Upf3b est la principale copie utilisée dans le NMD (Chan et al., 2007). Upf3a est un inhibiteur du NMD (Shum et al., 2016), un des domaines clés de la protéine, permettant l'interaction avec l'EJC est modifié et entraîne la séquestration d'Upf2 loin des cibles ARNs. La régulation du NMD par Upf3a est indispensable lors de la spermatogenèse.

Dans le modèle NMD dépendant de l'EJC, chez les eucaryotes multicellulaires, Upf3 (Upf3b chez l'Homme) est considéré comme un facteur presque à part entière de l'EJC, il interagit

via un domaine de liaison à l'EJC (EBM : *EJC Binding Motif*) à la sous-unité Y14. Upf3 fait la navette entre le noyau, où il se lie à l'EJC, et le cytoplasme. Sur les ARNs avec PTCs, Upf3 permet de faire le pont entre Upf1, situé au PTC, et l'EJC situé en aval. Dans le cas des modèles de NMD indépendants de l'EJC, son rôle est moins clair. Des études chez la levure *S. cerevisiae* ont montré que Upf2 perdait sa liaison aux polysomes (ARN en traduction) avec la délétion d'Upf3 (Atkin et al., 1997) ce qui suggère que Upf3 même en l'absence d'EJC permet de lier Upf2 aux ARNs. Des données récentes confirment cette hypothèse et proposent que Upf3, en liant directement le ribosome, ralentit la terminaison de la traduction et fournit une plateforme pour lier Upf2 et Upf1 (Neu-Yilik et al., 2017) afin d'initier le NMD.

8. Le cycle de phosphorylation d'Upf1

Upf1 (Smg2) a été décrite pour la première fois comme une phospho-protéine chez le nématode *C. elegans* (Page et al., 1999). Les sites de modifications sont localisés au niveau des extrémités C-terminales et N-terminales peu conservées (**Figure 6B**) de la protéine ainsi qu'en position médiale. L'état de phosphorylation d'Upf1 est modulé par les autres protéines Smgs ; Smg1, Upf2 (Smg3) et Upf3 (Smg4) augmentent la phosphorylation tandis que Smg5, Smg6 et Smg7 la diminue. Smg1 est une sérine/thréonine kinase de la famille des *phosphatidyl Inositol (PI) 3-kinase-related kinase* (PIKK). Cette protéine est organisée autour de deux régions (**Figure 8A et B**, Arias-Palomo et al., 2011); une première région composée de répétition HEAT (Huntington, EF3A, ATM, TOR ; Andrade and Bork, 1995) ; et une deuxième région contenant, dans l'ordre, un domaine FAT (FRAP, ATM, and TRRAP), une région catalytique homologue à la PI-3-kinase, une insertion de fonction peu claire et une extrémité C-terminale FATC (FAT C-terminal). La région HEAT forme une protrusion mobile dans la protéine Smg1, surnommée « Arm » (bras), c'est une interface pour l'interaction avec d'autres protéines (**Figure 8B**). Le reste de Smg1 forme une structure globulaire appelée « Head » (tête), cette partie contient la région catalytique de la protéine (**Figure 8B**).

Smg1 a été caractérisé comme la kinase responsable de la phosphorylation d'Upf1 (Denning et al., 2001; Grimson et al., 2004; Page et al., 1999; Yamashita et al., 2001) au niveau des résidus sérines et thréonines des motifs S/T-Q (Sérine/Thréonine suivi d'une glutamine (Q)) en N- et C-terminale d'Upf1.

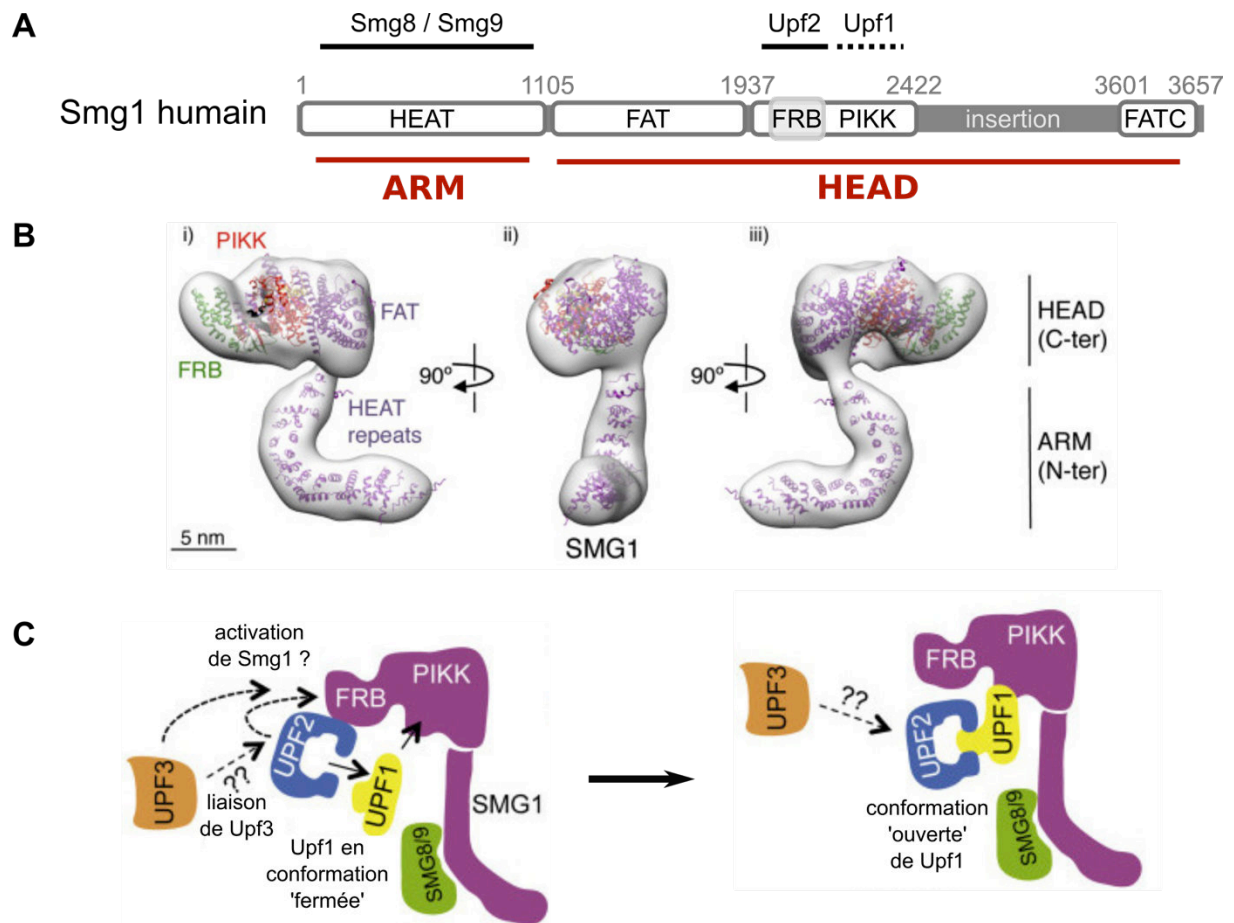


Figure 8 : Architecture de la protéine Smg1 et ses interactions

A. Organisation de la protéine Smg1 humaine. L'insertion en position C-terminale correspond à une région de fonction peu claire. **B.** Structure en cryo-microscopie électronique de la protéine Smg1 humaine assemblée avec les structures en cristallographie des répétitions HEAT de la protéine kinase DNA-PKcs (PDB 3kgv) et de la région catalytique de mTOR (PDB 4jsp). **C.** Modèle pour l'interaction de Upf1, Upf2 et Upf3 au niveau de Smg1 permettant l'activation de la phosphorylation. (Adapté de Melero et al., 2014)

La phosphorylation d'Upf1 par Smg1 est probablement régulée. Smg8 et Smg9, deux autres facteurs de la famille des Smgs ont été décrits comme régulateurs de son activité (Arias-Palomo et al., 2011; Deniaud et al., 2015; Yamashita et al., 2009). Ces deux protéines collaborent pour recruter Smg1 au niveau de l'ARN et inhiber son activité. En effet, la liaison de Smg8/9 au niveau de la région N-terminale HEAT de Smg1 entraîne une modification de sa conformation qui pourrait moduler son activité kinase (Arias-Palomo et al., 2011). Une fois Smg1 recruté au niveau des substrats de NMD, la phosphorylation est activée par l'interaction avec Upf2 (Kashima et al., 2006). Un modèle structural de cette activation (**Figure 8C**)

propose que Smg1 interagisse avec Upf2 dans un premier temps au niveau du domaine FRB (Melero et al., 2014). Puis, dans un deuxième temps Smg1 interagirait avec Upf1 au niveau du domaine PIKK. La proximité des régions d'interaction présumées d'Upf1 et Upf2 pourrait induire une gêne stérique importante résolue par le transfert d'Upf2 au niveau du CH d'Upf1. Cette dernière étape aurait pour conséquence d'activer la phosphorylation d'Upf1 et l'activité hélicase. D'autres études ont montré que l'activité hélicase d'Upf1 est aussi importante pour la phosphorylation (Page et al., 1999). Chez le ver par exemple, des mutations du domaine hélicase inhibent la phosphorylation. Tous ces résultats montrent que la phosphorylation d'Upf1 chez les eucaryotes multicellulaires et l'activation de l'hélicase sont des événements très liés et interconnectés. Des expériences de compétition *in vitro* pour l'interaction d'Upf1 et Upf2 avec Smg1 montrent d'ailleurs une perte d'Upf1 suite à l'augmentation de la concentration en Upf2 (Melero et al., 2014) ce qui semble indiquer que la liaison d'Upf2 diminue l'affinité entre Upf1 et Smg1. Cela pourrait être une autre conséquence de l'activation de l'hélicase d'Upf1.

Upf1 phosphorylé agit ensuite comme une plateforme de recrutement pour trois autres protéines Smgs : Smg5, Smg6 et Smg7 (Chakrabarti et al., 2014; Fukuhara et al., 2005; Ohnishi et al., 2003; Okada-Katsuhata et al., 2012). Ces protéines possèdent des domaines similaires au domaine 14-3-3 (**Figure 9A**) qui ont de l'affinité pour les motifs S/T-Q phosphorylés (Fukuhara et al., 2005; Gardino et al., 2006). Smg7 et Smg5 forment un hétérodimère qui est efficacement recruté au niveau d'Upf1 (Jonas et al., 2013). Smg7 a une forte affinité pour Upf1 phosphorylé. Smg5 n'a pas cette capacité (Okada-Katsuhata et al., 2012) mais permet d'augmenter la liaison et la spécificité de l'interaction Upf1/Smg7 (Chakrabarti et al., 2014; Jonas et al., 2013) au niveau de l'extrémité C-terminale d'Upf1 majoritairement et plus légèrement au niveau de l'extension N-terminale (**Figure 9B**, Ohnishi et al., 2003; Okada-Katsuhata et al., 2012). Smg6, lui n'a pas été décrit comme un dimère, son domaine 14-3-3 est similaire à celui de Smg5 et peut réaliser des interactions faibles avec les motifs S/T-Q phosphorylés d'Upf1 (Okada-Katsuhata et al., 2012). L'interaction dominante entre Smg6 et Upf1 est médiée par une région de faible complexité de Smg6, la région USR qui interagit avec la région hélicase d'Upf1, cette interaction ne dépend pas de la phosphorylation (**Figure 9A**, Chakrabarti et al., 2014; Okada-Katsuhata et al., 2012).

Les protéines Smg5 et Smg7 ont ensuite un rôle dans la déphosphorylation d'Upf1, elles permettraient de recruter la phosphatase PP2A (**Figure 9B**, Anders et al., 2003; Chiu et al., 2003; Ohnishi et al., 2003; Okada-Katsuhata et al., 2012).

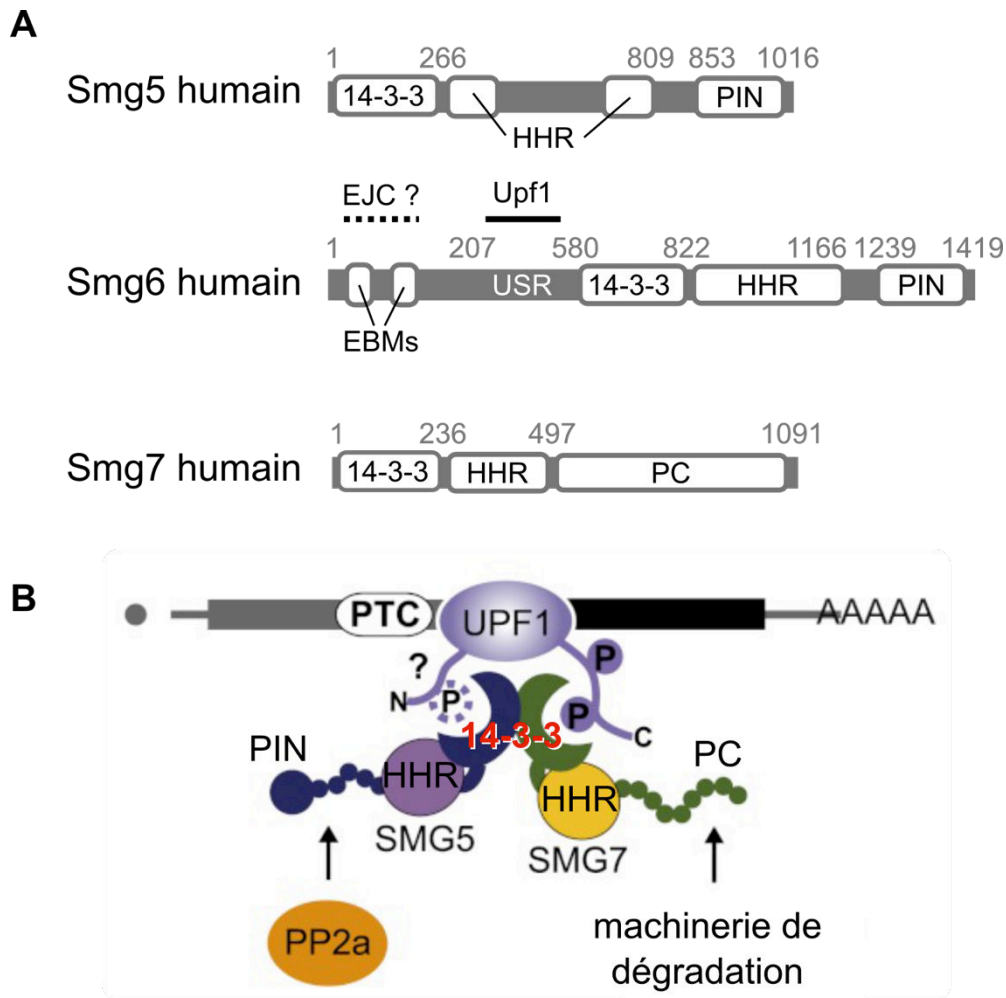


Figure 9 : Rôle de Smg5-7 dans le cycle de phosphorylation de Upf1

A. Organisation des protéines Smg5, 6 et 7 humaines. Les domaines 14-3-3 sont des 14-3-3-like, HHR est pour *Helical Hairpin Repeat*, PC est pour *Proline Rich domain*, USR est pour *UnStructured Region*. Le point d’interrogation à côté de l’EJC pour la liaison aux EBMs (*EJC Binding Motifs*) de Smg6 signifie que cette interaction n’est pas clairement élucidée. **B.** Modèle issu de Jonas et al., 2013 de l’interaction de l’hétérodimère Smg5-7 au niveau d’Upf1. Le point d’interrogation signifie que l’interaction entre le domaine 14-3-3 de Smg5 et l’extension N-terminale d’Upf1 humain est mal caractérisée et incertaine.

La présence d’Upf1 hypo- ou hyper-phosphorylé est corrélée à un défaut du NMD ce qui suggère une importance pour ce cycle de phosphorylation. En effet, le recrutement de Smg5-6-7 par Upf1 phosphorylé permet d’initier la dégradation des transcrits et la déphosphorylation permet ensuite de recycler Upf1 pour un nouveau cycle de NMD (Durand et al., 2007). Cependant, l’importance de ces phosphorylations dépend des espèces. La délétion de *SMG1* ou la mutation des résidus phosphorylés d’Upf1, abolit complètement le

NMD dans les cellules de mammifères ou le ver *C. elegans*. Au contraire, chez la drosophile, *SMG1* n'est pas requis pour le NMD (Chen et al., 2005). Chez la levure *S. cerevisiae*, les régions N- et C-terminales sont très peu conservées et ont perdu la majorité des sites de phosphorylation (Causier et al., 2017; Lasalde et al., 2013), de plus il n'y a pas d'équivalent connu de *SMG1* chez cette espèce.

9. Le recrutement de Smg5, Smg6 et Smg7 déclenche la dégradation des ARNs

Des études réalisées en 2004 montrent que l'ancrage artificiel de Smg7 ou de sa région C-terminal sur un ARN active sa dégradation par un mécanisme dépendant de Dcp2 et Xrn1 (Unterholzner and Izaurralde, 2004). Cela suggère donc que Smg7 a un rôle dans l'accélération de la dégradation par la voie 5' vers 3'. Plus récemment, une interaction entre la région C-terminale de Smg7 et la protéine Pop2 du complexe de dé-adénylation Ccr4/Pop2/Not a été décrite (Loh et al., 2013). Cette nouvelle donnée suggère donc que Smg7 active la voie de dégradation par hydrolyse de la coiffe dé-adénylation dépendante après son recrutement par Upf1 phosphorylé (Loh et al., 2013). Smg5 en activant l'interaction de Smg7 avec Upf1 est aussi crucial pour la dégradation efficace des ARNs (Chakrabarti et al., 2014; Jonas et al., 2013).

La dégradation des ARNs sensibles au NMD fait aussi intervenir le recrutement direct ou indirect du complexe d'hydrolyse de la coiffe (Dcp1/Dcp2). Le recrutement direct aurait lieu au niveau du domaine CH d'Upf1 et a été décrit chez la levure *S. cerevisiae* (Ford et al., 2006; He and Jacobson, 1995, 2015; Swisher and Parker, 2011; Tarassov et al., 2008) et chez les eucaryotes multicellulaires (Fenger-Grøn et al., 2005; Lykke-Andersen, 2002). Le recrutement indirect se ferait par l'intermédiaire de Pnrc2, une protéine interagissant avec Dcp1 et Smg5 (Cho et al., 2009, 2013), préférentiellement au niveau des Upf1 hyper-phosphorylés (Lai et al., 2012). Cependant, le rôle de Pnrc2 dans le NMD a été récemment remis en question. Pnrc2, Upf1 et Dcp1 pourraient être impliqués dans un mécanisme de GMD (*Glucocorticoid receptor-Mediated mRNA Decay*), distinct du NMD, et concernant certaines cibles spécifiques (Cho et al., 2015). Une autre étude (Nicholson et al., 2018) dément l'existence d'une interaction entre Pnrc2 et Smg5. L'équipe propose un rôle général de Pnrc2 dans le décapping mais pas de rôle particulier dans le NMD bien qu'une interaction Pnrc2/Upf1 puisse exister.

Ce type de recrutement des enzymes de décapping favorisé ou non par la phosphorylation d'Upf1 initie une dégradation des ARNs de l'extrémité 5' vers 3' indépendante de la dé-

adénylation (Hatfield et al., 1996; Muhlrاد and Parker, 1994). Cette voie est différente de celle dé-adénylation-dépendante. D'ailleurs certaines mutations de Dcp1 ont un impact fort sur la dégradation d'un ARN rapporteur non-substrat du NMD alors qu'elles n'ont que peu ou pas d'effet sur la voie indépendante de la dé-adénylation impliquée dans le NMD de levure (Tharun and Parker, 1999). Si cette voie de dégradation est favorisée chez la levure *S. cerevisiae*, chez les eucaryotes multicellulaires, elle semble moins présente (Lykke-Andersen et al., 2014a).

Des études, initialement chez la drosophile, ont mis en évidence un troisième mécanisme de dégradation des substrats du NMD par coupure endonucléolytique au niveau des PTCs. Cette coupure est ensuite suivie de la dégradation de l'ARN par l'exosome cytoplasmique pour les fragments situés en amont de la coupure avec une extrémité 3' non protégée et par Xrn1 pour les fragments situés en aval ayant une extrémité 5'-P (Eberle et al., 2009; Gatfield and Izaurralde, 2004; Huntzinger et al., 2008). Smg6 est l'endonucléase responsable de cette dégradation. Cette protéine contient un domaine PIN (*PilT N-terminus*) actif qui a une activité d'endonucléase (Clissold and Ponting, 2000; Glavan et al., 2006). Elle possède, contrairement à Smg5, tous les acides-aminés formant la poche catalytique (Matelska et al., 2017). La mutation de ces sites abolit complètement la dégradation des substrats de NMD par coupure endonucléolytique (Eberle et al., 2009; Glavan et al., 2006; Huntzinger et al., 2008; Kashima et al., 2010; Nicholson et al., 2014). Le recrutement de Smg6 sur Upf1 n'est pas claire, en effet des études montrent un mécanisme de recrutement dépendant de l'interaction entre Upf1 phosphorylé et le domaine 14-3-3-like quand d'autres identifient une nouvelle région de Smg6 (USR, **Figure 9A**) impliquée dans une interaction forte avec le domaine hélicase d'Upf1 indépendante de la phosphorylation. Par ailleurs, des études récentes ont aussi identifié des motifs EBMs sur Smg6 similaires à ceux d'Upf3 qui sont cruciaux pour le NMD (Kashima et al., 2010) et qui permettraient une interaction directe avec l'EJC. Cependant, cette dernière interaction n'est pas strictement requise pour le recrutement de Smg6 et la coupure endonucléolytique des ARNs (Boehm et al., 2014).

La coupure par Smg6 est une voie majeure de dégradation des ARNs sensibles au NMD chez les eucaryotes multicellulaires (Lykke-Andersen et al., 2014a). Elle est activée par la présence d'un EJC et par les longs 3'-UTR. Il existe une redondance de fonction entre la voie Smg6 et la voie Smg7 qui ont les mêmes cibles (Colombo et al., 2016) ce qui permettrait une meilleure efficacité du NMD. Cependant, leur efficacité diffère en fonction des classes de substrat du NMD (Ottens et al., 2017), ce qui laisse penser que les deux mécanismes ont des cibles

préférentielles mais qu'ils sont aussi capables de dégrader les autres types de substrats en cas de besoin.

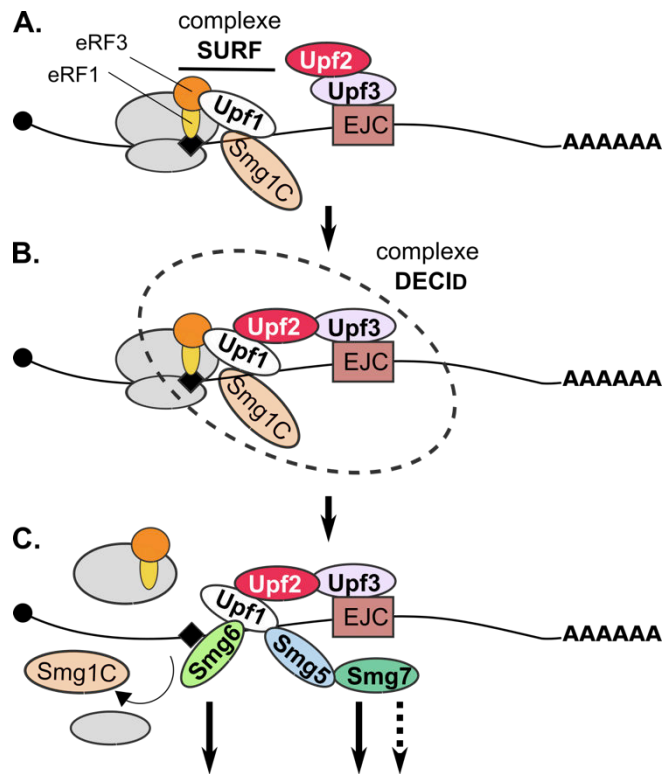
En parallèle du recrutement direct ou indirect de protéines effectrices de la dégradation, Upf1 phosphorylée interagit avec le facteur d'initiation de la traduction eIF3 (Isken et al., 2008). Cette interaction pourrait inhiber l'initiation de la traduction et alors promouvoir l'hydrolyse de la coiffe et la dégradation par la voie 5' vers 3'.

10. Modèle EJC-dépendant du NMD : SURF-DECID

Chez les eucaryotes multicellulaires (mammifères, nématodes, etc.), les données présentées précédemment ont abouti à la construction d'un modèle du NMD appelé SURF-DECID (**Figure 10**, Kashima et al., 2006). Dans ce modèle, Upf1 interagit avec les facteurs de terminaison de la traduction eRF1 et eRF3 et forme avec Smg1, le complexe SURF (Smg1-Upf1-eRF1 & 3). Smg8 et Smg9 sont des cofacteurs de Smg1 qui maintiennent son activité inhibée (Yamashita et al., 2009). À ce stade, Upf1 n'est pas phosphorylé. Parallèlement à la formation du SURF, Upf2 est recruté au niveau des ARNs substrats du NMD par l'intermédiaire d'Upf3. Upf3 est déjà présent sur ces mêmes ARNs car il est lié à l'EJC en aval du PTC (voir introduction partie II.3. Activation du NMD). Upf2 interagit ensuite avec Upf1 et Smg1 et active la phosphorylation d'Upf1 (résumé dans Yamashita, 2013). La réunion du SURF et de l'EJC par Upf2 forme un nouveau complexe protéique DECID (*DECay InDucing*). La phosphorylation d'Upf1 permet la liaison de l'hétérodimère Smg5/7 qui recrute alors des protéines effectrices de la dégradation. Smg6 est aussi recruté au niveau d'Upf1 phosphorylé pour initier la dégradation des substrats de NMD par coupure endonucléolytique (Boehm et al., 2014; Lykke-Andersen et al., 2014a).

Figure 10 : Modèle SURF-DECID

A. Formation du complexe SURF. Smg1C est composé de Smg1 ainsi que Smg8-9 qui sont associés au niveau de la région “arm” de Smg1. **B.** Formation du complexe DECID via Upf2 qui fait le pont entre le SURF et Upf3 associé à l’EJC. **C.** Initiation de la dégradation des ARNs par trois voies possibles d’importance différente. Smg6 induit la coupure endonucléolytique. Smg5/7 permet le recrutement de la machinerie de dé-adénylation ou des protéines Dcp1 et Dcp2 pour le décapping.



11. Modèle EJC-indépendant du NMD

Le mécanisme du NMD pour les substrats ne faisant pas intervenir l’EJC n’est pas clairement établi. Pour l’instant, il consiste en la reconnaissance des PTCs en suivant le modèle du « Faux 3’-UTR » décrit précédemment qui aboutit à la dégradation des cibles par la voie de l’hydrolyse de la coiffe sans dé-adénylation préalable chez la levure ou en suivant les voies Smg-dépendantes décrites dans le modèle SURF-DECID pour les eucaryotes multicellulaires.

La voie EJC-indépendante du NMD est peu étudiée par rapport à celle dépendante de l’EJC, pourtant elle semble être un mécanisme de base conservé chez tous les eucaryotes. En effet, les protéines impliquées sont toutes conservées, au contraire de l’EJC. De plus, même dans les espèces possédant un EJC-like, ce complexe n’est pas toujours impliqué dans la reconnaissance des cibles du NMD. C’est le cas du cilié *Tetrahymena* (Tian et al., 2017), de la paramécie (Contreras et al., 2014), de la levure *S. pombe* (Wen and Brogna, 2010) ou du ver *C. elegans* (Gatfield et al., 2003) alors que de multiples cibles ont des jonctions exon-exon en aval des PTCs. Par ailleurs, plusieurs cibles physiologiques du NMD ne dépendent pas de la

présence d'un EJC en aval du PTC chez les mammifères. Un exemple marquant est le transcrit GADD45A (Chapin et al., 2014; Lykke-Andersen et al., 2014b) qui est essentiel pour le développement des embryons de la mouche et humains (Nelson et al., 2016).

Le rôle principal de l'EJC dans le NMD serait le recrutement d'Upf3 puis d'Upf2 pour interagir ensuite avec Upf1 et l'activer. Cependant, certaines mutations de l'EJC empêchent l'interaction d'Upf3 sans perturber le NMD dans des cellules humaines (Gehring et al., 2005). Dans ces cas, il doit exister un autre mécanisme de recrutement d'Upf3 sur les substrats de NMD par exemple par un recrutement actif par les facteurs de terminaison de la traduction et/ou le ribosome comme proposé dans cette étude récente (Neu-Yilik et al., 2017). La dynamique de ce dernier mécanisme reste à élucider. Il semble donc que le modèle SURF-DECID soit incomplet car il donne des rôles centraux à des facteurs ou des étapes du NMD non conservés. L'EJC par exemple, comme vu précédemment, ou Smg1 qui est absent chez la levure *S. cerevisiae*, la plante *A. thaliana* et dont l'inactivation chez la drosophile entre autre n'a que très peu d'impact sur le NMD et le développement des embryons (Chen et al., 2005).

IV. Fonctions physiologiques et patho-physiologiques du NMD et de ses facteurs

La délétion des gènes *UPF1*, *UPF2*, *SMG1* et *SMG6* codant les facteurs du NMD est létale pour les mammifères à différents stades du développement. Chez la souris, la délétion complète d'*UPF1* ou *SMG6* empêche le développement au-delà des étapes de pré-implantation de l'embryon dans l'utérus (Li et al., 2015; Medghalchi et al., 2001). La délétion d'*UPF2* entraîne un décès de l'embryon après 3,5 à 7,5 jours (E3,5 – E7,5) de développement (Weischenfeldt et al., 2008). La délétion de *SMG1* a un effet plus modéré mais elle n'est plus viable après E8,5 et montre de sévères défauts du développement (McIlwain et al., 2010). Les cellules déplétées en *SMG1*, *UPF2* et *SMG6* montrent aussi un profil transcriptomique proche d'*upf1Δ* avec beaucoup de cibles connues du NMD stabilisées par rapport à une souche sauvage. Il semblerait alors que le NMD soit crucial à différents stades du développement. Notamment, il a été montré que la dégradation du transcrit GADD45 par le NMD était essentielle pour la viabilité des embryons de drosophile et de souris (Nelson et al., 2016). Par ailleurs, Upf1, Smg1 et Smg6 ont aussi des rôles spécifiques décrits dans d'autres voies métaboliques de la cellule, qui pourraient aussi expliquer les effets de ces protéines pendant le développement. Ainsi Upf1 serait impliqué dans la maintenance des télomères ou la réplication. Le laboratoire du Dr. Azzalin a montré une localisation mineure de Upf1 dans le

noyau et décrit des effets de la délétion de *UPF1* et de *SMG6* sur la réplication des télomères (Azzalin, 2012; Azzalin and Lingner, 2006; Lew et al., 1998) ainsi que des interactions entre Upf1, Smg1 ou Smg6 avec la télomérase (Azzalin et al., 2007; Chawla et al., 2011; Reichenbach et al., 2003). Les observations décrites dans ces articles demanderaient à être validées par d'autres laboratoires car les effets observés pourraient être indirects. Par exemple, il a été montré que la délétion d'*UPF1* affecte la quantité de la protéine Stn1 qui est impliquée dans la réplication des télomères (Addinall et al., 2011). Upf1 a aussi un rôle dans la dégradation des ARNm des histones (résumé dans Isken and Maquat, 2008) et dans un autre mécanisme de surveillance des ARNs, le *Staufen-Mediated mRNA Decay* (SMD, résumé dans Gong et al., 2013).

En plus de son rôle classique dans le contrôle qualité des ARNs aberrants, le NMD régule de multiples voies métaboliques cellulaires en affectant la stabilité d'ARNs très variés codant pour des protéines effectrices. Chez les mammifères, 3-10% des transcrits sont affectés par le NMD (Mendell et al., 2004; Schweingruber et al., 2013; Tani et al., 2012; Wittmann et al., 2006). La majorité est due à des variations dans l'épissage (AS-NMD, *Alternative Splicing-NMD*) qui introduit ou fait disparaître des PTCs en réponse à des signaux cellulaires spécifiques. Dans ce cas, le NMD est un outil de régulation dans les voies métaboliques. Par exemple, les protéines Ptp1 et Ptp2 (*Polypyrimidine-Tract-Binding Proteins 1 and 2*) sont des régulateurs d'épissage (Keppetipola et al., 2012). Ptp1 s'autorégule *via* le NMD en modifiant l'épissage alternatif de son propre ARNm (Spellman et al., 2005; Wollerton et al., 2004). Par ailleurs, dans les progéniteurs neuronaux, on observe une régulation particulière de PTBP2 par Ptp1. Ptp1, fortement exprimé dans ce type cellulaire agit au niveau de l'épissage de son homologue ARN PTBP2 en favorisant l'exclusion de l'exon 10. Cet épissage particulier induit un changement de phase au niveau de l'ARN et introduit un PTC. L'ARN PTBP2 est alors dégradé dans le cytoplasme par le NMD. Après la différenciation des progéniteurs en neurones, il y a une diminution du niveau de PTBP1 via un ARN non codant, miR-124, ciblant spécifiquement cet ARN. En conséquence, l'ARN PTBP2 est correctement épissé et il est stable dans le cytoplasme. La balance entre PTBP1 et 2 est importante dans la différenciation neuronale et fait intervenir le NMD (Boutz et al., 2007; Makeyev et al., 2007). D'autres voies font intervenir le NMD, comme la guidance des axones (Colak et al., 2013), la réponse aux stress (Karam et al., 2015), etc. De plus, il y a des variations de l'efficacité du NMD en fonction des cellules ou des tissus (Bruno et al., 2011; Linde et al., 2007; Schoenberg and Maquat, 2012). Par exemple l'efficacité du NMD diminue au cours de la différenciation des myoblastes en myotubes.

Certaines cibles du NMD ont la particularité d'être des facteurs proto-oncogènes ou suppresseurs de tumeur. Dans ce cas, le NMD est impliqué dans les processus de cancérogenèse (Lindeboom et al., 2016). Par exemple, *CYR61* (*CY*stein *R*ich *61*) est induit en condition d'hypoxie et agit comme un facteur angiogénique pro-tumoral. Il existe sous deux isoformes dues à un mécanisme de rétention d'intron (Hirschfeld et al., 2009). L'intron inclus introduit probablement un PTC au niveau de l'ARN qui est alors dégradé par le NMD. Dans les cancers du sein, l'équipe de Hirschfeld a montré que la rétention de l'intron de *CYR61* est inhibée ce qui favorise l'expression unique de la forme complète et active de *Cyr61* pro-tumorale. D'autres cas de cancers montrent une inhibition directe du NMD, par exemple il a été montré que *Upf1* possédait de multiples mutations dans les tumeurs du carcinome adénosquameux du pancréas (Liu et al., 2014).

Les PTCs sont aussi impliqués dans les maladies génétiques, 30% des maladies génétiques rares seraient la conséquence de mutations non-sens ou de changements de phase de lecture au niveau des ARNm (Miller and Pearce, 2014). Les mutations non-sens en particulier sont responsables de plus de 20% des maladies associées à des mutations d'une seule base dans le génome (Mort et al., 2008). Ces mutations inactivent en général l'expression du gène concerné car le NMD reconnaît alors les PTCs et conduit à la dégradation des ARNm correspondants. Ces mutations sont impliquées dans des pathologies comme le rétinoblastome, la dystrophie de Duchenne ou certains cancers associés à *TP53* ou à *BRCA1* (Holbrook et al., 2004). Le NMD est un important modulateur de la sévérité de ces maladies. Dans certains cas, la dégradation de l'ARN aberrant empêche l'expression d'une protéine tronquée mais active qui serait suffisante pour rétablir un fonctionnement quasi-normal de la cellule (dystrophies). Dans d'autres cas, son inactivation entraîne l'expression de facteurs pro-tumoraux, par exemple une version dominant-négative de *p53* dans le carcinome adénosquameux du pancréas (Liu et al., 2014). Des stratégies thérapeutiques existent et combinent des drogues permettant la trans-lecture des PTCs et l'inhibition du NMD (Ataluren par exemple, Kerem et al., 2014). Pour l'instant, l'effet de ces molécules dans les essais cliniques sont discutés et les effets secondaires souvent trop importants (Aslam et al., 2017).

Un dernier point que je souhaite aborder par rapport au rôle physiologique du NMD est son implication dans la réponse aux virus. En première ligne les virus à ARN positif qui sont directement ciblés par le NMD lors de la phase de traduction active et qui ont par leur organisation particulière des phases de lecture avec des longs 3'-UTR ciblées par le NMD (Garcia et al., 2014). Chez les plantes *Arabidopsis thaliana* et *Nicotiana clevelandii*, la

surexpression d'une version dominant-négative d'*UPF1* entraîne une meilleure réplication du virus X de la pomme de terre (PVX : *Potato Virus X*) et du virus du rabougrissement du navet (TCV : *Turnip Crinkle Virus*) (Garcia et al., 2014). De la même manière, un crible de siRNA réalisé dans des cellules humaines (Balistreri et al., 2014) a montré que la déplétion d'UPF1, SMG5 ou SMG7 activait la réplication du virus Sindbis (SINV : *SINdbis Virus*) et du virus de la forêt de Semliki (SFV : *Semliki Forest Virus*).

Les autres types de virus, rétrovirus, virus à ADN ou à ARN négatif peuvent aussi générer des ARNs substrats de NMD mais eux ont développé des mécanismes de défense. Un exemple est l'existence de motif RSE (*RNA Stability Element*) qui sont présents sur l'ARN du rétrovirus du sarcome de Rous (RSV : *Rous Sarcoma Virus*) et qui inhibent directement le NMD en cis (Weil and Beemon, 2006; Withers and Beemon, 2010). D'autres virus agissent directement sur l'efficacité du NMD dans les cellules infectées (Fiorini et al., 2018; Mocquet et al., 2015; Nakano et al., 2013). Récemment, des chercheurs ont identifiés un nouveau système de protection contre le NMD chez le virus Zika (Fontaine et al., 2018), un virus à ARN positif. Il semblerait que l'interaction entre la capsid du virus et la protéine Upf1 entraîne la diminution de la quantité d'Upf1 dans la cellule. Dans un premier temps, cet effet pourrait empêcher la dégradation de l'ARN virale par le NMD. Dans un deuxième temps, la diminution de l'efficacité du NMD provoquée par la baisse d'Upf1 stabiliserait des transcrits impliqués dans l'apoptose et le cycle cellulaire. Ces deux voies métaboliques sont impliquées dans la microcéphalie congénitale, une caractéristique phénotypique retrouvée chez les fœtus infectés par le virus Zika (Li et al., 2016).

Objectifs de la thèse

La recherche sur le NMD est très intense depuis les débuts, c'est le mécanisme de contrôle qualité des ARNs le plus étudié. Cette activité intense a permis de proposer différents modèles pour la reconnaissance des cibles du NMD, son activation et les différentes voies de dégradation utilisées. Chez les eucaryotes multicellulaires la voie EJC-dépendante est mieux caractérisée que la voie EJC-indépendante. Pour cette dernière, les étapes de la dégradation des substrats ne sont pas connues tout comme les mécanismes moléculaires sous-jacents à ses étapes. Ce point est un obstacle à la compréhension de la prise en charge de la grande variété d'ARNs régulés par le NMD et sur l'évolution de ces mécanismes à travers les eucaryotes.

Pendant mes quatre années de thèse, je me suis concentrée sur l'étude de la voie du NMD EJC-indépendante et j'ai tenté de mieux comprendre la dynamique de ce mécanisme. Pour cela, j'ai travaillé dans le modèle *S. cerevisiae* qui ne possède pas de complexes EJC mais où les autres facteurs Upf1, Upf2 et Upf3 sont très conservés par rapport aux eucaryotes multicellulaires. J'ai utilisé des approches de transcriptomique et de protéomique pour permettre une compréhension du mécanisme en tenant compte de tous les substrats de la levure. Les trois axes de mes recherches ont été les suivants :

1. Mieux comprendre comment et où est recruté Upf1 au niveau des ARNs cibles du NMD.
2. Quels sont les facteurs impliqués dans le NMD de levure ? Est-ce qu'il existe de nouveaux facteurs ?
3. Quelle est la dynamique d'action de ces facteurs impliqués dans le mécanisme du NMD EJC-indépendant ?

Dans un premier temps, je décrirais les développements méthodologiques qui ont été nécessaires pour les analyses bio-informatiques des résultats de protéomique de ce travail de thèse. Ensuite je présenterai les résultats obtenus sur la localisation d'Upf1 sur les ARNs et les tentatives pour préciser les sites de recrutement et les facteurs associés à cette étape. Je décrirai ensuite en détail les avancées principales que notre groupe a réalisé sur le mécanisme du NMD de levure en présentant mon article de thèse. Je conclurai enfin sur les nouvelles perspectives que nos résultats apportent à la communauté et sur les expériences préliminaires que j'ai réalisé pour approfondir le modèle que nous décrivons.

Développement technologique

Analyse des données de spectrométrie de masse

Une partie importante de mes résultats de thèse repose sur l'analyse quantitative en spectrométrie de masse des protéines précipitées après purifications de différents facteurs du NMD. Nous avons réalisé plus d'une centaine de purifications différentes en utilisant 12 protéines appât différentes et plusieurs conditions d'expériences.

L'enjeu était alors de comparer ces purifications réalisées de 2013 à 2017, par trois manipulateurs.

D'un point de vue expérimental, la procédure était très bien maîtrisée au laboratoire afin de limiter les biais. Tous les manipulateurs ont utilisé le même protocole de purification (voir Matériel et Méthode de l'article) avec la même proportion de billes magnétiques couplées aux immunoglobulines G (IgG) et de protéase TEV par rapport à la quantité de cellule de départ, les mêmes conditions de culture des levures, le même protocole de précipitation des protéines, etc.

Concernant la spectrométrie de masse, nous avons utilisé le même matériel pour toutes nos expériences. De plus, à chaque passage, nous avons calibré la machine en utilisant une solution de BSA (*Bovin Serum Albumin*) et vérifié l'efficacité de détection des peptides de ce contrôle. La spectrométrie de masse permet d'identifier et de quantifier les peptides contenus dans un échantillon. Des étapes en amont de l'acquisition permettent de digérer les protéines de l'échantillon en peptides par action de protéase (**Figure 11A**). La trypsine par exemple est une protéase présente dans les sucs pancréatiques, elle hydrolyse les liaisons peptidiques du côté carboxy-terminal d'un résidu lysine ou arginine, deux acides-aminés basiques. Après digestion des protéines entières, les peptides sont séparés en fonction de leur polarité dans une colonne nano-HPLC (*High-Performance Liquid Chromatography*) puis injectés dans le spectromètre de masse. La chromatographie avant l'analyse en spectrométrie de masse permet de séparer les différents peptides pour une meilleure acquisition. En effet, si trop de peptides sont injectés en même temps dans le spectromètre, certains pics ne seront pas analysés. Les peptides sont ionisés au moment de leur injection par électro spray. Une première acquisition MS permet d'identifier la masse sur charge (m/z) des différents peptides du mélange (**Figure 11B**) et de mesurer leur intensité. Chaque pic correspondant à un peptide est ensuite fragmenté et les m/z des fragments sont mesurées dans une deuxième acquisition MS (MS/MS, **Figure 11C**). Un logiciel analyse ensuite les spectres de MS/MS afin de reconstruire la séquence ordonnée des peptides, en effet la m/z de chaque résidu est unique et

dépend des résidus voisins. Une fois les peptides identifiés, la comparaison avec une base de donnée de peptide générée *in silico* permet de définir les protéines correspondantes (**Figure 11D**). L'intensité des peptides mesurés dans la première MS est aussi récupérée et permet de calculer l'intensité du signal pour chaque protéine.

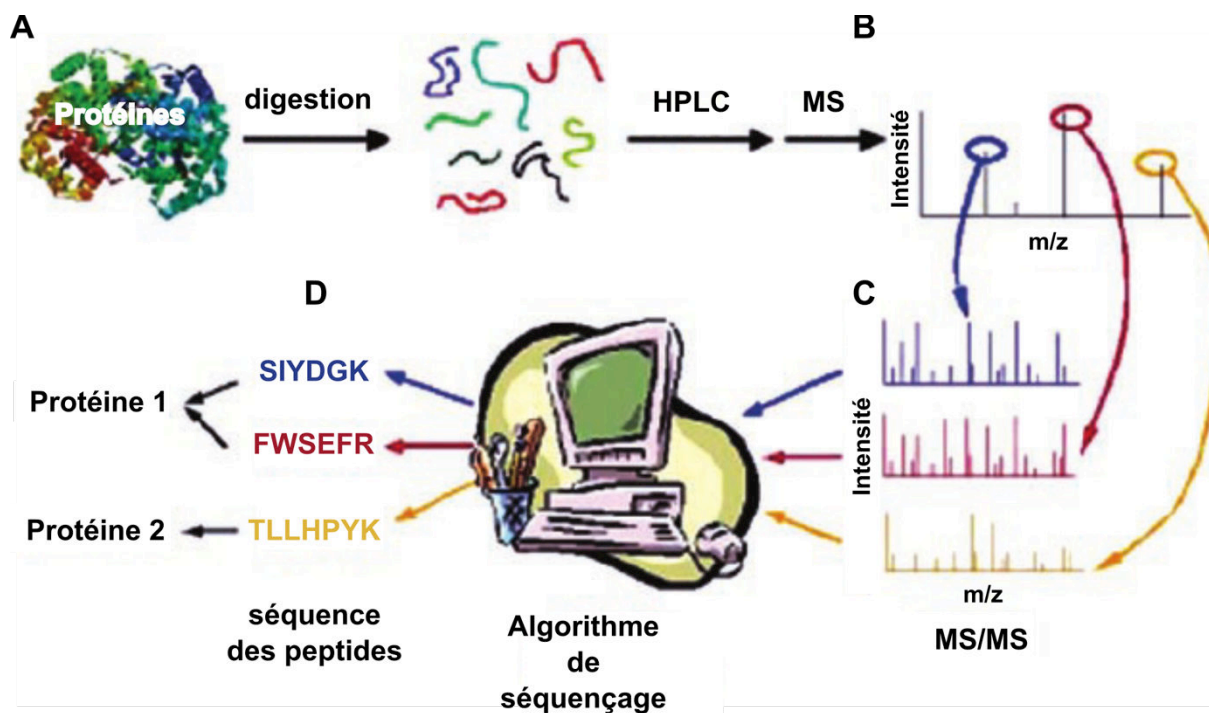


Figure 11: Schéma de l'analyse en spectrométrie de masse en tandem

A. Digestion des échantillons de protéines par l'action des protéases trypsine et Endoprotéase Lys-C. **B.** Nano-HPLC pour séparer les peptides, injection dans spectromètre et première acquisition MS. Identification de la masse/charge (m/z) des peptides et mesure de leur intensité. **C.** Pour chaque peptide identifié, fragmentation et deuxième acquisition MS (MS/MS) pour identifier la m/z et l'intensité des différents fragments. **D.** Analyse bio-informatique pour reconstruire la séquence des peptides à partir des spectres de m/z des MS/MS et identifier les protéines correspondantes. Les intensités des protéines sont aussi calculées en fonction des intensités des peptides mesurés pendant la première acquisition MS.

À chaque étape décrite ci-dessus, que ce soit pendant la préparation des échantillons ou les acquisitions MS, des irrégularités sont possibles. Celles-ci doivent être gommées pour pouvoir comparer les expériences entre elles. Une partie de mon travail de thèse a consisté à développer des méthodes de normalisation des résultats, à différentes phases de l'analyse bio-

informatique, pour corriger ces irrégularités. Le pipeline que nous avons utilisé est représenté dans la **Appendix Fig. S1**. Il se découpe en plusieurs étapes que je vais vous décrire maintenant.

Étape 1 : Analyse des spectres MS/MS et reconstitution des séquences des peptides identifiés. Les séquences des peptides permettent ensuite de rechercher les protéines correspondantes dans une base de donnée. Cette étape a été réalisée par Abdelkader Namane en utilisant le logiciel MaxQuant (version 1.5.5.1, Cox and Mann, 2008) et la base de donnée Andromeda associée (Cox et al., 2011). Le résultat obtenu est sous la forme de deux fichiers, un premier contenant les intensités des peptides et un deuxième contenant les intensités des protéines obtenues à partir du premier fichier. Ces fichiers sont liés.

Étape 2 : Les algorithmes de MaxQuant ne permettent pas de personnaliser les analyses bio-informatiques. Pour plus de flexibilité, nous avons poursuivi par une analyse utilisant le langage R (sur R studio) en construisant nos propres scripts. Nous avons commencé par filtrer les peptides quantifiés en éliminant tous les peptides pouvant correspondre à plusieurs groupes de protéines ; un groupe de protéines étant un groupe composé de une ou plusieurs protéines de séquence très proche et ne pouvant pas être distinguées. Nous avons aussi éliminé les peptides dont l'identification, par les acquisitions MS, était aberrante, une base de donnée de ce type de peptide existe et comprend des peptides issus d'une inversion des séquences des protéines par exemple (Elias and Gygi, 2007).

Étape 3 : Une des erreurs possibles lors du processus de préparation des échantillons pour la spectrométrie de masse est la digestion incomplète de certains peptides. Cette situation aboutit à l'identification de plusieurs peptides différents en spectrométrie de masse mais qui sont superposés au niveau de la séquence de la protéine. L'intensité mesurée pour ces peptides ne reflète pas l'intensité réelle de la protéine. En effet, dans une situation sans oubli de coupure, la présence d'une protéine dans le mélange aboutit à la production d'un exemplaire de chaque peptide pour la protéine ; en cas d'oubli de coupure, tous les peptides ne sont pas représentés et en fonction de la fréquence de coupure leurs intensités peuvent varier et alors mésestimer l'intensité réelle de la protéine entière. Pour contrecarrer cette irrégularité nous avons recalculé les intensités des peptides en regroupant les peptides superposés. Les intensités des peptides superposés ont été sommées et attribuées à un méta-peptide correspondant au peptide le plus grand formé par les peptides superposés. Cette uniformisation permet d'éviter des biais dus à la variation de l'efficacité de coupure des protéines des échantillons.

Étape 4 : Après le filtrage des intensités des peptides et la correction des oublis de coupure, nous avons calculé des valeurs d'intensités pour les protéines. Plusieurs méthodes existent pour cela, le LFQ (*Label Free Quantification*, Cox et al., 2014), le comptage des spectres des peptides (Liu et al., 2004), le top3 (Silva et al., 2006), etc.

Le LFQ est basé sur un algorithme qui compare, dans les différents échantillons, les intensités de chaque peptide commun. À partir de ces comparaisons, il calcule des ratios entre les échantillons, ce qui permet de quantifier les différences entre les échantillons et de déduire des intensités relatives pour les protéines. La seule condition pour l'utilisation de cet algorithme est la conservation d'une majorité de protéines communes dans les différents échantillons. Au vu de cette dernière information, le LFQ ne semble pas indiqué pour la quantification des protéines co-purifiées avec différents appâts ou dans des conditions différentes car nous ne pouvons pas être assurés de cette conservation. Par ailleurs l'algorithme LFQ est plus efficace avec des échantillons comprenant un très grand nombre de protéines, ce qui n'est pas le cas des échantillons purifiés. Finalement, les valeurs obtenues dépendent des échantillons qui ont été analysés en même temps. Les valeurs LFQ varient donc d'une analyse à une autre pour le même échantillon.

Une autre méthode d'estimation des quantités relatives des protéines est le Top3, cette méthode consiste à faire la moyenne des trois peptides les plus intenses afin de déterminer, après transformation logarithmique, l'intensité de la protéine associée. Dans des extraits totaux, le calcul du Top3 est très corrélé à l'abondance des protéines (Ahrné et al., 2013), en outre, il est assez efficace pour estimer une différence d'expression de protéines entre deux conditions (Cox et al., 2014), contrairement au comptage des spectres d'intensité des peptides. Par ailleurs, le Top3 ne dépend pas des valeurs des intensités des peptides dans les autres échantillons et il est en ce sens plus indiqué pour comparer des expériences très différentes comme peuvent l'être des purifications de protéines variées. Ces caractéristiques du Top3 sont illustrées par le graphique de l'intensité des peptides pour l'appât Upf1 quantifiés dans plusieurs expériences. On observe, parmi la grande variabilité des intensités mesurées, une différence significative de l'intensité des peptides les plus forts entre les conditions où Upf1 est exprimé de manière endogène et les conditions où Upf1 est surexprimé (**Figure 12**).

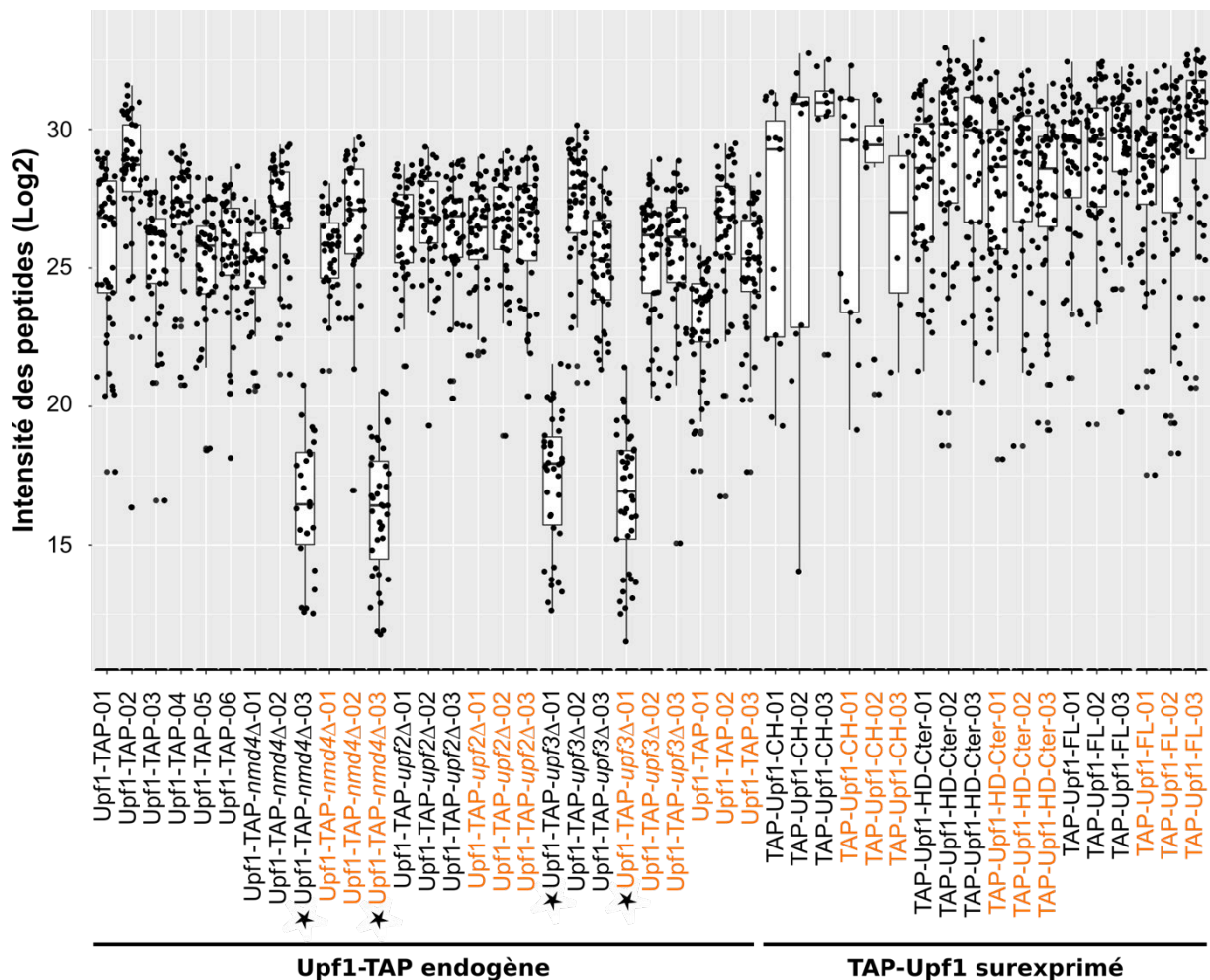


Figure 12 : Graphique de l'intensité des peptides quantifiés pour Upf1 purifié dans différentes conditions et souches.

Toutes les expériences présentées ici sont des purifications d'Upf1 dans différentes souches de levure et différentes conditions. Chaque point représente un peptide, les boîtes à moustaches représentent la variation des intensités des peptides pour chaque expérience, la ligne au milieu de chaque boîte est la médiane. Les expériences indiquées en orange ont subi un traitement RNase au cours de la purification. Les expériences marquées d'une étoile noire ont été éliminées des analyses car les intensités des peptides pour Upf1 étaient très basses par rapport aux autres expériences et aurait entraînées une augmentation du bruit de fond. Les numéros des expériences sont les numéros des répliques.

La représentation des intensités des peptides permet aussi de repérer des répliques atypiques comme ceux indiqué par une étoile (**Figure 12**). Il semblerait que les intensités pour ces répliques soient plus faibles pour l'appât que dans les autres répliques. Nous avons décidé

d'écarter ce type de répliques où un problème a dû survenir lors de la précipitation des protéines après purification ou lors de l'acquisition MS.

La variabilité de l'intensité des peptides peut s'expliquer par leurs différences de propriétés chimiques, ils n'ont alors pas le même comportement dans le spectromètre de masse (Ruth et al., 2010) ce qui peut affecter leur détection. En fonction du nombre de peptide quantifié par protéine, l'impact des « mauvais » peptides est variable. Le Top3 permet de se focaliser uniquement sur les peptides les plus souvent identifiés, plus représentatif de l'intensité réelle de la protéine correspondante. Pour les petites protéines, ayant peu de peptides, l'utilisation du Top3 peut sous-estimer leur quantité, le risque est alors de manquer une protéine parmi les partenaires. De plus, certaines protéines vraiment petites comme Lsm2-7, par exemple, peuvent avoir moins de trois peptides identifiés, rendant le calcul du Top3 traditionnel impossible. Dans ce cas, nous avons développé une nouvelle formule calculant le Top3 lorsque trois peptides étaient quantifiés mais autorisant l'utilisation de deux peptides seulement si moins de trois peptides avaient été identifiés. Ce score est appelé le **LTOP2** (Logarithme pour le Top3 ou 2 si besoin).

Étape 5 : Après calcul des scores d'intensités des protéines dans chaque expérience, nous avons filtré les résultats en fonction de la reproductibilité des identifications. Pour cela, nous avons appliqué un script permettant d'éliminer dans chaque groupe de purification les protéines identifiées moins de deux fois parmi les répliques. Au sein de chaque groupe de purification nous avons ensuite normalisé les distributions en utilisant la méthode des quantiles (Bolstad et al., 2003).

Après cette étape, nous avons séparé les échantillons correspondant à des purifications et les échantillons correspondant à des extraits totaux.

Les extraits totaux vont être comparés à un jeu de données récapitulant les abondances des protéines de la levure (**Figure 13 et Fig. 1D de l'article**). On a observé une très bonne corrélation entre le calcul du LTOP2 pour les protéines quantifiées dans nos extraits totaux et l'abondance des protéines de levure (Ho et al., 2018). Le coefficient de régression est de 0,82. La régression linéaire a permis de calculer un facteur de correction des LTOP2 par rapport à la véritable abondance des protéines.

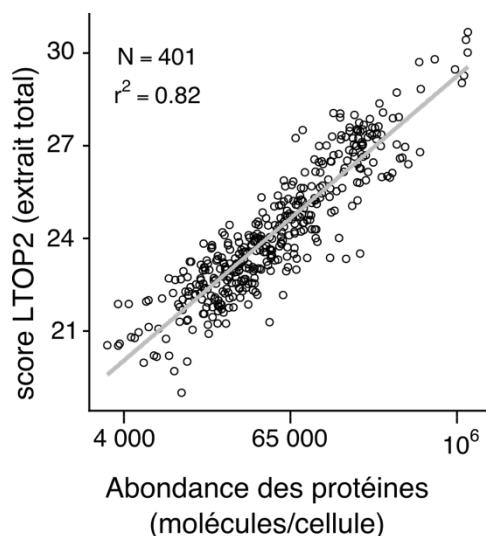


Figure 13 : Comparaison des LTOP2 des protéines quantifiées dans les extraits totaux avec l'abondance connue de ces protéines

N est le nombre de protéines en commun entre les trois extraits totaux et la base de donnée d'abondance publié par Grant W Brown et collègues (Ho et al., 2018) qui ont servi de base pour le calcul de la régression linéaire et du facteurs de correction entre le score LTOP2 et les abondances. r^2 est le coefficient de corrélation calculé avec la méthode de Pearson.

Étape 6 : La quantification des protéines pour les purifications est une étape cruciale pour nos analyses. Nous avons utilisé plusieurs appâts et des conditions d'expériences parfois très différentes. Nous savons que l'efficacité des purifications dépend de tous ces paramètres intrinsèques et aussi des irrégularités du manipulateur et des acquisitions en spectrométrie de masse. Pour éliminer ces irrégularités nous avons effectué une normalisation des purifications par rapport à la quantité de protéase TEV mesurée en MS. La quantité de protéase TEV ajoutée en excès, au moment de l'élution des purifications, est identique dans toutes les purifications (**Figure 14**). Au moment de l'analyse bio-informatique, nous avons donc appliqué un facteur de correction afin d'égaliser la quantification de la TEV dans tous les échantillons.

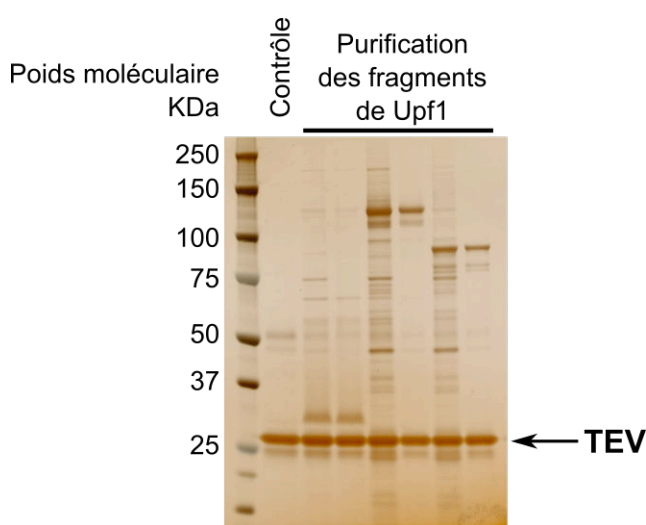


Figure 14 : Coloration à l'argent du résultat de sept purifications différentes

La protéase TEV est indiquée par une flèche. Le contrôle correspond à une purification sans protéine taguée.

Étape 7 : Pour déterminer si une protéine est spécifiquement associée à la protéine appât, il faut savoir si sa quantité est enrichie par rapport à l'extrait total de départ. Pour pouvoir comparer les purifications et les extraits totaux nous avons appliqué le facteur de correction entre abondance et LTOP2, calculé précédemment, aux purifications. Nous avons ensuite calculé la proportion de chaque protéine dans toutes les expériences de purification et les extraits totaux. Le rapport des proportions dans les purifications et les extraits totaux pour chaque protéine a permis de déterminer l'enrichissement selon les calculs suivants (**Figure 15**).

Calcul des proportions :

$$P(\text{protéine } X) = \frac{LTOP2(X)}{\sum LTOP2(\text{toutes les protéines})} \quad \text{dans les purifications}$$

$$T(\text{protéine } X) = \frac{\text{abondance}(X)}{\sum \text{abondance}(\text{toutes les protéines})} \quad \text{dans les extraits totaux}$$

Calcul des enrichissements :

$$E(\text{protéine } X) = \frac{P(X) \text{ purification}}{T(X) \text{ extraits totaux}}$$

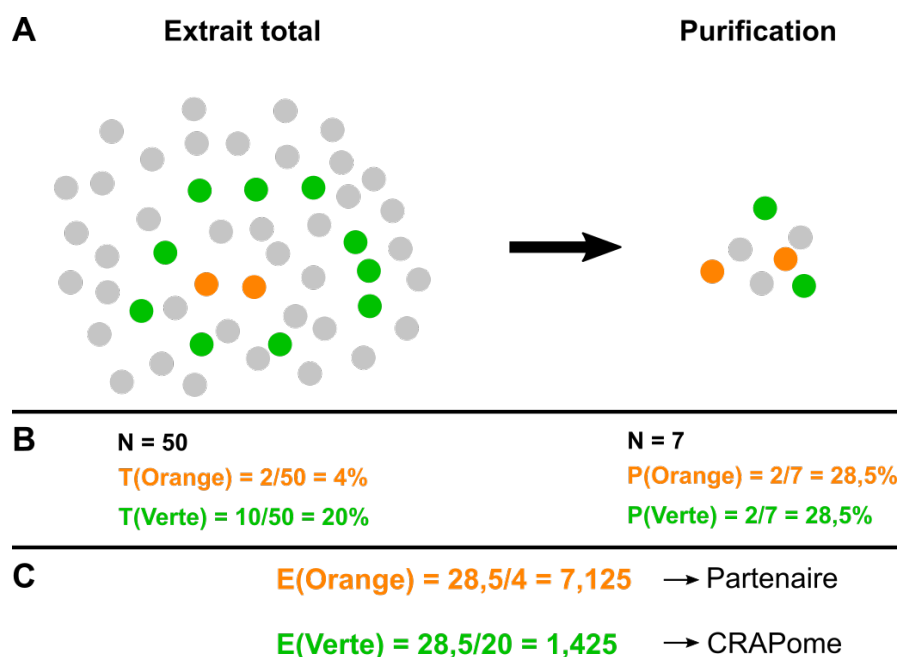


Figure 15 : Principe du calcul de l'enrichissement

A. Schéma d'une purification, chaque bille est une protéine. Les billes orange correspondent à un type de protéine, les billes vertes sont un autre type de protéine abondante et les billes grises sont d'autres protéines. **B.** Calcul de la proportion des protéines orange et vertes. **C.** Calcul de l'enrichissement des protéines orange et vertes dans la purification par rapport à l'extrait total.

Le calcul de l'enrichissement permet d'éliminer des partenaires potentielles les protéines abondantes des extraits totaux qui sont retrouvées dans les purifications du fait de leur abondance initiale (**Figure 15**). Ce type de protéine constitue le CRAPome (Mellacheruvu et al., 2013).

Pour les analyses des résultats de spectrométrie de masse des purifications, il est important de garder les informations de l'enrichissement et du LTOP2. L'enrichissement évalue la spécificité et la stabilité de l'interaction entre les partenaires. Le LTOP2 dépend directement de la mesure en spectrométrie de masse et permet d'évaluer la quantité absolue de protéine dans les différents extraits. Nous avons pris l'habitude de représenter nos résultats sous forme d'un graphique représentant en abscisse le score LTOP2 et en ordonnée l'enrichissement (**Figure 16** pour exemple). Les interactions les plus spécifiques et robustes seront celles ayant un fort enrichissement et un score LTOP2 pas trop faible. En effet, lorsque le LTOP2 est faible, sa valeur pour estimer la quantité d'une protéine en spectrométrie de masse est moins précise. Sur ce graphique, les protéines considérées comme le CRAPome ont un enrichissement faible et un LTOP2 élevé. Il faut cependant être prudent concernant les protéines très abondantes qui peuvent ne pas être enrichies dans les purifications et pourtant avoir une interaction spécifique avec l'appât plus ou moins élevée.

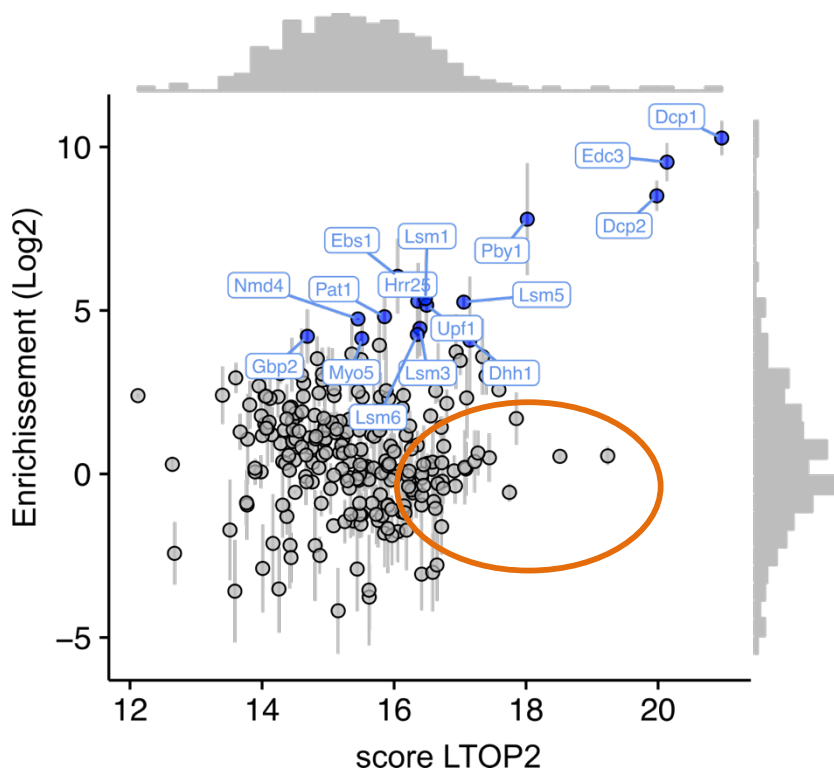


Figure 16 : Représentation des résultats de la purification de Dcp1

Le LTOP2 est représenté en abscisse, l'enrichissement en ordonnée. L'enrichissement est en log2 dans cette figure. La distribution des LTOP2 est aussi dessinée en haut de la figure, celle de l'enrichissement est à droite. Les protéines localisées dans le cercle orange sont probablement du CRAPome.

La purification de Dcp1 est un exemple qui valide notre approche pour identifier les partenaires d'une protéine. En effet, grâce au calcul de l'enrichissement, nous récapitulons les interactions connues de Dcp1 comme Dcp2, Edc3, Pby1, le complexe LSM cytoplasmique, Pat1, Dhh1. L'utilisation du LTOP2 seul, n'aurait pas permis une identification facile d'autant de partenaires. Pat1 par exemple ou le complexe Lsm1-7 ont un score LTOP2 d'environ 16,5 et aurait été « noyés » dans le CRAPome (**Figure 16**). On a aussi observé que Dcp1, Dcp2 et Edc3 étaient enrichis de la même manière, ces trois facteurs font partie de la forme dite active pour l'hydrolyse de la coiffe (Charenton et al., 2016).

Résultats

I. Recrutement d'Upf1 au niveau des ARNs cibles

1. Upf1 interagit avec les 3'-UTR de centaines d'ARNs

Comme indiqué en introduction, deux hypothèses s'opposent pour expliquer l'accumulation d'Upf1 au niveau des 3'-UTR (Shigeoka et al., 2012; Zünd et al., 2013) des cibles du NMD (Hogg and Goff, 2010; Johansson et al., 2007). La première hypothèse consiste en le recrutement d'Upf1 tout le long des transcrits dans un premier temps. Puis, les ribosomes en traduction décrocheraient les protéines Upf1 des régions codantes. La deuxième hypothèse propose un recrutement d'Upf1 au niveau du codon Stop puis le déplacement de la protéine sur la région 3'-UTR. Le mécanisme de recrutement n'est donc pas élucidé. Upf1 est une hélicase très processive qui a la possibilité *in vitro* de se déplacer sur les acides nucléiques. Si on imagine que Upf1 garde cette capacité *in vivo* alors la situation à l'équilibre de la position d'Upf1 ne révélera en rien la position du recrutement de la protéine. Pour répondre à cette interrogation, il faudrait inactiver la fonction hélicase d'Upf1 sans empêcher la liaison à l'ARN ce qui permettrait de figer les protéines Upf1 à leur point de recrutement. Plusieurs mutants du domaine hélicase ont été décrits (Weng et al., 1996a, 1996b, 1998). La substitution de la lysine en position 436 de Upf1 par une glutamine (K436Q chez la levure, K498Q chez l'humain, Kashima et al., 2006) empêche la fixation de l'ATP nécessaire à l'activité de l'hélicase. Le mutant des résidus tyrosine et arginine en position 800 remplacés par deux alanines (TR800AA pour *S. cerevisiae*) empêche la fixation à l'ARN *in vitro*. Les substitutions des résidus aspartate et glutamate en position 572, remplacés par deux alanines (DE572AA pour *S. cerevisiae*, DE637AA pour Upf1 humain) empêchent l'hydrolyse de l'ATP. Ces mutants de l'hélicase d'Upf1 inactivent complètement le mécanisme du NMD chez la levure et chez d'autres espèces (Kashima et al., 2006; Weng et al., 1996a, 1996b, 1998).

Chez la levure *S. cerevisiae*, les mutants décrits précédemment n'ont pas ou ont très peu de défaut de croissance. La levure est donc un bon modèle pour tester la localisation sur l'ARN des versions mutantes d'Upf1. Dans un premier temps nous avons identifié chez la levure les sites de fixation d'Upf1 sauvage sur les ARNs. Pour cela nous avons utilisé la technique de CRAC (Cross Link et Analyse d'ADNc, Granneman et al., 2009). Cette technique a été développée dans le laboratoire par Gwenaël Badis et se déroule en plusieurs étapes (**Figure 17**). La protéine étudiée, ici Upf1, est marquée en position C-terminale avec une longue étiquette composée d'une répétition de six histidines (6His), d'un site de clivage à la protéase

TEV (dérivée du virus de la gravure du tabac, *Tobacco Etch Virus*) et d'une séquence codant la protéine A (tag HTP : His-TEV-ProtA, **Figure 17A**). L'étiquette permet l'immuno-précipitation d'Upf1 et de ses ARNs associées.

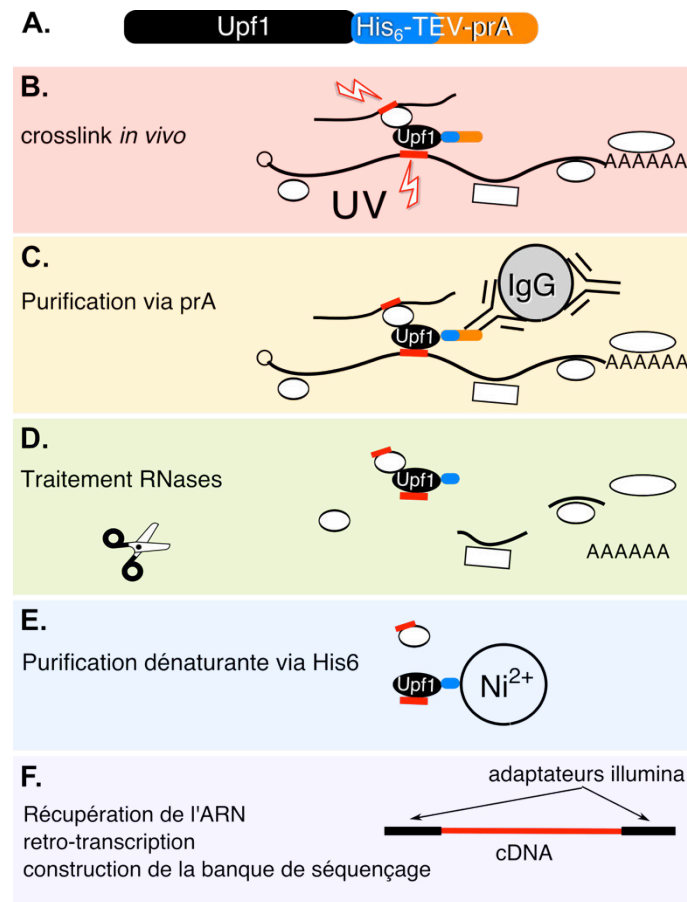


Figure 17 : Schéma de la méthode CRAC

A. Schéma de l'étiquette HTP utilisée pour les purifications successives d'Upf1 pour les expériences de CRAC. **B.** Établissement de liaisons covalentes entre les protéines et les ARNs par application de rayons UV à une longueur d'onde de 254nm. **C.** Purification d'Upf1 par la partie protéine A de l'étiquette HTP et élution des complexes ARNs/protéines par clivage au niveau du site de reconnaissance de la protéase TEV. **D.** Grignotage de l'ARN autour des régions protégées par les protéines par les RNases non spécifiques A et T1. **E.** Purification en condition dénaturante *via* la partie 6His de l'étiquette. **F.** Construction d'une banque de fragments retro-transcrits en ADNc pour le séquençage à haut débit. Ligation des adaptateurs sur le fragment d'ARN, sélection par la taille des fragments liés à Upf1, digestion d'Upf1 avec de la protéinase K puis retro-transcription et amplification par PCR de la banque en utilisant des amorces dans les adaptateurs Illumina®.

Dans une première étape, les cellules de levure (provenant de 4L à 8L de culture liquide) sont soumises à l'action de rayons ultraviolets (254nm) pour créer des liaisons covalentes entres

les protéines et les ARNs (**Figure 17B**) *in vivo*. Les cellules sont ensuite lysées en utilisant une méthode mécanique douce afin de préserver les ARNs. Une première purification en utilisant des billes magnétiques couplées aux IgG, reconnaissant l'étiquette protéine A, permet d'enrichir les complexes ARNs-protéines associés à Upf1 (**Figure 17C**). Les échantillons sont ensuite traités avec de la RNase A et T1 afin de grignoter les ARNs et ne garder que les fragments protégés par les protéines (**Figure 17D**). La particularité du CRAC par rapport à d'autre méthode tel le CLIP (CrossLink et Immuno-PréciPitation, Ule et al., 2003), est l'existence d'une deuxième étape de purification en condition dénaturante *via* l'étiquette 6His. Cette étape permet d'éliminer les protéines interagissant avec Upf1 et qui pourrait lier covalamment un autre fragment d'ARN (**Figure 17E**). Les étapes résumées **Figure 17F** sont la retro-transcription des fragments d'ARN en ADN complémentaire (ADNc), la liaison des adaptateurs Illumina® permettant ensuite le séquençage de la banque de fragments et l'amplification de la banque par PCR. Une fois les fragments séquencés, une analyse bio-informatique permet de cartographier les fragments sur le génome de la levure. L'analyse des positions des fragments séquencés permet ensuite de déterminer les régions de fixation de la protéine immuno-précipitée, ici Upf1-HTP, sur les ARNs.

L'étude de certains transcrits typiques des cibles du NMD a montré une accumulation de fragments provenant des régions 3'-UTR (**Figure 18**). Par exemple, le transcrit RPL8A (**Figure 18A**) existe sous deux formes, une première forme avec un court 3'-UTR et une deuxième forme avec un long 3'-UTR sensible au NMD qui est dû à l'existence d'un site alternatif de terminaison de la transcription (TTS : *Transcription Termination Site*). Le signal d'Upf1 a été localisé sur la forme longue uniquement du transcrit. NMD4 (**Figure 18B**) est un autre substrat physiologique du NMD, son 3'-UTR est très long (439 nucléotides, Pelechano et al., 2013) alors que la taille moyenne des 3'-UTR de levure est d'environ 91 à 140 bases (David et al., 2006; Tuller et al., 2009; Wilkening et al., 2013).

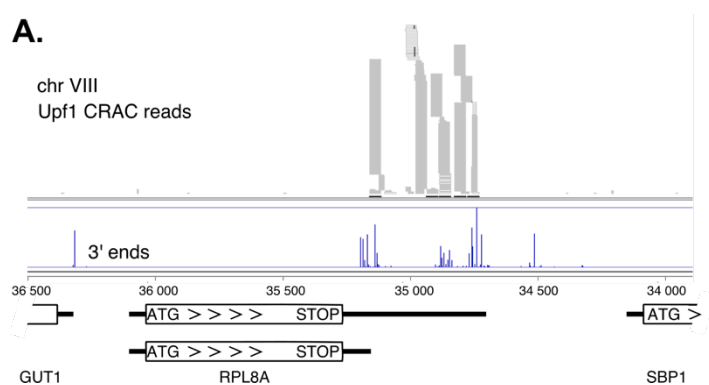
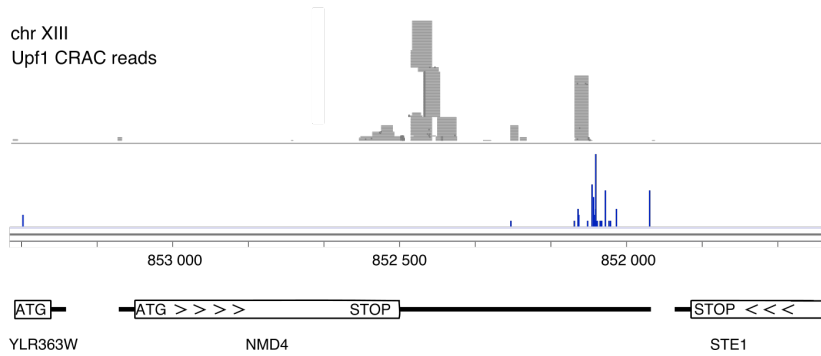


Figure 18 : Exemples de la localisation d'Upf1 sur deux transcrits typiques des substrats de NMD

Un isoforme de RPL8A (**A.**) et NMD4 (**B.**).

B.



Nous avons ensuite tenté de résumer les résultats à l'échelle du transcriptome de levure. Pour cela nous avons analysé une fenêtre de 300 nucléotides autour du codon d'initiation de la traduction (AUG) et autour du codon Stop canonique pour 5077 phases codantes annotées de la levure *S. cerevisiae*. Pour éviter les effets forts de certains transcrits très présents dans la cellule et pour lesquels il y avait un grand nombre de séquences, nous avons compté pour chaque nucléotide des deux fenêtres le nombre de gènes pour lesquels on avait un signal pour l'expérience et le contrôle (**Figure 19A**). Le contrôle a consisté en une expérience sans liaison covalente des protéines aux ARNs par les UV. Il permet de déterminer le bruit du séquençage de l'expérience, car il n'y a en théorie pas de maintien des interactions spécifiques ARNs/protéines sans la liaison covalente au cours de l'expérience de CRAC. Les résultats pour l'ensemble des transcrits « codants » (**Figure 19B et C**) ont montré une accumulation spécifique d'Upf1 au niveau du 3'-UTR par rapport au reste de la séquence codante et au 5'-UTR pour des centaines d'ARNs. La répartition du signal montre un pic entre 100 et 150 bases après le codon Stop, correspondant à la variation de la longueur des 3'-UTR des ARNm de levure.

L'expérience a été réalisée deux fois de manières similaires avec les mêmes résultats globaux. On observe cependant une différence de profondeur entre les deux répliques. Dans la première expérience, l'enrichissement dans les 3'-UTR a concerné 1000 transcrits alors qu'il fut visible dans seulement la moitié dans l'expérience 2. Cet écart pourrait aussi expliquer la différence de signal au niveau de la région du codon Start (**Figure 19B**) où on observe un enrichissement d'Upf1 au début de la séquence codante pour l'expérience 1 mais pas pour l'expérience 2. Ce dernier signal correspond probablement aux transcrits avec des uORFs et dont le pseudo long 3'-UTR se superpose à la séquence codante (visible aussi avant le codon Stop dans **Figure 19C**). Dans les deux expériences, le signal obtenu pour le contrôle négatif sans UV (ligne grise, **Figure 19B et C**) était proche de zéro, ce qui a permis de valider les enrichissements

spécifiques d'Upf1 décrits plus haut. Enfin, la comparaison entre le groupe d'ARNs accumulant Upf1 et le groupe d'ARNs stabilisés dans le mutant de délétion d'*UPF1* (*upf1Δ*, Malabat et al., 2015) a montré une bonne corrélation (résultats non montrés) signifiant que Upf1 s'accumule au niveau du 3'-UTR des substrats de NMD.

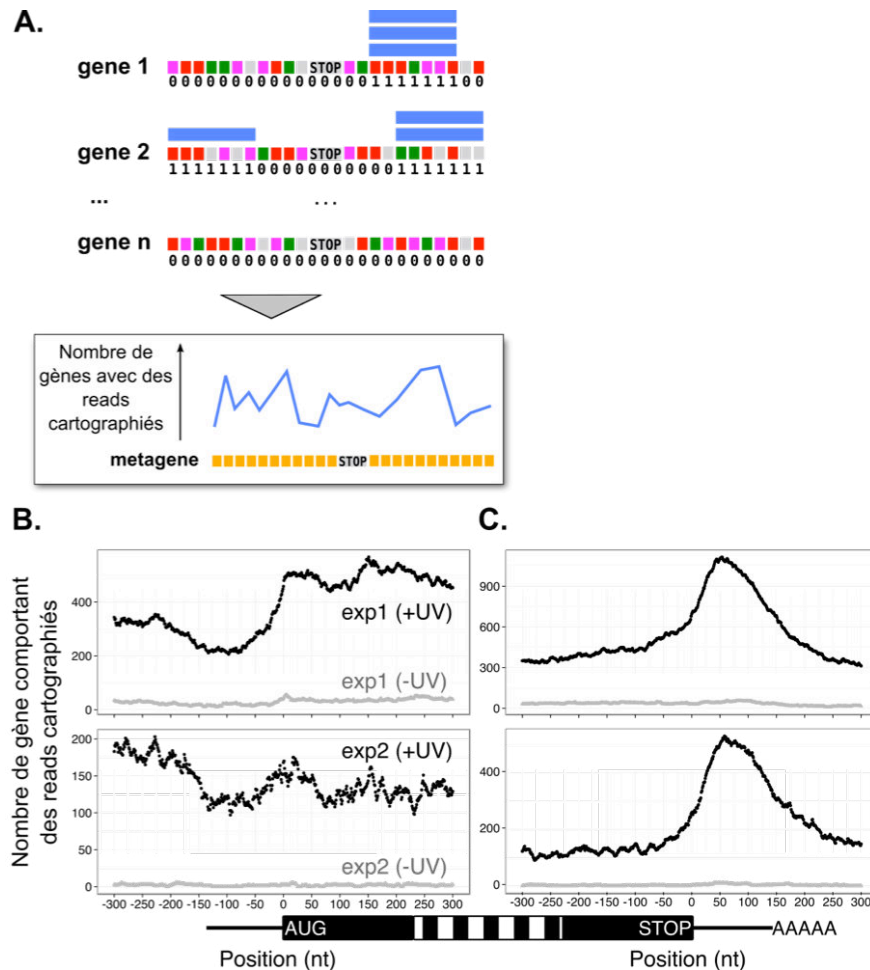


Figure 19 : Localisation d'Upf1 sur les transcrits de levure

A. Méthode d'intégration du signal de séquençage au niveau des transcrits. **B. et C.** Cartographie du nombre de gènes pour lesquels les ARNm ont montré des régions de liaisons avec Upf1 autour du codon START (**B.**) et STOP (**C.**) annotés des transcrits de levure. Les deux expériences indépendantes sont représentées avec en gris le signal pour le contrôle sans UV et en noir le signal pour l'expérience de CRAC Upf1.

Avec l'aide de Laurence Decourty, nous avons ensuite construit plusieurs souches de levure contenant une version mutante d'Upf1 comme seule source de la protéine. Nous avons testé les mutations K436Q et DE572AA affectant la région ATPase de la protéine, la mutation TR800AA affectant *in vitro* la liaison à l'ARN et la mutation C122S qui affecte la structure

du domaine CH d'Upf1 et la liaison à Upf2. Les souches mutées dans Upf1 ont aussi une étiquette HTP ou TAP (**Table 1**) qui permet de les détecter en western blot et de les purifier. Les mutants K436Q, TR800AA et C122S ont été construits au même locus que Upf1 sauvage, sur le chromosome ; en revanche le mutant DE572AA est exprimé à partir d'un plasmide centromérique sous le contrôle d'un promoteur répressible tetO7. Les versions taguées d'Upf1 sauvage et mutées s'expriment bien dans la cellule (**Figure 20A et B**). Les tests en RT-qPCR ont confirmé par ailleurs l'invalidation du NMD dans ces souches mutées (**Figure 20C**, exemple de TR800AA, C122S et DE572AA).

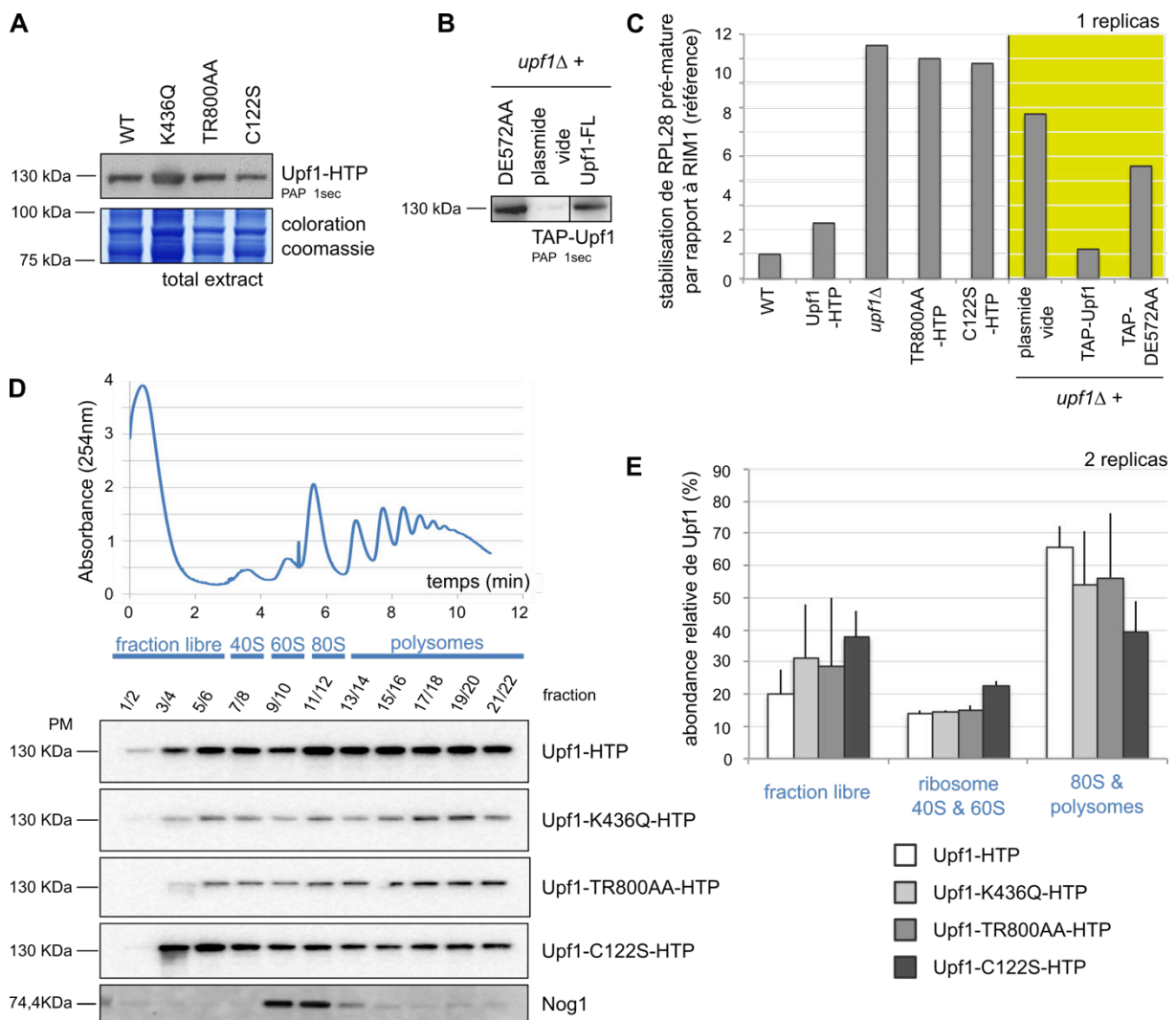


Figure 20 : Caractérisation des mutants d'Upf1

A. Contrôle du niveau d'expression des versions mutées K436Q, TR800AA et C122S d'Upf1-HTP par rapport au sauvage par western-blot à partir d'extraits totaux de levure exprimant les différentes versions de manière endogène. Un extrait d'une coloration au bleu de Coomassie en dessous a servi de contrôle de charge. **B.** Contrôle de l'expression des protéines Upf1 sauvage et mutée DE572AA,

taguée en position N-terminale et exprimée à partir d'un plasmide centromérique dans une souche *upf1Δ*. **C.** Mesure du niveau de stabilisation de l'ARN pré-messager de RPL28 sensible au NMD par rapport à un transcrit de référence non sensible RIM1 dans les souches de levure BY4741 (souche de référence), *upf1Δ*, Upf1-HTP, Upf1-TR800AA-HTP, Upf1-C122S-HTP et *upf1Δ* transformée avec le plasmide possédant la séquence codant TAP-Upf1-DE572AA ou TAP-Upf1-FL (version sauvage de Upf1) ou un plasmide vide. Les résultats de la partie surlignée en jaune sont issus de RT-qPCR indépendantes par rapport à la partie blanche. **D.** western-blot des fractions des gradients de polysomes réalisés pour les souches Upf1-HTP, Upf1-K436Q-HTP, Upf1-TR800AA-HTP et Upf1-C122S-HTP. Immuno-marquage avec l'anticorps PAP reconnaissant la partie protéine A de l'étiquette HTP et avec l'anticorps contre la protéine Nog1 permettant d'identifier la fraction 60S dans le gradient. **E.** Quantification relative des pourcentages d'Upf1-HTP dans les fractions libres, ribosomales (40S et 60S) et polysomiales (dont 80S) pour chaque souche. Les barres d'erreur représentent l'écart type.

Les mutants K436Q, TR800AA et C122S ont été testés *in vivo* pour leur liaison à l'ARN en analysant leur sédimentation dans un gradient de sucrose. Ce type de gradient permet de séparer la fraction de protéine cytoplasmique libre, des complexes ribosomiaux et des polysomes (**Figure 20D et E**). Upf1-TAP a sédimenté principalement dans les régions lourdes du gradient, les fractions polysomiales et monosomiales. Une fraction de la protéine est également retrouvée dans les autres fractions plus légères. Les mutants K436Q et TR800AA se sont comportés de manière intermédiaire, leur profil de sédimentation dans le gradient de sucrose a montré une répartition légèrement plus basse dans les polysomes, compensée dans la fraction libre. En revanche, le mutant C122S a perdu complètement l'enrichissement dans les polysomes. Ces données indiquent un affaiblissement des interactions entre les mutants K436Q et TR800AA avec les ARNs traduits et une perte pour le mutant C122S.

Ces résultats ont été complétés par l'étude des ARNs co-précipitants avec Upf1 sauvage et ses versions mutantes en utilisant la technique du RIP (*RNA Immuno Precipitation*) suivie d'un séquençage des ARNs ou d'une analyse en qPCR. L'expérience de séquençage n'a pas fonctionné suffisamment bien pour réaliser une analyse à l'échelle du transcriptome mais a permis de confirmer la liaison à l'ARN du mutant K436Q sur quelques transcrits fortement stabilisés. Les mutants C122S et TR800AA semblaient finalement perdre leur interaction avec l'ARN. Nous avons alors décidé de tester en CRAC le mutant ATPase K436Q en comparant avec Upf1 sauvage et Upf1-TR800AA. Malheureusement, nous n'avons pas réussi à avoir des résultats convaincants. La méthode CRAC avait besoin d'optimisation afin d'être utilisée de manière efficace avec nos échantillons. En effet, Upf1 est une protéine moyennement

exprimée et ses ARNs cibles sont rapidement dégradés en condition sauvage, ce qui a compliqué l'expérience. Pour les mutants qui stabilisent fortement les ARNs substrats, le problème aurait pu être résolu. Cependant, des résultats préliminaires de RIP-qPCR ont suggéré une perte de spécificité des mutants K436Q, TR800AA et C122S d'Upf1 pour les cibles du NMD (**Figure 21**). La partie bleue représente la stabilisation d'un transcrite sensible au NMD, le pré-messager RPL28, dans des extraits totaux d'une souche de levure contenant Upf1 sauvage tagué ou une version mutante. Ces mesures ont servi de base de normalisation pour calculer l'enrichissement spécifique de ce substrat, par rapport à un transcrite non cible du NMD (ici, la forme mature de RPL28), après immuno-précipitation d'Upf1 (partie verte). On peut observer une baisse de l'enrichissement spécifique de RPL28 pré-messager avec les mutants par comparaison avec Upf1 sauvage. Ce résultat est d'ailleurs corrélé avec l'étude récente montrant que l'activité hélicase d'Upf1 chez l'Homme serait cruciale pour la reconnaissance spécifique des cibles (Lee et al., 2015). Il faudrait donc réussir à avoir une couverture du séquençage encore meilleure pour espérer observer le signal de liaison d'Upf1 si celui-ci est réparti sur tout le transcriptome.

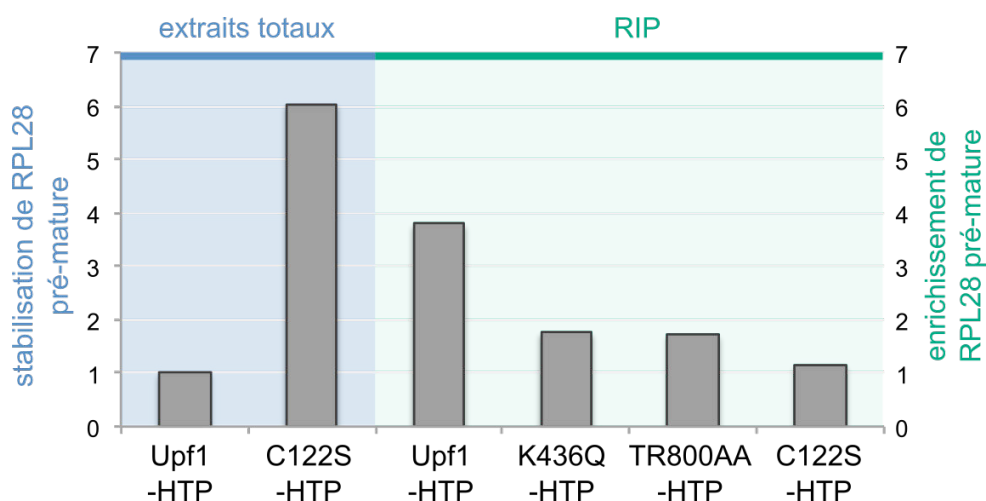


Figure 21 : Les mutants d'Upf1 perdent la spécificité de liaison aux transcrits RPL28 immatures, cibles du NMD

Les mesures de la stabilisation de RPL28 immature dans les extraits totaux ont été réalisées comme dans la **Figure 20C** mais j'ai utilisé l'ARN mature de *RPL28* résistant au NMD comme référence. Pour les RIP-qPCR, j'ai normalisé la quantité relative de RPL28 immature par rapport à celle dans les extraits totaux d'Upf1-HTP pour le contrôle positif et d'Upf1-C122S-HTP pour les trois mutants.

Depuis ces expériences préliminaires, Gwénaél Badis du laboratoire a développé un CRAC nouvelle génération en s'aidant de l'expérience utilisateur des laboratoires de Domenico Libri

et de Hervé Le Hir. Les résultats pour Upf1 sont meilleurs et le projet sera continué. Comme la liaison non spécifique d'Upf1 à l'ARN ne suit pas obligatoirement les mêmes règles que le recrutement spécifique sur les cibles du NMD, il faudra être très prudent avec les résultats obtenus.

2. Les mutations du domaine hélicase d'Upf1 n'affectent pas ses interactions avec les autres partenaires.

En parallèle des analyses en CRAC de la position d'Upf1, j'ai aussi analysé la composition des particules ribonucléiques (RNPs) en situation sauvage pour Upf1 et mutante aux différentes positions décrites précédemment. Nous avons purifié Upf1 et les versions mutantes dans des conditions similaires puis nous avons calculé les enrichissements des protéines comme décrit dans le chapitre précédent sur le développement méthodologique. L'expérience a été réalisée cinq fois pour Upf1 sauvage et au moins deux fois pour les mutants K436Q, TR800AA et C122S (**Table 1**). Pour les versions d'Upf1 surexprimées, trois réplicas ont été réalisés pour TAP-Upf1 et un seul pour le mutant DE572AA. Pour certains mutants, j'ai aussi réalisé des purifications avec un traitement RNase (Voir Matériels et Méthodes de l'article) afin d'éliminer les interactions non directes médiées par l'ARN.

Expérience	NMD	Niveau d'expression	Nb réplicas	Nb réplicas + RNase
Upf1-TAP	+	endogène	5	3
Upf1-K436Q-HTP	-	endogène	3	1
Upf1-TR800AA-HTP	-	endogène	3	0
Upf1-C122S-HTP	-	endogène	2	1
TAP-Upf1	+	surexpression	3	0
TAP-Upf1-DE572AA	-	surexpression	1	0

Table 1: Caractéristiques des souches Upf1 sauvage et mutantes purifiées

J'ai procédé dans un premier temps à une analyse comparative des différents réplicas entre eux par un regroupement hiérarchique (**Figure 22A**). Nous avons observé que les réplicas des différentes conditions de purification sont plus proches entre eux que des autres expériences, ce qui a montré une bonne reproductibilité. On a remarqué une différence nette entre les expériences avec traitement RNase et celles sans le traitement, qui sont séparées dans deux branches bien distinctes de l'arbre. Seul le profil d'interaction du mutant C122S était plus proche des profils de purification avec traitement RNase. Ce résultat est corrélé avec les

expériences précédentes montrant une perte de l'interaction avec l'ARN de ce mutant. Parmi les expériences sans traitement RNase, les purifications des versions surexprimées d'Upf1 étaient groupées dans la même branche, le mutant K436Q avait un profil similaire à ceux des expériences avec Upf1 sauvage et le mutant TR800AA était dans une branche séparée plus proche des profils pour Upf1 sauvage et K436Q que des autres conditions.

Pour permettre une analyse plus précise des purifications, j'ai détaillé les interactions identifiées pour Upf1 sauvage puis je les ai comparées avec celles trouvées dans les conditions mutantes. Les résultats pour Upf1 sauvage (**Figure 22B**) sont représentés sous forme d'un graphique avec en abscisse l'intensité des protéines quantifiées en spectrométrie de masse et en ordonné l'enrichissement de ces protéines par rapport à un extrait total de levure. On a observé, comme attendu, un enrichissement des deux autres facteurs cœurs Upf2 et Upf3 parmi les facteurs les plus enrichis ce qui validait la méthode utilisée pour la spectrométrie de masse et son analyse. En outre, les deux protéines les plus enrichies avec Upf1-TAP ont été Nmd4 et Ebs1. Nmd4 a été décrit précédemment comme interagissant avec Upf1 dans un crible double hybride (He and Jacobson, 1995) puis très peu caractérisé. C'est une petite protéine de 25 KDa environ qui consiste presque uniquement en un domaine PIN. Ce domaine peut avoir une activité de liaisons à l'ARN, d'endonucléase ou être une interface pour des interactions protéine/protéine (Schneider et al., 2009). Ebs1 a été caractérisé en 2007, il est homologue de la protéine Smg7 humaine et pourrait avoir un rôle dans le NMD (Ford et al., 2006) ou dans la régulation de la traduction (Luke et al., 2007). Les protéines Dcp1, Dcp2 ainsi que Edc3, qui font parties du complexe d'hydrolyse de la coiffe ont aussi été retrouvées fortement associées avec Upf1-TAP purifié. Les autres facteurs enrichis étaient pour la plupart des protéines des liaisons aux ARNs. Parmi elles, on a remarqué quelques protéines particulières comme des caséines kinases (le complexe CK2 et la protéine Hrr25), la E3-ubiquitine ligase Ubr1 et le complexe Lsm1-7/Pat1.

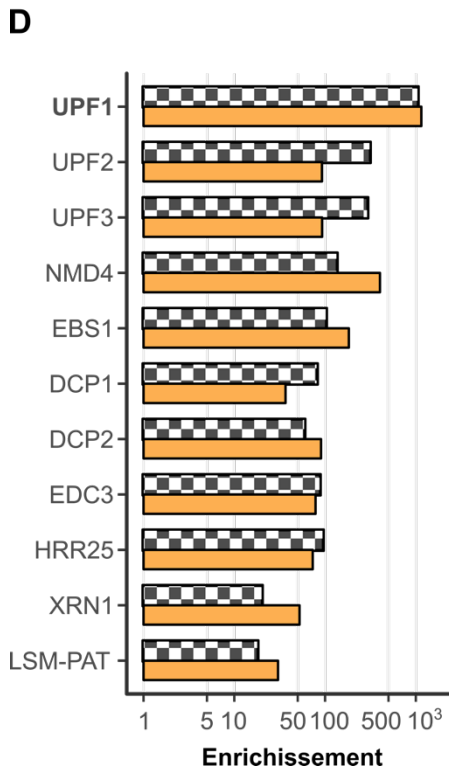
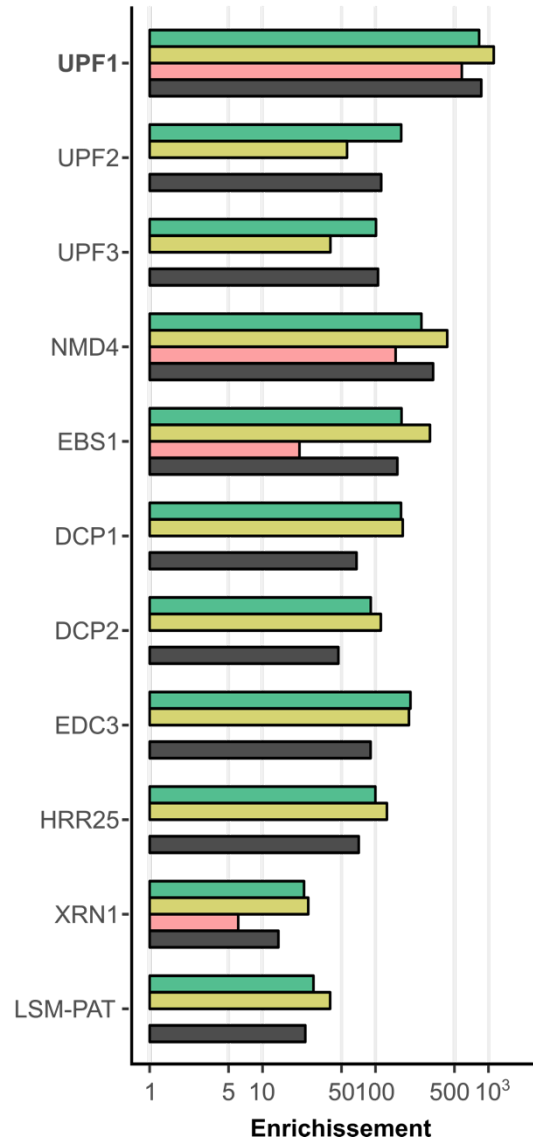
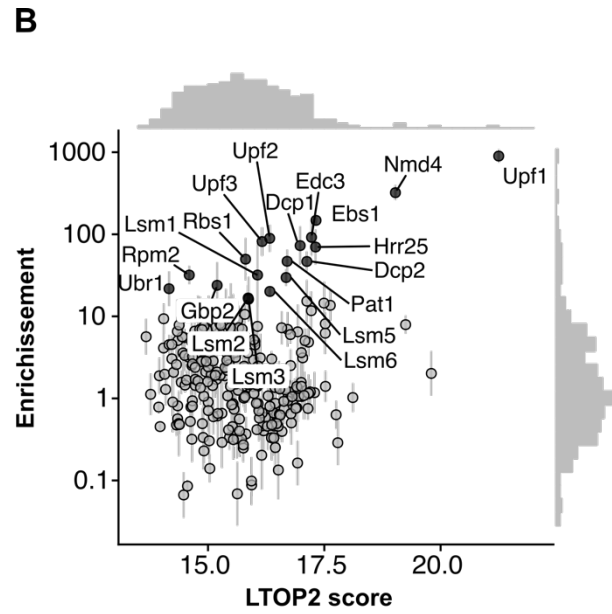
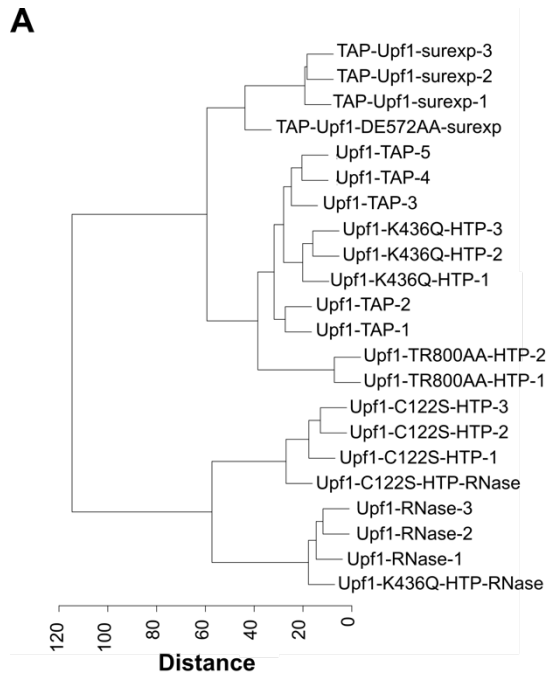


Figure 22 : Upf1 muté au niveau de l'hélicase reste associé à ses partenaires

A. Regroupement hiérarchique des répliques des expériences réalisées pour cette étude sur les mutants d'Upf1. TAP-Upf1-surexp et TAP-Upf1-DE572AA-surexp sont des échantillons dans lesquels la seule version d'Upf1, mutée ou non est exprimée à partir d'un plasmide centromérique sous le contrôle d'un promoteur tetO7 induisant une surexpression de 50 fois environ. J'ai utilisé une transformation de Ward qui minimise la variance entre les échantillons de chaque groupe pour réaliser le regroupement.

B. Combinaison des données d'intensité des protéines quantifiées en spectrométrie de masse (abscisse) par rapport à l'enrichissement des protéines dans la purification par rapport à un extrait total (ordonné). Les axes utilisent une échelle logarithmique. **C.** et **D.** Graphique représentant pour chaque version d'Upf1 l'enrichissement des protéines d'intérêts. L'histogramme **C.** représente Upf1-TAP (gris), Upf1-K436Q (vert clair), Upf1-TR800AA (vert foncé) et Upf1-C122S (saumon). L'histogramme **D.** représente TAP-Upf1 (damier blanc) et le mutant Upf1-DE572AA (orange). La protéine utilisée pour la purification est représentée en gras sur l'axe des ordonnées. L'axe des abscisses utilise une échelle logarithmique.

Pour comparer les enrichissements entre les purifications d'Upf1 sauvage et les mutants, j'ai restreint mon analyse aux protéines les plus enrichies avec Upf1 sauvage identifiées ci-dessus. J'ai regroupé les protéines Lsm1 à 7 et Pat1, sous la dénomination LSM-PAT, leurs profils dans les différentes purifications étaient similaires et elles font partie d'un complexe déjà défini. Les résultats des comparaisons (**Figure 22C et D**) ont montré peu de différences entre Upf1 sauvage et les mutants K436Q, DE572AA et TR800AA pour le groupe de protéines étudiés. Il semble que ces trois mutants continuent d'interagir avec les mêmes protéines que la version sauvage. En revanche, le mutant C122S s'est comporté différemment. Il a perdu toutes ses interactions avec Upf2, Upf3 et la machinerie d'hydrolyse de la coiffe. Les interactions avec les protéines LSM-PAT sont aussi abolies ce qui semble se corrélérer avec la perte de l'interaction aux ARNs de la protéine. Seul l'enrichissement de Nmd4 n'a pas été affecté tandis que l'interaction avec Ebs1 a été perturbée. Ces résultats ciblés sur quelques protéines confirment l'analyse globale des données par clustering montrant que les mutants K436Q et la version WT d'Upf1 ont des profils très similaires au contraire du mutant C122S. Le mutant TR800AA est très similaire à Upf1 sauvage pour les interactions les plus fortes, comme K436Q, cependant une analyse étendue des facteurs sensibles à la mutation en utilisant Gene Ontology (GOterm) montre une perte plus importante des facteurs ribosomiaux dans le mutant TR800AA que K436Q qui suggère un affaiblissement de la liaison à l'ARN pour le mutant TR800AA.

En condition stationnaire, il n'y a pas de différences majeures d'associations des protéines impliquées directement dans le NMD avec Upf1 sauvage ou mutée dans le domaine hélicase ARN. C'est assez surprenant puisque ces mutations abolissent complètement le NMD. La perte de fonction des versions mutantes pourrait alors être due à un défaut dans la dynamique d'action des différents facteurs autour d'Upf1. Une autre possibilité est que l'activité ATPase soit requise pour une étape tardive du mécanisme, par exemple le recyclage de Upf1 (Franks et al., 2010). Cela n'affecterait pas l'interaction avec les autres facteurs mais bloquerait les protéines Upf1 sur un ARN en empêchant leur réutilisation dans un autre événement de NMD. Pour comprendre le comportement de ces mutants, nous manquons de connaissances sur l'enchaînement des étapes conduisant à la dégradation des ARNs après reconnaissance des PTCs et sur le rôle exact de l'activité ATPase dans le mécanisme. L'équipe de Hervé Le Hir, avec qui nous collaborons, travaille sur la caractérisation de la fonction ATPase et hélicase de Upf1. Pendant une deuxième partie de ma thèse, je me suis donc intéressée à la dynamique d'assemblage des complexes autour d'Upf1.

II. Complexes protéiques impliqués dans le NMD et leur dynamique

1. Stratégie utilisée et introduction aux résultats

L'étude par CRAC des complexes ARN/protéines a montré les limites de la technique, qui est particulièrement puissante pour trouver les sites d'interaction d'une protéine ribosomique avec l'ARNr, par exemple, mais s'avère plus difficile quand il s'agit d'une protéine d'abondance moyenne, comme Upf1, qui lie des milliers d'ARNs différents.

De manière étonnante, les mutants de l'hélicase d'Upf1 n'ont pas montré un effet fort sur les interactions avec les partenaires protéiques. Nous avons donc décidé de poursuivre l'étude du NMD avec une méthode exhaustive et systématique : purifier les différents partenaires dans plusieurs conditions afin d'étudier les relations entre eux et leur dynamique.

J'ai résumé **Figure 23**, l'état des connaissances en 2016 pour les interactions d'Upf1, Upf2 et Upf3. Ce réseau a été construit à partir des interactions physiques recensées chez la levure *S. cerevisiae* dans la base de donnée Biogrid (<https://thebiogrid.org/>, Stark et al., 2006). Le trio de facteurs cœur Upf1, Upf2 et Upf3 est bien caractérisé notamment les interactions entre Upf1 et Upf2 et en moindre mesure entre Upf1 et Upf3. Des interactions ont aussi été décrites entre Upf1 et les facteurs de terminaison de la traduction eRF1 et eRF3 (respectivement Sup45 et Sup35) impliqués dans la reconnaissance des cibles dans le modèle du « Faux 3'-

UTR » ; avec la machinerie d'hydrolyse de la coiffe Dcp1 et Dcp2 et avec Ebs1. En plus de ces interactions qui sont supportées par plusieurs expériences, d'autres situées en haut de la figure n'ont pas été reproduites et posent questions. La plupart des facteurs sont reliés à la biologie de l'ARN et pourrait donc avoir un rôle même indirect dans le mécanisme du NMD qui n'est pas connu aujourd'hui. Pour certains, cependant, nous savons maintenant que l'association était certainement indirecte et sans fonction spécifique dans le NMD, c'est le cas de Nmd3 qui est un facteur nucléaire de la biogenèse des ribosomes (Gadal et al., 2001; Ho et al., 2000).

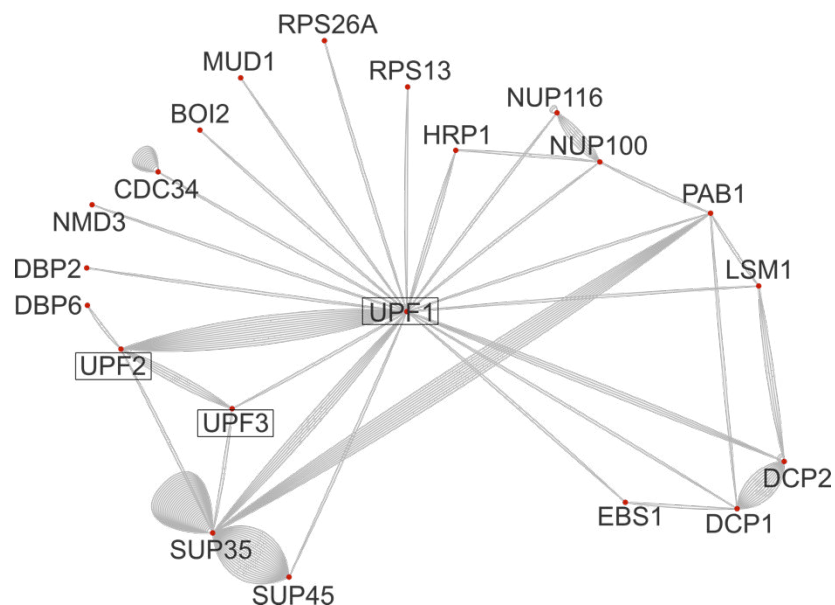


Figure 23 : Réseau des interactions des protéines Upf1, Upf2 et Upf3

Les interactions représentées ici sont celles recensées chez la levure *S. cerevisiae* sur la base de donnée Biogrid en 2016. Cela correspond à 188 interactions provenant de 100 publications. SUP35 et SUP45 sont respectivement eRF3 et eRF1.

Le problème de ce réseau d'interactions est qu'il n'est pas suffisamment précis pour permettre d'identifier les interactions directes des facteurs autour de Upf1-3 et les relations entre les différentes interactions. Par ailleurs, la plupart de ces interactions sont issues de données de cribles en double-hybride chez la levure ou de co-purifications à partir de protéines recombinantes purifiées qui ne reproduisent pas les conditions de l'interaction *in vivo*. De plus, comme vu dans la partie de développement méthodologique, l'utilisation de la spectrométrie de masse pour identifier et quantifier des protéines purifiées n'est pas encore

utilisée en routine et les analyses réalisées jusqu'à présent (pour exemple, Brannan et al., 2016) ne permettaient pas d'identifier sans à priori les facteurs du NMD.

Les questions que nous nous sommes posées à partir de ce réseau d'interactions sont :

- Pouvons-nous identifier de nouveaux facteurs impliqués dans le NMD ?
- Comment les différents facteurs sont coordonnés afin de permettre la reconnaissance des cibles et la dégradation des ARNs ?
- Est-ce-que cette dynamique est conservée à travers les espèces ?

Pour répondre à ces questions nous avons développé une stratégie exploratoire par purification des différents facteurs connus du NMD et analyse croisée de leur profils d'interaction. Nous avons réalisé plus d'une centaine de purifications différentes classées en quatre catégories (**Figure 24**). La purification d'Upf1 est la référence pour nos analyses. Son étude détaillée présentée dans le chapitre précédent et dans l'article de thèse a fait apparaître un paysage d'interactions autour de Upf1 assez différent de celui représenté **Figure 23**. L'importance des facteurs cœurs Upf2 et Upf3 a notamment était modéré par l'enrichissement supérieur d'autres facteurs tel Nmd4 et Ebs1. Certains facteurs comme eRF1 et eRF3 n'ont pas été retrouvés dans nos conditions, tout comme les interactions avec Hrp1, Nmd3 et toutes les protéines situées dans la partie supérieure du réseau (**Figure 23**). À l'inverse, de nouveaux facteurs ont été identifiés comme Ubr1, Gbp2, Hrr25, Lsm1-7 et Pat1. Parmi les quatre catégories de purification réalisée nous avons débuté par celles de tous les partenaires principaux identifiés dans nos purifications de Upf1 (**Figure 24A et Fig. 2 de l'article**). Une grande quantité de protéines liant l'ARN était retrouvée dans toutes nos purifications, nous avons donc réalisés des répliques dans des conditions de traitement à la RNase (**Figure 24B et Fig. 2 de l'article**). Un des objectifs de notre travail était de caractériser plus précisément l'organisation des différents partenaires directs de Upf1 autour de lui et pour cela nous avons réalisé des purifications de fragments de Upf1 (**Figure 24C et Fig. 3 de l'article**). Enfin, nous avons purifiés quelques un des facteurs du NMD dans des conditions mutantes pour certains autres facteurs afin de connaître les liens entre les différents partenaires pour leur liaison à Upf1 (**Figure 24D et Fig. 4 & 6 de l'article**).

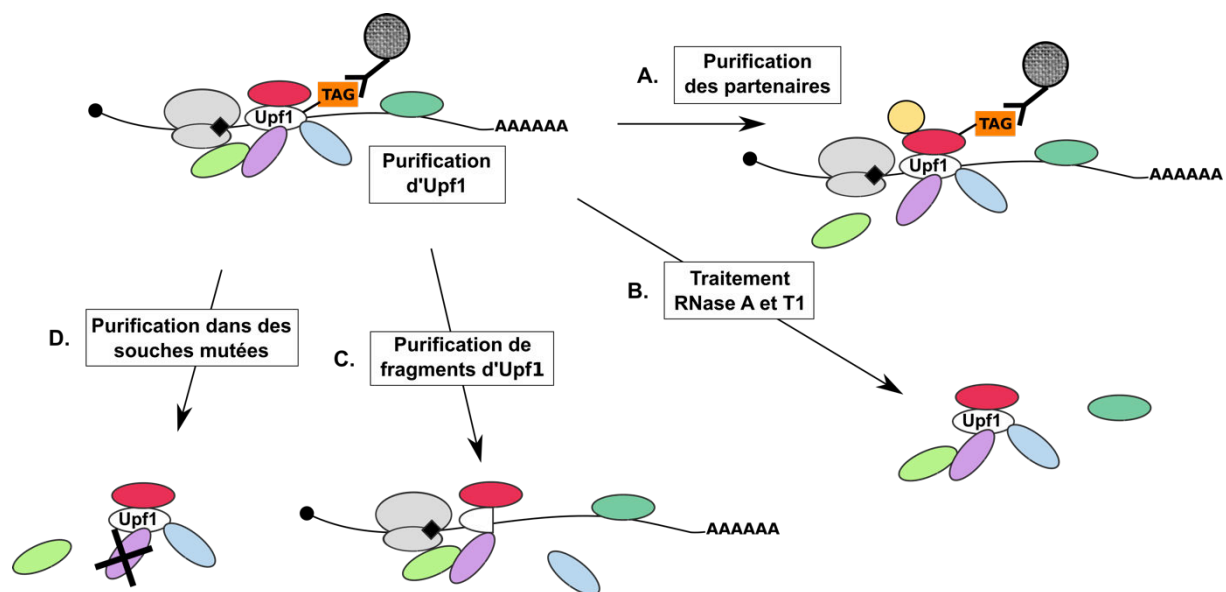


Figure 24 : Stratégie pour la caractérisation des partenaires d’Upf1 par purification d’affinité et spectrométrie de masse quantitative

A. Purification des partenaires principaux identifiés et comparaison des profils d’interaction. **B.** Purifications dans des conditions de traitement astringent des échantillons avec les RNases A et T1 afin d’éliminer les interactions indirectes des facteurs avec la protéine purifiée *via* l’ARN. Le traitement est répété deux fois au cours de la purification pour augmenter son efficacité. **C.** Purification de plusieurs fragments d’Upf1. **D.** Purification de plusieurs facteurs principaux dans des conditions où un autre facteur est délété (ex. Upf1-TAP/*upf2Δ*).

Notre analyse des interactions autour d’Upf1 est probablement une des plus exhaustives qui existe actuellement sur le NMD chez la levure. Les résultats obtenus nous ont permis de montrer l’existence de deux complexes distincts associés à Upf1 (**Figure 25**). Un premier complexe, nommé “Effecteur”, est composé des facteurs Upf1-3 et serait impliqué dans la reconnaissance des cibles du NMD. Le deuxième complexe, nommé “Décteur”, contient Upf1, Nmd4, Ebs1, Dcp1, Dcp2 et Edc3 et permettrait la dégradation rapide des ARNs par hydrolyse rapide de la coiffe en 5’ et grignotage de l’ARN par Xrn1. Ces deux complexes agissent successivement au niveau des ARNs cibles du NMD (**Figure 25**).

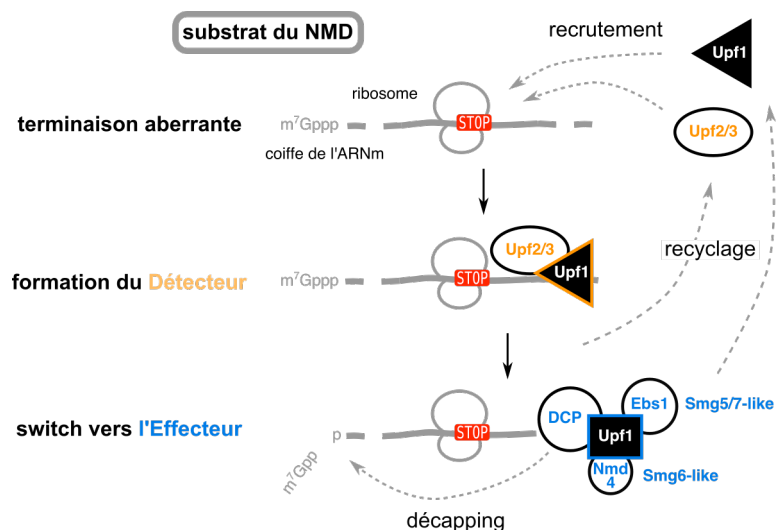


Figure 25 : Modèle proposé pour le mécanisme du NMD EJC-indépendant chez la levure *S. cerevisiae*.

Cette étude a aussi permis de mieux caractériser les rôles de Ebs1 et Nmd4 dans le mécanisme du NMD de levure. En effet, nous avons montré un effet additif des délétions de *NMD4* et *EBS1* sur la dégradation d'un ARNs rapporteur du NMD ainsi qu'au niveau du transcriptome. Nous avons aussi trouvé des conditions dans lesquels *NMD4* et *EBS1* deviennent essentiels pour le NMD. Nous avons précisé les régions d'interactions de ces deux protéines au niveau d'Upf1. Nmd4 interagit directement au niveau de la région hélicase (208-853) et l'interaction avec Ebs1 est très stabilisée par l'extrémité C-terminale (853-971). L'interaction de Nmd4 au niveau de l'hélicase pourrait avoir un rôle sur l'association à l'ARN de Upf1 comme le suggère des expériences de co-précipitation d'Upf1 *in vitro* par l'intermédiaire d'un ARN synthétique et montrant une plus forte association en présence de Nmd4 (**Fig. 4E**).

Les données complémentaires sur Ebs1 présentées dans notre article ont permis de renforcer l'hypothèse de l'homologie entre Ebs1 et la protéine Smg7 impliquée dans le NMD des organismes multicellulaires. Comme pour Smg7, Ebs1 serait impliqué dans la phase de dégradation des ARNs ciblés par le NMD *via* une interaction au niveau de la région C-terminale d'Upf1 non conservée. Pour Nmd4, l'homologie de séquence entre son domaine PIN et ceux de Smg5-6 avait déjà été discutée auparavant (Clissold and Ponting, 2000). Nous montrons en plus que Nmd4 pourrait être un homologue fonctionnel de Smg6. En effet, Nmd4 interagit directement avec Upf1 au niveau de la région hélicase qui correspond à un domaine de liaison de Smg6 à Upf1 humain (Chakrabarti et al., 2014). De plus, il intervient au niveau des étapes effectrices de dégradation des ARNs tout comme Smg6.

Les homologies fonctionnelles entre les facteurs accessoires Nmd4 et Ebs1 et des facteurs présents chez les eucaryotes multicellulaires suggèrent une conservation du mécanisme que nous proposons, que nous discutons dans l'article.

2. Article de thèse

Nonsense-mediated mRNA Decay Involves Distinct and Successive Upf1-bound Complexes

Marine Dehecq^{1,2}, Laurence Decourty¹, Abdelkader Namane¹, Caroline Proux³, Joanne Kanaan⁴, Hervé Le Hir⁴, Alain Jacquier¹, Cosmin Saveanu^{1*}

1 – Génétique des Interactions Macromoléculaires, Genomes and Genetics Department, Institut Pasteur, 25-28 rue du docteur Roux 75015 Paris, France

2 – Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France

3 – Transcriptome and Epigenome, CITECH, Institut Pasteur, 25-28 rue du docteur Roux 75015 Paris, France

4 – Expression des ARN messagers eucaryotes, CNRS UMR8197, Inserm U1024, Biology department, Institut de Biologie de l'Ecole Normale Supérieure, 46 rue d'Ulm 75005 Paris, France

* – Corresponding author

Abstract

Nonsense-mediated mRNA decay (NMD) is a translation-dependent RNA degradation pathway involved in many cellular pathways and crucial for telomere maintenance and embryo development. Core NMD factors Upf1, Upf2 and Upf3 are conserved from yeast to mammals, but a universal NMD model is lacking. We used affinity purification coupled with mass spectrometry and an improved data analysis protocol to characterize the composition and dynamics of yeast NMD complexes in yeast (112 experiments). Unexpectedly, we identified two distinct complexes associated with Upf1: *Detector* (Upf1/2/3) and *Effector*. *Effector* contained the mRNA decapping enzyme, together with Nmd4 and Ebs1, two proteins that globally affected NMD and were critical for RNA degradation mediated by the Upf1 C-terminal helicase region. The fact that Nmd4 association to RNA was dependent on *Detector* components and the similarity between Nmd4/Ebs1 and mammalian Smg5-7 proteins suggest that NMD operates through successive Upf1-bound *Detector* and *Effector* complexes in other species. This model can be extended to accommodate steps that are missing in yeast, and thus serve for mechanistic studies of NMD in all eukaryotes.

Running title: The *Detector/Effector* NMD model

Keywords: Affinity purification / NMD / Quantitative mass-spectrometry / RNA decay / *Saccharomyces cerevisiae*

Introduction

Nonsense-mediated mRNA decay (NMD) is a major surveillance pathway defined by its importance for the fast decay of mRNAs with premature termination codons (PTC). Rapid translation-dependent RNA degradation prevents the production of truncated proteins (Hall & Thein, 1994) and reduces the loss of ribosomes “locked” on RNA at aberrant termination sites (Ghosh *et al*, 2010; Serdar *et al*, 2016). NMD factors are essential for the development of metazoans (Medghalchi *et al*, 2001) and participate to telomere maintenance and telomerase activity regulation in yeast (Addinall *et al*, 2011) and human cells (Azzalin *et al*, 2007). As a quality control pathway, NMD shapes the transcriptome of eukaryotic cells, both under normal and pathologic conditions, including cancer (Lindeboom *et al*, 2016). While NMD substrates were initially thought to be rather rare, single-nucleotide resolution RNA sequencing in yeast revealed that genome-wide transcription generates thousands of RNAs that are degraded by NMD (Malabat *et al*, 2015).

Large-scale results confirmed and validated previous studies on individual NMD reporters that described two classes of NMD substrates: one in which a long 3' untranslated region (UTR) extension follows a short coding sequence and another that includes an exon-exon junction, bound by the exon-exon junction complex (EJC), downstream a stop codon. While long 3'-UTRs destabilize RNAs in all the eukaryotes studied to date, EJC-enhanced NMD is absent from *S. cerevisiae*, *S. pombe* (Wen & Brogna, 2010) or *C. elegans* (Gatfield *et al*, 2003). Most of mRNAs physiologically regulated by NMD belong to the long 3'-UTR category (Yepiskoposyan *et al*, 2011) but many mRNAs generated by point mutation in neoplastic cells belong to the EJC class (Lindeboom *et al*, 2016). Degradation of both types of NMD substrates requires three universally conserved “core” factors: Upf1, Upf2 and Upf3 (Leeds *et al*, 1992, 1991; Cui *et al*, 1995; Pulak & Anderson, 1993; Perlick *et al*, 1996). Independent of the initial triggering feature, RNA degradation in NMD occurs through activation of general RNA degradation pathways: decapping and endonucleolytic cleavage. Final degradation of RNA depends on the major 5' to 3' exonuclease Xrn1 (He & Jacobson, 2001) and the cytoplasmic exosome 3' to 5' exonucleolytic activity (Eberle *et al*, 2009).

How the NMD machinery differentiates normal termination codons from PTC has been a long-standing question (He & Jacobson, 2015). For long 3'-UTR NMD substrates, the prevalent model proposes that detection of a stop codon as premature occurs when the stop is far from the poly(A) tail ('faux 3'-UTR' model, Amrani *et al*, 2004). In this model, the polyA-binding protein, Pab1 in yeast and PABPC1 in mammals, affects translation termination efficiency. Slow termination allows recruitment of Upf1, Upf2, Upf3, which leads to rapid decapping and degradation of NMD substrates (Muhlrad & Parker, 1994, 1999). A long distance from the PTC to PABPC1 triggers NMD in all studied species (Bühler *et al*, 2006; Eberle *et al*, 2008; Singh *et al*, 2008; Huang *et al*, 2011).

EJC-enhanced NMD does not require a long 3'-UTR and depends on an exon-exon junction located at more than 50 nucleotides downstream the PTC (Holbrook *et al*, 2004; Maquat, 2004). EJC positioning is an important element of the prevalent model for the first steps in

mammalian NMD, called SURF/DECID. This model involves the formation of an initial complex that contains a ribosome stalled at a termination codon, the translation termination factors eRF1 and eRF3, Upf1 and the protein kinase Smg1 (SURF). Recruitment of the Upf2 and Upf3 proteins to SURF *via* a proximal EJC, leads to formation of the DECID complex in which Upf1 N and C-terminal regions are phosphorylated by Smg1 (Kashima *et al*, 2006). Phosphorylated Upf1 binds the Smg6 endonuclease and the Smg5-7 heterodimer, which, in turn, activates RNA deadenylation and decapping. Recruitment of a phosphatase ensures the return of Upf1 to its initial state for another NMD cycle (Ohnishi *et al*, 2003).

While the SURF/DECID model offers an explanation for EJC-enhanced NMD, it does not apply to organisms that do not have EJC components, such as the budding yeast, and does not explain how NMD works on RNAs with long 3'-UTRs. Understanding how such substrates are degraded through NMD is relevant for mammalian cells since several physiologically important NMD substrates do not depend on an EJC downstream the termination codon. A salient example is the mRNA of the GADD45A gene (Chapin *et al*, 2014; Lykke-Andersen *et al*, 2014), a transcript whose destabilization through NMD is essential for the development of fly and mammal embryos (Nelson *et al*, 2016). It is possible that EJC-enhanced and EJC-independent NMD are two versions of the same, not yet understood, generic NMD mechanisms. Such a mechanism is unlikely to rely on a SURF/DECID-like sequence of events for several reasons. First, the SURF/DECID model proposes crucial roles for factors or events that are not conserved in all eukaryotes, even if the key NMD proteins are present from yeast to humans. For example, the protein kinase Smg1, a central component of the SURF complex, has no known equivalent in *S. cerevisiae* and its absence has little impact on NMD in *D. melanogaster* embryos (Chen *et al*, 2005). Second, a major assumption of the SURF/DECID model, the role of EJC in bringing Upf3 in the proximity of Upf1 (Le Hir *et al*, 2001; Chamieh *et al*, 2008), was challenged by very recent biochemical analyses that show interactions of Upf3 with translation termination complexes independent of EJC components (Neu-Yilik *et al*, 2017). Third, while all the RNA degradation mechanisms can play a role in NMD, the model leaves open the question of the mechanisms of recruitment for the RNA degradation factors. Major roles have been attributed to the decapping machinery (Muhlrad & Parker, 1994; Lykke-Andersen, 2002), endonucleolytic cleavage (Huntzinger *et al*, 2008; Eberle *et al*, 2009) or deadenylation (Loh *et al*, 2013), depending on the reporter RNA used. The molecular mechanisms responsible for the redundancy and interplay between these different degradation pathways (Metze *et al*, 2013; Colombo *et al*, 2016) remain unclear. Hence, a re-evaluation of the available data and new biochemical results on the involved complexes are critical to uncover a universally conserved mechanism of NMD in eukaryotes. Such a mechanism could then serve as the basis to understand the evolution of NMD, its variations across species and additional levels of complexity.

The enumerated weaknesses of current NMD models originate, at least in part, from the lack of a global view of the biochemistry of the NMD process. We show here that fast affinity purification of yeast NMD complexes coupled with high-resolution quantitative

mass spectrometry (LC-MS/MS) allows a global and unprecedented view of NMD complexes. Our results uncover the existence of two distinct and successive protein complexes that both contain the essential NMD factor Upf1. We identified two yeast accessory NMD factors that interact with Upf1 and become essential for NMD under limiting conditions. Their similarity with Smg6 and Smg5/7 implies that comparable molecular mechanisms for NMD have been conserved from yeast to humans.

Furthermore, our results suggest that Upf1 might suffer major conformation changes to accommodate a switch from a “Detector” complex to an “Effector” complex, that triggers RNA decapping, the first step in NMD substrates degradation.

Results

Quantitative mass spectrometry reveals the global composition of Upf1 associated complexes

Previous large-scale attempts to describe the composition of Upf1 associated complexes in yeast failed to identify known Upf1 partners, such as Upf2 or Upf3 (Gavin *et al*, 2006; Krogan *et al*, 2006; Collins *et al*, 2007) even if these studies provided high-quality interaction data on other proteins. This negative result could be due to the relatively low abundance of Upf2 and Upf3, estimated to 1470 and 1439 molecules per cell (Ho *et al*, 2018), which places them in the category of the 25% least abundant proteins in yeast (**Fig. 1A**). The instability of the complexes during the purification procedure could also play a role. To overcome these problems we used a combination of fast affinity purification (Oeffinger *et al*, 2007) with the current high sensitivity of mass-spectrometry methods for protein identification. TAP-tagged (Rigaut *et al*, 1999) Upf1 was isolated from cells extracts obtained from 6L of yeast culture grown in exponential phase. The total cell extract was used as a reference and an untagged strain as a negative control. Four hundred proteins were reliably identified in the total extract and corresponded, as expected, to abundant yeast proteins (the frequency distribution of the abundance for the identified proteins is shown as dark grey bars in **Fig. 1B**). Many abundant proteins were also identified in the Upf1-TAP purified fraction (distribution in dark grey, **Fig. 1C**). In contrast to the input fraction, less abundant proteins were also identified in the purified fraction (distribution in light grey, **Fig. 1C**).

To assess to what extent the identified proteins were specific to the Upf1 purification, we used a label-free quantification method based on the intensity of the most abundant peptides for each of the identified proteins (Silva *et al*, 2006), a method that accurately estimates protein levels (Ahrné *et al*, 2013). For each identified protein, we derived a peptide intensity-based score that we called LTOP2 - “L”, because it includes a log₂ transformation, and “TOP2” for the minimum number of required peptides (see **Materials and Methods**). The LTOP2 score of proteins identified in the total protein extract was well correlated with protein abundance estimates (Ho *et al*, 2018), with a Pearson determination coefficient of 0.82 (**Fig. 1D**).

Upf1 had the highest LTOP2 score in the Upf1-TAP purification, as expected, validating both the use of this MS score for protein quantitation and the efficiency of the affinity purification. Most of the other proteins with high LTOP2 scores in the purification, such as the plasma membrane proton pump Pma1, or the glyceraldehyde-3-phosphate dehydrogenase Tdh3 (**Fig. 1E**), were unrelated to Upf1 or NMD, but are among the 10% most abundant proteins in yeast. NMD factors Upf2 and Upf3, the best-known partners of Upf1, ranked low in the list of 270 LTOP2 scores, with position 80 and 91 respectively (**Dataset EV1**). Thus, many of the proteins present in the affinity purified sample were contaminants, known to pollute affinity purifications and complicate the conclusions of such experiments (Mellacheruvu *et al*, 2013). To filter this type of contamination, we chose to calculate an enrichment score, which reflects the ratio between the proportion of a protein in a sample and its proportion in a total extract

(Ho *et al.*, 2018). Top enrichment scores corresponded to the tagged protein Upf1, Upf2 (rank 5) and Upf3 (rank 6), validating enrichment as an excellent noise removal strategy in the identification of specific partners in affinity-purified samples (**Fig. 1F**, and **Dataset EV1**). To simultaneously visualize both the amount of protein found in a purified sample and its enrichment score, we combined LTOP2 and enrichment in a graphical representation that is similar to the intensity versus fold change “MA” plots used for the analysis of differential gene expression (**Fig. 1G**). The workflow for mass-spectrometry data analysis is presented in **Appendix Fig. S1**.

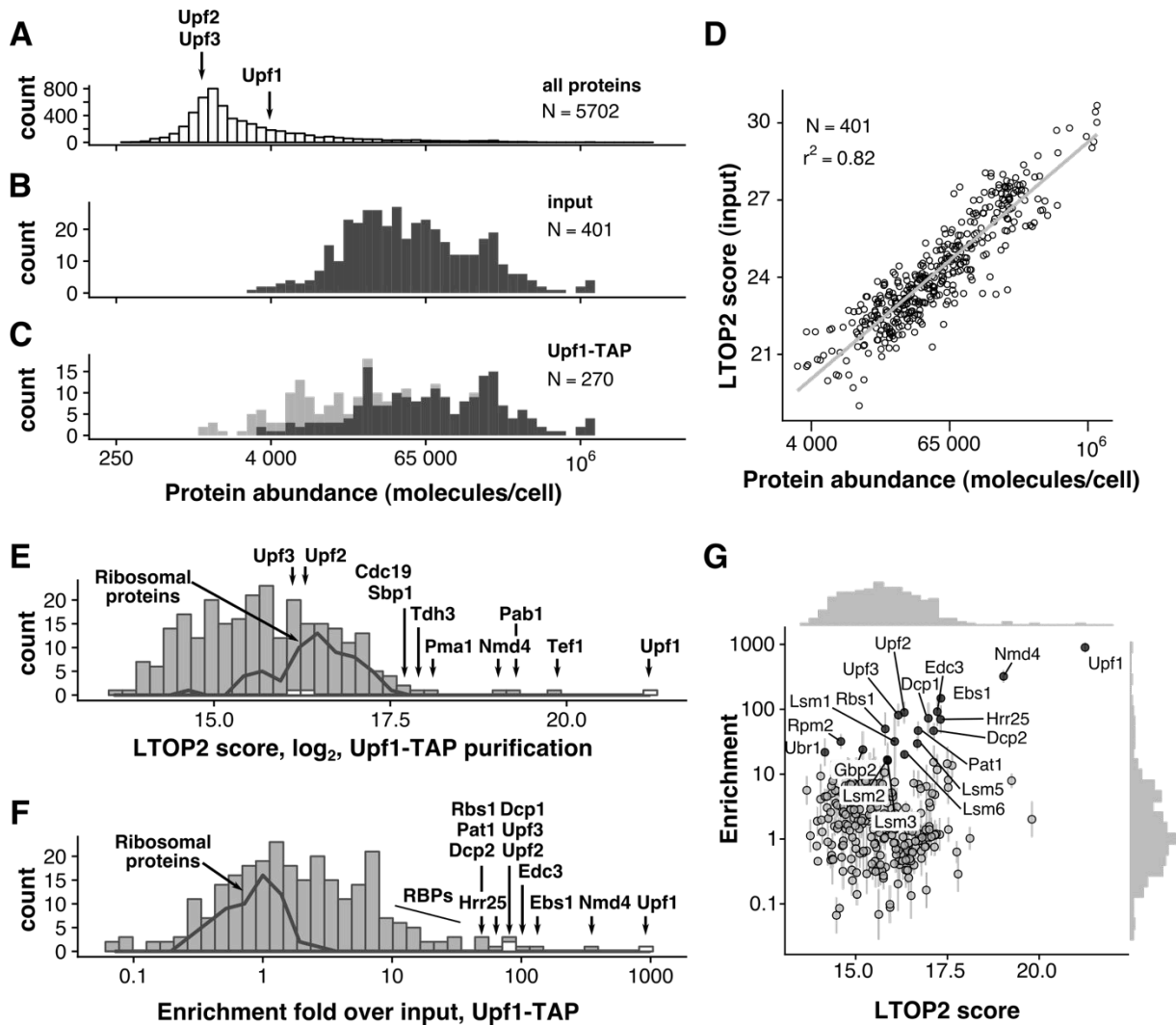


Fig. 1: Enrichment analysis accurately describes Upf1-associated proteins.

A. Positioning of Upf1, Upf2 and Upf3 in the abundance distribution of all yeast proteins, based on the data compiled by Ho *et al.*, 2018. The distribution of abundance for the proteins identified by mass spectrometry in input (**B**), and purified samples (**C**), show the expected bias for proteins present in high copy numbers. Light grey colour highlights the positions of proteins that were only identified in the Upf1-TAP samples (6 replicates) and not in the total extract (3 replicates). **D.** Correlation of LTOP2-based protein level estimates in a total yeast protein extract with published protein abundance data (Ho *et al.*, 2018). **E.** Distribution of the intensity for the proteins quantified in association with Upf1 (LTOP2 score, \log_2 scale). Positioning of Upf1, Upf2 and Upf3 are highlighted as white

rectangles. The dark grey line indicates the region of intensities corresponding to ribosomal proteins. **F.** Distribution of the protein enrichment values for Upf1 associated proteins based on LTOP2 scores and known protein abundance in a total yeast extract. Upf1, Upf2 and Upf3 positions are also highlighted as white rectangles. **G.** A combination of the data presented in (E) and (F) as a scatter plot, to see both the amount and enrichment of proteins in Upf1-TAP purified samples. The horizontal axis represents the LTOP2 score, with the vertical axis showing enrichment over input values. Both axes use \log_2 transformed values.

In addition to the canonical Upf1-Upf2-Upf3 interactions (He *et al.*, 1996, 1997), our enrichment strategy identified the RNA decapping enzyme Dcp2, its cofactor Dcp1, and decapping activators, Edc3, Lsm1-7, and Pat1 as Upf1 partners. This result was correlated with the requirement for decapping for the degradation of NMD substrates in yeast (Muhlrad & Parker, 1994). Direct or indirect interactions of Upf1 with these different factors have been previously reported, mostly through yeast two-hybrid, or affinity purification coupled with immunoblots (He & Jacobson, 2015). Unexpectedly, two poorly characterized proteins, Nmd4 and Ebs1 ranked second and third in the Upf1 purification. Both proteins were previously linked with NMD. Nmd4 was identified in yeast Upf1 two-hybrid screen but was not further studied (He & Jacobson, 1995). Ebs1, identified through sequence similarity as a potential yeast equivalent of human Smg7, a component of mammalian NMD, was shown to co-purify with Upf1 and to affect yeast NMD (Luke *et al.*, 2007). Other proteins enriched in the Upf1-associated fraction by a factor of 10 or more (Rbs1, Hrr25, Gbp2, Gar1, Cbf5, Sbp1, Ssd1, Lsm12, Lhp1 and Rbg1) have functions related with RNA metabolism and could be indirectly associated with the Upf1 RNA-containing complexes. None of the proteins found enriched with Upf1 by a factor of 8 or higher were detected in the control purification (**Dataset EV1**). These results indicated that Upf1 has more protein partners than previously thought and validated quantitative mass-spectrometry of purified complexes as a powerful strategy to characterize yeast NMD complexes.

Upf1-associated proteins are grouped in two mutually exclusive complexes.

Since our data indicated that Upf1 is associated with a relatively large number of factors, and since Upf1 directly binds RNA (Weng *et al.*, 1998) in large polysomal complexes (Atkin *et al.*, 1997), some identified interactions could be mediated by RNA. To establish which interactions were RNA dependent, we repeated the Upf1-TAP purification, in a protocol including two RNase A and RNase T1 treatments. Ribosome constituents and RNA binding proteins were lost following this harsh RNase treatment (p-values associated with GO term enrichment analysis between 10^{-7} and 10^{-30} for the corresponding terms, **Appendix Table S1**). RNase treatment was efficient, suggesting that the proteins that remained specifically enriched were bound to the Upf1 complex through direct protein-protein interactions. Among these proteins, we identified the decapping enzyme Dcp2 and its co-factors Dcp1 and Edc3, Nmd4, Ebs1, and the protein kinase Hrr25, a homologue of mammalian casein kinase delta, CSNK1D (**Fig. 2A**). Upf2 and Upf3 interactions with Upf1-TAP were sensitive to RNase. In

view of these results, and to obtain a global view of NMD complexes in yeast, we affinity purified TAP-tagged versions of the components of Upf1-associated complexes identified here with or without RNase treatment: Upf2, Upf3, Nmd4, Ebs1, Dcp1, and Hrr25 (**Dataset EV1**). All the used tagged strains showed close to wild-type levels of the RPL28 unspliced pre-mRNA, a well-known NMD substrate in yeast (**Fig. EV1, A and B**), indicating that the TAP tag had a minimal impact on the function of the corresponding proteins.

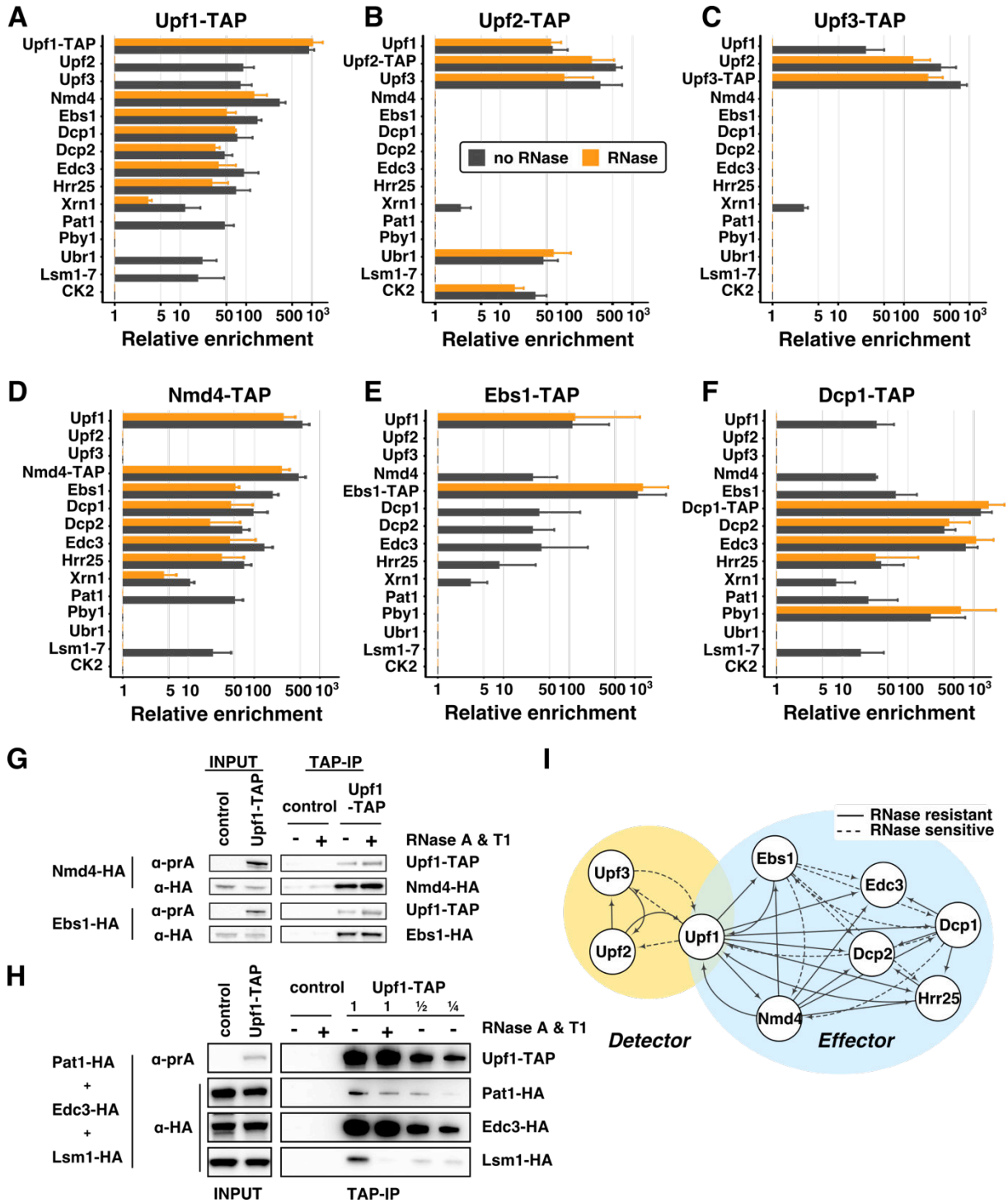


Fig. 2: Purification of NMD factors reveals two distinct complexes containing Upf1.

Enrichment values for proteins identified in tagged Upf1 (**A**), Upf2 (**B**), Upf3 (**C**), Nmd4 (**D**), Ebs1

(E), and Dcp1 (F) purifications. Tagged proteins are indicated in bold for each experiment. Black bars correspond to enrichment values obtained in purifications done without RNase and orange bars with an RNase A and RNase T1 treatment. Only proteins enriched by a factor of 16 or more in one of the purifications presented here are shown. For clarity, groups of related proteins were combined (the group of Lsm1 to Lsm7 is marked as “Lsm1-7” and the values correspond to the mean of all the values found in purifications; CK2 group corresponds to Cka1, Cka2, Ckb1 and Ckb2). **G.** Immunoblot validation of Upf1-TAP interactions with HA tagged Nmd4 and Ebs1. Control strains did not express Upf1-TAP. **H.** Immunoblot validation of Upf1-TAP interaction with Pat1-HA, Lsm1-HA and Edc3-HA. The purification was performed with a mix of three strains, expressing Upf1-TAP or not (control). **I.** Representation of binary interactions identified by our experiments: dashed lines correspond to interactions that are sensitive to RNase treatment; the arrows start at the tagged protein and indicate the enriched factor.

One of the most surprising results of the extensive purification data was the difference between the sets of proteins enriched with Upf2-TAP (**Fig. 2B**) or Upf3-TAP (**Fig. 2C**) in comparison with Upf1-TAP (**Fig. 1G** and **2A**). Many of the proteins present in the Upf1 complex, such as the decapping factors, Ebs1, and Nmd4 were absent in the Upf2-TAP and Upf3-TAP purifications. Conversely, the purifications done with Nmd4-TAP, Ebs1-TAP, Dcp1-TAP or Hrr25-TAP, did not co-purify Upf2 or Upf3, even if Upf1 was specifically enriched with or without RNase treatment (**Fig. 2D, E, F** and **Dataset EV1**). Nmd4-TAP shared the same purification profile with Upf1-TAP except for Upf2 and Upf3 (**Fig. 2, A** compared with **D**). Ebs1-TAP, also purified the same set of proteins, but the enrichment of most of the factors, except Upf1, was dependent on RNA (**Fig. 2E**). Dcp1 co-purified with its known partners Dcp2 and Edc3, but also with Upf1, Nmd4, Ebs1, and Hrr25 (**Fig. 2F**). In addition to Upf1-associated proteins, purifications with Dcp1-TAP and Hrr25-TAP revealed their presence in other protein complexes. For example, Dcp1 purification also showed the enrichment of another decapping-associated factor, Pby1 (**Fig. 2F**). The results, including the variability of biological replicates, are presented in **Dataset EV1** and can be explored at hub05.hosting.pasteur.fr/NMD_complexes.

To confirm the mass-spectrometry quantitative results by a different detection method, we purified Upf1-TAP in strains in which Nmd4, Ebs1, Pat1, Edc3 and Lsm1 were tagged with 3 repeats of the HA epitope at the C-termini. Immunoblots on purified fractions showed that Nmd4-HA and Ebs1-HA were enriched with Upf1 with or without RNase treatment, as expected (**Fig. 2G**). RNase treatment reduced the amounts of Pat1-HA co-purified with Upf1 by a factor of two and reduced the amounts of Lsm1-HA more than four fold, while Edc3-HA was not affected, in an experiment in which a mix of the three strains was used as the input fraction (**Fig. 2H**).

The obtained results suggest that Upf1 is part of two mutually exclusive complexes, one composed of Upf1, Upf2 and Upf3 and the other containing Upf1, Nmd4, Ebs1 and decapping factors (**Fig. 2I**). Since the second complex contains the decapping machinery, known to trigger the degradation of NMD substrates (Muhlrad & Parker, 1994), we refer to this complex as the “**Effector**”. It is likely that the action of Upf2 and Upf3 precedes decapping

and ensures specificity for NMD substrates, so we decided to call the Upf1-Upf2-Upf3 complex, the “**Detector**”. Altogether, these data suggest that Upf1 coordinates the recognition and degradation of NMD by binding to two mutually exclusive sets of proteins.

The Upf1 cysteine-histidine rich N-terminal domain is an interaction « hub » *in vivo*

To find out how the NMD complexes are organised and to investigate why the Upf1-binding proteins in Detector and Effector were mutually exclusive, we tested the importance of Upf1 domains in the interaction with the associated factors. We built plasmids for the expression of tagged Upf1 fragments (**Fig. 3A**) to test the cysteine-histidine rich N-terminal domain (Upf1-CH, 2-208), the C-terminal helicase domain (Upf1-HD-Cter, 208-971), as well as variants of these constructs, comprising either the helicase domain alone (Upf1-HD, 208-853) or a truncated Upf1 lacking the C-terminal extension (Upf1-CH-HD, 2-853). Strains deleted for endogenous *UPF1* were transformed with plasmids expressing N-terminal TAP-tagged full length Upf1 (Upf1-FL) or the various fragments. TAP-Upf1-FL was expressed to levels ten to fifty folds higher than the C-terminal tagged Upf1 expressed from the chromosomal locus, as estimated by immunoblots on total protein cell extracts. The various fragments were also stably expressed to high levels at the expected size (**Fig. EV2A**).

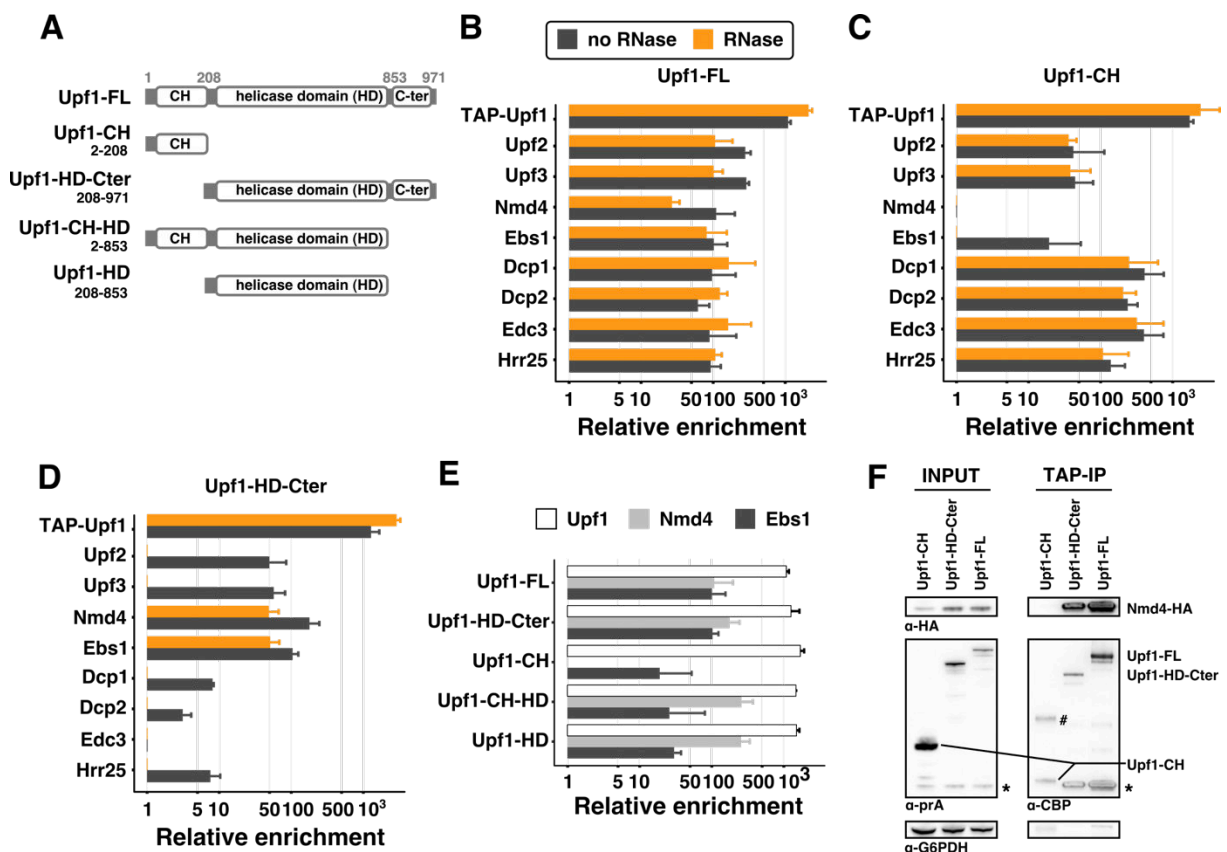


Fig. 3: Nmd4 and Ebs1 are the only Detector components that interact with Upf1 independent of the N-terminal CH domain.

A. Upf1 fragments tested: FL is for the full-length protein; HD-Cter for the region containing the helicase domain and the C-terminal part of Upf1 (208-971); CH for the N-terminal domain (2-208) that contains the N-terminal unstructured region and the Cysteine-Histidine rich domain; CH-HD for a version a full-length Upf1 lacking the C-terminal region 854-971; and HD for the helicase domain of Upf1 alone (not to scale). N-terminal TAP-tagged Upf1 versions were expressed from a single copy vector under an inducible tetO7 promoter. **B. to D.** Results of purifications using Upf1-FL (**B**), Upf1-CH (**C**) and Upf1-Cter (**D**) as bait with the x-axis showing the enrichment value (\log_2 scale) for Effector and Detector components. The colour of the bars illustrates the treatment used during the purification, black without RNase, and orange with an RNase treatment. **E.** Comparison of the enrichment of Upf1, Nmd4 and Ebs1 in the purification of the Upf1 fragments. The Upf1 C-terminal region (854-971) affected the association with Nmd4 or Ebs1 (Student t-test with a one-sided alternative hypothesis). We compared the enrichment of each of Ebs1 and Nmd4 between purifications of Upf1 fragments having this region (Upf1-FL and Upf1-HD-Cter, 6 experiments) and purifications of Upf1 fragments lacking the extension (Upf1-CH-HD and Upf1-HD, 4 experiments). The C-terminal region had no effect on Nmd4 enrichment (p -value ≈ 0.98), whereas there was significantly less associated Ebs1 on this small region (p -value ≈ 0.0013). **F.** Western blot of input and eluates of Upf1 domains purification in a Nmd4-HA strain. The band with the # might corresponds to a dimer of Upf1-CH, bands marked with a star correspond to residual signal with the anti-HA antibodies (Nmd4). Fragments in the eluate have a smaller size because the protein A part of the tag was removed by digestion with the TEV protease. G6PDH served as a loading control in the input samples.

Enrichment of the various proteins in the purification done with overexpressed TAP-Upf1-FL was well correlated with the enrichment of the same proteins in the previous purifications done with the protein tagged at its C-terminus and expressed from the endogenous *UPF1* locus (**Fig. EV2B** and **Dataset EV1**), confirming that neither overexpression nor tag position had major effects on Detector and Effector composition (see **Fig. 2A** compared with **Fig. 3B**). The only marked difference between the results with the overexpressed N-terminal Upf1 and the C-terminal tagged Upf1 was the persistence of Upf2 and Upf3 among the enriched proteins after RNase treatment (**Fig. 3B**, orange bars). Since the Upf1 interaction was also preserved when the Detector complex was purified using tagged Upf2 (**Fig. 2B**), but not when using tagged Upf3 (**Fig. 2C**), this variability in Detector sensitivity to RNase indicates the importance of RNA for the stability of the Upf1-Upf2-Upf3 complex. In contrast, the interactions of Upf1 with the Effector complex proteins, and especially with Nmd4, were insensitive to RNase in all the studied situations.

Tagged Upf1-CH, recovered a majority of the proteins associated with the full-length protein, with the notable exception of Nmd4 (not detected) and of Ebs1, which was detected only in the sample not treated with RNase, albeit to a lower intensity than in the full-length Upf1 purification (**Fig. 3B, C**). The observation that the other proteins of the Effector and Detector complexes were highly enriched in this purification and the fact that the interactions were not sensitive to RNase (**Fig. 3C**), raise the possibility that Upf2 and Upf3 compete with the decapping factors for binding to Upf1-CH domain. Thus, the CH domain seems to be crucial for both Detector and Effector organisation.

Nmd4 and Ebs1 are tightly associated with Upf1-HD-Cter

Nmd4 and Ebs1 were the only proteins interacting specifically with the HD-Cter domain of Upf1 independently of RNA (**Fig. 3D**). To establish if the C-ter domain of Upf1 could play a role in these interactions, we purified Upf1-CH-HD and Upf1-HD and evaluated the relative enrichment levels for Ebs1 and Nmd4 in these conditions. While Nmd4 was recovered with both Upf1 fragments, Ebs1 levels were substantially lower in the absence of the C-terminal extension (**Fig. 3E**, t-test p-value 0.0013, n=4). These results suggest that the C-terminal extension of Upf1 has a crucial effect on binding Ebs1, and does not affect binding of Nmd4.

To confirm the observed strong Nmd4-Upf1 interactions, we built Nmd4-HA strains and tested the presence of the tagged protein in fractions enriched with overexpressed tagged Upf1-FL, Upf1-CH and Upf1-HD-Cter. The purification of Upf1-FL and Upf1-HD-Cter co-enriched Nmd4-HA, while the Upf1-CH region alone did not. Unexpectedly, overexpression of Upf1-FL or Upf1-HD-Cter also led to increased levels of Nmd4-HA in the total extracts (**Fig. 3F**), suggesting that Nmd4-HA protein stability can be affected by its interaction with Upf1. To test the importance of the observed interaction between Nmd4 and Upf1 for the association with other factors of the Effector complex, we purified Upf1-TAP in the absence of *NMD4*. We observed no major effects on the various associated proteins (**Fig. 4A**, **Appendix Fig. S2** and **Dataset EV1**). In the reciprocal experiment, in which we purified the Nmd4-associated complexes in the absence of *UPF1*, all the specific components of the Effector complex were lost (**Fig. 4B**).

These results, together with the strong and RNase insensitive enrichment of Nmd4 in Upf1 complexes, suggested that Upf1 directly interacts with Nmd4. To validate this hypothesis, we expressed CBP-Nmd4 as a recombinant protein in *E. coli* and tested its association *in vitro* with recombinant purified Upf1 helicase domains of yeast and human origin (yUpf1-HD 220-851 and hUpf1-HD 295-914, **Fig. 4C**). Yeast Upf1, but not human Upf1 co-purified with CBP-Nmd4, showing that the interaction is direct and specific (**Fig. 4D**). To test whether this interaction could affect Upf1 function, we used a short 3' biotinylated RNA fragment to purify Upf1, as previously described (Fiorini *et al.*, 2012). The concomitant presence of Nmd4 led to a stronger binding of Upf1 to RNA, especially visible at higher ionic strength (**Fig. 4E**). These results designate Nmd4 as a tightly bound Upf1 co-factor, specific to the conformation of Upf1 present in the Effector complex, and affecting the affinity of Upf1 for RNA.

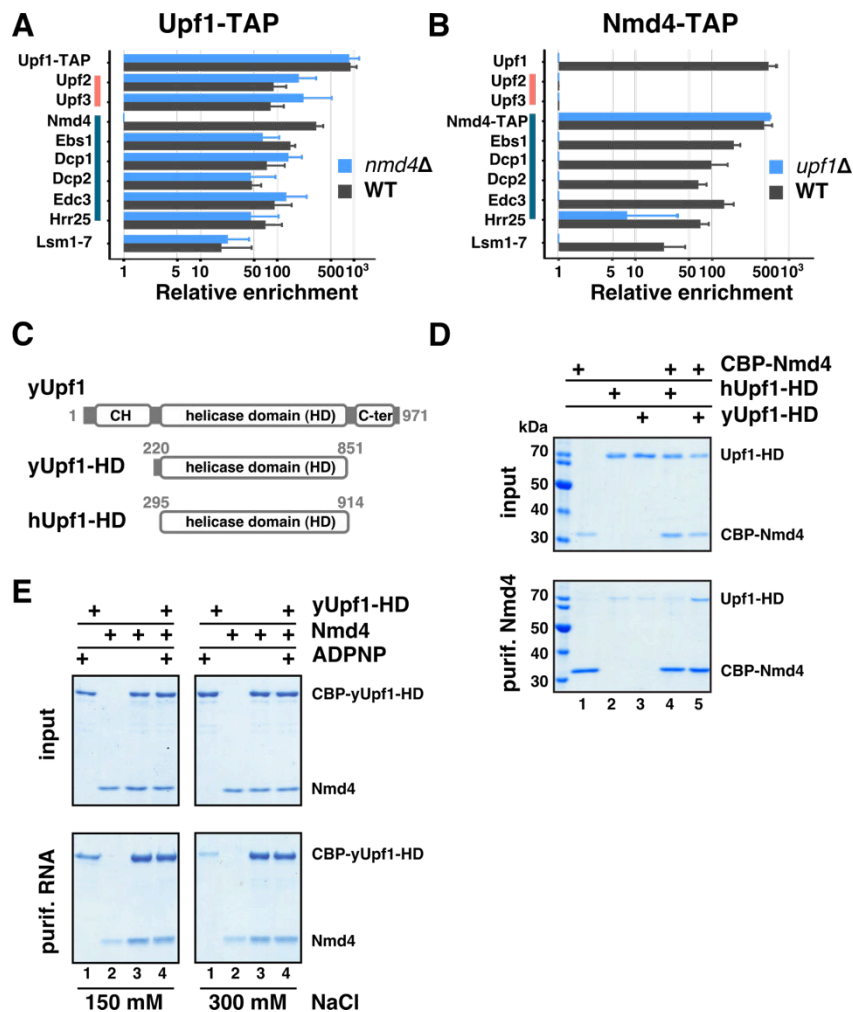


Fig. 4: Nmd4 interaction with the Effector complex is direct and mediated by Upf1.

A. Enrichment of Detector and Effector components in purified Upf1-TAP in the presence (black bars) and absence (blue bars) of *NMD4*. Pink and dark blue vertical lines highlight proteins of the Detector and Effector complexes respectively. **B.** Enrichment of Effector components in Nmd4-TAP in the presence (black bars) and absence (blue bars) of *UPF1*. **C.** Schematics of Upf1 fragments used for the *in vitro* interaction assay; yUpf1-HD is the helicase domain (220-851) of the yeast Upf1 protein, hUpf1-HD is the helicase domain (295-914) of human Upf1 (not to scale). **D.** CBP-Nmd4 was mixed with hUPF1-HD or yUPF1-HD (all the proteins overexpressed in *E. coli* and purified). Protein mixtures before (input, 20% of total) or after purification on calmodulin affinity beads were separated on 10% SDS-PAGE (w/v) acrylamide gels. **E.** Upf1 interaction with a 30-mer 3' biotinylated RNA fragment was tested by mixing purified Upf1 helicase domain with Nmd4 and testing the fraction of recovered protein on streptavidin beads after washes with 150 mM and 300 mM NaCl containing buffer solution.

Potential Smg5-6-7 homologs can be essential for NMD efficiency under limiting conditions

The association of Nmd4 and Ebs1 to Upf1 support the hypothesis that they are the yeast functional equivalents of human Smg6 and Smg5/7. In line with this idea, Nmd4 contains an endonuclease-like region from the PIN domain family (Clissold & Ponting, 2000), like Smg6 and Smg5 (Fig. 5A). Ebs1 has strong similarities with the N-terminal 14-3-3 domains of Smg5, Smg6 and Smg7 (Luke *et al.*, 2007), with similar percentages of identity in the aligned sequences for this domain (Fig. 5A, Appendix Fig. S3). The interactions that we described here indicate that Nmd4 and Ebs1 are components of the Effector complex and co-factors of Upf1. While a role for Ebs1 in the degradation of an NMD reporter RNA has been previously reported (Luke *et al.*, 2007), no data were available about a potential global role of *NMD4* and *EBS1* in NMD. To investigate the impact of *NMD4* and *EBS1* on NMD on a large scale, we performed strand-specific RNASeq experiments in strains deleted for each of the two genes.

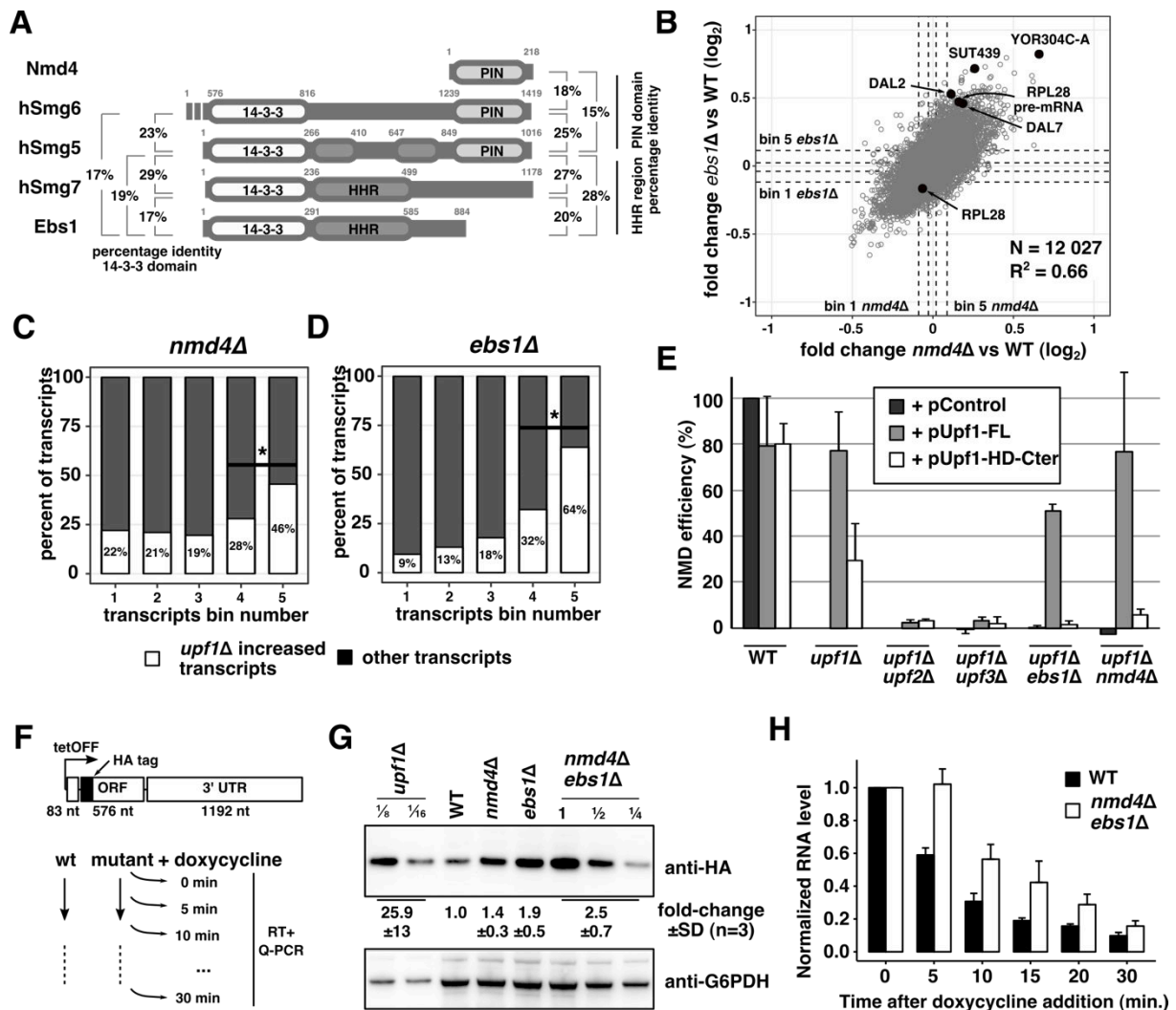


Fig. 5: Nmd4 and Ebs1 are essential for NMD elicited by the overexpression of the helicase domain of Upf1.

A. Schematic representation of the domain structure of Nmd4 and Ebs1 from *S. cerevisiae* compared with human (h) Smg6, Smg5 and Smg7. 14-3-3, HHR and PIN domains were defined based on literature data (Fukuhara *et al*, 2005; Luke *et al*, 2007). Boundaries of the PIN domains were chosen based on a Mafft alignment of Smg5/6 of several species, while for Nmd4 the entire protein sequence was used. Drawing is not to scale. Identity percentages among the different domains of Smg proteins, Nmd4 and Ebs1 are indicated. Values represent percent of identical residues as a fraction of all the aligned residues. **B.** Scatter plot of the mean fold change of transcripts in *nmd4Δ* RNAseq experiment against transcript mean fold change in *ebs1Δ* (Pearson correlation, $r = 0.66$, $p\text{-value} < 2.10^{-16}$). Black dots correspond to NMD substrate examples for which individual traces are shown in **Fig. EV3**. Dashed lines represent boundaries of five bins of equal numbers of transcripts (2046 transcripts or notable features per bin), with bin 5 containing the transcripts that were most increased in the mutant condition compared with wild type. **C.** Percentage of transcripts affected by UPF1 deletion (increase by at least 1.4 fold) among *nmd4Δ* bins, as defined in **B**. Differences between percentages of *upf1Δ* affected transcripts in bin 4 and 5 and bin 3 and 4 were significant (binomial test, $p < 10^{-24}$). **D.** Same as in **C**, for *ebs1Δ*. Differences between distribution of transcripts in bin n and bin $n+1$ were significant (binomial test, $p < 10^{-9}$). **E.** NMD efficiency of WT, *upf1Δ* or double mutant strains complemented with Upf1-FL, Upf1-HD-Cter or an empty plasmid. The NMD efficiency for each strain is based on reverse-transcription followed by quantitative PCR for RPL28 pre-mRNA. A wild-type strain has 100% NMD efficiency and a *upf1Δ* strain has 0% NMD efficiency, by definition. **F.** Schematic drawing of an NMD reporter whose transcription can be repressed by doxycycline addition to the culture. **G.** Changes in the steady-state level of the reporter encoded HA-tagged protein in *upf1Δ*, *nmd4Δ*, *ebs1Δ* and the *nmd4Δ/ebs1Δ* strains were estimated by immunoblot from three independent experiments, with one example shown. **H.** RNA decay in a wild-type and *nmd4Δ/ebs1Δ* strains was tested by reverse-transcription quantitative PCR specific to the 5' end of the reporter RNA at different times after transcription shut-off. The quantifications represent mean RNA amounts and standard deviation of the average (three independent experiments).

The obtained RNASeq data allowed us to quantify variations of RNA levels for 12 027 transcripts (mapping statistics available in **Appendix Table S2** and quantifications in **Dataset EV2**). DESeq2 (Love *et al*, 2014) was used to adjust, normalize and compute the changes in expression levels in *ebs1Δ* and *nmd4Δ* strains in comparison with a wild-type strain. Even if the amplitude of change in transcript levels was modest (examples shown in **Fig. EV3A-E**), the observed variations were highly correlated between the two strains (**Fig. 5B**, $N = 12\ 027$; Pearson correlation coefficient = 0.66; $p\text{-value} < 2.2 \cdot 10^{-16}$). To see to what extent the transcripts enriched in these mutant strains were related with transcripts stabilized in the absence of *UPF1*, we performed a similar RNAseq experiment by comparing *upf1Δ* to a wild-type strain. 3 271 transcripts showed an increase of more than 1.4 fold in the absence of *UPF1* and were used as a reference set for NMD-sensitive RNAs (**Fig. EV3F**). To estimate the presence of such RNAs in the population of transcripts affected by the absence of *NMD4* or *EBS1*, we divided the set of obtained values in five bins of equal size ($N = 2\ 406$ per bin, **Fig. 5B**). We found a strong enrichment of *UPF1*-affected RNA in bin 5, which contains the RNAs most increased in the absence of *NMD4* (**Fig. 5C**) or *EBS1* (**Fig. 5D**) ($p\text{-value} < 10^{-74}$, binomial test performed with one-sided alternative null hypothesis, $N = 2\ 406$).

To establish how *NMD4* and *EBS1* could affect NMD substrates, we searched for conditions that would simplify our phenotypic analysis. We chose to focus on the helicase and C-terminal domain of Upf1, important for binding of both Nmd4 and Ebs1 (**Fig. 3D**). It has been previously observed that overexpression of this region of Upf1 can complement the deletion of the *UPF1* gene (Weng *et al*, 1996). We thus tested whether *NMD4* and *EBS1* were required for this complementation. First, we validated the effect of overexpressing full-length Upf1 or its different domains (depicted in **Fig. 3A**) on the levels of RPL28 (also known as CYH2) pre-mRNA, a well-studied NMD substrate. To facilitate the visualization of the complementation levels, we devised an NMD efficiency measure that takes into account RPL28 pre-mRNA levels in WT (100% NMD) and *upf1* Δ (0% NMD) strains (**Materials and Methods** and **Appendix**). While overexpression of full-length *UPF1* led to the destabilization of RPL28 pre-mRNA, with an estimated reconstituted NMD efficiency of 80%, overexpression of Upf1-HD-Cter allowed recovery of 30% of NMD in a *upf1* Δ strain (**Fig. 5E** and **Fig. EV4A**). This partial complementation of NMD required classical NMD factors since it was abolished in strains lacking *UPF2* and *UPF3*, in strains lacking the 5' to 3' exonuclease *XRNI*, and in a strain in which the decapping enzyme Dcp2, required for yeast NMD, was depleted (**Fig. 5E** and **Fig. EV4A** and **B**). Deletion of either *EBS1* or *NMD4* abolished the partial complementation effect (**Fig. 5E**, **Appendix Figure S4**), suggesting that under limiting conditions, *EBS1* and *NMD4* become essential for NMD. These results correlated with the inability of the Upf1-HD fragment, which showed decreased binding of Ebs1 (**Fig. 3E**), to complement NMD (**Fig. EV4A**). In line with these observations, *EBS1* deletion also decreased the NMD efficiency of Upf1-FL (**Fig. 5E**). Altogether, these results indicate that the mechanism by which Upf1-HD-Cter destabilizes NMD substrates depends on the presence of Nmd4 and Ebs1, the only factors that specifically interact with this region of Upf1 independent on RNA (**Fig. 3D**).

While the biochemical association, RNASeq data and the NMD complementation assay strongly suggested a specific role for Nmd4 and Ebs1 in NMD, we wanted to test the impact of the corresponding genes in a reporter system that allows measuring RNA half-life changes. To this end, we used an NMD reporter that codes for a HA tagged protein fragment and whose transcription can be shut off by the addition of doxycycline (**Fig. 5F**). Immunoblots with anti-HA antibodies showed a consistent increase in the amounts of produced protein in strains lacking either *NMD4*, *EBS1* or both genes (**Fig. 5G**). Since the strongest effect was seen in the double mutant strain, we tested the decrease in the reporter NMD RNA levels at different times after transcription shut-off in this condition. Compared with the wild type, the *nmd4* Δ /*ebs1* Δ strain showed a marked delay in the degradation of the reporter RNA, best seen at 5 minutes after doxycycline addition (**Fig. 5H**). These data suggest that Ebs1 and Nmd4 are NMD factors that affect the degradation of NMD substrates through their specific interactions with the Upf1 helicase and C-terminal domain.

Detector forms by binding of a Upf2/Upf3 heterodimer to Upf1 *in vivo*

While the results presented in the previous paragraphs identified a role for Ebs1 and Nmd4 linked with the helicase domain of Upf1, the CH domain of the protein binds the best-known NMD factors Upf2 and Upf3. Among the results that stood out from the purifications of Detector complex components, one of the most striking was the loss of Upf2 and Upf3 in the Upf1 complex after RNase treatment (**Fig. 2A**). This result was unexpected, since a complex can be reconstituted with domains of the three proteins in the absence of RNA (Chamieh *et al*, 2008). The reciprocal experiments using Upf2-TAP and Upf3-TAP as baits showed that Upf1 could be efficiently recovered in the presence or absence of RNase with Upf2-TAP (**Fig. 2B**), but its interaction with Upf3-TAP was sensitive to RNase (**Fig. 2C**). Thus, the role played by RNA in the stability of the purified Upf1-Upf2-Upf3 complex suggested that Detector is assembled on RNA. To investigate the assembly process, we analysed the composition of complexes purified with Upf1-TAP, Upf2-TAP and Upf3-TAP in the absence of *UPF1*, *UPF2* or *UPF3*. Deletion of either *UPF2* or *UPF3* had a major impact on the enrichment of Upf3 and Upf2 with Upf1 (**Fig. 6A, B**). In the absence of *UPF2*, we could no longer detect Upf3 (**Fig. 6A**), which correlates with previously published data that described human Upf2 as bridging Upf1 and Upf3 (Chamieh *et al*, 2008). Unexpectedly, in the absence of Upf3, we could not detect Upf2 in the Upf1-associated complex (**Fig. 6B**). Thus, in the absence of *UPF2*, Upf3 no longer stably associated with Upf1 and in the absence of *UPF3*, Upf2 was lost from the Upf1-associated factors. We verified that these changes were not due to changes in the total levels of Upf3-TAP and Upf2-TAP in the absence of *UPF2* (**Fig. 6C**) and *UPF3* (**Fig. 6D**). The absence of Upf1 had no impact on the formation of the Upf2-Upf3 complex, independent of which of the two proteins was tagged (**Fig. 6E** and **Dataset EV1**). Altogether, these results suggest that Upf2 and Upf3 form a heterodimer independent of Upf1 and that both proteins are required for a stable association with Upf1 in the Detector complex.

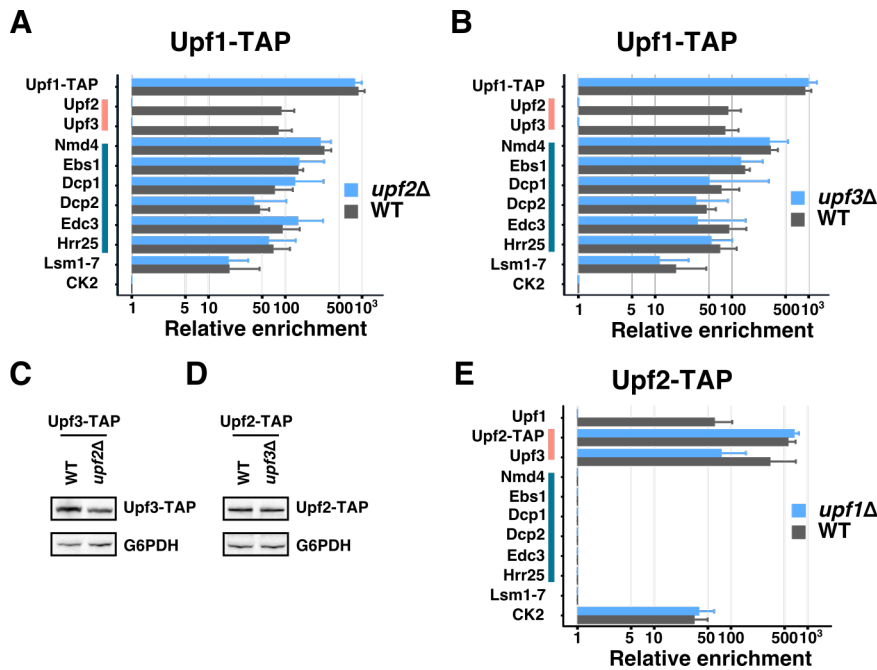


Fig. 6: Upf2 and Upf3 function as a heterodimer in the formation of Detector.

Comparison between enrichment values for Effector and Detector components, Lsm proteins and CK subunits in Upf1-TAP purified samples in the presence (grey bars) or absence (blue bars) of *UPF2* (A) or *UPF3* (B). Evaluation of the effects of *UPF2* deletion on Upf3-TAP levels (C) and of *UPF3* deletion on Upf2-TAP levels (D) in total extracts, in comparison with a loading control (G6PDH) was done by immunoblot. E. Comparison between enrichment values for Effector and Detector components, Lsm proteins and CK subunits in Upf2-TAP purified samples in the presence (grey bars) or absence (blue bars) of *UPF1*. The levels of expression of Upf2-TAP and Upf3-TAP proteins have been verified in this condition (Appendix Fig. S2).

Binding of Effector to NMD substrates depends on Upf2

The formation of Detector on NMD substrates could be followed by a switch to Effector, for RNA decapping and initiation of degradation. Surprisingly, the absence of Upf2 or Upf3 did not alter the protein composition of the purified Effector complex (Fig. 6A, B). This puzzling result made us wonder if Effector exists in two forms, one that associates to NMD substrates, and another that would correspond to Upf1 and the decapping machinery in the process of being recycled as RNA-free complexes. To explore this hypothesis we took advantage of the fact that Upf1 binds preferentially to RNAs that are degraded through NMD (Johansson *et al*, 2007). To distinguish between the Effector-bound and Detector-bound Upf1, we purified Effector *via* Nmd4-TAP (Fig. 2D). The 6-fold enrichment of RPL28 pre-mRNA as compared with total RNA in the input fraction (Fig. 7A) indicated that a fraction of the Effector complex binds NMD substrates.

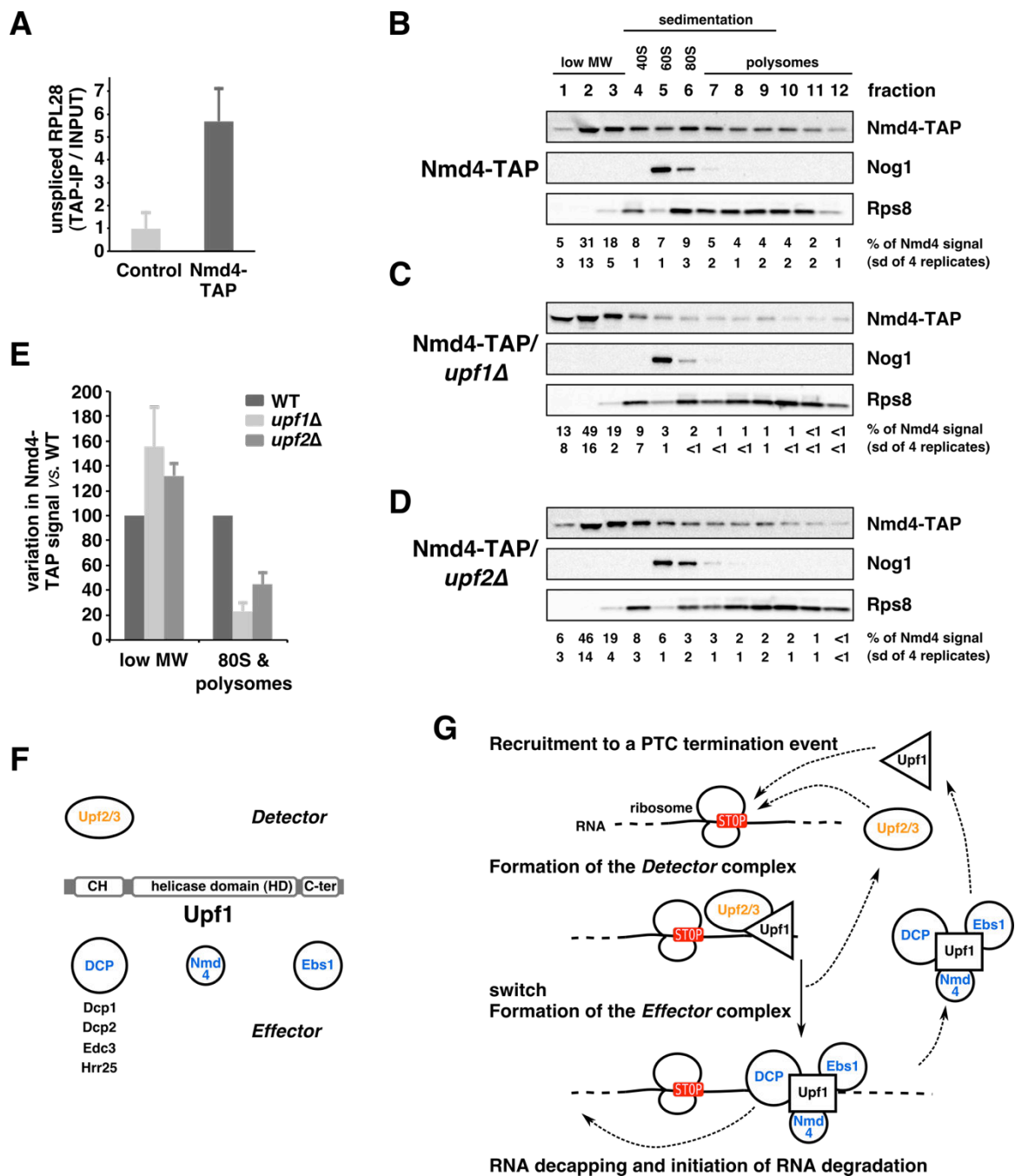


Fig. 7: Association of Nmd4 to NMD substrates depends on Detector.

A. Unspliced RPL28, an NMD substrate, was enriched in Nmd4-TAP purification in comparison with a control untagged strain, as measured by reverse transcription followed by quantitative PCR. Bars correspond to the mean of pre-RPL28 enrichment for 3 independent experiments, as compared with RIM1, a non-NMD mRNA. Error bars correspond to the standard deviation. **B.** to **D.** Distribution of Nmd4-TAP, Nog1 (marker of free 60S subunits) and Rps8 (marker of 40S, 80S and polysome fractions) in wild type (**B**), *upf1Δ* (**C**) and *upf2Δ* (**D**) strains, tested by immunoblot. Fractions 1, 2, 3 are light fractions, fractions 4, 5 and 6 correspond respectively to ribosomal 40S, 60S and 80S positions, fractions 7 to 12 correspond to polysomes. The percent of total signal for Nmd4-TAP in three independent replicates along with standard deviation of the values are indicated for each

fraction. **E.** Quantified relative changes in Nmd4-TAP levels in light and polysomal fractions in *upf1* Δ and *upf2* Δ strains (100% correspond to levels in the wild type strain). **F.** Summary of the Upf1 domains interaction with Detector and Effector components, based on the data presented in this work. **G.** Proposed Detector/Effector sequence of events for yeast NMD.

We next wanted to know if Effector binding to RNA was dependent on Detector components, Upf2 and Upf3, which would be suggestive of a mandatory sequence of association of the NMD complexes to the RNA substrates. We could not test this hypothesis directly, because deletion of *UPF2* or *UPF3* leads to a massive stabilization of NMD-sensitive RNAs and renders RNA enrichment calculations and comparisons with the wild-type situation unreliable. To overcome this problem, we looked at the association of the Effector complex to RNA by measuring the distribution of Nmd4-TAP in a polysome gradient in the absence or presence of *UPF2*. While most of Nmd4 sedimented in the upper part of the gradient, a fraction of the protein was found in the polysomal and monosomal fractions (**Fig. 7B**). The monosomal region concentrates the majority of NMD substrates (Heyer & Moore, 2016). In the absence of *UPF1*, which mediates all the interactions between Nmd4 and the other components of the effector complex (**Fig. 4B**), Nmd4 was lost from the monosome and polysome fractions (**Fig. 7C**). In the absence of *UPF2*, an essential component of the Detector complex, we also observed a similar change in the distribution of Nmd4 (**Fig. 7D**). Compared with the wild type situation, the absence of Upf1 or Upf2, led to an increase of Nmd4-TAP levels in the lower molecular weight fractions and a decrease in the polysome fractions (**Fig. 7E**). These changes were specific for Nmd4, since the gradient distribution of Rps8, a component of the 40S ribosomal subunits used as a control, was not altered, neither in the *upf1* (**Fig. 7C**) nor in the *upf2* Δ strain (**Fig. 7D**).

The association of Nmd4, a specific component of the Effector complex, to NMD substrates, and its redistribution into lower molecular weight complexes in the absence of the Detector formation, suggest that Detector and Effector are obligatory successive complexes on RNA, and allows us to propose a succession of events compatible with our observations for yeast NMD (**Fig. 7F, G**). This model includes two forms of Effector: one that is RNA-bound and depends on Detector components Upf2 and Upf3, and another that is RNA-free.

The fact that Nmd4 was bound to a region of Upf1 not required for other interactions and still could only be enriched in Effector complexes suggests that Upf1 has a different conformation in Detector and Effector. A switch from the Detector to the Effector conformation described here would permit recruitment of decapping factors, required to degrade NMD substrates (Muhlrad & Parker, 1994), and could be an essential step of the NMD mechanism.

Discussion

The current study addresses the composition and dynamics of NMD complexes in yeast. We identified two distinct, mutually exclusive (**Fig. 2**) and successive (**Fig. 7**) complexes containing Upf1. We propose that the Detector complex, composed of Upf1, Upf2 and Upf3, is recruited on NMD substrates. A complete re-organization of the initial complex leads to the replacement of the Upf2/3 heterodimer with Nmd4, Ebs1 and the decapping factors Dcp1, Dcp2 and Edc3. Nmd4 and Ebs1 are accessory factors for NMD in yeast and could be functional homologues of human Smg6 and Smg5/7 respectively (**Fig. 4** and **5**). Both the decapping machinery and Upf2/3 interacted with the CH domain of Upf1 (**Fig. 3**), in support of the hypothesis that decapping competes with Upf2-3 to form the Effector. This competition is likely to occur on RNA, since Nmd4 was lost from large RNA-associated complexes in the absence of *UPF2* (**Fig. 7**). Thus, recruitment of decapping on NMD substrates is likely to depend on this switch on RNA, which could be a crucial, albeit not yet explored step in NMD.

How exactly the switch from the “Detector” to an “Effector” complex occurs is unclear, but similar events could be also important for recruiting RNA degradation factors for NMD in other organisms. Since all the factors that we describe in yeast NMD complexes have mammalian equivalents, and since our results are compatible with a large body of experimental evidence in various other organisms, we propose an extension of the canonical model of NMD (**Appendix Fig. S5**). We discuss how a revised NMD model, called “Detector/Effector” fits the available experimental data in yeast and other species and to what extent it can serve as a basis to build NMD models that incorporate RNA splicing and Upf1 phosphorylation, two molecular events that do not affect NMD in *S. cerevisiae*.

Affinity purification strategies for NMD

Before entering into the description of the Detector/Effector model, it is important to see how the results presented here extend and improve those obtained in previous studies. Affinity purification followed by mass-spectrometry is the most effective way to find functional association of proteins and surpasses in precision all the other large-scale interaction methods in yeast (Benschop *et al*, 2010). However, despite the excellent results obtained with this method on hundreds of different complexes, large-scale affinity purification failed to assign specific interactions of Upf1, Upf2 or Upf3 with other yeast proteins (Gavin *et al*, 2006). As a result, most of what was known on the composition of yeast NMD complexes was based on two-hybrid experiments and the use of co-purification and immunoblotting for previously identified factors (e.g. He & Jacobson, 1995; Czapinski *et al*, 1998; Swisher & Parker, 2011). The situation is similar in other species and can be illustrated by the results obtained with the recently published SONAR approach (Brannan *et al*, 2016), which analysed the proteins co-purified with tagged human Upf1 (hUpf1). Human Upf2 was found as the 112th most enriched factor in the list of Upf1-associated proteins and only reached position 28 when the sample was treated with RNase A. Except for Upf3b, ranked second, no other specific NMD factors were identified in the first 100 proteins associated with hUpf1. Thus, without previous

knowledge of the identity of NMD factors, the number of false positive results precludes the use of these data to define NMD complexes in human cells.

To obtain a global and high-resolution view of the composition and dynamics of the NMD complexes we used a combination of affinity isolation with an innovative computational workflow. Fast affinity purification preserves transient interactions (Oeffinger *et al*, 2007) and the use of surface-coated magnetic beads gives access to the large RNA-protein complexes in which NMD takes place (Zhang *et al*, 1997) without the bias induced by porous chromatography media (Halbeisen *et al*, 2009). To get rid of abundant contaminants, we used the relative enrichment of proteins, based on mass-spectrometry quantitations (see **Materials and Methods**), which efficiently highlighted specific interactions.

These robust quantitative data identified novel interactions, including two protein kinases associated with Upf1 (Hrr25; **Fig. 1F, G, 2A** and **Dataset EV1**) and Upf2 (Casein Kinase 2 complex; **Fig. 2B** and **Dataset EV1**). Besides the ability to show the presence of interactions, quantitative mass spectrometry provided a way to define lost or absent proteins in a given complex. This led to the identification of mutually exclusive complexes, illustrated by the comparison of Detector components purified *via* Upf2-TAP with Effector complexes associated with Nmd4-TAP (**Fig. 2B, D**). The existence of these two distinct complexes in yeast NMD was not known but is compatible with previously published results, like the two-hybrid interactions of Upf1 with Upf2 and Upf3 (He *et al*, 1997), the co-purification of human Dcp1 and Dcp2 with Upf1 (Lykke-Andersen, 2002) or the two-hybrid interactions between yeast Edc3 and Upf1 (Swisher & Parker, 2011). Our results, showing the importance of the CH domain in the interactions between Upf1 and other components of Detector and Effector (**Fig. 3**), are also compatible with previous two-hybrid data showing the importance of this domain in the interactions with Upf2 and Upf3 (He *et al*, 1997) and with Dcp2 (Swisher & Parker, 2011; He & Jacobson, 2015).

Altogether, our enrichment approach and the combination of RNase treated and native complexes analyses drastically reduced the number of NMD-associated candidate proteins to a short list that only contains highly specific top hits and is devoid of factors that are not related to NMD (**Fig. 2I** and **Dataset EV1**). The extensive measurements of the dynamics of NMD complexes in different mutant conditions bound or not to RNA, support a new hierarchy of the molecular events required for yeast NMD, that we call the Detector/Effector model (**Fig. 7G**).

Detector formation is both similar and distinct from SURF/DECID

The proposed Detector/Effector yeast mechanism (**Fig. 7G, Appendix Fig. S5**) can be extended to other species to include elements of the canonical SURF/DECID model that was, until now, restricted to mammalian EJC-enhanced NMD. Similar unified NMD models that include EJC-enhanced and EJC-independent NMD as two aspects of the same molecular mechanism have been proposed earlier (Stalder & Mühlemann, 2008), but differ from the

extended Detector/Effector model in terms of the order of interactions during NMD substrate detection and of the importance of Smg1 in the process.

The first step in the model proposed here involves the recognition of an aberrant translation termination event (Amrani *et al*, 2004; Behm-Ansmant *et al*, 2007) and the formation of a complex containing Upf1, Upf2 and Upf3. Its formation depends on an aberrant translation termination event and could occur through an interaction of Upf2/3 with the translation termination factors eRF1/eRF3 or a terminating ribosome. This type of mechanism has been recently proposed as an alternative to SURF formation in recruiting Upf1 to NMD substrates in human cells (Neu-Yilik *et al*, 2017). A less detailed, but similar mechanism, has been proposed previously, based on experiments on human cells in culture (Lykke-Andersen, 2002).

Our data support the formation of a Upf1, Upf2, Upf3 complex through the interaction of a Upf2-3 heterodimer (**Fig. 6**) with Upf1 on RNA (**Fig. 2A, C**). The importance of heterodimerisation for the function of Upf2-3 was also observed in *C. elegans*, where co-purification of Upf2/Smg3 with Upf1/Smg2 is strongly reduced in the absence of Upf3/Smg4 (Johns *et al*, 2007). The hypothesis of a recruitment of Upf1 through a Upf2-3 heterodimer is also compatible with the fact that Upf2 and Upf3 co-sediment in yeast polysomal fractions independent on the presence of *UPF1*, while Upf2 loses its association with heavy complexes in the absence of *UPF3* (Atkin *et al*, 1997)

The importance of the C-terminal Upf1 domain in yeast and mammalian NMD

We show here that the presence of the C-terminal region of Upf1 is important for NMD in yeast, as demonstrated by its effect on the ability of Upf1 fragments to complement the deletion of the gene (**Fig. 5E** and **Fig. EV4A**) and by its role in binding Ebs1 (**Fig. 3E**). Thus, the C-terminal extension of yeast Upf1 might function like the SQ C-terminal extension in other species, even in the absence of the characteristic SQ/TQ motifs and phosphorylation. The C-terminal extension of yeast Upf1 can thus be the functional equivalent of the C-terminal Upf1 extension in other species. It is notable that Smg1, the protein kinase involved in the phosphorylation of Upf1 in many species, is absent from red and brown algae, excavates such as *Trypanosoma brucei* and *Giardia lamblia* (Lloyd & Davies, 2013) and ciliates, such as *Tetrahymena thermophila* (Tian *et al*, 2017). However, these species present active NMD mechanisms that depend on equivalents of Upf1, Upf2 and Upf3, suggesting that phosphorylation of the C-terminal region is not an absolute requirement for NMD. In line with this observation, no phosphorylated residues have been reported in yeast Upf1 C-terminal region, even if phosphorylated residues have been detected in Upf1 (Lasalde *et al*, 2013). These observations indicate that the C-terminal domain *per se* is important for NMD and that the main role of the C-terminal extension, with or without phosphorylation, is the recruitment of later factors and the formation of Effector-like complexes. In support to this idea, a chimeric Upf1 where the helicase domain of the yeast protein was replaced with the similar region of human Upf1 partially complements the deletion of *UPF1* in yeast as long as it preserves the fungal-specific N and C-terminal extensions (Perlick *et al*, 1996).

The similarity between the binding of Ebs1 to the C-terminal domain of yeast Upf1 and the recruitment of Smg5/7 to phosphorylated SQ residues in metazoan NMD also extends to Nmd4. Nmd4 is a specific marker of the Effector complex that directly binds the helicase domain of Upf1 (**Fig. 4**) and consists of a PIN domain similar with the PIN domain of mammalian Smg6 (**Fig. 5A**). Nmd4 binding to the helicase domain of yeast Upf1 correlates with the known phosphorylation-independent binding of Smg6 to a region of Upf1 helicase domain (Chakrabarti *et al*, 2014; Nicholson *et al*, 2014). Thus, the architecture and function of Upf1 and the associated factors seem to be more conserved than previously thought.

A molecular switch around Upf1 for the Detector to Effector transition

The defining feature of the proposed NMD model is the major change in composition of Upf1-bound complexes from the Detector to the Effector. Upf1 conformation is likely to be dramatically different between the two complexes since: Upf2 and Upf3 were completely absent from complexes purified using Nmd4-TAP (**Fig. 2D**); Nmd4 was bound to a region of Upf1 that is different from the N-terminal CH domain, to which decapping factors and Upf2-3 bind (**Fig. 3C, D**); structures of Upf1 with and without a Upf2 C-terminal fragment show a different organisation of the Upf1 helicase sub-domains 1B and 1C (Clerici *et al*, 2009); all the interactions of Nmd4 with Effector components depend on Upf1 (**Fig. 4B**). This change was not apparent in previous studies that lacked quantitative estimations of all the components of purified complexes at the same time.

The main result of the switch is the recruitment of the decapping enzyme, specifically the core decapping proteins Dcp1, Dcp2 and Edc3 (**Fig. 2A, D, H**). Decapping is the first and major step in the degradation of NMD substrates in yeast (Muhlrad & Parker, 1994) and is responsible for the degradation of about a third of NMD substrates in human cells (Lykke-Andersen *et al*, 2014). The other two-thirds of NMD substrates in human cells are likely to be degraded through a pathway that depends on the presence of catalytically active Smg6 (Eberle *et al*, 2009; Huntzinger *et al*, 2008). In view of the modest increase in NMD substrates levels in a *nmd4Δ* mutant (**Fig. 5** and **Fig. EV3**), it is unlikely that the Nmd4 PIN domain has an endonucleolytic activity in yeast NMD. However, this hypothesis cannot be completely ruled out, since the sequence of Nmd4 contains a conserved Asp residue in a position equivalent with Asp1392 of hSmg6, replaced by a Gly residue in the inactive PIN of hSmg5 (**Appendix Fig. S6**). The tight association of Nmd4 with Upf1 (**Fig. 2A, D** and **4D**), its impact on Upf1 binding to RNA (**Fig. 4E**), and its effects on NMD triggered by overexpression of a truncated Upf1 (**Fig. 5E**) and on a reporter RNA (**Fig. 5G, H**), suggest that Nmd4 is a Upf1 co-factor that assists the helicase in yeast NMD. The importance of Nmd4 in fungal NMD is supported by the presence of orthologs in many species, including *Yarrowia lipolytica* (Pryszcz *et al*, 2011), a yeast that is as distant from *S. cerevisiae* as the urochordate *Ciona intestinalis* is from humans (Dujon *et al*, 2004). Thus, it is likely that different organisms use the same basic molecular machineries to degrade NMD substrates, but the balance between decapping, endonucleolysis or other alternative degradation processes is variable depending on each species.

Conclusion

Our data and the extended Detector/Effector model provide a solution to the unclear situation of the conservation of NMD molecular mechanisms among eukaryotes and will allow future work on yet unsolved issues: how aberrant termination is recognized and leads to Detector formation, where on RNA are positioned Detector and Effector complexes, how Effector and Detector components are recycled and at what steps ATP binding and hydrolysis by Upf1 play crucial roles.

Materials and Methods

Yeast and bacterial strains

Saccharomyces cerevisiae strains were derived from BY4741 (mat a) and BY4742 (mat alpha) strains. Deletion strains were part of the systematic yeast gene deletion collection (Giaever *et al*, 2002) distributed by EuroScarf (<http://www.euroscarf.de/>) or were built by transformation of BY4741 strain with a cassette containing a selection marker cassette flanked by long recombination arms located upstream and downstream the open reading frame of the gene. Deletions were tested by PCR amplification of the modified locus. The list of used strains is provided as **Appendix Table S3**.

E. coli strain NEB 10-beta (New England Biolabs) was used for construction of plasmids by Gibson assembly and their multiplication. *E. coli* BL21(DE3) cells were used for expression of recombinant Nmd4 and Upf1 proteins.

Plasmids construction for yeast experiments

To build single copy plasmids for the expression of N-terminal TAP-tagged Upf1 fragments in yeast we used pCM189 (Gari *et al*, 1997) as a start, since it provides a Tet-OFF promoter regulation system. Cloning of Upf1 fragments into plasmids was done by a modified version of Gibson assembly (Fu *et al*, 2014) using NotI digested pCM189-NTAP (plasmid 1233, Saveanu *et al*, 2007) and the PCR product obtained with oligonucleotides CS1359 and CS1364 (UPF1-FL), CS1359 and CS1361 (UPF1-CH), CS1362 and CS1364 (UPF1-Cter), CS1359 and CS1393 (UPF1-NoCex) or CS1362 and CS1393 (UPF1-Cter-NoCex) on template pDEST14-NAM7 (pl. 1350) or pDONR201-NAM7 (pl. 1330).

The NMD reporter plasmid pTG189-HA-ALA1-KAN was generated from a short-lived NMD reporter used previously (Decourty *et al*, 2014) and which consists of a fragment of the yeast *ALAI* coding sequence (576 nt), tagged by a HA tag at its N-terminus, and followed by the KanMX6 cassette ending with a TEF' terminator. The RNA has a long 3' UTR of about 1192 nucleotides, which induces degradation of the transcript through NMD and its transcription is repressed by addition of doxycycline in the culture medium.

The list of plasmids and oligonucleotides used in this study is provided as **Appendix Table S4** and **S5**.

Plasmids construction for *in vitro* interaction studies

Coding sequences of yeast full length *NMD4*, yeast *UPF1* helicase domain, and human *UPF1* helicase domains (Uniprot accession codes Q12129, P30771, and Q92900-2 respectively) were cloned between NheI/NotI, NheI/XhoI, and NdeI/XhoI in variants of pET28a (Novagen) for the *in vitro* interaction tests.

TAP affinity purifications

TAP-tagged proteins were purified using a one step purification method. Briefly, frozen cell pellets of 2 to 8L culture were resuspended in lysis buffer (20mM Hepes K pH7.4, 145mM KCl, Protease inhibitor, 1X Vanadyl-Ribonucleoside Complex, 40u·mL⁻¹RNasin) and lysed using a French press (2 passages at 1200 PSI). The lysate was cleared at 4°C for 20min in a JA-25.50 rotor (Beckman) at 15 000rpm. Magnetic beads (Dynabeads M-270 epoxy) coupled to IgG were added to the protein extract and incubated for 1 hour at 4° (Oeffinger *et al*, 2007). Beads were collected with a magnet and extensively washed three times with a first buffer (HKI: 20mM Hepes K pH7.4, 145mM KCl, 40u·mL⁻¹ RNasin, 0.1% Igepal) and twice with a second buffer (HK: 20mM Hepes K pH7.4, 145mM KCl, 40u·mL⁻¹ RNasin, 1mM DTT). The lack of detergent in the second buffer ensured compatibility with mass spectrometry. After washing, complexes were eluted by incubation in elution buffer (HK + recombinant purified AcTEV, Thermo Fisher Scientific) for 1 hour at 16°C on a rotating wheel. Supernatant that contained purified complexes was collected and precipitated using the methanol/chloroform method (Wessel & Flügge, 1984) before further analysis by silver staining, western blot or mass spectrometry.

In case of RNase treatment, lysed cells were divided in two equal parts, one part was treated with RNase A and RNase T1 and the other part was not. Purifications were done in parallel. For treated samples, clarified lysate were treated a first time with 300u·mL⁻¹ of RNase T1 (Thermo scientific) and 167.10⁵ u·mL⁻¹ of RNase A (Thermo scientific), 20 minutes on ice. The sample was also treated a second time, after adsorption of proteins on beads and washing, using a first wash with HK buffer with RNase T1 at 83u·mL⁻¹ of the clarified lysate previously used and RNase A at 17u·mL⁻¹ of the clarified lysate previously used for 20 minutes at 25°C. When purified complexes were tested by immunoblot, the TEV elution was replaced with elution in 1% SDS (incubated 10 minutes at 65°C).

Mass spectrometry, sample preparation, acquisition and data analysis

Protein samples were treated with Endoprotease Lys-C (Wako) and Trypsin (Trypsin Gold Mass Spec Grade; Promega). Peptide samples were desalted using OMIX C18 pipette tips (Agilent Technologies). The peptides mixtures were analyzed by nano-LC-MS/MS using an Ultimate 3000 system (Thermo Fisher Scientific) coupled to an LTQ-Orbitrap Velos mass spectrometer or LTQ-Orbitrap XL (Thermo Fisher Scientific). Peptides were desalted on-line using a trap column (C18 Pepmap100, 5µm, 300µm×5mm, Thermo Scientific) and then separated using 120min RP gradient (5–45% acetonitrile/0.1% formic acid) on an Acclaim PepMap100 analytical column (C18, 3µm, 100Å, 75µm id x 150mm, Thermo Scientific) with a flow rate of 0.340µL·min⁻¹. The mass spectrometer was operated in standard data dependent acquisition mode controlled by Xcalibur 2.2. The instrument was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 60,000 at m/z 400, with the top 20 most abundant multiply-charged (2+ and higher) ions subjected to CID fragmentation in the linear ion trap. An FTMS target value of 1e6 and an ion trap MSn target value of 10000

were used. Dynamic exclusion was enabled with a repeat duration of 30s with an exclusion list of 500 and exclusion duration of 60s. Lock mass of 445.12002 was enabled for all experiments.

Raw data files from MS were used for spectra matching with the MaxQuant/Andromeda search engine (version 1.5.5.1 Cox & Mann, 2008; Cox *et al.*, 2011) against the Uniprot *Saccharomyces cerevisiae* proteome database (downloaded 12/02/2016; total 6,721 entries) completed with the 22 virus protein sequences from yeast, and a list of commonly observed contaminants supplied by MaxQuant. Searches allowed either trypsin specificity with two missed cleavages. Cysteine carbamidomethylation was selected as fixed modification, and oxidation of methionine and N-terminal acetylation were searched as variable modifications. Peptide identification was performed using 6 ppm as mass tolerance at the MS level for FT mass analyser and 0.5Da at the MS/MS level for IT analyzer. MaxQuant/Andromeda used seven amino acid minimum for peptide length, with 0.01 false discovery rate for both protein and peptide identification. The protein identification was considered valid only when at least one unique or “razor” peptide was present. Following protein identification, the intensity for each identified protein was calculated using peptide signal intensities. Identification transfer protocol (“match-between-runs” feature in MaxQuant) was applied.

Only protein identifications based on a minimum of two peptides were selected for further quantitative studies. Bioinformatics analysis of the MaxQuant/Andromeda workflow output and the analysis of the abundances of the identified proteins were performed using R (workflow in **Appendix Fig. S1**). A score called LTOP2 was computed for each protein group in the different experiments. This score is similar with the “top three” average described (Silva *et al.*, 2006), with several differences. First, we built meta-peptide intensities, based on the intensity of overlapping peptides with missed cleavages (minimum 4 residues overlapping). Next, we took the 3 most intense peptides (top3), or the top 2 if only two peptide intensities were available, and calculated the average of log₂-transformed intensities. Finally, LTOP2 between experiments were normalized using the TEV protease as a spike in, as this protease is added in each sample during the purification process in equivalent quantities. To overcome some limitations of the score, and specifically the contamination by abundant yeast proteins, we calculated enrichments for each protein. Enrichment values were obtained by dividing the relative ratio of a purified protein with the ratio of the protein in a total extract.

Reverse transcription and quantitative PCR

To measure the enrichment of NMD substrates such as the pre-messenger RPL28 or the DAL7 mRNA in different experiments, we performed reverse transcription followed by qPCR. After RNA extraction and DNase treatment, 500 ng of RNA were reverse-transcribed using Superscript III (Invitrogen) and a mix of reverse qPCR oligonucleotides, CS888 (RPL28-premature), CS889 (RPL28), CS1077 (RIM1), CS1430 (DAL7); for RNA immunoprecipitation, the purified and input samples were used for reverse-transcription with transcript-specific primers. Eight fold dilutions of cDNA were tested by quantitative PCR.

The amplification was done in a Stratagene MX3005P and corresponding software (MXpro quantitative PCR system), with step 1 (95°C for 5min) and step 2 (40 cycles of 95°C for 20s and 60°C for 1min).

NMD efficiency calculation

Complementation of an NMD deficient strain leads to a decrease in the levels of an NMD-sensitive transcript like the unspliced precursor of the RPL28 mRNA. To be able to judge how the efficiency of NMD affects the level of a transcript and to simplify the interpretation of complementation experiments, we computed NMD efficiency using the maximum (no NMD) and minimum (NMD fully active) steady-state levels of a given transcript. We assumed that RNA synthesis is constant and that the degradation of RNA occurs by two competitive pathways, with different first-order rate constants.

Polysome fractionation

Polysome extracts were obtained from 150mL of mid-log phase ($A_{600} \approx 0.6$) yeast cells culture treated with $50\mu\text{g}\cdot\text{ml}^{-1}$ cycloheximide for 5 minutes at 4°C and broken with glass beads using a MagNA lyser (three times for 60s at 3000 rpm) in breaking buffer (20mM Tris-HCl, pH 7.4, 50mM KCl, 10mM MgCl_2 , $50\mu\text{g}\cdot\text{ml}^{-1}$ cycloheximide, 1mM DTT, $0.1\text{mg}\cdot\text{ml}^{-1}$ Heparin, and EDTA-free protease inhibitors from Roche). 10 to 15 A_{260} of clarified cell lysates were loaded on 10–50% sucrose gradients. Gradients were centrifuged 3h15 at 36000 rpm with slow acceleration in a SW41-Ti rotor (Beckman) in a Optima L-80 ultracentrifuge. Half of collected fractions were precipitated with TCA and the precipitate was resuspended in 50 μL of sample buffer. 10 μL were loaded on polyacrylamide NuPAGE Novex 4-12% Bis-Tris gels (Life technologies). After transfer to nitrocellulose membrane with a semi-dry system, proteins were detected by hybridization with appropriate antibodies.

For polysome analysis, western blot signals were quantified using Image Lab software (version 5.2.1 build 11 - Biorad). We used the square volume tool to define boxes of identical size at the expected position of the protein band we wanted to quantify. We also defined a series of blank boxes for each lane for background definition. To normalize the results, we calculated the percentage of signal over all the fractions of a run for both Nmd4 and Rps8 (control). Statistical significance of the signal differences between conditions were assessed by the paired Student t-test. Stars on plots indicate that the p-value associated with the test was less than 0.05.

Acknowledgments

We thank Magalie Duchateau, Julia Chamot-Rooke and Mariette Matondo of the Mass Spectrometry for Biology UtechS at the Pasteur Institute for access to the mass-spectrometry facility, Jean-Yves Coppée, from the Transcriptome and Epigenome, Biomics platform for access to the sequencing facility. We thank Giorgio Dieci, University of Parma, Italy, for the

Rps8 antibodies. We thank members of the Genetics of Macromolecular Interactions Unit: Gwenael Badis for providing RNA sequencing results for the *upf1*Δ strain, helpful discussions and critical reading of the manuscript, Antonia Doyen for performing initial affinity purification experiments and for building strains, Frank Feuerbach for providing strains, reagents, critical suggestions, advice and support in writing the manuscript, and Micheline Fromont-Racine for discussions and critical reading of the manuscript. We thank Christophe Malabat from the Bioinformatics and Biostatistics hub, Pasteur Institute, for help with data analysis and support in making the results available through a web interface. The study was supported by the ANR CLEANMD grant (ANR-14-CE10-0014) from the French Agence Nationale de la Recherche to C.S., A.J. and H.L-H. and by the Ministère de l'Enseignement Supérieur et de la Recherche and the "Fondation ARC pour la recherche sur le cancer" PhD fellowships to M.D. and J.K., and by continuous financial support from the Institut Pasteur and Centre National de Recherche Scientifique, France.

Author Contributions

M.D., A.N. and C.S. conceived the project. M.D., L.D., A.N., C.P., J.K., H.L-H., A.J., and C.S. designed experiments. M.D., L.D., A.N., J.K. and C.S. performed experiments. A.N., L.D, M.D., H.L-H. and C.S. analyzed the data. M.D. and C.S. wrote the manuscript. All authors discussed results and participated to the corrections of the manuscript.

Conflict of interest

The authors declare not having any conflict of interest.

References

- Addinall SG, Holstein E-M, Lawless C, Yu M, Chapman K, Banks AP, Ngo H-P, Maringele L, Taschuk M, Young A, Ciesiolka A, Lister AL, Wipat A, Wilkinson DJ & Lydall D (2011) Quantitative fitness analysis shows that NMD proteins and many other protein complexes suppress or enhance distinct telomere cap defects. *PLoS Genet.* **7**: e1001362
- Ahrné E, Molzahn L, Glatter T & Schmidt A (2013) Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **13**: 2567–2578
- Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S & Jacobson A (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**: 112–118
- Atkin AL, Schenkman LR, Eastham M, Dahlseid JN, Lelivelt MJ & Culbertson MR (1997) Relationship between yeast polyribosomes and Upf proteins required for nonsense mRNA decay. *J. Biol. Chem.* **272**: 22163–22172
- Azzalin CM, Reichenbach P, Khoriantuli L, Giulotto E & Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**: 798–801
- Behm-Ansmant I, Gatfield D, Rehwinkel J, Hilgers V & Izaurralde E (2007) A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* **26**: 1591–1601
- Benschop JJ, Brabers N, van Leenen D, Bakker LV, van Deutekom HWM, van Berkum NL, Apweiler E, Lijnzaad P, Holstege FCP & Kemmeren P (2010) A Consensus of Core Protein Complex Compositions for *Saccharomyces cerevisiae*. *Molecular Cell* **38**: 916–928
- Brannan KW, Jin W, Huelga SC, Banks CAS, Gilmore JM, Florens L, Washburn MP, Van Nostrand EL, Pratt GA, Schwinn MK, Daniels DL & Yeo GW (2016) SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Molecular Cell* **64**: 282–293
- Bühler M, Steiner S, Mohn F, Paillusson A & Mühlemann O (2006) EJC-independent degradation of nonsense immunoglobulin- μ mRNA depends on 3' UTR length. *Nat Struct Mol Biol* **13**: 462–464
- Chakrabarti S, Bonneau F, Schüssler S, Eppinger E & Conti E (2014) Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5–SMG7 and SMG6. *Nucl. Acids Res.:* gku578
- 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
- Chapin A, Hu H, Rynearson SG, Hollien J, Yandell M & Metzstein MM (2014) In Vivo Determination of Direct Targets of the Nonsense-Mediated Decay Pathway in *Drosophila*. *G3 (Bethesda)* **4**: 485–496
- Chen Z, Smith KR, Batterham P & Robin C (2005) Smg1 Nonsense Mutations Do Not Abolish Nonsense-Mediated mRNA Decay in *Drosophila melanogaster*. *Genetics* **171**: 403–406
- Clerici M, Mourao A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M & Cusack S (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J* **28**: 2293–2306
- Clissold PM & Ponting CP (2000) PIN domains in nonsense-mediated mRNA decay and RNAi. *Current Biology* **10**: R888–R890

- Collins SR, Kemmeren P, Zhao X-C, Greenblatt JF, Spencer F, Holstege FCP, Weissman JS & Krogan NJ (2007) Toward a Comprehensive Atlas of the Physical Interactome of *Saccharomyces cerevisiae*. *Mol Cell Proteomics* **6**: 439–450
- Colombo M, Karousis ED, Bourquin J, Bruggmann R & Mühlemann O (2016) Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA*: ma.059055.116
- Cox J & Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26**: 1367–1372
- Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV & Mann M (2011) Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**: 1794–1805
- Cui Y, Hagan KW, Zhang S & Peltz SW (1995) Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev.* **9**: 423–436
- Czaplinski K, Ruiz-Echevarria MJ, Paushkin SV, Han X, Weng Y, Perlick HA, Dietz HC, Ter-Avanesyan MD & Peltz SW (1998) The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12**: 1665–1677
- Decourty L, Doyen A, Malabat C, Frachon E, Rispal D, Séraphin B, Feuerbach F, Jacquier A & Saveanu C (2014) Long open reading frame transcripts escape nonsense-mediated mRNA decay in yeast. *Cell Rep* **6**: 593–598
- Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J-M, Beyne E, et al (2004) Genome evolution in yeasts. *Nature* **430**: 35–44
- Eberle AB, Lykke-Andersen S, Mühlemann O & Jensen TH (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16**: 49–55
- Eberle AB, Stalder L, Mathys H, Orozco RZ & Mühlemann O (2008) Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol.* **6**: Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18447580> [Accessed September 18, 2012]
- Fiorini F, Bonneau F & Le Hir H (2012) Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. *Meth. Enzymol.* **511**: 255–274
- Fu C, Donovan WP, Shikapwashya-Hasser O, Ye X & Cole RH (2014) Hot Fusion: An Efficient Method to Clone Multiple DNA Fragments as Well as Inverted Repeats without Ligase. *PLoS One* **9**: Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281135/> [Accessed September 11, 2017]
- 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. *Molecular Cell* **17**: 537–547
- Garí E, Piedrafita L, Aldea M & Herrero E (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* **13**: 837–848
- Gatfield D, Unterholzner L, Ciccarelli FD, Bork P & Izaurralde E (2003) Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J* **22**: 3960–3970

- Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier M-A, Hoffman V, Hoefert C, Klein K, Hudak M, Michon A-M, Schelder M, Schirle M, Remor M, et al (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 631–636
- Ghosh S, Ganesan R, Amrani N & Jacobson A (2010) Translational competence of ribosomes released from a premature termination codon is modulated by NMD factors. *RNA* **16**: 1832–1847
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–391
- Halbeisen RE, Scherrer T & Gerber AP (2009) Affinity purification of ribosomes to access the translome. *Methods* **48**: 306–310
- Hall GW & Thein S (1994) Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. *Blood* **83**: 2031–2037
- He F, Brown AH & Jacobson A (1996) Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast. *RNA* **2**: 153–170
- He F, Brown AH & Jacobson A (1997) Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol Cell Biol* **17**: 1580–1594
- He F & Jacobson A (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev.* **9**: 437–454
- He F & Jacobson A (2001) Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. *Mol. Cell. Biol.* **21**: 1515–1530
- He F & Jacobson A (2015) Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu. Rev. Genet.* **49**: 339–366
- Heyer EE & Moore MJ (2016) Redefining the Translational Status of 80S Monosomes. *Cell* **164**: 757–769
- Ho B, Baryshnikova A & Brown GW (2018) Unification of Protein Abundance Datasets Yields a Quantitative *Saccharomyces cerevisiae* Proteome. *Cell Systems* Available at: <http://www.sciencedirect.com/science/article/pii/S240547121730546X> [Accessed February 5, 2018]
- Holbrook JA, Neu-Yilik G, Hentze MW & Kulozik AE (2004) Nonsense-mediated decay approaches the clinic. *Nat Genet* **36**: 801–808
- Huang L, Lou C-H, Chan W, Shum EY, Shao A, Stone E, Karam R, Song H-W & Wilkinson MF (2011) RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. *Mol Cell* **43**: 950–961
- Huntzinger E, Kashima I, Fauser M, Sauliere J & Izaurralde E (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**: 2609–2617
- Johansson MJO, He F, Spatrick P, Li C & Jacobson A (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc Natl Acad Sci U S A* **104**: 20872–20877
- Johns L, Grimson A, Kuchma SL, Newman CL & Anderson P (2007) *Caenorhabditis elegans* SMG-2 Selectively Marks mRNAs Containing Premature Translation Termination Codons. *Mol Cell Biol* **27**: 5630–5638
- 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
- Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, et al (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**: 637–643
- Lasalde C, Rivera AV, León AJ, González-Feliciano JA, Estrella LA, Rodríguez-Cruz EN, Correa ME, Cajigas IJ, Bracho DP, Vega IE, Wilkinson MF & González CI (2013) Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. *Nucleic Acids Res.*
- Le Hir H, Gatfield D, Izaurralde E & Moore MJ (2001) The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* **20**: 4987–4997
- Leeds P, Peltz SW, Jacobson A & Culbertson MR (1991) The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **5**: 2303–2314
- Leeds P, Wood JM, Lee BS & Culbertson MR (1992) Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**: 2165–2177
- Lindeboom RGH, Supek F & Lehner B (2016) The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* **48**: 1112–1118
- Lloyd JPB & Davies B (2013) SMG1 is an ancient nonsense-mediated mRNA decay effector. *Plant J* **76**: 800–810
- Loh B, Jonas S & Izaurralde E (2013) The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27**: 2125–2138
- Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550
- Luke B, Azzalin CM, Hug N, Deplazes A, Peter M & Lingner J (2007) *Saccharomyces cerevisiae* Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. *Nucleic Acids Res.* **35**: 7688–7697
- Lykke-Andersen J (2002) Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay. *Mol Cell Biol* **22**: 8114–8121
- Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A & Jensen TH (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* **28**: 2498–2517
- 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**:
- Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* **5**: 89–99
- Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM & Dietz HC (2001) Rnt1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* **10**: 99–105
- Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY, Halim VA, Bagshaw RD, Hubner NC, al-Hakim A, Bouchard A, Faubert D, Fermin D, Dunham WH, Goudreault M, Lin Z-Y, et al (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Meth* **10**: 730–736
- Metze S, Herzog VA, Ruepp M-D & Mühlemann O (2013) Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation

- pathways. *RNA* **19**: 1432–1448
- Muhlrad D & Parker R (1994) Premature translational termination triggers mRNA decapping. *Nature* **370**: 578–581
- Muhlrad D & Parker R (1999) Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**: 1299–1307
- Nelson JO, Moore KA, Chapin A, Hollien J & Metzstein MM (2016) Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. *eLife* **5**: Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4848089/> [Accessed November 20, 2017]
- Neu-Yilik G, Raimondeau E, Eliseev B, Yeramala L, Amthor B, Deniaud A, Huard K, Kerschgens K, Hentze MW, Schaffitzel C & Kulozik AE (2017) Dual function of UPF3B in early and late translation termination. *The EMBO Journal*: e201797079
- Nicholson P, Josi C, Kurosawa H, Yamashita A & Mühlemann O (2014) A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res* **42**: 9217–9235
- Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD & Rout MP (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods* **4**: 951–956
- Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A, Hachiya T, Hentze MW, Anderson P & Ohno S (2003) Phosphorylation of hUPF1 Induces Formation of mRNA Surveillance Complexes Containing hSMG-5 and hSMG-7. *Molecular Cell* **12**: 1187–1200
- Perlick HA, Medghalchi SM, Spencer FA, Kendzior RJ & Dietz HC (1996) Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc Natl Acad Sci U S A* **93**: 10928–10932
- Pryszcz LP, Huerta-Cepas J & Gabaldón T (2011) MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res* **39**: e32
- Pulak R & Anderson P (1993) mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes Dev.* **7**: 1885–1897
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M & Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol* **17**: 1030–1032
- Saveanu C, Rousselle J-C, Lenormand P, Namane A, Jacquier A & Fromont-Racine M (2007) The p21-activated protein kinase inhibitor Skb15 and its budding yeast homologue are 60S ribosome assembly factors. *Mol. Cell. Biol.* **27**: 2897–2909
- Serdar LD, Whiteside DL & Baker KE (2016) ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nature Communications* **7**: 14021
- Silva JC, Gorenstein MV, Li G-Z, Vissers JPC & Geromanos SJ (2006) Absolute Quantification of Proteins by LCMSE A Virtue of Parallel ms Acquisition. *Mol Cell Proteomics* **5**: 144–156
- 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**:
- Stalder L & Mühlemann O (2008) The meaning of nonsense. *Trends Cell Biol.* **18**: 315–321
- Swisher KD & Parker R (2011) Interactions between Upf1 and the Decapping Factors Edc3 and Pat1 in *Saccharomyces cerevisiae*. *PLoS One* **6**: Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3204985/> [Accessed July 12, 2017]
- Tian M, Yang W, Zhang J, Dang H, Lu X, Fu C & Miao W (2017) Nonsense-mediated

- mRNA decay in *Tetrahymena* is EJC independent and requires a protozoa-specific nuclease. *Nucleic Acids Res.* **45**: 6848–6863
- Wen J & Brogna S (2010) Splicing-dependent NMD does not require the EJC in *Schizosaccharomyces pombe*. *EMBO J* **29**: 1537–1551
- Weng Y, Czaplinski K & Peltz SW (1996) Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Mol Cell Biol* **16**: 5491–5506
- Weng Y, Czaplinski K & Peltz SW (1998) ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA* **4**: 205–214
- Wessel D & Flügge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**: 141–143
- Yepiskoposyan H, Aeschmann F, Nilsson D, Okoniewski M & Mühlemann O (2011) Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17**: 2108–2118
- Zhang S, Welch EM, Hogan K, Brown AH, Peltz SW & Jacobson A (1997) Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in *Saccharomyces cerevisiae*. *RNA* **3**: 234–2

Expanded view Figure

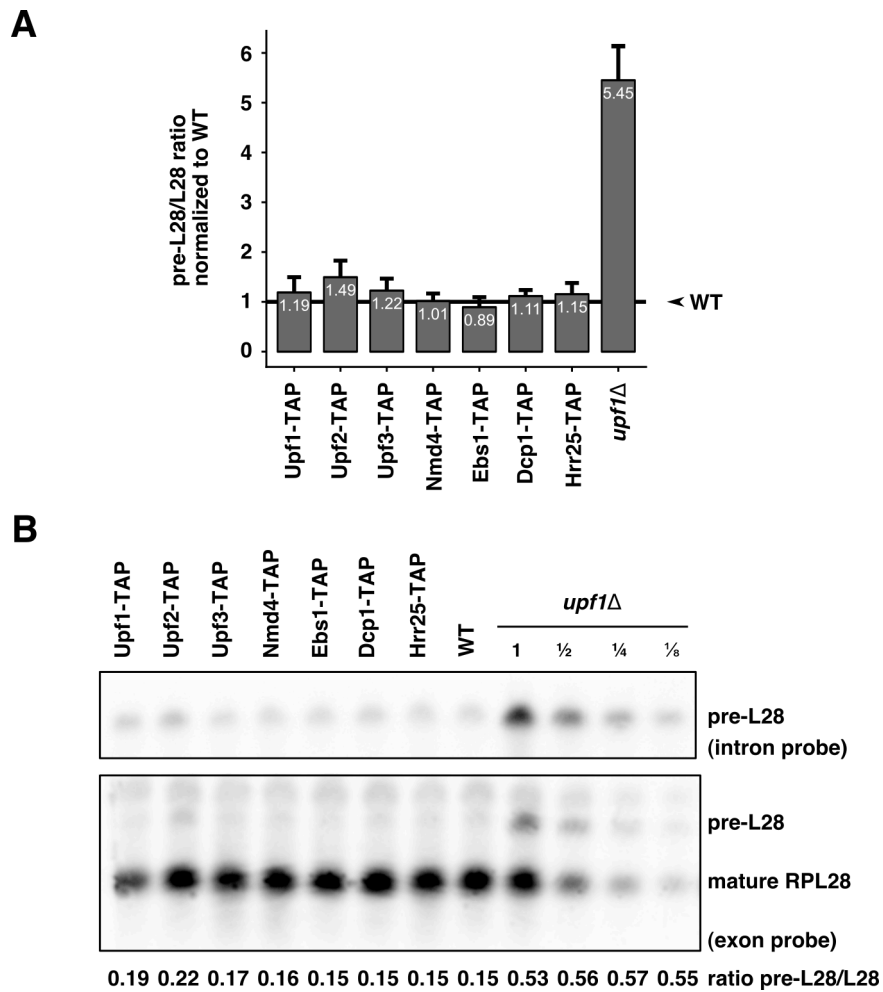


Figure EV1: Tagged proteins are functional for NMD.

A. RNA from the indicated tagged strains and *upf1*Δ, as control, were quantified by reverse-transcription followed by quantitative PCR for the unspliced precursor of RPL28 mRNA and for a normalization RNA, insensitive for NMD (RIM1). Mean values and standard deviations are depicted.

B. The stabilization of the endogenous NMD target pre-RPL28 was tested in comparison with the mature form (RPL28) by northern blot using dsDNA probes and chemiluminescent detection.

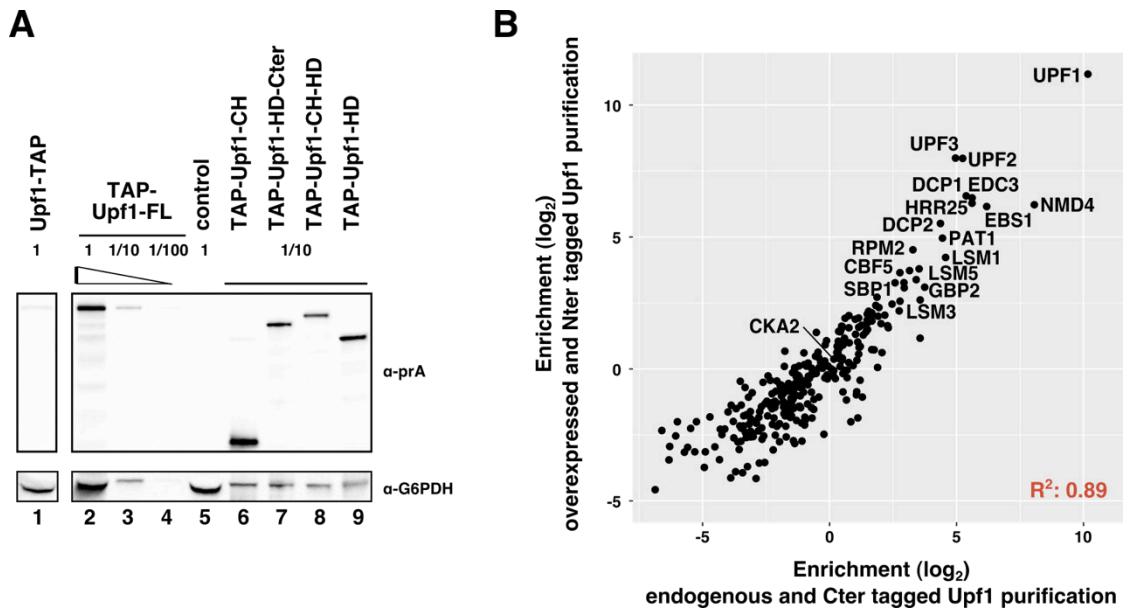


Figure EV2: N-terminal and C-terminal tagged Upf1 enrich similar sets of specific proteins.

A. Estimation of the levels of overexpression for N-terminal tagged Upf1 fragments, in comparison with chromosomally C-terminal tagged protein. G6PDH was used as a loading control. Serial dilutions were used to test the ability of the immunoblot signal to estimate protein levels. **B.** Enrichment values for purifications done with chromosomally C-terminal tagged Upf1 (x axis) and N-terminally TAP tagged Upf1.

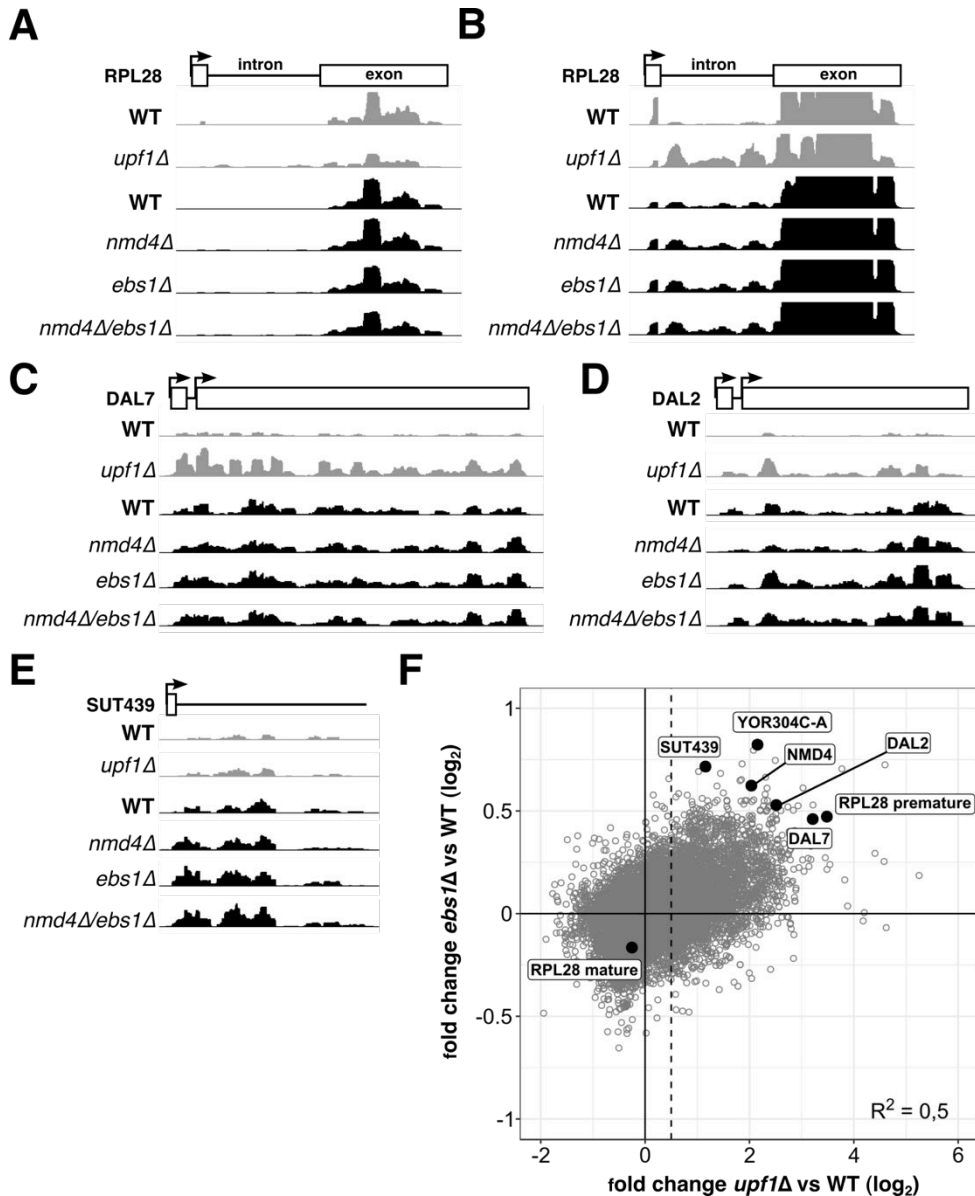


Figure EV3: Deletion of *NMD4* and *EBS1* stabilize a set of transcripts that is also stabilized in the absence of *UPF1*.

(A) to (E) Examples of NMD substrates sequencing profiles in WT, *upf1*Δ, *nmd4*Δ, *ebs1*Δ and *nmd4*Δ/*ebs1*Δ experiments. NMD substrates belong to different classes, intron containing (RPL28; A and B, different scaling to show intron signal), uORF (DAL7, DAL2; C and D), and non-coding RNA (SUT439; E). Profiles were normalized using the samples median counts. F. Scatter plot of transcript \log_2 fold change in *upf1*Δ against *ebs1*Δ. The dashed line represents the limit over which RNA were considered as stabilized, with a \log_2 value of 0.5 (1.4 fold change).

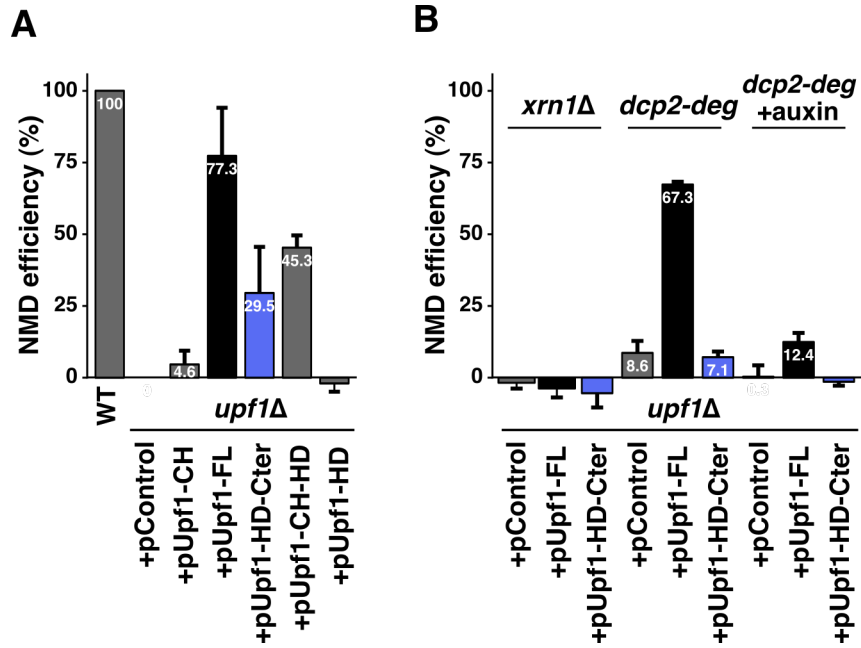


Figure EV4: The helicase domain of Upf1 alone can destabilize RPL28 pre-mRNA, an NMD substrate.

A. Total RNA from wild-type or a *upf1Δ* strain transformed with an empty plasmid (pControl) or plasmids expressing various Upf1 fragments (see **Fig. 3A**) was tested by reverse transcription and quantitative PCR. The levels of RPL28 pre-mRNA were normalized using an NMD-insensitive transcript (RIM1) and an NMD efficiency score was calculated based on the difference between a wild type (100%) and the *upf1Δ* (0% NMD) strain. **B.** Complementation of *UPF1* deletion was tested in combination with the deletion of *XRN1*, and in strains with a degron regulated Dcp2 protein, when protein degradation was not induced (*dcp2-deg*) or was induced by addition of auxin (*dcp2-deg + auxin*) and incubation for 1 hour. Values represent average and standard deviation for at least three independent experiments.

Appendix Material and Methods

Access to primary data (mass-spectrometry and RNASeq):

The accession number for the sequencing data reported in this paper is **GEO: GSE102099**.

The MS proteomics data that support the findings of this study have been deposited in the ProteomeXchange Consortium *via* the PRIDE repository (Vizcaíno et al, 2016) with the dataset identifier **PXD007159**.

In addition to the access to raw data, enrichment and intensity data are available as an interactive web application at: https://hub05.hosting.pasteur.fr/NMD_complexes.

Media and growth conditions

Yeast cells were grown in YPD (20g·L⁻¹ glucose, 10g·L⁻¹ yeast extract, 20g·L⁻¹ bacto-peptone, 20g·L⁻¹ bacto-agar for plates only) and in synthetic media without uracil to select transformants and maintain URA3 plasmids. All yeast strains were freshly thawed from frozen stocks and grown at 30°C. Bacterial strains were grown in LB media, supplemented with antibiotics when necessary, at 37°C.

Yeast strains construction

Yeast strains used in this paper are listed in **Appendix Table S3**. C-terminal TAP-tagged strains originated from the collection of systematically built strain (Ghaemmaghami *et al*, 2003). For some strains, we modified the classical TAP tag (CBP-TEV cleavage site-protein A) and added an additional 6-His tag between CBP and the TEV cleavage site. To do that we have used the pCRBlunt-CRAP(6-HisTAP) plasmid that can replace the original tag by homologous recombination. TAP or CRAP (CBP-6His-TEV-proteinA) tagged versions of proteins were used for our experiments and gave similar results; both versions bear the “TAP” name in the text, as only the protein A part of the tag and the TEV cleavage site were used in our experiments.

Plasmids construction

For Gibson assembly, we mixed 30 to 50ng of linearized vector with 60 to 300ng of amplified PCR insert(s) in 200μL microtubes containing 10μL of 2x Hot Fusion Buffer: 2x pre-assembly buffer 5x (0.5M Tris pH 7.5, 50mM MgCl₂, 1mM each dNTP, 50mM DTT, 25% PEG-8000) with 0.0075u·μL⁻¹ of T5 exonuclease and 0.05u·μL⁻¹ of Phusion Hot Start DNA polymerase. Distilled water was added to the tube for a final volume of 20μL. Tubes were incubated in a thermocycler for 1 hour at 50°C, then slowly ramped down to 20°C in 5 minutes (0.1°C per second), and held at 10°C. The Hot Fusion reaction was used for transformation or stored at -20°C if not used immediately. Otherwise, 1μL of the reaction was transformed in *E. coli* NEB 10-beta competent cells. Insertion of *UPF1* fragments into vector was verified by restriction enzyme digestion and sequencing.

Coding sequences of yeast full length *NMD4*, yeast *UPF1* helicase domain, and human *UPF1* helicase domains (Uniprot accession codes Q12129, P30771, and Q92900-2 respectively) were cloned between *NheI/NotI*, *NheI/XhoI*, and *NdeI/XhoI* in variants of pET28a (Novagen). In these vectors the *NcoI-NdeI* cassette was either deleted by mutating the *NcoI* site to an *NdeI* site or replaced by the coding sequence for a His-tag or a CBP tag followed by a TEV protease cleavage site. In the vector without the *NcoI-NdeI* cassette, a TEV protease cleavage site was engineered in front of the C-terminal His-tag. Fragments of human *UPF1* (helicase domain, 195-914), yeast *UPF1* (helicase domain, residues 220–851), and *NMD4* (Full length, residues 1–883) were amplified by PCR with the appropriate restriction sites, using oligonucleotide pairs HLH725/HLH726, HLH2705/HLH2706 and MD99/MD100 respectively. PCR products were purified using PCR cleanup Qiaquick (*NMD4*) or Promega (human and yeast *UPF1*) kits, digested for 1 hour at 37°C using NEB Cutsmart buffer and the corresponding enzyme couples, then further gel purified using Qiaquick kit (*NMD4*) or Promega gel purification kit (human and yeast *UPF1*). Plasmids pHL5 and pHL4 were digested in parallel using *NheI/NotI* (pHL5), *NheI/XhoI* (pHL4) or *NdeI/XhoI* (pHL4) and purified on 0.8% agarose gels. Digested PCR products were mixed with corresponding plasmids in a 1:3 molar ratio and ligated overnight at 16°C in a 15µl reaction using T4 DNA ligase. Ligase was heat-inactivated 10 minutes at 65°C, then ligation products were used to transform *E. coli* NEB 5-alpha competent cells. Insertion of *NMD4* and *UPF1* fragments into vectors was verified by restriction enzyme digestion and sequencing.

Total protein extracts and western blots

Total protein extracts were prepared from 5 A_{600} of exponential culture with a fast method using alkaline treatment (Kushnirov, 2000). Cells were incubated in 200µL of 0.1M NaOH for 5 min at room temperature, collected by 3 min centrifugation and resuspended in 50µL of sample buffer containing DTT (0.1M). Proteins were denatured for 3 min at 95°C, and cellular debris were pelleted by centrifugation. 10µL of supernatant or diluted supernatant (for quantification scale) were loaded on acrylamide NuPAGE Novex 4-12% Bis-Tris gels (Life technologies). After transfer to nitrocellulose membrane with a semi-dry fast system (Biorad trans-blot) with discontinuous buffer (BioRad technote 2134), proteins were detected by hybridization with appropriate antibodies.

RNA extraction

For RT-qPCR and RNAseq, cells were first grown in YPD to log phase and collected. Total RNA was extracted using the hot phenol extraction method and precipitated using ammonium acetate and ethanol. The extracted RNA samples were treated with DNase I (Ambion TURBO DNA-free kit) before reverse-transcription (RT) and library preparation.

Affinity purification for RNA analysis

We used 4L yeast culture at exponential phase, A_{600} 0.6 to 0.8 and processed with the same method as TAP immuno-purification but with more precaution to work fast, on ice and in an

RNase free environment. In addition, buffers were freshly prepared, lysis buffer contained 5mM MgCl₂ to maintain mRNP integrity, washing steps were reduced to three washes with HKI + DTT buffer and elution was done in HKI + DTT + AcTEV. After elution, RNA extraction was done by 3 steps of acid phenol/chloroform followed by overnight ammonium acetate precipitation at -20°C. RNA pellet were resuspended in water and processed for DNase treatment (Ambion TURBO DNA-free kit) and RT-qPCR analysis.

At the first RNA extraction step, we also collected proteins that were at the interface between aqueous phase (RNA) and organic phase. Proteins were precipitated by adding 100% ethanol, incubating at -20°C for 1h and centrifugation. Washed and dried protein pellets were used for western blotting to verify purification efficiency.

Dcp2 depletion strain

The Dcp2-degron strain derives from BY4741, except that the *DCP2* ORF is followed by the polyG-miniAID-KANMX-TIR1 cassette (adapted from Nishimura *et al*, 2009; Kubota *et al*, 2013). The addition of the cassette leads to the inducible poly-ubiquitination of Dcp2 protein and its degradation. The cassette has been added by transformation of the BY4741 strain with two PCR products; a first one amplifying the polyG-miniAID part with half of the kanamycin resistance cassette from the plasmid 1451 and the second one amplifying the *OsTIR1* gene and the second half of the kanamycin resistance cassette from the plasmid 1367. The polyG is used as a spacer between the Dcp2 ORF and the miniAID element. The miniAID comes from *A. thaliana* AID element (AtIAA17 amino acids 65–132), it is the auxin/IAA inducing degron element. *OsTIR1* is the plant-specific gene from *Oryza sativa*, in our construction it is transcribed from the same chromosomal region as Dcp2 under the control of the ADE2 promoter. *TIR1* gene product interacts with the miniAID fused to Dcp2 in one hand and with a group of protein including an ubiquitin ligase in the other hand. The ubiquitin ligase will poly-ubiquitinate the miniAID domain triggering the degradation of Dcp2-miniAID by the proteasome.

The strain has been tested by western blot with or without IAA at 100 µM for 1 hour and by growth tests on YPD+/- IAA plates. Depletion of Dcp2 was complete after 1h of auxin (IAA, Indole-3-acetic acid, ref: SERVA 26181.01) treatment. We also observed a strong growth defect after auxin treatment.

For the qPCR experiment, 50mL of YPD were inoculated with LMA4423. Once the culture has reached an optical density of 0.5 we divided the culture in two equal parts, and added IAA at a concentration of 100µM to one of the cultures. After 1 hour the cells were collected by centrifugation, and we proceeded with RNA extraction, DNase treatment and RT-qPCR.

Protein expression and purification from *E. coli*

Plasmids pHL1484 (CBP-*NMD4* full length-6His), pHL1301 (yeast *UPF1* helicase domain-6His) and pHL201 (human *UPF1* helicase domain-6His) were used to transform BL21 (DE3)

Codon Plus competent cells. After plating and overnight cell growth on Kanamycin LB plates, 3-4 colonies were inoculated in 25 ml LB media supplemented with Kanamycine (50mg.L⁻¹). After 6 hours of incubation at 37°C, each starter culture was used to inoculate a 1L LB culture supplemented with Kanamycine (50mg.L⁻¹) and Chloramphenicol (34mg.L⁻¹). Cultures were first incubated at 37°C to a 0.6 optical density at 600nm wavelength, then transferred and incubated at 16°C. When optical density reached 0.8, protein expression was induced using IPTG (0.5mM).

After overnight incubation at 16°C, cells were harvested at 6000 rpm for 10 minutes, washed once with cold 1x PBS, then collected at 5000 rpm for 10 minutes.

Cells were mixed in a lysis buffer (1.5x PBS, 20mM Imidazole, 0.1% Igepal, 10% Glycerol, 1mM MgCl₂) supplemented with protease inhibitors (Aprotinin 2µg.mL⁻¹, Leupeptin 1µg.mL⁻¹, Pepstatin 1µg.mL⁻¹, PMSF 50µg.mL⁻¹) then lysed using sonication for 4 minutes on ice, and centrifuged at 18000 rpm for 25 minutes at 4°C.

Supernatants were mixed with 500µl Ni-NTA agarose resin (=500µl of 50% slurry, Clontech) pre-equilibrated in lysis buffer, and incubated for 2 hours in 50ml falcons on a rotator at 4°C. Beads were collected and washed with lysis buffer then transferred on Bio-Spin columns (Biorad) pre-washed in lysis buffer. Columns were further washed with lysis buffer followed by a wash buffer (1.5x PBS, 50mM Imidazole, 250mM NaCl, 0.1% Igepal, 10% Glycerol, 1mM MgCl₂), before protein elution in 800 ul fractions (1.5x PBS, 150mM Imidazole, 10% Glycerol, 1mM MgCl₂).

Nmd4 was further dialyzed against calmodulin binding buffer (1x PBS, 100mM NaCl, 0.1% Igepal, 10% Glycerol, 1mM MgCl₂, 4mM CaCl₂, 1mM DTT) overnight at 4°C in Spectrapor-4 (12-14 MWCO), then mixed with 500µl Calmodulin Affinity Resin (= 1ml of 50% slurry, Agilent) pre-equilibrated in binding buffer. After 1 hour incubation into a Bio-Spin column on a rotator at 4°C, beads were washed twice with binding buffer, before protein elution (1x PBS, 100mM NaCl, 0.05% Igepal, 10% Glycerol, 1mM MgCl₂, 20mM EGTA, 1mM DTT).

Proteins were finally dialyzed against 1.5x PBS, 150mM NaCl, 10% (w/v) glycerol, 1mM DTT and 1mM MgCl₂ in Spectrapor-4 (12-14 MWCO) then stored at -80°C.

In vitro pull-down assays

Pull-down was performed using preblocked calmodulin affinity beads (Agilent).

Preblocking beads

Briefly, in order to preblock beads, 1 ml calmodulin sepharose beads (50% Slurry) were spun, and resuspended in 20mM Hepes, 500mM NaCl, 0.1% Igepal, 0.08mg.ml⁻¹ glycogen carrier, 0.08mg.ml⁻¹ tRNA and 0.8mg.ml⁻¹ BSA. After 2 hours at 4°C, beads were washed 3 times (20mM Hepes, 150mM NaCl, 0.1% Igepal) then resuspended in 500µl 1x binding buffer

250/10 (20mM Hepes, 250mM NaCl, 0.1% Igepal, 2mM MgAc₂, 2mM CaCl₂, 10% glycerol, 1mM DTT).

In vitro pull-down assay

Recombinant CBP-Nmd4, yeast Upf1 and human Upf1 proteins were thawed on ice. Each mix contained 2µg of each protein in 30µl reaction mixes. NaCl and glycerol concentrations were adjusted to 150mM and 15% respectively. Five microliters (1/6) aliquots were taken out of each mix to load on Input gel. Each mix was further supplemented with 5µl of water, and 30µl of a NaCl/glycerol buffer to reach a 60µl reaction volume containing 125mM NaCl and 12.5 % glycerol.

Mixes were incubated for 20 minutes at 30°C. To perform pull-downs, 12µl of preblocked calmodulin beads were added to each mix, along with 200µl 1x binding buffer 150/10 (20mM Hepes, 150mM NaCl, 0.1% Igepal, 2mM MgAc₂, 2mM CaCl₂, 10% glycerol, 1mM DTT), then rotated 1 hour at 4°C. Beads were further washed 3 times with 1x binding buffer 150/10, then dried using a thinned Pasteur pipet. The retained complexes were eluted using 20µl elution buffer (10mM Tris pH=7.5, 150mM NaCl, 1mM MgAc₂, 2mM Imidazole, 20mM EGTA, 0.1% Igepal, 14% glycerol, 10mM β-mercaptoethanol) while shaking 5 minutes at 1400 rpm, 30°C.

Eluates were collected after spinning, transferred to fresh tubes, concentrated 30 minutes in a Speed-vac then loaded on 10% SDS-PAGE gels.

Pull-down on streptavidin beads using biotinylated-RNA

The experiments were performed as previously described for similar assays (Fiorini *et al*, 2012). To preblock beads, 300 ul of Dynabeads® MyOne Strept (Invitrogen, 10 mg/ml) were spun, rinsed (20 mM Hepes, 150 mM NaCl, 0.1% Igepal) and resuspended in 20 mM Hepes, 500 mM NaCl, 0.1% Igepal, 0.08 mg/ml glycogen carrier, 0.08 mg/ml tRNA and 0.8 mg/ml BSA. After 2 hours at 4°C, beads were washed 2 times (20 mM Hepes, 150 mM NaCl, 0.1% Igepal) then resuspended in 300 ul 1x storage buffer (10 mM Hepes, 250 mM NaCl, 1mM EDTA).

Recombinant NMD4 and yeast UPF1 HD proteins were thawed on ice. Each mix contained 2 ug of each protein in 30 µl reaction mixes. NaCl and glycerol concentrations were adjusted to 150 mM and 15% respectively. 5 µl (1/6) aliquots were taken out of each mix to load on Input gel. Each mix was further supplemented with 1 µl biotinylated RNA (1 µl at 10 µM) and 4 µl of water, and 30 µl of a NaCl/glycerol buffer to reach a 60 µl reaction volume containing 125 mM NaCl and 12.5 % glycerol.

Mixes were incubated for 20 minutes at 30°C. To perform pull-downs, 6 ul of preblocked streptavidin beads were added to each mix, along with 200 ul 1x binding buffer 150/10 (20

mM Hepes, 150 mM NaCl, 0.1% Igepal, 2 mM MgAc₂, 10% glycerol, 1 mM DTT), then rotated 1 hour at 4°C. Beads were further washed 3 times with 1x binding buffer 150/10 or 300/10 (same composition, but 300 mM NaCl instead 150 mM). Since the beads are magnetic, supernatant is removed during each wash by placing the tubes on a magnetic tube holder.

The complexes retained on the beads via the biotinylated RNA were eluted using 7.5 µl 1x SDS loading dye added directly on the beads. Tubes were flicked to suspend the beads, then spun 1 min at 3000 rpm. Eluates were carefully collected without the beads and transferred to fresh tubes, then boiled 3 min at 94°C. Eluates and input samples were loaded on 10% SDS-PAGE gels.

NMD efficiency calculation details

Complementation of an NMD deficient strain leads to a decrease in the levels of an NMD-sensitive transcript like RPL28. To be able to judge how the efficiency of NMD affects the level of a transcript and to simplify the interpretation of complementation experiments, we developed a formula to calculate NMD efficiency using the maximum and minimum steady-state levels of a given transcript. We assume that RNA synthesis is constant and that the degradation of RNA occurs by two competitive pathways, with different first-order rate constants: k_{NMD} and k_{base} . Thus, the variation in the RNA levels with time could be expressed as:

$$\frac{d[RNA]}{dt} = T - (k_{NMD} + k_{base}) \times [RNA]$$

where:

T is transcription rate (synthesis, and export),

[RNA] is RNA concentration,

k_{NMD} is the global rate constant for degradation through NMD,

k_{base} is the base rate constant for degradation of the RNA, independent of NMD.

At steady state, the change in RNA concentration is null. A given steady-state level can be obtained either by both high transcription and high degradation or low transcription and low degradation. Steady state levels of a given RNA would be:

$$\frac{d[RNA]}{dt} = T - (k_{NMD} + k_{base}) \times [RNA]$$

Since most of the time we do not know the transcription rate, or the degradation constants, we need to remove one of the variables. Let us consider two situations in which transcription is supposed invariable, a wt and a mutant strain devoid of NMD:

$$[RNA]_{wt} = \frac{T}{k_{NMD} + k_{base}}$$

$$[RNA]_{mut} = \frac{T}{k_{base}}$$

Thus, the ratio between the mutant that has no NMD and wild type, N, is:

$$(1) \quad \frac{[RNA]_{mut}}{[RNA]_{wt}} = N = \frac{k_{NMD} + k_{base}}{k_{base}} = \frac{k_{NMD}}{k_{base}} + 1$$

If the RNA is not an NMD substrate, k_{NMD} is 0 and the ratio between the mutant and wt becomes 1.

If the RNA is exclusively degraded through NMD, the ratio would be infinite.

At a fractional NMD efficiency, noted α , the degradation through NMD would be $\alpha \times k_{NMD}$ and the ratio of RNA towards wt in this mutant, P, becomes:

$$(2) \quad \frac{[RNA]_{obs}}{[RNA]_{wt}} = P = \frac{k_{base} + k_{NMD}}{k_{base} + \alpha k_{NMD}} = \frac{1 + \frac{k_{NMD}}{k_{base}}}{1 + \alpha \frac{k_{NMD}}{k_{base}}}$$

N is always larger than P.

We can substitute k_{NMD}/k_{base} in (2) with $(N-1)$, from (1):

$$\frac{[RNA]_{obs}}{[RNA]_{wt}} = P = \frac{k_{base} + k_{NMD}}{k_{base} + \alpha k_{NMD}} = \frac{1 + \frac{k_{NMD}}{k_{base}}}{1 + \alpha \frac{k_{NMD}}{k_{base}}}$$

$$\frac{N}{P} = 1 + \alpha(N - 1)$$

Thus the fraction of NMD, used to define efficiency, can be calculated from:

$$\alpha = \frac{\frac{N}{P} - 1}{N - 1}$$

Where N is the enrichment of RNA in the NMD null mutant and P the enrichment in the tested strain.

Example: if the RNA is enriched 8 fold in an NMD mutant and 4 fold in a "partial" NMD strain over WT, the fractional efficiency of NMD would be 1/7, 0.14 or 14%. This example shows that the correlation between observed ratios and NMD efficiency is not linear. The formula maps values obtained by RT-QPCR to a percentage of "NMD efficiency" and works fine for extreme values. If a mutant has no NMD defect, P will be equal to 1 and α becomes 1. If the mutant is 100% NMD deficient, P will approach N and, α becomes 0. Due to experimental error, P might be superior to N, but in that case α should be considered 0.

Protein sequence alignment and percentage of identity calculation

We used metaPhOr (Pryszcz *et al*, 2011) to obtain sequences of orthologues for the studied proteins. We aligned entire proteins or fragments using Mafft software called as a web service from Jalview. PIN domain boundaries for Smg6, Smg5 have been determined by sequence homology with orthologous proteins from *D. melanogaster*, *C. elegans*, *M. musculus* (positions 1239-1419 for hSmg6, 849-1016 for hSmg5). For Nmd4, we used all the 1-218 amino acid sequence for the alignment, even if the last 60 amino acids are not part of the PIN domain. Identity was computed as the percentage of identical aligned residues over the total number of aligned residues. To strengthen alignment reliability, we also performed delta-BLAST alignment (Boratyn *et al*, 2012) on the *S. cerevisiae* protein database with the hSmg6 and hSmg5 PIN domains and hSmg7 14-3-3 domain as query, using E-values as significance scores.

GO Term analysis and statistics

We used the GO Term Finder tool associated with SGD database (Cherry *et al*, 2012) to search for common function of protein list extracted from MS analysis; for example, the group of proteins removed by RNase treatment and group of proteins of a given enrichment; default settings were used. We calculated the significance of the accumulation of certain protein classes using hypergeometric distribution test.

RNA libraries preparation and sequencing

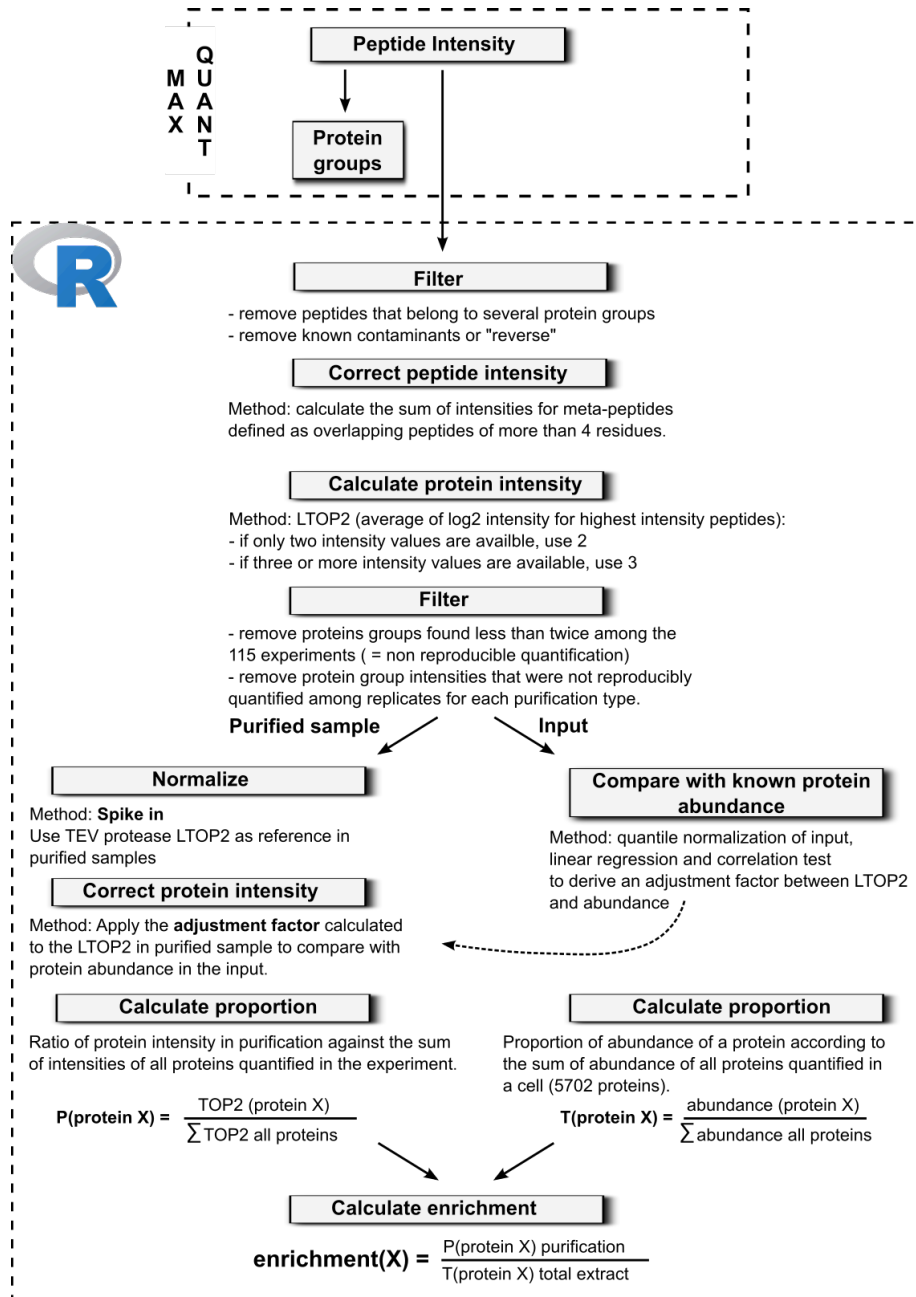
RNA, extracted and treated with DNase, were subjected to a RiboZero treatment to remove ribosomal RNAs. Libraries of mRNA were prepared with the TruSeq Stranded kit (Illumina). Libraries were sequenced on the Illumina HiSeq 2500 for 50 bases (for *upf1Δ* and BY4741) and 65 bases (for *nmd4Δ*, *ebs1Δ*, *nmd4Δ/ebs1Δ* and BY4741). Sequenced reads were aligned to yeast genome version sacCer3 (version R64-2-1) using STAR (Dobin *et al*, 2013) with the default parameters except for -s 0 -0. We used "featureCounts" (Liao *et al*, 2014) from the subread package (version 1.5.0-p3-Linux-x86_64) to count the number of reads per features. For the features, we used a custom list of coordinates where unstable transcripts, as well as introns could be specifically counted. Three independent experiments were performed for each condition.

RNAseq statistical analysis

Expression of gene in the different samples and replicates were normalized to WT strain using DESeq2 (Love *et al*, 2014). We analyzed the three replicates independently and used the mean for figures. During DESeq2 workflow, we removed features identified by zero reads to avoid problems with the logarithm transformation. We determined NMD substrates from *upf1Δ* RNAseq results and fixed the threshold for these RNA to a minimum of 1.4 fold increase. For the bin analysis, we used a custom script to divide RNA sequenced in *nmd4Δ* and *ebs1Δ* experiments into 5 bins containing the same number of transcript and calculate the

percentage of NMD substrates for each bin. Statistical significance of the differences between bins was assessed using a binomial test.

Appendix Figures



Appendix Fig. S1: Workflow for quantitative analysis of MS/MS results from affinity purified complexes

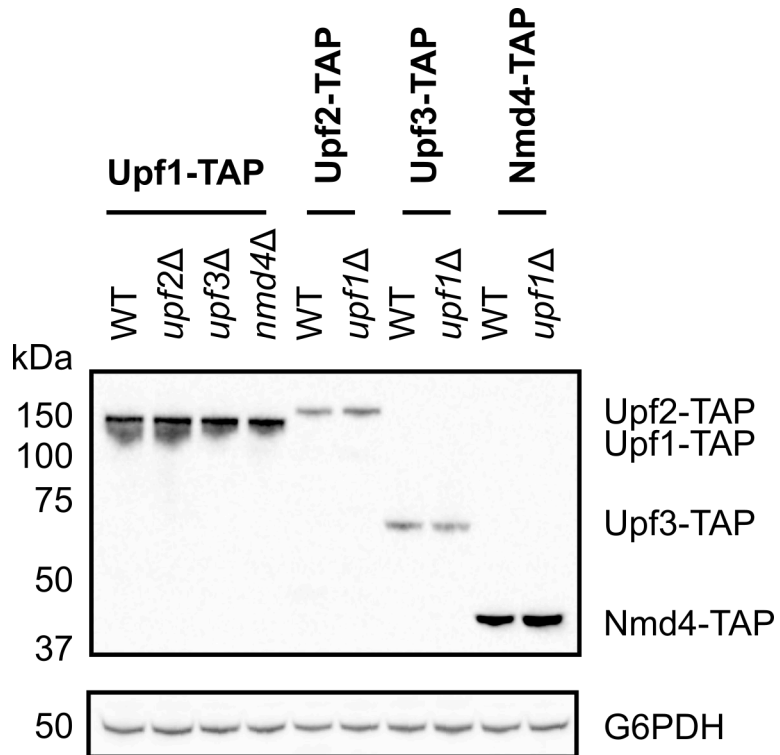
Notes about the depicted procedure: 1. MaxQuant output for peptide intensities and their association with a given protein was used as the main input for computing LTOP2 scores and enrichment. To calculate the false discovery rate (FDR) of the MS/MS analysis, MaxQuant builds reverse sequence « artificial » proteins that serve as negative controls for the identification procedure. Reversed

sequences and common contaminants (trypsin, keratins) were removed from further analyses in the early steps of the analysis.

2. A protein group corresponds to a single protein or several proteins with very high sequence similarity that cannot be discriminated by peptide analysis. For further analyses, we used the identity of the protein of the group with most coverage.

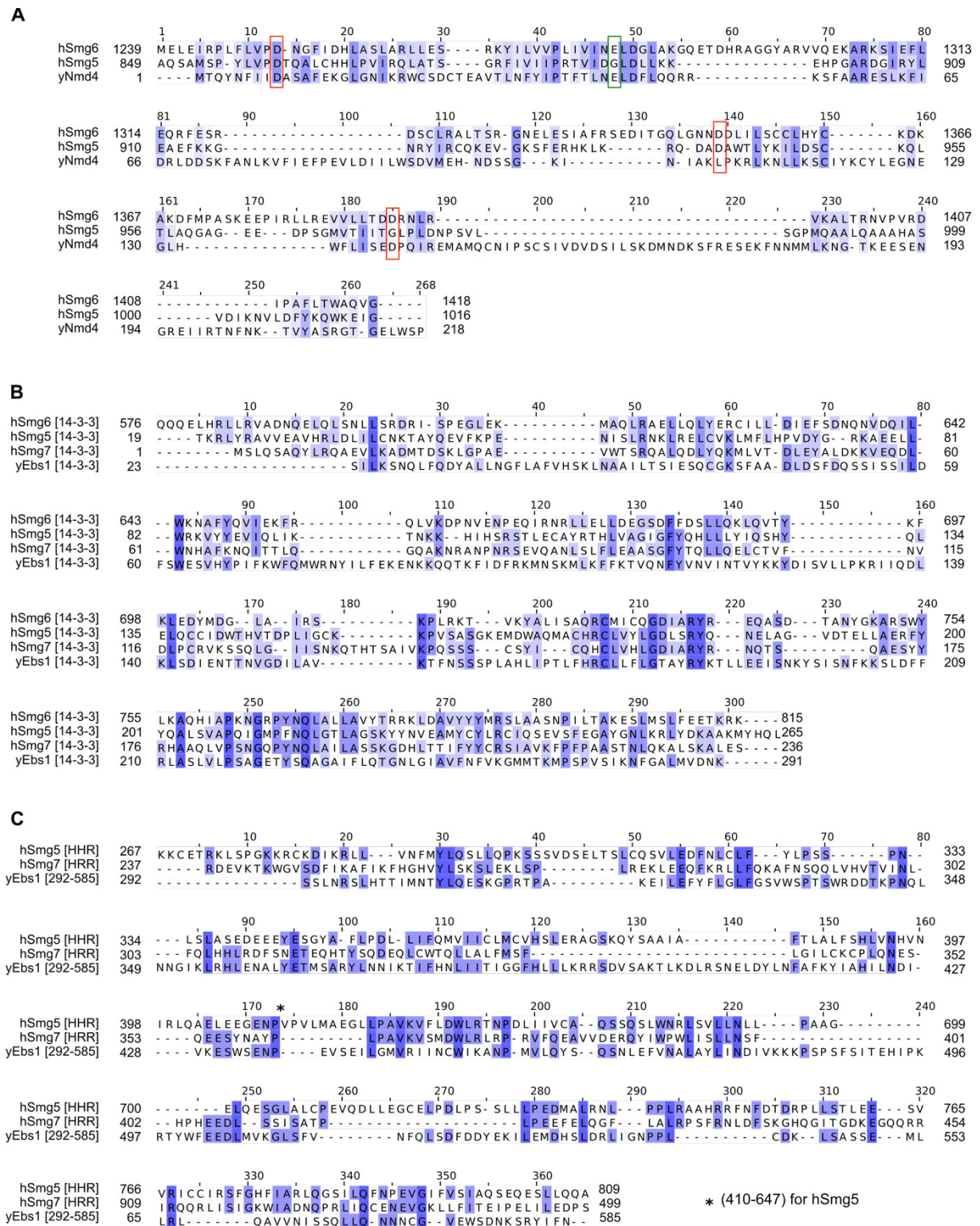
3. The TEV protease was added in each purification experiments with the same relative amount to elute complexes from beads. This step is important to be able to compare replicates between them and the different purification types.

4. The comparison of our input LTOP2 with the abundance data from Ho et al. 2017, allowed to calculate a factor that served to adjust LTOP2 values and make them compatible with the dynamic range and scale of published abundance values.



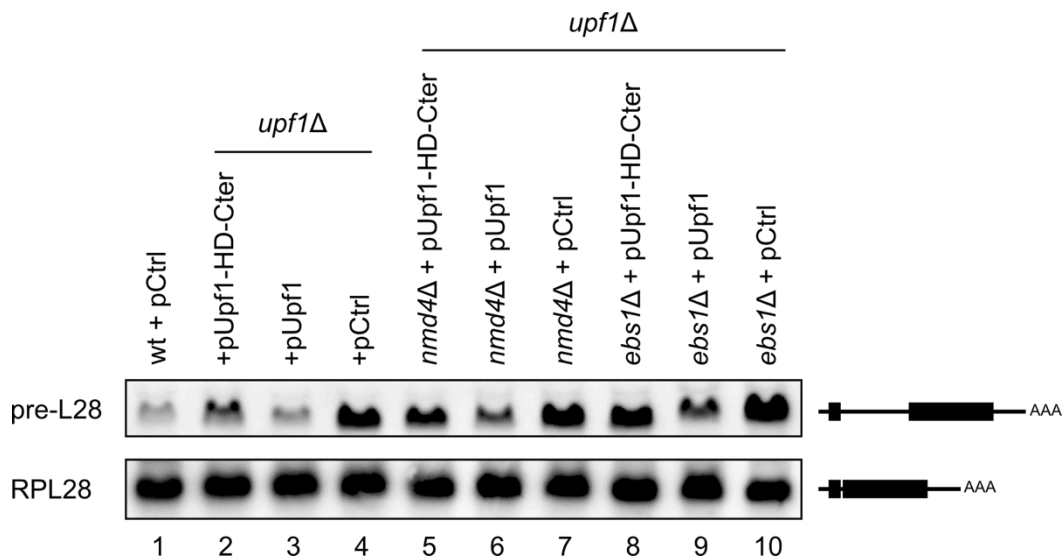
Appendix Fig. S2: Controls of total tagged protein levels in the presence or absence of other NMD components

Total protein extracts from the described strains were tested by immunoblot to detect the protein A part of the TAP tag. G6PDH signal was used as a loading control.



Appendix Fig. S3: Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences

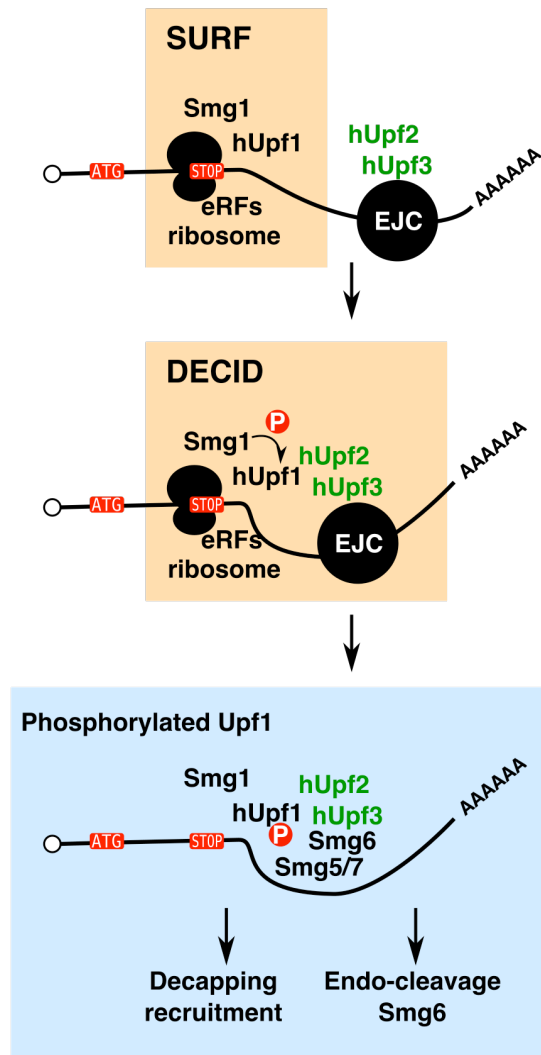
Alignment of the PIN domains of hSmg6, hSmg5 and ScNmd4 (A), the 14-3-3 domains of hSmg6, hSmg5, hSmg7 and Ebs1 (B) and of the helical hairpin region (HHR) of hSmg5, hSmg7 and Ebs1 (292-585) (C). These alignments were obtained with the algorithm Mafft with default parameters; colour represents percentage of identity or similarity (BLOSUM62).



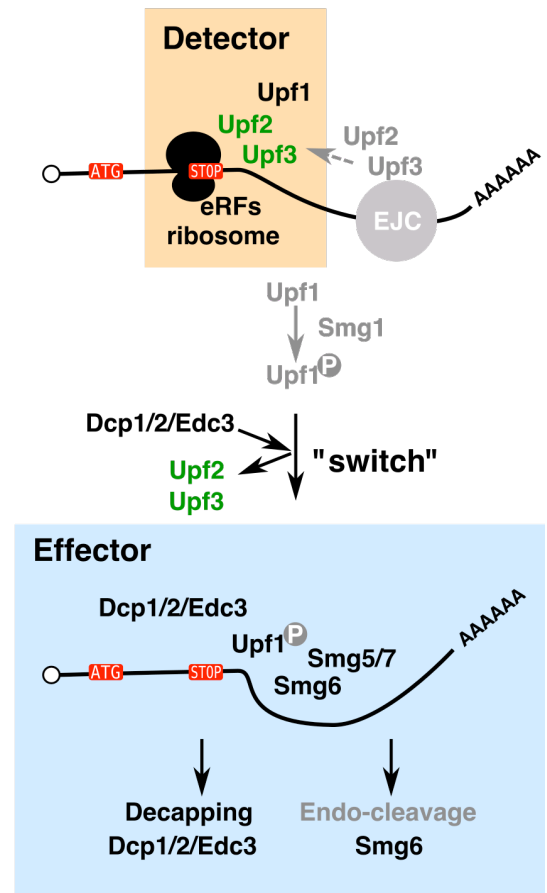
Appendix Fig. S4: Effect of NMD4 and EBS1 in the destabilization of an NMD substrate by the Upf1-HD-Cter domain

Northern blot estimation of the complementation of *upf1Δ* and *upf1Δ/nmd4Δ* and *upf1Δ/ebs1Δ* strains with a control plasmid (pCtrl), full-length Upf1 (pUpf1) or the HD-Cter fragment (pUpf1-HD-Cter). Digoxigenin containing PCR probes specific to the intron of RPL28 mRNA precursor and for the mature form were used with anti-digoxigenin peroxidase coupled detection of the two RNA forms.

A SURF/DECID NMD model

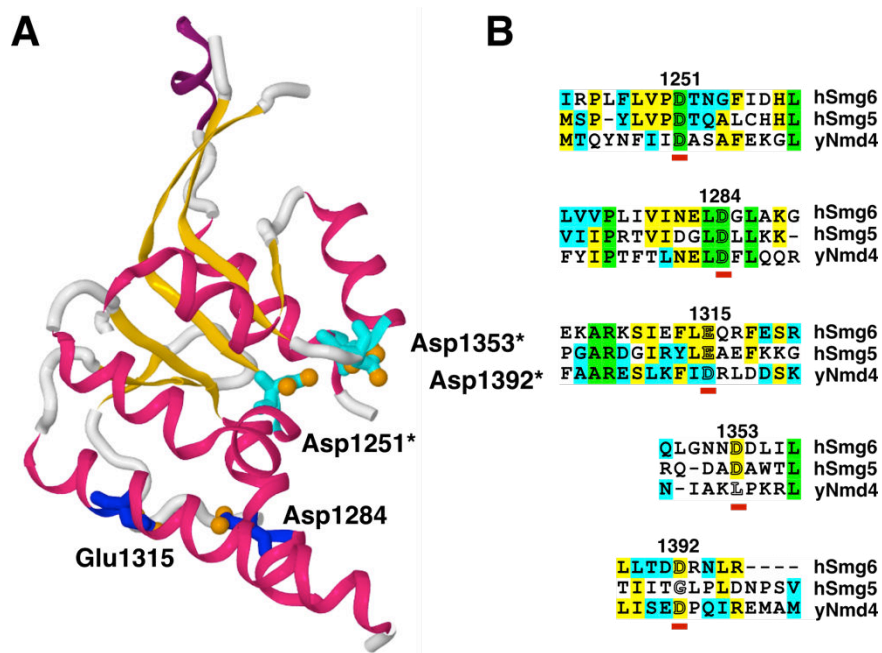


B Extended Detector/Effector model



Appendix Fig. S5: Comparison between the canonical SURF/DECID model features (A) and our extended Detector/Effector model (B) for NMD.

Orange and blue squares mark equivalent steps in both models. Light grey elements in the revised model represent optional steps that can further enhance the NMD process under certain conditions and in specific organisms.



Appendix Fig. S6. Similarities and differences between yeast Nmd4 PIN and PIN domains of human SMG6 and SMG5

(A) Structure of the PIN domain of hSmg6 (Glavan *et al*, 2006, PDB entry 2HWW), highlighting conserved aspartate and glutamate residues. The Asp residues required for catalysis (Asp1251, 1353, 1392) are depicted with an asterisk. (B) Zones of sequence alignment between the human sequences and the PIN domain of yeast Nmd4 centered on conserved acidic residues are underlined. Unlike the PIN domain of hSmg5, with no catalytic activity, the PIN domain of Nmd4 retains a conserved Asp in a position equivalent with D1392 in hSmg6. Residue numbers correspond to positions in the human Smg6 protein (Uniprot Q86US8).

Appendix Tables

over-representation
(hypergeometric text)

Experiment	strain	number of replicates with/without RNase		number of quantified proteins		number of RNase sensitive proteins	constituents of ribosome (231)		RNA binding (847)	
		NO	YES	NO	YES		count	p-value	count	p-value
BY4741	WT	5	4	9	0	/	/	/	/	/
Upf1-TAP	WT	6	3	270	45	225	47	5E-27	80	7E-22
Upf2-TAP	WT	6	3	172	62	111	30	2E-21	44	8E-15
Upf3-TAP	WT	5	3	217	70	151	25	1E-12	62	4E-21
Nmd4-TAP	WT	4	3	121	32	89	34	7E-30	37	1E-13
Ebs1-TAP	WT	3	3	195	133	79	19	3E-13	24	2E-06
Dcp1-TAP	WT	3	3	255	84	175	19	7E-07	66	6E-20
Hrr25-TAP	WT	4	3	203	116	99	15	8E-08	38	1E-12
TAP-Upf1-2-971	upf1Δ	3	3	457	214	247	24	3E-07	74	1E-15
TAP-Upf1-2-853	upf1Δ	2	0	223	/	/	/	/	/	/
TAP-Upf1-208-853	upf1Δ	2	0	235	/	/	/	/	/	/
TAP-Upf1-2-208	upf1Δ	3	3	208	59	149	37	2E-24	60	6E-20
TAP-Upf1-208-971	upf1Δ	3	3	429	104	325	41	3E-15	99	6E-21
Upf1-TAP	upf2Δ	3	3	200	42	161	46	4E-33	56	3E-15
Upf1-TAP	upf3Δ	3	3	242	50	192	37	3E-20	71	1E-20
Upf2-TAP	upf1Δ	2	3	242	86	157	20	2E-08	45	2E-09
Upf3-TAP	upf1Δ	2	0	235	/	/	/	/	/	/
Upf1-TAP	nmd4Δ	3	3	140	35	107	44	6E-40	41	2E-13
Nmd4-TAP	upf1Δ	2	2	235	80	165	32	9E-18	64	4E-20

Appendix Table S1: Summary of the number of replicates, number of proteins robustly quantified and number of proteins lost after RNase treatment.

The over-representation of RNA binding factors in the 'lost' fraction was computed using the hypergeometric distribution.

Experiment	number of replicates	replicates	total reads	number of uniquely mapped reads	reads assigned to known features
BY4741(GB)	3	BY4741_GB_rep1	30 670 583	25 581 713	25 401 415
		BY4741_GB_rep2	25 220 407	20 672 565	20 528 503
		BY4741_GB_rep3	29 245 177	25 083 674	24 903 966
upf1Δ(GB)	3	upf1-delta_rep1	24 224 426	20 728 624	20 501 571
		upf1-delta_rep2	30 152 228	24 816 935	24 548 407
		upf1-delta_rep3	28 395 388	21 711 919	21 468 870
BY4741	3	BY4741_rep1	21 764 071	17 906 929	17 733 252
		BY4741_rep2	29 781 174	13 645 348	13 502 966
		BY4741_rep3	40 044 726	28 578 567	28 287 975
nmd4Δ	3	nmd4-delta_rep1	23 282 316	19 201 253	19 003 405
		nmd4-delta_rep2	33 352 825	27 228 434	26 935 166
		nmd4-delta_rep3	32 569 200	25 866 315	25 600 752
ebs1Δ	3	ebs1-delta_rep1	25 183 381	20 795 767	20 583 050
		ebs1-delta_rep2	30 637 150	25 237 059	24 978 404
		ebs1-delta_rep3	29 713 531	22 452 058	22 197 467
nmd4Δ/ ebs1Δ	3	nmd4-ebs1-delta_rep1	23 253 638	19 428 559	19 219 804
		nmd4-ebs1-delta_rep2	23 635 820	20 419 649	20 199 665
		nmd4-ebs1-delta_rep3	37 828 620	30 145 648	29 809 901

Appendix Table S2: RNA sequencing raw data analysis summary

strain	original strain	Genotype	Reference
LMA2154 (BY4741)		<i>MATa ura3Δ0 his3Δ1 leu2Δ0 met15Δ0</i>	Brachmann et al. 1998
LMA2155 (BY4742)		<i>MATa ura3Δ0 his3Δ1 leu2Δ0 lys2Δ0</i>	Brachmann et al. 1998
LMA2194 (UPF1-TAP)	BY4741	NAM7-TAP::HIS3MX	Ghaemmaghami et al., 2003
LMA3730 (UPF1-TAP)	BY4741	NAM7-CRAP::URA3	This study
LMA2192 (UPF2-TAP)	BY4741	NMD2-TAP::HIS3MX	Ghaemmaghami et al., 2003
LMA3731 (UPF2-TAP)	BY4741	NMD2-CRAP::URA3	This study
LMA2193 (UPF3-TAP)	BY4741	UPF3-TAP::HIS3MX	Ghaemmaghami et al., 2003
LMA4263 (UPF3-TAP)	BY4741	UPF3-CRAP::URA3	This study
LMA3317 (NMD4-TAP)	BY4741	NMD4-TAP::HIS3MX	Ghaemmaghami et al., 2003
LMA3728/3729 (NMD4-TAP)	BY4741	NMD4-CRAP::URA3	This study
LMA4264 (DCP1-TAP)	BY4741	DCP1-TAP::HIS3MX	Ghaemmaghami et al., 2003
(EBS1-TAP)	BY4741	EBS1-CRAP::URA3	This study
LMA3312 (HRR25-TAP)	BY4741	HRR25-TAP::HIS3MX	Ghaemmaghami et al., 2003
LMA3849 (UPF1-TAP/NMD4-HA)	BY4741	NAM7-CRAP::URA3 / NMD4-HA::KANMX6	This study
LMA3851 (UPF1-TAP/EBS1-HA)	BY4741	NAM7-CRAP::URA3 / EBS1-HA::KANMX6	This study
LMA3852 (UPF1-TAP/EDC3)	BY4741	NAM7-CRAP::URA3 / EDC3-HA::KANMX6	This study
LMA1667 (<i>upf1Δ</i>)	BY4741	NAM7::KANMX6	Giaever et al., 2002
LMA1669 (<i>upf2Δ</i>)	BY4741	NMD2::KANMX6	Giaever et al., 2002
LMA1671 (<i>upf3Δ</i>)	BY4741	UPF3::KANMX6	Giaever et al., 2002
LMA3732 (<i>nmd4Δ</i>)	BY4741	NMD4::KANMX6	Giaever et al., 2002
LMA4112 (UPF2-TAP/ <i>upf1Δ</i>)	BY4741	NAM7::KANMX6 / NMD2-CRAP::URA3	This study
LMA4113 (UPF3-TAP/ <i>upf1Δ</i>)	BY4741	NAM7::KANMX6 / UPF3-CRAP::URA3	This study
LMA4114 (NMD4-TAP/ <i>upf1Δ</i>)	BY4741	NAM7::KANMX6 / NMD4-CRAP::URA3	This study
LMA3739 (UPF1-TAP/ <i>upf2Δ</i>)	BY4741	NMD2::KANMX6 / NAM7-CRAP::URA3	This study
LMA4701/4702 (NMD4-TAP/ <i>upf2Δ</i>)	BY4741	NMD2::KANMX6 / NMD4-CRAP::URA3	This study
LMA3735 (UPF1-TAP/ <i>upf3Δ</i>)	BY4741	UPF3::KANMX6 / NAM7-CRAP::URA3	This study
LMA3736 (UPF1-TAP/ <i>nmd4Δ</i>)	BY4741	NMD4::KANMX6 / NAM7-CRAP::URA3	This study
LMA4523 (<i>upf1Δ/upf2Δ</i>)	BY4741	NAM7::HIS3MX / NMD2::KANMX6	This study
LMA4524 (<i>upf1Δ/upf3Δ</i>)	BY4741	NAM7::HIS3MX / UPF3::KANMX6	This study
LMA4525 (<i>upf1Δ/ebs1Δ</i>)	BY4741	NAM7::HIS3MX / EBS1::KANMX6	This study
LMA4678 (<i>upf1Δ/nmd4Δ</i>)	BY4741	NAM7::HIS3MX / NMD4::KANMX6	This study
LMA3853/3854 (<i>nmd4Δ/ebs1Δ</i>)	BY4742	NMD4::ProMFA α 2NAT / EBS1::KANMX6	This study

Appendix Table S3 (strains)

Plasmid	Description and ID (yeast marker)	Reference
pRS316	pRS316 (URA3)	Sikorski & Hieter, 1989
pCM189-NTAP	pl.1233 (H1) (URA3)	This study
pCM189-NTAP-NAM7-FL	pl.1442 (TAP-UPF1-FL) (URA3)	This study
pCM189-NTAP-NAM7-CH	pl.1443 (TAP-UPF1-CH) (URA3)	This study
pCM189-NTAP-NAM7-NoCH	pl.1444 (TAP-UPF1-Cter) (URA3)	This study
pCM189-NTAP-Upf1 2-853	pl.1521 (TAP-UPF1-NoSQ) (URA3)	This study
pCM189-NTAP-Upf1 208-853	pl.1522 (TAP-UPF1-Cter-NoSQ) (URA3)	This study
pDEST14-NAM7	pl.1350 – Gateway destination vector	This study
pDONR201-NAM7	pl.1330 – Gateway source vector	This study
pCRBlunt-CRAP(6-HisTAP)	pl.1287 – TAP to CRAP cassette vector	This study

Appendix Table S4 (plasmids)

Oligonucleotide	Usage	Sequence
CS887_fw_RPL28intron	real time PCR forward primer	CCATCTCACTGTTGAGACGG
CS888_rv_RPL28intron	real time PCR reverse primer	CTCAGTTTGCGATGGAAGAG
CS889_rv_RPL28exon2	real time PCR reverse primer	ATGTTGACCACCGGCCATAC
CS946_RPL28_fw_exonoverlapQPCR	real time PCR forward primer	TCACGTCTCAGCCGGTAAAG
CS1076_fw_RIM1Qex1ex2	real time PCR forward primer	GTTAGAAAAGGCGCTTTGGTATATG
CS1077_rv_RIM1QRTex2	real time PCR reverse primer	AACCGTCGTCTCTCTCGAAG
CS1429_DAL7_fwQ	real time PCR forward primer	TGAAACTTTGCCAGCGGCCTTC
CS1430_DAL7_rvQ	real time PCR reverse primer	TCCCAACGACCACAGTTCAAACC
CS1359_fw_NAM7_2_pTM189Not	construction oligonucleotide	ttaagaaaatctcatctccggggcacttGATgcgGTCGGTTCCGGTTCTCACAC
CS1361_rv_NAM7_208_pTM189Notv1	construction oligonucleotide	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTAAT TGGATCTCCATTTGCCTC
CS1362_fw_NAM7_s208_pTM189Not	construction oligonucleotide	ttaagaaaatctcatctccggggcacttGATgcgAATAAAGACGCTACAAT TAATGATATTGACG
CS1364_rv_NAM7_971_pTM189Notv2	construction oligonucleotide	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTATA TTCCCAAATTGCTGAAGTC
CS1393_rv_UPF1_853STOP_pTM189Notv2	construction oligonucleotide	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTActgaggacgaactaattgaac

Appendix Table S5 (oligonucleotides)

Appendix references

- Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ & Madden TL (2012) Domain enhanced lookup time accelerated BLAST. *Biol Direct* **7**: 12
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, et al (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**: D700-705
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21
- Fiorini F, Bonneau F & Le Hir H (2012) Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. *Meth. Enzymol.* **511**: 255–274
- Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O’Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425**: 737–741
- Glavan F, Behm-Ansmant I, Izaurralde E & Conti E (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J.* **25**: 5117–5125
- Kubota T, Nishimura K, Kanemaki MT & Donaldson AD (2013) The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. *Mol. Cell* **50**: 273–280
- Kushnirov VV (2000) Rapid and reliable protein extraction from yeast. *Yeast* **16**: 857–860
- Liao Y, Smyth GK & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930
- Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T & Kanemaki M (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**: 917–922
- Pryszcz LP, Huerta-Cepas J & Gabaldón T (2011) MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res* **39**: e32
- Vizcaino JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu Q-W, Wang R & Hermjakob H (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**: D447-456

Discussion générale et perspectives

I. Résumé des résultats

Ce travail de thèse a permis d'explorer les interactions protéines-protéines au cours du mécanisme de NMD. Comme discuté dans l'article, l'utilisation de la spectrométrie de masse quantitative ainsi que l'optimisation de son analyse a permis non seulement d'identifier de nouveaux partenaires mais aussi de disséquer les relations entre eux. Nous avons caractérisé deux complexes protéiques distincts : le **Détecteur** composé d'Upf1, Upf2 et Upf3 et l'**Effecteur**, rassemblant Upf1, Nmd4, Ebs1 et la machinerie de décapping Dcp1, Dcp2 et Edc3. Nous avons montré que l'action de ces deux complexes était successive sur les ARNs substrats du NMD. Le Détecteur est impliqué dans des étapes précoces du NMD, il se forme au niveau de l'ARN par l'assemblage de l'hétérodimère Upf2-Upf3 et d'Upf1 recruté indépendamment (**Fig. 3 et 6 de l'article**, Atkin et al., 1997). L'Effecteur est assemblé sur l'ARN après la dissociation du Détecteur. Upf1 est alors associé à des cofacteurs, stabilisant Upf1 sur les ARNs pour Nmd4 et de fonction inconnue pour Ebs1, qui participent à l'activation de la dégradation des ARNs. La machinerie de décapping recrutée au niveau de l'Effecteur permettrait l'hydrolyse rapide de la coiffe pour permettre l'action de Xrn1.

La question du recrutement d'Upf1 n'a pas abouti à une réponse claire. Nous avons montré que la version sauvage était localisée au niveau des régions 3'-UTR de centaines d'ARNs sensibles au NMD chez la levure. Mais l'utilisation des mutants de l'hélicase d'Upf1 n'a pas encore permis, pour des raisons techniques, de préciser les sites de recrutement de la protéine. Les expériences réalisées ont tout de même précisé les caractéristiques des mutants *in vivo*. Nous avons observé que le mutant TR800AA pouvait interagir faiblement avec l'ARN, que le mutant K436Q conservait de l'affinité pour les ARNs traduit alors que le mutant C122S perdait toutes les interactions du domaine CH ainsi que la liaison à l'ARN. Concernant les interactions protéiques des versions mutantes d'Upf1, excepté pour le mutant C122S, les trois mutants testés dans la région hélicase ont montré un profil d'interactions similaire à la version sauvage.

Les prochaines sections vont discuter ces résultats, en particulier les rôles des nouveaux partenaires potentiels du NMD identifiés en spectrométrie de masse, les modèles pour le switch des complexes Détecteur et Effecteur et le lien entre notre travail et celui de nos collaborateurs de l'équipe « Expression des ARN messagers eucaryotes » à l'ENS.

Pour une bonne compréhension du référencement des figures dans la discussion, les figures issues de l'article sont citées en utilisant la nomenclature **Fig. XA** tandis que les autres figures de la thèse sont citées comme **Figure XA**. Les résultats de spectrométrie de masse de la partie

Il sont visualisables sur une application R/Shiny en ligne à l'adresse suivante : hub05.hosting.pasteur.fr/NMD_complexes, n'hésitez pas à explorer les données.

II. Discussion et perspectives

1. Ubr1 est associé spécifiquement à Upf2

L'analyse des purifications par spectrométrie de masse quantitative nous a permis d'identifier plusieurs nouveaux facteurs potentiels associés aux protéines du NMD. Par exemple, nous avons observé une interaction spécifique et forte entre Upf2 et Ubr1 qui ne dépend pas de l'ARN (**Figure 26**). En absence d'Upf1, un complexe Upf2/Upf3/Ubr1 persiste, indépendamment de l'ARN.

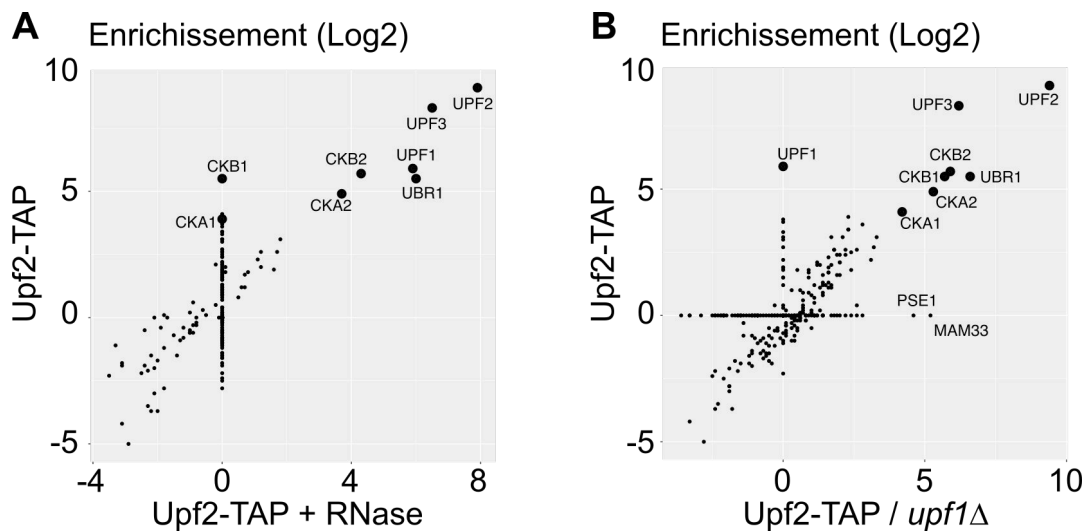


Figure 26 : Précision sur les interactions d'Upf2

A. Impact du traitement RNase sur les interactions d'Upf2-TAP. **B.** Impact de la délétion d'*UPF1* sur les interactions d'Upf2-TAP. Attention, les valeurs indiquées pour l'enrichissement sont en log2 dans cette figure.

Ubr1 est une E3 Ubiquitine ligase (E3) possédant un domaine RING (*Really Interesting New Gene*). Ce type de protéine permet de sélectionner certaines protéines de la cellule et de coordonner leur dégradation par poly-ubiquitylation et adressage au protéasome. Dans le cas des E3-RING, les ubiquitines sont ajoutées par une enzyme de conjugaison de l'ubiquitine E2 sur la protéine sélectionnée par l'E3. Il existe de multiples protéines E3 associées à différentes cibles préférentielles. Ubr1 est impliqué dans plusieurs voies de dégradation : la

voie ERAD (*Endoplasmic-Reticulum Associated Degradation*) et la voie de *N-end rule*. C'est un des facteurs majeurs du contrôle cytoplasmique des protéines (CytoQC : *Cytoplasmic protein Quality Control*) chez la levure.

La voie ERAD prend en charge les protéines mal-repliées dans le réticulum endoplasmique. Celles-ci sont transloquées dans le cytoplasme et, dans des conditions normales, les E3 Hrd1 et Doa10 permettent la fixation des ubiquitines (Ismail and Ng, 2006). Dans ce mécanisme, Ubr1 est redondant avec Hrd1 et Doa10 et prend le relais en cas d'inactivation des deux premières E3 ou en cas de stress important augmentant l'activité ERAD (Stolz et al., 2013). Cette redondance est peut être due au mécanisme récemment décrit, le SHRED (*Stress-induced Homeostatically REgulated protein Degradation*), qui changerait la spécificité de Ubr1 pour les protéines en situation de stress (Szoradi et al., 2018). Ubr1 a aussi un rôle dans l'ubiquitylation des protéines possédant des acides-aminés déstabilisants en N-terminale (*N-end rule*, Hwang et al., 2010a; Xia et al., 2008).

Au vu de la fonction d'Ubr1 dans les cellules, son association à Upf2 pourrait permettre la dégradation active du peptide tronqué traduit à partir de l'ARN aberrant ciblé par le NMD. Le devenir de ces peptides n'est pas clair. La terminaison de la traduction, bien que retardée dans le NMD, utilise les mêmes facteurs que la terminaison normale (**Figure 4C et D**), contrairement au NGD et NSD. Dans ces conditions, le peptide néo-synthétisé d'un ARN aberrant ciblé par le NMD devrait être relâché avant le recyclage normal des ribosomes. Il existe cependant une différence notable entre la terminaison normale et la terminaison à un PTC. Dans le NMD, la distance entre le PTC et la queue poly(A) empêcherait l'interaction entre Pab1 et eRF3, or cette interaction est importante pour l'activation de eRF1 et donc pour l'hydrolyse de la liaison de l'ARNt avec le dernier résidu du peptide néo-synthétisé. L'absence d'interaction Pab1/eRF3 retarde la terminaison (Amrani et al., 2006; Hoshino et al., 1999; Kononenko et al., 2010; Uchida et al., 2002), son impact direct sur le décrochage du peptide néo-synthétisé n'a pas été étudié mais on peut penser que les deux évènements sont liés. On peut alors imaginer un modèle dans lequel Upf2, Upf3 et Ubr1 resteraient associés après le switch d'Upf1 dans l'Effecteur et la dissociation du ribosome. Ubr1, en tant que E3 de la famille des RING, aurait de l'affinité pour le peptide néo-synthétisé et pourrait interagir avec celui-ci alors qu'il n'est pas encore décroché. Ubr1 permettrait ensuite le recrutement des enzymes de conjugaison de l'ubiquitine E2 à proximité pour l'ajout d'une extension poly-ubiquitine et la dégradation du peptide (**Figure 27**).

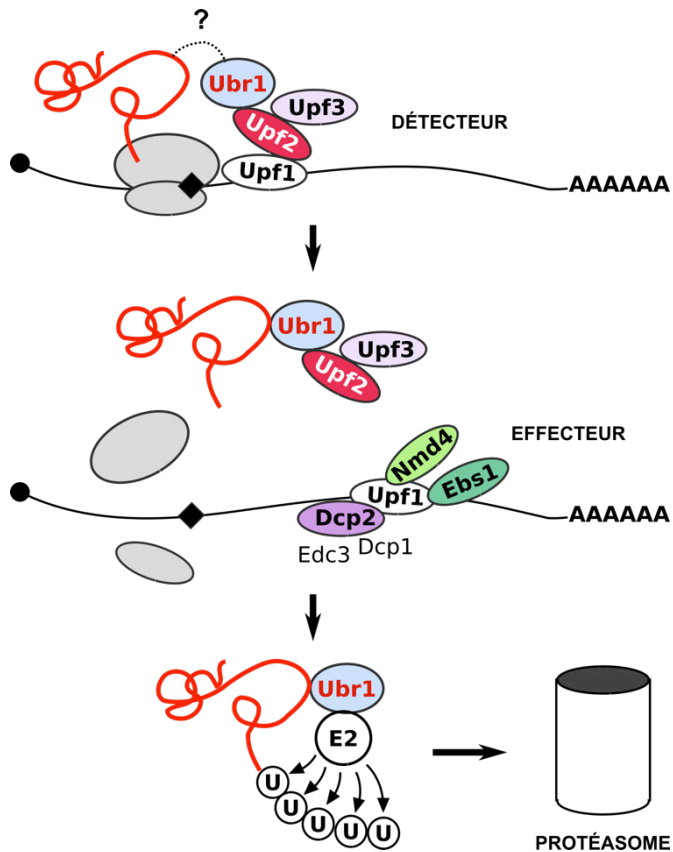


Figure 27 : Modèle de l'adressage au protéasome des peptides néo-synthétisés, par Ubr1

Étape 1 : Ubr1 est associé au complexe détecteur après reconnaissance des PTCs. Le point d'interrogation signifie que Ubr1 pourrait interagir avec le peptide néo-synthétisé. La terminaison de la traduction a lieu normalement. Le peptide néo-synthétisé est représenté en rouge.

Étape 2 : Après le switch de Upf1 et la dissociation du ribosome, Upf2-3 et Ubr1 restent associés avec le peptide naissant. La dégradation de l'ARN continue avec l'Effecteur.

Étape 3 : Ubr1 recrute l'enzyme de conjugaison de l'ubiquitine, E2, qui transfère l'ubiquitine sur le peptide ciblé par l'E3-RING (Rieser et al., 2013).

Ce modèle de la prise en charge des peptides néo-synthétisés permettrait de seconder la voie de dégradation des protéines mal repliées (UPR : *Unfolded Protein Response*). En effet, il est probable qu'une partie des peptides tronqués n'arrivent pas à adopter une bonne conformation et soient reconnues par le système UPR.

2. Le complexe CK2, responsable de la phosphorylation de Upf2 ?

Une autre interaction spécifique d'Upf2 a été décrite dans nos expériences, le complexe caséine kinase 2 (CK2) composé de Cka1, Cka2, Ckb1 et Ckb2 (**Figure 26**). Cette interaction était stabilisée par l'ARN (**Figure 26A**) et indépendante d'Upf1 (**Figure 26B**).

Le complexe CK2 est composé de deux sous-unités catalytiques Cka1 et Cka2 et de deux sous-unités régulatrices Ckb1 et Ckb2 (Bidwai et al., 1994). Il a une activité Sérine/Thréonine kinase (Bojanowski et al., 1993). CK2 est impliqué dans différentes voies assez éloignées du NMD telle que la régulation de la transcription par les polymérase I et III, le vieillissement des cellules ou la réponse à des stress osmotiques provoquant des dommages à l'ADN. Il est

capable de phosphoryler des protéines très variées. Par ailleurs, nous savons que la protéine Upf2 de levure est phosphorylée au niveau de résidus sérine (Wang et al., 2006). Ces données seraient donc compatibles avec la phosphorylation de Upf2 par le complexe CK2.

L'association entre Upf2 et Cka2-Ckb2 résiste au traitement RNase tandis que Cka1 et Ckb1 sont perdus, ce qui suggère une stabilisation de l'interaction avec l'ensemble du CK2 par l'ARN (**Figure 26A**). La phosphorylation pourrait avoir lieu au moment de la reconnaissance des cibles au niveau du complexe Détecteur, sur l'ARN. La phosphorylation est importante pour l'interaction avec Hrp1 impliqué dans la dégradation de certains transcrits (González et al., 2000; Wang et al., 2006). Cependant, dans nos conditions, l'enrichissement du complexe CK2 co-purifiant avec Upf2 est similaire à celui d'Upf1, ce qui suggère qu'il est un composant de tous les Détecteurs. La phosphorylation pourrait donc avoir un rôle dans tous les événements de NMD.

3. Hrr25 fait partie de l'effecteur

Parmi les nouveaux facteurs identifiés dans notre étude, Hrr25 est assez remarquable. Hrr25 interagit fortement avec Upf1 indépendamment de la présence de l'ARN (**Fig. 2A**). Il interagit aussi avec les autres partenaires de l'Effecteur Nmd4, Ebs1 et Dcp1, il est un composant à part entière de ce complexe (**Fig. 2D, E, F**). Hrr25 interagit avec la machinerie de décapping dans une purification de Dcp1 (**Fig. 2F**), en revanche la purification de Hrr25 ne permet pas de retrouver l'interaction avec Dcp1/Dcp2 (**Figure 28**). Cette configuration est le signe qu'une petite partie seulement des protéines Hrr25 interagit avec la machinerie de décapping. La majeure partie de Hrr25 interagit avec l'élongator, Kti12, Elp2-3, Iki3 ; avec des cofacteurs de la biogenèse des sous-unités 40S (Ghalei et al., 2015) et 60S (Ray et al., 2008) ainsi que des précurseurs 20S du ribosome, Rio2, Tsr1, YMR295C, Dim1, Nob1. Nous avons aussi identifié une interaction forte avec Esl1 et Esl2, deux protéines ressemblant aux protéines Est1 et Smg5-7 humaine et ayant un rôle dans la réponse des cellules aux stimuli de l'environnement (Lai et al., 2013). L'interaction de Dcp2 avec Hrr25 est cependant très importante, elle entraîne Hrr25 dans des granules d'ARN appelé p-bodies (Zhang et al., 2016) et serait cruciale pour l'accomplissement de la méiose (Zhang et al., 2018). Hrr25 est une sérine/thréonine et tyrosine kinase.

Dans ce contexte, nous nous sommes demandé quel pouvait être le rôle de Hrr25 au sein de l'Effecteur et s'il participait à la phosphorylation d'Upf1 ou de Dcp2 (Albuquerque et al., 2008; Swaney et al., 2013; Yoon et al., 2010). Cependant, la multiplicité des fonctions de

Hrr25 dans la biologie de l'ARN et dans d'autres mécanismes rend difficile la différenciation des effets de mutation dans ce gène sachant qu'ils peuvent être liés. De plus, Hrr25 est un gène essentiel dans les souches S288C de levure que nous utilisons au laboratoire (Giaever et al., 2002).

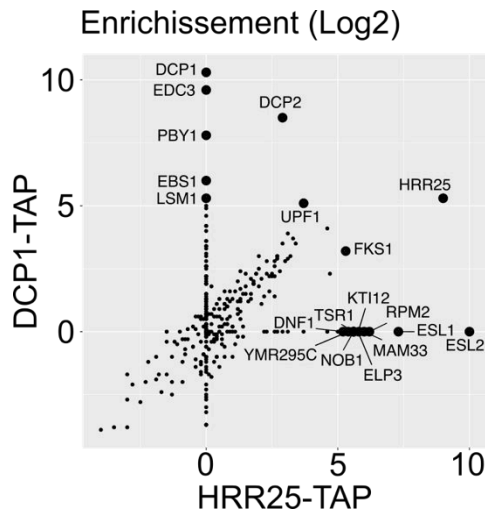


Figure 28 : Comparaison des partenaires enrichis dans les purifications de Dcp1 et Hrr25

Les protéines indiquées sur le graphe ont un enrichissement en log supérieur ou égal à 5. Attention les valeurs indiquées pour l'enrichissement sont en log2.

Mon équipe d'accueil envisage maintenant de reconstituer le complexe Effecteur *in vitro* afin de tester l'activité de l'hydrolyse de la coiffe au sein de ce complexe. Ce système isolé devrait ensuite permettre d'étudier l'effet direct de Hrr25 et d'autres cofacteurs sur le décappage associé à Upf1.

4. Plusieurs hypothèses existent pour le switch

Le résultat majeur de ce travail de thèse est la découverte d'un mécanisme du NMD chez la levure en deux étapes accomplies par deux complexes protéiques distincts et successifs au niveau des ARNs cibles. Ce modèle implique ainsi un échange (switch) entre les protéines du Détecteur et celle de l'Effecteur.

4.1. Échange des deux complexes ou échanges des facteurs autour de Upf1 ?

Deux solutions existent pour expliquer le switch. La première hypothèse possible est l'échange complet du Détecteur avec l'Effecteur après reconnaissance des cibles. En effet, les interactions au sein des complexes sont assez fortes et de plus certaines interactions occupent les mêmes régions d'Upf1. Cependant, ce modèle nécessiterait un décrochage de l'ARN de Upf1-Upf2-Upf3 et le recrutement de l'Effecteur sur les mêmes ARNs. On a vu que l'interaction entre Upf1, Upf2 et Upf3 était stabilisée au niveau de l'ARN (**Fig. 2A**), en outre,

Upf1 a une liaison intrinsèque extrêmement forte pour les ARNs et son activité de translocase est très processive (Kanaan et al., in revision). Il semble alors peu probable que Upf1 se décroche de l'ARN avant la fin du processus NMD. Par ailleurs le mécanisme du NMD est très rapide, l'ajout d'un long 3'-UTR à un ARN peut faire chuter sa demi-vie de moitié selon les cas (Wen and Brogna, 2010; pour exemple). Cela nous amène à la deuxième hypothèse pour le switch, l'échange des facteurs Upf2-3 au profit de Nmd4-Ebs1 et la machinerie de décapping. Ce modèle semble plus efficace pour assurer la continuité entre reconnaissance des cibles et dégradation rapide.

Frank Feuerbach a tenté une approche originale (Fairhead and Howarth, 2015) pour évaluer le lien entre Détecteur et Effecteur mais malheureusement sans résultats probants pour le moment. Il a construit des souches de levure possédant un domaine catalytique permettant la biotinylation (domaine BirA) fusionné à *UPF2* qu'il a croisé avec des souches où *UPF1* était fusionné à un motif biotinylable (AviTag) et où *NMD4* était fusionné à une étiquette TAP pour la purification. Après culture de la souche dans un milieu sans biotine, il a ajouté pendant un temps court de la biotine dans le milieu. Après ajout de la biotine, le motif AviTag de Upf1 était phosphorylé par BirA positionné au niveau d'Upf2. L'objectif était ensuite de purifier Nmd4 et d'observer l'état de biotinylation de la protéine Upf1 associée. Si celle-ci était biotinylée, cela signifiait qu'il avait d'abord été en contact avec Upf2. Cette technique était très séduisante en théorie mais nous manquions d'efficacité de biotinylation ce qui a rendu les conclusions peu convaincantes.

4.2. Dimérisation de Upf1

Pendant ma thèse, nous avons aussi soulevé une autre hypothèse pour le switch ; la dimérisation de Upf1. Le modèle était qu'après reconnaissance des cibles par le Détecteur, un autre Upf1 potentiellement accompagné des partenaires de l'Effecteur puisse interagir avec le Détecteur et permette alors la suite du mécanisme. Ophélie Lautier, qui a réalisé son stage de validation de BTS dans notre équipe, a testé cette hypothèse. Les résultats préliminaires qu'elle a obtenus ont suggéré la dimérisation possible d'Upf1 en présence d'ARN *in vivo*. *In vitro*, il a été précédemment montré par la technique de filtration sur gel que Upf1 ne formait pas de dimère en solution (Cheng et al., 2007). Par ailleurs, il est probable que plusieurs Upf1 puissent se retrouver sur le même ARN (Hogg and Goff, 2010). L'observation d'Ophélie reflèterait alors une interaction indirecte due à la cohabitation de plusieurs Upf1 sur les ARNs.

4.3. Rôle de la phosphorylation d'Upf2

Nous avons décrit précédemment, une interaction du complexe CK2 avec Upf2 qui pourrait être responsable de la phosphorylation de ce dernier. Dans beaucoup de cas, les modifications post-traductionnelles permettent d'interagir avec de nouveaux partenaires ou de changer de conformation. Dans notre cas, la phosphorylation d'Upf2 pourrait ainsi participer à l'affaiblissement des interactions avec Upf1 et permettre l'échange avec la machinerie de décapping.

4.4. Ebs1 comme intermédiaire entre le Détecteur et l'Effecteur

Le signal déclenchant le switch des deux complexes par l'un ou l'autre des modèles est la pièce manquante de notre puzzle. Vu l'efficacité et la rapidité du mécanisme, nous pensons que certaines protéines du Détecteur ou de l'Effecteur pourraient être impliquées directement. Ebs1 est un candidat :

- sa délétion a un impact sur le NMD visible en situation d'équilibre (**Fig. 5**) ce qui est compatible avec un rôle dans une étape à priori centrale du mécanisme du NMD comme le switch.
- son profil d'interaction en spectrométrie de masse est différent de celui de Nmd4, il interagit spécifiquement avec Upf1 mais son interaction avec les autres partenaires de l'Effecteur est moins forte (**Fig. 2D et E**). Cela serait compatible avec un rôle intermédiaire entre le Détecteur et l'Effecteur.
- Ebs1 conserve une association avec le domaine CH d'Upf1 (**Fig. 3C**) qui dépend de l'ARN alors que sa région d'interaction, l'extrémité C-ter (**Fig. 3E**) est absente. Cela pourrait indiquer une interaction complémentaire au niveau de l'extrémité N-ter comme décrit pour Smg5 chez l'Homme (Ohnishi et al., 2003).
- La région d'interaction d'Ebs1 a un effet sur l'interaction d'Upf2 et Upf3. Upf2 et Upf3 peuvent interagir avec le domaine HD-Cter de Upf1 en présence d'ARN (**Fig. 3D**), mais pas avec le domaine HD seul (**Figure 29**). Cela suggère que le C-ter pourrait aider l'interaction avec Upf2-3.

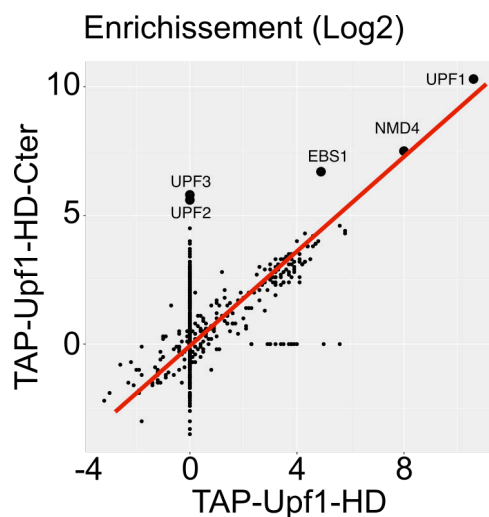


Figure 29 : Impact de l'extrémité C-terminale d'Upf1 sur les interactions protéiques

Les fragments d'Upf1 comme décrit Fig. 3A, sont surexprimés à partir de plasmides centromériques sous le contrôle d'un promoteur TetO7. La ligne rouge marque la corrélation parfaite entre les deux purifications. Attention les valeurs indiquées pour l'enrichissement sont en log2 pour cette figure.

Ces éléments suggèrent que l'extrémité C-ter d'Upf1 qui interagit avec Ebs1 dans l'Effecteur pourrait stabiliser l'interaction avec Upf2 et Upf3 sur l'ARN lorsqu'il est libre dans le Détecteur. Ainsi l'interaction de Ebs1 sur le C-ter et peut-être le N-ter pourrait affaiblir l'interaction de Upf1 avec Upf2 et initier le switch.

Afin de tester cette hypothèse j'ai réalisé plusieurs expériences qui sont actuellement en cours. J'ai réalisé des purifications de Upf1-TAP dans des souches où Ebs1 était délété pour évaluer l'impact d'Ebs1 sur la balance Détecteur / Effecteur. J'ai aussi testé l'impact d'une interaction continue entre Upf1 et Ebs1 sur le NMD. Pour cela, j'ai construit un plasmide centromérique avec une fusion de la séquence codant Upf1, flanquée d'une étiquette TAP en position N-terminale, et de la séquence codant Ebs1 (pCM189-TAP-Upf1-Ebs1). Ce type de fusion a déjà été réalisé, notamment pour réaliser une super polymérase ou dans le système de double hybride de levure. Par ailleurs nous savons que l'ajout d'une étiquette TAP de 21KDa en position N-terminale ou C-terminale d'Upf1 ne gêne pas sa fonction. La fusion s'exprime à un niveau similaire à TAP-Upf1, elle est surexprimée environ 50 fois par rapport à la version endogène. La protéine de fusion continue à interagir avec les ARNs traduits, une part de la fusion se retrouve dans les fractions lourdes d'un gradient de sucrose. La protéine de fusion interagit avec Nmd4 et la machinerie de décapping à des niveaux similaires à TAP-Upf1. Il semble donc que la fusion conserve la structure d'Upf1 permettant les interactions et sa fonction de liaison à l'ARN. Cependant, le NMD n'est pas efficace dans ces cellules. Si notre hypothèse est bonne, l'abolissement du NMD pourrait être du à la perte des interactions avec Upf2 et Upf3, que nous observons, permettant le ciblage spécifique des ARNs avec des PTCs. Les résultats sont encore très préliminaires et l'effet observé sur Upf3 est faible. De plus la

fusion génère beaucoup de fragment de dégradation lors des extractions protéiques, elle n'est pas très stable. Pour compléter ces expériences, nous voulons surexprimer Upf2 et Upf3 dans les cellules exprimant la fusion pour essayer de forcer l'interaction et rétablir le NMD.

4.5. Perspectives pour une meilleure compréhension du switch

D'autres pistes sont envisagées pour l'étude du switch, notamment *in vitro* en poursuivant notre collaboration avec le laboratoire d'Hervé Le Hir à l'ENS. Nous souhaitons par exemple faire des expériences de compétition pour l'interaction des facteurs Upf2, Nmd4 et Ebs1 avec Upf1 afin de mieux comprendre la dynamique des échanges entre les facteurs au niveau de l'ARN et dans le cytoplasme.

5. Quelle est la fonction de l'Effecteur non fixé aux ARNs ?

Il existe un autre switch des facteurs autour d'Upf1 pendant son recyclage. En effet, on observe dans notre étude que la quantité de complexe Effecteur libre est très importante. Upf1 doit donc pouvoir s'extraire de cette Effecteur libre pour recommencer un nouveau cycle de NMD (**Figure 25** et **Fig. 7G**). Le mécanisme moléculaire sous-jacent à ce recyclage est complètement inconnu.

L'abondance du complexe libre est assez étonnante, cela pourrait être une forme de stockage d'Upf1, Nmd4 et Ebs1 qui sont beaucoup plus abondant dans la cellule que les facteurs Upf2 et Upf3. Cet Effecteur libre pourrait aussi avoir un autre rôle dans la cellule, indépendant de Upf2 et Upf3, cela expliquerait notamment la fraction de Nmd4 restant associée aux polysomes lorsque *UPF2* est délété alors que nous avons une perte complète de l'association en *upf1Δ* (**Fig. 7**).

L'abondance de la forme libre de l'Effecteur peut aussi expliquer les résultats des purifications des mutants qui ne perdent aucune des interactions des facteurs du NMD. En effet, les mutants de l'hélicase pourraient continuer à interagir avec les partenaires dans des formes libres de stockage/recyclage. La perte du NMD dans ces cellules serait alors due à l'absence de recrutement de ces Effecteurs au niveau des ARNs substrats du NMD ou alors à un blocage du NMD en cours de processus comme c'est le cas pour DE572AA (Serdar et al., 2016). Pour tester cette hypothèse, nous pourrions tester l'association de Nmd4, signature du complexe Effecteur, aux polysomes et aux substrats du NMD dans un contexte où Upf1 est muté dans l'hélicase. Cette expérience permettrait aussi de nous renseigner sur le rôle de la fonction hélicase dans le switch entre les deux complexes.

6. Lsm1-7 et Pat1 sont présents sur les mêmes ARNs que Upf1

L'hypothèse d'une nouvelle fonction pour le complexe Effecteur est appuyée par l'identification du complexe Lsm cytoplasmique ainsi que Pat1 associés à Upf1 (**Fig. 2A et D**). Ce résultat était très surprenant car ces protéines sont impliquées dans la voie de dégradation des ARNs classiques par hydrolyse de la coiffe suivant la dé-adénylation (**Figure 2**) alors que la dégradation des substrats du NMD suit majoritairement une voie de dégradation par hydrolyse de la coiffe rapide court-circuitant la dé-adénylation chez la levure. Upf1 et ses partenaires de l'Effecteur pourraient-ils alors jouer un rôle dans la voie classique de dégradation des ARNs de 5' vers 3' ? Upf1 peut se lier non-spécifiquement à une grande variété d'ARN (Hogg and Goff, 2010; Lee et al., 2015) sans entraîner leur dégradation par le NMD. Pour l'instant, aucune fonction n'a été proposée pour cette association. Il est possible que Upf1 seul ou accompagné d'une partie de l'Effecteur soit impliqué dans un nouveau mécanisme, comme c'est le cas pour le SMD chez les eucaryotes multicellulaires (résumé dans Gong et al., 2013).

Une autre explication pour la co-précipitation de Lsm/Pat1 avec Upf1 serait une interaction protéine-protéine indirecte *via* Dcp1-2 ou Xrn1 qui ferait le pont entre Lsm/Pat1 et Upf1. Une possibilité est aussi la co-localisation de ces facteurs dans des p-bodies (*processing bodies*).

7. La question récurrente du lien entre les p-bodies et le NMD

Le complexe Effecteur libre pourrait aussi dans des situations de stress participer à la condensation des protéines et des ARNs dans les p-bodies. Les p-bodies sont des amas condensés de protéines et d'ARN dans le cytoplasme. Ils ont été mis en évidence la première fois dans des cellules de mammifère (Bashkirov et al., 1997; van Dijk et al., 2002; Ingelfinger et al., 2002). Initialement appelés Dcp-bodies du fait de la présence de Xrn1, Dcp1 et Dcp2 dans ces granules, une étude chez la levure les a renommés p-bodies (Sheth and Parker, 2003) après la caractérisation de leur contenu. Chez la levure, Upf1, Dcp1, Dcp2, Edc3, Nmd4 et Ebs1, soit tout le complexe Effecteur, sont co-localisés dans les p-bodies (Luke et al., 2007; Nadler-Holly et al., 2012; Parker and Sheth, 2007; Sheth and Parker, 2003). Les protéines Upf2 et Upf3 sont aussi recrutées dans les p-bodies mais seulement lorsque la voie de dégradation des ARNs 5' vers 3' est mutée (Sheth and Parker, 2006).

Après la découverte de ces granules d'ARNs et de protéines dans le cytoplasme, la première hypothèse pour leur fonction moléculaire était qu'elles servaient d'usine de dégradation pour

les ARNs. Au fur et à mesure des recherches, les hypothèses ont oscillé entre dégradation et stockage des ARNs. Des données récentes chez les mammifères proposent aujourd'hui que les p-bodies soient des zones de stockage réversibles des régulons d'ARN. Les ARNs seraient conservés dans ces corps denses sous forme non-traduites mais avec leurs protéines associées (régulons) puis, sous certaines conditions, les ARNs pourraient ressortir des p-bodies afin d'être traduits à nouveau ou dégradés (résumé dans Standart and Weil, 2018). Certaines protéines sont des précurseurs de la formation des p-bodies, c'est le cas d'Edc3 chez la levure. On a vu qu'Edc3 était aussi un composant de l'Effecteur, cela pourrait expliquer la présence de l'Effecteur dans les granules, par un recrutement passif dû à l'affinité des protéines. Par ailleurs, les protéines de l'Effecteur sont en large excès par rapport à Upf2 et Upf3. Les p-bodies pourraient permettre de conserver une réserve d'Upf1 et de machinerie de décapping en cas de besoin important et soudain. En outre, j'ai mentionné précédemment que Upf2 et 3 rejoignaient les granules dans une situation d'inactivation du décapping, cela pourrait être une forme de stockage de ces facteurs en attendant que l'activité normale de Dcp1-2 et Xrn1 soit revenue. Une autre alternative serait qu'en l'absence du décapping, les ARNs substrats du NMD reconnus par le Détecteur soient stockés dans les granules pour ne pas s'accumuler dans le cytoplasme et utiliser des ribosomes.

8. Nmd4, régulateur de l'association d'Upf1 à l'ARN

Le facteur le plus enrichi avec Upf1 dans toutes nos purifications était Nmd4. Nous avons montré une interaction entre ces deux protéines au niveau de la région hélicase d'Upf1. Apparemment, les résidus K436, DE572 et TR800 ne sont pas impliqués dans cette interaction (**Figure 22C et D**), c'est une autre région qui serait importante.

Concernant la fonction moléculaire de cette interaction, des expériences de co-précipitation d'un ARN synthétique *in vitro* ont montré une augmentation de l'affinité d'Upf1 avec cet ARN en présence de Nmd4 (**Fig. 4E**). Ces résultats suggèrent que Nmd4 pourrait moduler la force d'interaction d'Upf1 avec les ARNs. Une étude réalisée en parallèle par Joanne Kanaan chez la levure *S. cerevisiae* (Kanaan et al., in revision) a montré l'effet important des protrusions 1B et 1C de l'hélicase dans la processivité d'Upf1 et dans sa force de liaison à l'ARN. Trois résidus en particulier, A, K et S situés en position 484 à 486 et localisés dans la région d'interaction avec l'ARN, seraient impliqués dans cette forte association. Le mécanisme moléculaire sous-jacent à ce verrouillage d'Upf1 sur l'ARN n'est pas connu. Nmd4 semble donc avoir une fonction similaire à celle des résidus A, K et S *in vitro*.

De plus, des tests *in vivo* de l'efficacité du NMD en présence de la seule région hélicase d'Upf1 surexprimée (Upf1-HD-Cter : 30% d'efficacité, **Fig. 5E**) montre la perte complète du NMD dans les souches mutantes pour les résidus A, K et S (AKS487HPA) (Kanaan et al., in revision). Le même résultat est observé après la délétion de *NMD4* (**Fig. 5E**).

Les tests fonctionnels, à la fois *in vitro* et *in vivo*, pour le rôle des acides-aminée A, K et S et pour la fonction de Nmd4 dans le NMD ont donc donné des résultats similaires. Cela suggère un lien entre les deux éléments. Il faudrait maintenant cartographier plus précisément la région d'interaction de Nmd4 sur Upf1 et tester le rôle des trois résidus en questions dans cette interaction. Nmd4, en interagissant avec les résidus A, K et S pourrait permettre la réassociation forte de Upf1 sur les substrats assurant le recrutement efficace de la machinerie de décapping. Au vu de la conservation qui existe entre Nmd4 et les domaines PIN de Smg5 et Smg6, ces résultats ouvrent des perspectives intéressantes sur le rôle des domaines PIN dans le NMD.

Conclusion

La stratégie que nous avons suivie a consisté à enquêter, un peu comme le gardien ci-dessous sur les interactions des différentes protéines impliquées dans le NMD.

Quatre amis visitent un musée avec seulement 3 billets d'entrée. Ils rencontrent un gardien qui veut savoir celui qui n'a pas payé son entrée :

« - Ce n'est pas moi, dit Paul.

- C'est Jean, dit Jacques.

- C'est Pierre, dit Jean.

- Jacques a tort, dit Pierre. »

Sachant qu'un seul d'entre eux ment, quel est le resquilleur ?

Source : <http://www.maths-et-tiques.fr>

En confrontant ensuite les différents interrogatoires pour chaque protéine nous avons pu identifier les liens entre les protéines permettant de reconstituer le déroulement d'un événement de NMD. Notre étude du NMD chez la levure a permis d'identifier un remaniement complet des complexes autour d'Upf1 au cours du NMD qui n'avait pas été décrit auparavant. De plus, les partenaires principaux que nous décrivons chez la levure étant tous conservés chez les autres eucaryotes, ce travail suggère l'existence d'un modèle du mécanisme compatible chez tous les eucaryotes. En outre, nous avons identifié plusieurs interactions nouvelles qui pourraient assurer des étapes encore non résolues jusqu'à présent comme la dégradation du peptide tronqué synthétisé lors de la traduction d'un ARN portant un codon Stop précoce.

Références

Addinall, S.G., Holstein, E.-M., Lawless, C., Yu, M., Chapman, K., Banks, A.P., Ngo, H.-P., Maringele, L., Taschuk, M., Young, A., et al. (2011). Quantitative Fitness Analysis Shows That NMD Proteins and Many Other Protein Complexes Suppress or Enhance Distinct Telomere Cap Defects. *PLoS Genet.* 7.

Ahrné, E., Molzahn, L., Glatter, T., and Schmidt, A. (2013). Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* 13, 2567–2578.

Albuquerque, C.P., Smolka, M.B., Payne, S.H., Bafna, V., Eng, J., and Zhou, H. (2008). A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics MCP* 7, 1389–1396.

Alhusaini, N., and Collier, J. (2016). The deadenylase components Not2p, Not3p, and Not5p promote mRNA decapping. *RNA N. Y. N* 22, 709–721.

Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., and Pestova, T.V. (2006). In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. *Cell* 125, 1125–1136.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* 432, 112–118.

Amrani, N., Dong, S., He, F., Ganesan, R., Ghosh, S., Kervestin, S., Li, C., Mangus, D.A., Spatrick, P., and Jacobson, A. (2006). Aberrant termination triggers nonsense-mediated mRNA decay. *Biochem. Soc. Trans.* 34, 39–42.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for *C.elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J.* 22, 641–650.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.* 17, 1497–1506.

Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nat. Genet.* 11, 115–116.

Applequist, S.E., Selg, M., Raman, C., and Jäck, H.M. (1997). Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Res.* 25, 814–821.

Arias-Palomo, E., Yamashita, A., Fernández, I.S., Núñez-Ramírez, R., Bamba, Y., Izumi, N., Ohno, S., and Llorca, O. (2011). The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev.* 25, 153–164.

- Aronoff, R., Baran, R., and Hodgkin, J. (2001). Molecular identification of smg-4, required for mRNA surveillance in *C. elegans*. *Gene* 268, 153–164.
- Aslam, A.A., Higgins, C., Sinha, I.P., and Southern, K.W. (2017). Ataluren and similar compounds (specific therapies for premature termination codon class I mutations) for cystic fibrosis. *Cochrane Database Syst. Rev.* 1, CD012040.
- Atkin, A.L., Schenkman, L.R., Eastham, M., Dahlseid, J.N., Lelivelt, M.J., and Culbertson, M.R. (1997). Relationship between yeast polyribosomes and Upf proteins required for nonsense mRNA decay. *J. Biol. Chem.* 272, 22163–22172.
- Azzalin, C.M. (2012). UPF1: A leader at the end of chromosomes. *Nucleus* 3, 16–21.
- Azzalin, C.M., and Lingner, J. (2006). The double life of UPF1 in RNA and DNA stability pathways. *Cell Cycle Georget. Tex* 5, 1496–1498.
- Azzalin, C.M., Reichenbach, P., Khoraiuli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798–801.
- Badis, G., Saveanu, C., Fromont-Racine, M., and Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. *Mol. Cell* 15, 5–15.
- Bai, Y., Salvatore, C., Chiang, Y.C., Collart, M.A., Liu, H.Y., and Denis, C.L. (1999). The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.* 19, 6642–6651.
- Balistreri, G., Horvath, P., Schweingruber, C., Zünd, D., McInerney, G., Merits, A., Mühlmann, O., Azzalin, C., and Helenius, A. (2014). The Host Nonsense-Mediated mRNA Decay Pathway Restricts Mammalian RNA Virus Replication. *Cell Host Microbe* 16, 403–411.
- Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampé, R. (2011). 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. U. S. A.* 108, 3228–3233.
- Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M., and Heyer, W.D. (1997). A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell Biol.* 136, 761–773.
- Basquin, J., Roudko, V.V., Rode, M., Basquin, C., Séraphin, B., and Conti, E. (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. *Mol. Cell* 48, 207–218.
- Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret, D., and Namy, O. (2014). Genome-wide translational changes induced by the prion [PSI⁺]. *Cell Rep.* 8, 439–448.

- Bazzini, A.A., Del Viso, F., Moreno-Mateos, M.A., Johnstone, T.G., Vejnar, C.E., Qin, Y., Yao, J., Khokha, M.K., and Giraldez, A.J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* *35*, 2087–2103.
- Beelman, C.A., Stevens, A., Caponigro, G., LaGrandeur, T.E., Hatfield, L., Fortner, D.M., and Parker, R. (1996). An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature* *382*, 642–646.
- Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007a). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* *26*, 1591–1601.
- Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Saulière, J., Wittkopp, N., and Izaurralde, E. (2007b). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. *FEBS Lett.* *581*, 2845–2853.
- Beilharz, T.H., and Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA N. Y. N* *13*, 982–997.
- Belew, A.T., Advani, V.M., and Dinman, J.D. (2011). Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Res.* *39*, 2799–2808.
- Belloc, E., Piqué, M., and Méndez, R. (2008). Sequential waves of polyadenylation and deadenylation define a translation circuit that drives meiotic progression. *Biochem. Soc. Trans.* *36*, 665–670.
- Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G., and Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA* *6*, 1236–1247.
- Bianchin, C., Mauxion, F., Sentis, S., Séraphin, B., and Corbo, L. (2005). Conservation of the deadenylase activity of proteins of the Caf1 family in human. *RNA N. Y. N* *11*, 487–494.
- Bidwai, A.P., Reed, J.C., and Glover, C.V. (1994). Casein kinase II of *Saccharomyces cerevisiae* contains two distinct regulatory subunits, beta and beta'. *Arch. Biochem. Biophys.* *309*, 348–355.
- Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M.J., and Namy, O. (2015). New insights into stop codon recognition by eRF1. *Nucleic Acids Res.* *43*, 3298–3308.
- Boeck, R., Tarun, S., Jr, Rieger, M., Deardorff, J.A., Müller-Auer, S., and Sachs, A.B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.* *271*, 432–438.

Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N.H. (2014). 3' UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons. *Cell Rep.* *9*, 555–568.

Bojanowski, K., Filhol, O., Cochet, C., Chambaz, E.M., and Larsen, A.K. (1993). DNA topoisomerase II and casein kinase II associate in a molecular complex that is catalytically active. *J. Biol. Chem.* *268*, 22920–22926.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinforma. Oxf. Engl.* *19*, 185–193.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* *139*, 547–559.

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* *102*, 765–775.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., and Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* *21*, 1636–1652.

Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A.S., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwinn, M.K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol. Cell* *64*, 282–293.

Bresson, S., and Tollervey, D. (2018). Tailing Off: PABP and CNOT Generate Cycles of mRNA Deadenylation. *Mol. Cell* *70*, 987–988.

Brown, C.E., and Sachs, A.B. (1998). Poly(A) Tail Length Control in *Saccharomyces cerevisiae* Occurs by Message-Specific Deadenylation. *Mol. Cell. Biol.* *18*, 6548–6559.

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

Brown, C.E., Tarun, S.Z., Jr, Boeck, R., and Sachs, A.B. (1996). PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *16*, 5744–5753.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., et al. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* *42*, 500–510.

- Bühler, M., Steiner, S., Mohn, F., Paillusson, A., and Mühlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin- μ mRNA depends on 3' UTR length. *Nat. Struct. Mol. Biol.* *13*, 462–464.
- Cali, B.M., Kuchma, S.L., Latham, J., and Anderson, P. (1999). smg-7 is required for mRNA surveillance in *Caenorhabditis elegans*. *Genetics* *151*, 605–616.
- Carballo, E., Lai, W.S., and Blackshear, P.J. (1998). Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin. *Science* *281*, 1001–1005.
- Casadio, A., Longman, D., Hug, N., Delavaine, L., Vallejos Baier, R., Alonso, C.R., and Cáceres, J.F. (2015). Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. *EMBO Rep.* *16*, 71–78.
- Causier, B., Li, Z., De Smet, R., Lloyd, J.P.B., Van de Peer, Y., and Davies, B. (2017). Conservation of Nonsense-Mediated mRNA Decay Complex Components Throughout Eukaryotic Evolution. *Sci. Rep.* *7*, 16692.
- Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell* *41*, 693–703.
- Chakrabarti, S., Bonneau, F., Schüssler, S., Eppinger, E., and Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5–SMG7 and SMG6. *Nucleic Acids Res.* *gku578*.
- Chamieh, H., Ballut, L., Bonneau, F., and 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.
- Chan, W.-K., Huang, L., Gudikote, J.P., Chang, Y.-F., Imam, J.S., MacLean, J.A., and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. *EMBO J.* *26*, 1820–1830.
- Chang, C.-T., Bercovich, N., Loh, B., Jonas, S., and Izaurralde, E. (2014). The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res.* *42*, 5217–5233.
- Chang, J.H., Jiao, X., Chiba, K., Oh, C., Martin, C.E., Kiledjian, M., and Tong, L. (2012). Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. *Nat. Struct. Mol. Biol.* *19*, 1011–1017.
- Chapin, A., Hu, H., Rynearson, S.G., Hollien, J., Yandell, M., and Metzstein, M.M. (2014). In Vivo Determination of Direct Targets of the Nonsense-Mediated Decay Pathway in *Drosophila*. *G3 GenesGenomesGenetics* *4*, 485–496.

- Charenton, C., Taverniti, V., Gaudon-Plesse, C., Back, R., Séraphin, B., and Graille, M. (2016). Structure of the active form of Dcp1-Dcp2 decapping enzyme bound to m7GDP and its Edc3 activator. *Nat. Struct. Mol. Biol. advance online publication*.
- Charenton, C., Gaudon-Plesse, C., Fourati, Z., Taverniti, V., Back, R., Kolesnikova, O., Séraphin, B., and Graille, M. (2017). A unique surface on Pat1 C-terminal domain directly interacts with Dcp2 decapping enzyme and Xrn1 5'–3' mRNA exonuclease in yeast. *Proc. Natl. Acad. Sci. U. S. A.* *114*, E9493–E9501.
- Chawla, R., Redon, S., Raftopoulou, C., Wischniewski, H., Gagos, S., and Azzalin, C.M. (2011). Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication. *EMBO J.* *30*, 4047–4058.
- Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Rajmakers, R., Pruijn, G.J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* *107*, 451–464.
- Chen, Y.-H., Su, L.-H., and Sun, C.-H. (2008). Incomplete nonsense-mediated mRNA decay in *Giardia lamblia*. *Int. J. Parasitol.* *38*, 1305–1317.
- Chen, Z., Smith, K.R., Batterham, P., and Robin, C. (2005). Smg1 Nonsense Mutations Do Not Abolish Nonsense-Mediated mRNA Decay in *Drosophila melanogaster*. *Genetics* *171*, 403–406.
- Cheng, J., Belgrader, P., Zhou, X., and Maquat, L.E. (1994). Introns are cis effectors of the nonsense-codon-mediated reduction in nuclear mRNA abundance. *Mol. Cell. Biol.* *14*, 6317–6325.
- Cheng, J., Maier, K.C., Avsec, Ž., Rus, P., and Gagneur, J. (2017). Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. *RNA* *23*, 1648–1659.
- Cheng, Z., Muhrad, D., Lim, M.K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. *EMBO J.* *26*, 253–264.
- Cheng, Z., Saito, K., Pisarev, A.V., Wada, M., Pisareva, V.P., Pestova, T.V., Gajda, M., Round, A., Kong, C., Lim, M., et al. (2009). Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.* *23*, 1106–1118.
- Chiu, S.-Y., Serin, G., Ohara, O., and Maquat, L.E. (2003). Characterization of human Smg5/7a: A protein with similarities to *Caenorhabditis elegans* SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *RNA* *9*, 77–87.
- Cho, H., Kim, K.M., and Kim, Y.K. (2009). Human Proline-Rich Nuclear Receptor Coregulatory Protein 2 Mediates an Interaction between mRNA Surveillance Machinery and Decapping Complex. *Mol. Cell* *33*, 75–86.

Cho, H., Han, S., Choe, J., Park, S.G., Choi, S.S., and Kim, Y.K. (2013). SMG5–PNRC2 is functionally dominant compared with SMG5–SMG7 in mammalian nonsense-mediated mRNA decay. *Nucleic Acids Res.* *41*, 1319–1328.

Cho, H., Park, O.H., Park, J., Ryu, I., Kim, J., Ko, J., and Kim, Y.K. (2015). Glucocorticoid receptor interacts with PNRC2 in a ligand-dependent manner to recruit UPF1 for rapid mRNA degradation. *Proc. Natl. Acad. Sci. U. S. A.* *112*, E1540–E1549.

Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA N. Y. N* *13*, 998–1016.

Clerici, M., Mourao, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* *28*, 2293–2306.

Clerici, M., Deniaud, A., Boehm, V., Gehring, N.H., Schaffitzel, C., and Cusack, S. (2013). Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. *Nucleic Acids Res.*

Clissold, P.M., and Ponting, C.P. (2000). PIN domains in nonsense-mediated mRNA decay and RNAi. *Curr. Biol.* *10*, R888–R890.

Colak, D., Ji, S.-J., Porse, B.T., and Jaffrey, S.R. (2013). Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. *Cell* *153*, 1252–1265.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA N. Y. N* *7*, 1717–1727.

Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R., and Mühlemann, O. (2016). Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* *ma.059055.116*.

Contreras, J., Begley, V., Macias, S., and Villalobo, E. (2014). An UPF3-based nonsense-mediated decay in Paramecium. *Res. Microbiol.* *165*, 841–846.

Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R., and Grange, T. (1997). Messenger RNA deadenylation precedes decapping in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 5628–5633.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* *26*, 1367–1372.

- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* *10*, 1794–1805.
- Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics MCP* *13*, 2513–2526.
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature* *227*, 561–563.
- Crick, F.H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* *12*, 138–163.
- Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev.* *9*, 423–436.
- Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.* *13*, 207–214.
- Culbertson, M.R., Underbrink, K.M., and Fink, G.R. (1980). Frameshift suppression *Saccharomyces cerevisiae*. II. Genetic properties of group II suppressors. *Genetics* *95*, 833–853.
- Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and 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.
- Daugeron, M.C., Mauxion, F., and Séraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* *29*, 2448–2455.
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 5320–5325.
- Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 3262–3267.
- Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* *7*, 1632–1643.
- Decourty, L., Doyen, A., Malabat, C., Frachon, E., Rispal, D., Séraphin, B., Feuerbach, F., Jacquier, A., and Saveanu, C. (2014). Long open reading frame transcripts escape nonsense-mediated mRNA decay in yeast. *Cell Rep.* *6*, 593–598.

Defenouillère, Q., and Fromont-Racine, M. (2017). The ribosome-bound quality control complex: from aberrant peptide clearance to proteostasis maintenance. *Curr. Genet.* *63*, 997–1005.

Deniaud, A., Karuppasamy, M., Bock, T., Masiulis, S., Huard, K., Garzoni, F., Kerschgens, K., Hentze, M.W., Kulozik, A.E., Beck, M., et al. (2015). A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. *Nucleic Acids Res.*

Denning, G., Jamieson, L., Maquat, L.E., Thompson, E.A., and Fields, A.P. (2001). Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J. Biol. Chem.* *276*, 22709–22714.

Deshmukh, M.V., Jones, B.N., Quang-Dang, D.-U., Flinders, J., Floor, S.N., Kim, C., Jemielity, J., Kalek, M., Darzynkiewicz, E., and Gross, J.D. (2008). mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol. Cell* *29*, 324–336.

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Séraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* *21*, 6915–6924.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* *475*, 114–117.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* *440*, 561–564.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol. CB* *12*, 1060–1067.

Dunkley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J.* *18*, 5411–5422.

Dunkle, J.A., and Dunham, C.M. (2015). Mechanisms of mRNA frame maintenance and its subversion during translation of the genetic code. *Biochimie* *114*, 90–96.

Durand, S., Cougot, N., Mahuteau-Betzer, F., Nguyen, C.-H., Grierson, D.S., Bertrand, E., Tazi, J., and Lejeune, F. (2007). Inhibition of nonsense-mediated mRNA decay (NMD) by a new chemical molecule reveals the dynamic of NMD factors in P-bodies. *J. Cell Biol.* *178*, 1145–1160.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Mühlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol.* *6*.

- Eberle, A.B., Lykke-Andersen, S., Mühlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* *16*, 49–55.
- Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* *4*, 207–214.
- Fairhead, M., and Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol. Clifton NJ* *1266*, 171–184.
- Fan, S., Yuan, R., Ma, Y.X., Meng, Q., Goldberg, I.D., and Rosen, E.M. (2001). Mutant BRCA1 genes antagonize phenotype of wild-type BRCA1. *Oncogene* *20*, 8215–8235.
- Fatscher, T., Boehm, V., Weiche, B., and Gehring, N.H. (2014). The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay. *RNA N. Y. N* *20*, 1579–1592.
- Fenger-Grøn, M., Fillman, C., Norrild, B., and Lykke-Andersen, J. (2005). Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol. Cell* *20*, 905–915.
- Fiorini, F., Bonneau, F., and Le Hir, H. (2012). Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. *Methods Enzymol.* *511*, 255–274.
- Fiorini, F., Boudvillain, M., and Le Hir, H. (2013). Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res.* *41*, 2404–2415.
- Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat. Commun.* *6*.
- Fiorini, F., Robin, J.-P., Kanaan, J., Borowiak, M., Croquette, V., Le Hir, H., Jalinot, P., and Mocquet, V. (2018). HTLV-1 Tax plugs and freezes UPF1 helicase leading to nonsense-mediated mRNA decay inhibition. *Nat. Commun.* *9*.
- Fontaine, K.A., Leon, K.E., Khalid, M.M., Jimenez-Morales, D., Dunlap, M., Kaye, J.A., Shah, P.S., Finkbeiner, S., Krogan, N.J., and Ott, M. (2018). The cellular NMD pathway restricts Zika virus infection and is targeted by the viral capsid protein. *BioRxiv* 290296.
- Ford, A.S., Guan, Q., Neeno-Eckwall, E., and Culbertson, M.R. (2006). Ebs1p, a Negative Regulator of Gene Expression Controlled by the Upf Proteins in the Yeast *Saccharomyces cerevisiae*. *Eukaryot. Cell* *5*, 301–312.
- Fourati, Z., Roy, B., Millan, C., Coureux, P.-D., Kervestin, S., van Tilbeurgh, H., He, F., Usón, I., Jacobson, A., and Graille, M. (2014). A Highly Conserved Region Essential for NMD in the Upf2 N-Terminal Domain. *J. Mol. Biol.* *426*, 3689–3702.

- Franks, T.M., Singh, G., and Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell* *143*, 938–950.
- Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* *295*, 2258–2261.
- Frolova, L.Y., Tsivkovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I., and Kisselev, L.L. (1999). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA N. Y. N* *5*, 1014–1020.
- Fromont-Racine, M., Mayes, A.E., Brunet-Simon, A., Rain, J.C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J.D., et al. (2000). Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast Chichester Engl.* *17*, 95–110.
- Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 Is a 14-3-3-like Adaptor in the Nonsense-Mediated mRNA Decay Pathway. *Mol. Cell* *17*, 537–547.
- Gaba, A., Jacobson, A., and Sachs, M.S. (2005). Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol. Cell* *20*, 449–460.
- Gadal, O., Strauss, D., Kessl, J., Trumpower, B., Tollervey, D., and Hurt, E. (2001). Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell Biol.* *21*, 3405–3415.
- Garcia, D., Garcia, S., and Voinnet, O. (2014). Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell Host Microbe* *16*, 391–402.
- Gardino, A.K., Smerdon, S.J., and Yaffe, M.B. (2006). Structural determinants of 14-3-3 binding specificities and regulation of subcellular localization of 14-3-3-ligand complexes: a comparison of the X-ray crystal structures of all human 14-3-3 isoforms. *Semin. Cancer Biol.* *16*, 173–182.
- Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature* *429*, 575–578.
- Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J.* *22*, 3960–3970.
- Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and 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.

Ghalei, H., Schaub, F.X., Doherty, J.R., Noguchi, Y., Roush, W.R., Cleveland, J.L., Stroupe, M.E., and Karbstein, K. (2015). Hrr25/CK1 δ -directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J Cell Biol* *208*, 745–759.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* *418*, 387–391.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J.* *25*, 5117–5125.

Goldstrohm, A.C., Hook, B.A., Seay, D.J., and Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. *Nat. Struct. Mol. Biol.* *13*, 533–539.

Gong, C., Tang, Y., and Maquat, L.E. (2013). mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat. Struct. Mol. Biol.* *20*, 1214–1220.

González, C.I., Ruiz-Echevarría, M.J., Vasudevan, S., Henry, M.F., and Peltz, S.W. (2000). The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol. Cell* *5*, 489–499.

Gout, J.-F., Li, W., Fritsch, C., Li, A., Haroon, S., Singh, L., Hua, D., Fazelinia, H., Smith, Z., Seeholzer, S., et al. (2017). The landscape of transcription errors in eukaryotic cells. *Sci. Adv.* *3*.

Gowravaram, M., Bonneau, F., Kanaan, J., Maciej, V.D., Fiorini, F., Raj, S., Croquette, V., Le Hir, H., and Chakrabarti, S. (2018). A conserved structural element in the RNA helicase UPF1 regulates its catalytic activity in an isoform-specific manner. *Nucleic Acids Res.* *46*, 2648–2659.

Graber, J.H., Cantor, C.R., Mohr, S.C., and Smith, T.F. (1999). In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14055–14060.

Graille, M., Chaillet, M., and Tilbeurgh, H. van (2008). Structure of Yeast Dom34 A PROTEIN RELATED TO TRANSLATION TERMINATION FACTOR Erf1 AND INVOLVED IN No-Go DECAY. *J. Biol. Chem.* *283*, 7145–7154.

Grange, T., de Sa, C.M., Oddos, J., and Pictet, R. (1987). Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif. *Nucleic Acids Res.* *15*, 4771–4787.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 9613–9618.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. *Mol. Cell* *54*, 573–585.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in *Caenorhabditis elegans*. *Mol. Cell. Biol.* *24*, 7483–7490.

Groušl, T., Ivanov, P., Frydlová, I., Vašicová, P., Janda, F., Vojtová, J., Malínská, K., Malcová, I., Nováková, L., Janošková, D., et al. (2009). Robust heat shock induces eIF2 α -phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, *Saccharomyces cerevisiae*. *J Cell Sci* *122*, 2078–2088.

Guan, Q., Zheng, W., Tang, S., Liu, X., Zinkel, R.A., Tsui, K.-W., Yandell, B.S., and Culbertson, M.R. (2006). Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet.* *2*, e203.

Guydosh, N.R., and Green, R. (2017). Translation of poly(A) tails leads to precise mRNA cleavage. *RNA N. Y. N* *23*, 749–761.

Hagan, K.W., Ruiz-Echevarria, M.J., Quan, Y., and Peltz, S.W. (1995). Characterization of cis-acting sequences and decay intermediates involved in nonsense-mediated mRNA turnover. *Mol. Cell. Biol.* *15*, 809–823.

Halbach, F., Reichelt, P., Rode, M., and Conti, E. (2013). The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* *154*, 814–826.

Hall, G.W., and Thein, S. (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. *Blood* *83*, 2031–2037.

Harigaya, Y., Jones, B.N., Muhrad, D., Gross, J.D., and Parker, R. (2010). Identification and analysis of the interaction between Edc3 and Dcp2 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *30*, 1446–1456.

Hatfield, L., Beelman, C.A., Stevens, A., and Parker, R. (1996). Mutations in trans-acting factors affecting mRNA decapping in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *16*, 5830–5838.

He, F., and Jacobson, A. (1995). Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev.* *9*, 437–454.

- He, F., and Jacobson, A. (2015). Control of mRNA decapping by positive and negative regulatory elements in the Dcp2 C-terminal domain. *RNA*.
- He, F., Peltz, S.W., Donahue, J.L., Rosbash, M., and Jacobson, A. (1993). Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 7034–7038.
- He, F., Brown, A.H., and Jacobson, A. (1996). Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast. *RNA* *2*, 153–170.
- He, F., Brown, A.H., and Jacobson, A. (1997). Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.* *17*, 1580–1594.
- He, F., Li, C., Roy, B., and Jacobson, A. (2014). Yeast Edc3 targets RPS28B mRNA for decapping by binding to a 3' untranslated region decay-inducing regulatory element. *Mol. Cell. Biol.* *34*, 1438–1451.
- Hellen, C.U.T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harb. Perspect. Biol.*
- Hennigan, A.N., and Jacobson, A. (1996). Functional mapping of the translation-dependent instability element of yeast MATalpha1 mRNA. *Mol. Cell. Biol.* *16*, 3833–3843.
- Herrick, D., Parker, R., and Jacobson, A. (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *10*, 2269–2284.
- Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H. (2001). Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* *413*, 538–542.
- Hirschfeld, M., Hausen, A. zur, Bettendorf, H., Jäger, M., and Stickeler, E. (2009). Alternative Splicing of Cyr61 Is Regulated by Hypoxia and Significantly Changed in Breast Cancer. *Cancer Res.* *69*, 2082–2090.
- Ho, B., Baryshnikova, A., and Brown, G.W. (2018). Unification of Protein Abundance Datasets Yields a Quantitative *Saccharomyces cerevisiae* Proteome. *Cell Syst.* *6*, 192-205.e3.
- Ho, J.H., Kallstrom, G., and Johnson, A.W. (2000). Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J. Cell Biol.* *151*, 1057–1066.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* *415*, 180–183.
- Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A New Kind of Informational Suppression in the Nematode *Caenorhabditis Elegans*. *Genetics* *123*, 301–313.

- Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* *143*, 379–389.
- Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. *Nat. Genet.* *36*, 801–808.
- van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* *295*, 2262–2264.
- Hoshino, S., Hosoda, N., Araki, Y., Kobayashi, T., Uchida, N., Funakoshi, Y., and Katada, T. (1999). Novel function of the eukaryotic polypeptide-chain releasing factor 3 (eRF3/GSPT) in the mRNA degradation pathway. *Biochem. Biokhimiia* *64*, 1367–1372.
- Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast Chichester Engl.* *23*, 439–454.
- Hsu, C.L., and Stevens, A. (1993). Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* *13*, 4826–4835.
- Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Collier, J. (2009). Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* *461*, 225–229.
- Huang, L., Lou, C.-H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.-W., and Wilkinson, M.F. (2011). RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. *Mol. Cell* *43*, 950–961.
- Hug, N., and Cáceres, J.F. (2014). The RNA Helicase DHX34 Activates NMD by Promoting a Transition from the Surveillance to the Decay-Inducing Complex. *Cell Rep.* *8*, 1845–1856.
- Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* *14*, 2609–2617.
- Hwang, C.-S., Shemorry, A., Auerbach, D., and Varshavsky, A. (2010a). The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. *Nat. Cell Biol.* *12*, 1177–1185.
- Hwang, J., Sato, H., Tang, Y., Matsuda, D., and Maquat, L.E. (2010b). UPF1 Association With the Cap-binding Protein, CBP80, Promotes Nonsense-mediated mRNA Decay at Two Distinct Steps. *Mol. Cell* *39*, 396–409.
- Ingelfinger, D., Arndt-Jovin, D.J., Lührmann, R., and Achsel, T. (2002). The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrn1 in distinct cytoplasmic foci. *RNA N. Y. N* *8*, 1489–1501.

- Isken, O., and Maquat, L.E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* *9*, 699–712.
- Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E. (2008). Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. *Cell* *133*, 314–327.
- Ismail, N., and Ng, D.T.W. (2006). Have you HRD? Understanding ERAD is DOAble! *Cell* *126*, 237–239.
- Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., and 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.
- Izumi, N., Yamashita, A., Iwamatsu, A., Kurata, R., Nakamura, H., Saari, B., Hirano, H., Anderson, P., and Ohno, S. (2010). AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Sci. Signal.* *3*, ra27.
- Jensen, T.H., Patricio, K., McCarthy, T., and Rosbash, M. (2001). A Block to mRNA Nuclear Export in *S. cerevisiae* Leads to Hyperadenylation of Transcripts that Accumulate at the Site of Transcription. *Mol. Cell* *7*, 887–898.
- Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. *Mol. Cell* *52*, 473–484.
- Jiao, X., Xiang, S., Oh, C., Martin, C.E., Tong, L., and Kiledjian, M. (2010). Identification of a quality control mechanism for mRNA 5'-end capping. *Nature* *467*, 608–611.
- Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol. Cell* *50*, 104–115.
- Jin, H., Kelley, A.C., Loakes, D., and 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.
- Johansson, M.J.O., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007). Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 20872–20877.
- Jonas, S., Weichenrieder, O., and Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5–SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev.* *27*, 211–225.
- Joncourt, R., Eberle, A.B., Rufener, S.C., and Muhlemann, O. (2014). Eukaryotic Initiation Factor 4G Suppresses Nonsense-Mediated mRNA Decay by Two Genetically Separable Mechanisms. *PLoS ONE* *9*.

Kadlec, J., Guilligay, D., Ravelli, R.B., and Cusack, S. (2006). Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA* *12*, 1817–1824.

Kanaan, J., Saurabh, R., Decourty, L., Saveanu, C., Croquette, V., and Le Hir, H. (in revision). UPF1-like helicase grip on nucleic acids dictates processivity.

Karam, R., Lou, C.-H., Kroeger, H., Huang, L., Lin, J.H., and Wilkinson, M.F. (2015). The unfolded protein response is shaped by the NMD pathway. *EMBO Rep.* *16*, 599–609.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and 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.

Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N., and Izaurralde, E. (2010). SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev.* *24*, 2440–2450.

Keppetipola, N., Sharma, S., Li, Q., and Black, D.L. (2012). Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol.* *47*, 360–378.

Kerem, E., Konstan, M.W., De Boeck, K., Accurso, F.J., Sermet-Gaudelus, I., Wilschanski, M., Elborn, J.S., Melotti, P., Bronsveld, I., Fajac, I., et al. (2014). Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Respir. Med.* *2*, 539–547.

Kervestin, S., Li, C., Buckingham, R., and Jacobson, A. (2012). Testing the faux-UTR model for NMD: analysis of Upf1p and Pab1p competition for binding to eRF3/Sup35p. *Biochimie* *94*, 1560–1571.

Kononenko, A.V., Mitkevich, V.A., Atkinson, G.C., Tenson, T., Dubovaya, V.I., Frolova, L.Y., Makarov, A.A., and Haurlyuk, V. (2010). GTP-dependent structural rearrangement of the eRF1:eRF3 complex and eRF3 sequence motifs essential for PABP binding. *Nucleic Acids Res.* *38*, 548–558.

Kressler, D., de la Cruz, J., Rojo, M., and Linder, P. (1998). Dbp6p is an essential putative ATP-dependent RNA helicase required for 60S-ribosomal-subunit assembly in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *18*, 1855–1865.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* *440*, 637–643.

Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 3357–3362.

- Lai, T., Cho, H., Liu, Z., Bowler, M.W., Piao, S., Parker, R., Kim, Y.K., and Song, H. (2012). Structural basis of the PNRC2-mediated link between mrna surveillance and decapping. *Struct. Lond. Engl.* 1993 20, 2025–2037.
- Lai, X., Beilharz, T., Au, W.-C., Hammet, A., Preiss, T., Basrai, M.A., and Heierhorst, J. (2013). Yeast hEST1A/B (SMG5/6)-like proteins contribute to environment-sensing adaptive gene expression responses. *G3 Bethesda Md* 3, 1649–1659.
- Lasalde, C., Rivera, A.V., León, A.J., González-Feliciano, J.A., Estrella, L.A., Rodríguez-Cruz, E.N., Correa, M.E., Cajigas, I.J., Bracho, D.P., Vega, I.E., et al. (2013). Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. *Nucleic Acids Res.*
- Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* 20, 4987–4997.
- Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* 456, 993–996.
- Lee, H.H., Kim, Y.-S., Kim, K.H., Heo, I., Kim, S.K., Kim, O., Kim, H.K., Yoon, J.Y., Kim, H.S., Kim, D.J., et al. (2007). Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol. Cell* 27, 938–950.
- Lee, S.R., Pratt, G.A., Martinez, F.J., Yeo, G.W., and Lykke-Andersen, J. (2015). Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Mol. Cell* 59, 413–425.
- Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* 5, 2303–2314.
- Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 2165–2177.
- Leeuwen, F.W. van, Kleijn, D.P.V. de, Hurk, H.H. van den, Neubauer, A., Sonnemans, M.A.F., Sluijs, J.A., Köycü, S., Ramdjielal, R.D.J., Salehi, A., Martens, G.J.M., et al. (1998). Frameshift Mutants of β Amyloid Precursor Protein and Ubiquitin-B in Alzheimer's and Down Patients. *Science* 279, 242–247.
- Lew, J.E., Enomoto, S., and Berman, J. (1998). Telomere length regulation and telomeric chromatin require the nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.* 18, 6121–6130.
- Li, C., Xu, D., Ye, Q., Hong, S., Jiang, Y., Liu, X., Zhang, N., Shi, L., Qin, C.-F., and Xu, Z. (2016). Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. *Cell Stem Cell* 19, 672.

Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y.-S., Groth, M., Krueger, A., Platzer, M., et al. (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.*

Linde, L., Boelz, S., Neu-Yilik, G., Kulozik, A.E., and Kerem, B. (2007). The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells. *Eur. J. Hum. Genet. EJHG 15*, 1156–1162.

Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet. 48*, 1112–1118.

Liu, C., Karam, R., Zhou, Y., Su, F., Ji, Y., Li, G., Xu, G., Lu, L., Wang, C., Song, M., et al. (2014). The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenocarcinoma. *Nat. Med. 20*, 596–598.

Liu, H., Sadygov, R.G., and Yates, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem. 76*, 4193–4201.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev. 27*, 2125–2138.

Longman, D., Hug, N., Keith, M., Anastasaki, C., Patton, E.E., Grimes, G., and Cáceres, J.F. (2013). DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and *Caenorhabditis elegans*. *Nucleic Acids Res. 41*, 8319–8331.

López-Perrote, A., Castaño, R., Melero, R., Zamarro, T., Kurosawa, H., Ohnishi, T., Uchiyama, A., Aoyagi, K., Buchwald, G., Kataoka, N., et al. (2016). Human nonsense-mediated mRNA decay factor UPF2 interacts directly with eRF3 and the SURF complex. *Nucleic Acids Res.*

Losson, R., and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc. Natl. Acad. Sci. U. S. A. 76*, 5134–5137.

Luke, B., Azzalin, C.M., Hug, N., Deplazes, A., Peter, M., and Lingner, J. (2007). *Saccharomyces cerevisiae* Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. *Nucleic Acids Res. 35*, 7688–7697.

Lykke-Andersen, J. (2002). Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay. *Mol. Cell. Biol. 22*, 8114–8121.

Lykke-Andersen, J., Shu, M.-D., and 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., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014a). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* *28*, 2498–2517.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014b). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* *28*, 2498–2517.

Lynch, M., Ackerman, M.S., Gout, J.-F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* *17*, 704–714.

Mahadevan, S., Raghunand, T.R., Panicker, S., and Struhl, K. (1997). Characterisation of 3' end formation of the yeast HIS3 mRNA. *Gene* *190*, 69–76.

Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* *27*, 435–448.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *ELife* *4*.

Matelska, D., Steczkiewicz, K., and Ginalski, K. (2017). Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Res.* *45*, 6995–7020.

McIlwain, D.R., Pan, Q., Reilly, P.T., Elia, A.J., McCracken, S., Wakeham, A.C., Itie-Youten, A., Blencowe, B.J., and Mak, T.W. (2010). Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 12186–12191.

Meaux, S., van Hoof, A., and Baker, K.E. (2008). Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. *Mol. Cell* *29*, 134–140.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rnt1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* *10*, 99–105.

Melero, R., Uchiyama, A., Castaño, R., Kataoka, N., Kurosawa, H., Ohno, S., Yamashita, A., and Llorca, O. (2014). Structures of SMG1-UPFs Complexes: SMG1 Contributes to Regulate UPF2-Dependent Activation of UPF1 in NMD. *Structure* *22*, 1105–1119.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardi, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* *10*, 730–736.

Mendell, J.T., Medghalchi, S.M., Lake, R.G., Noensie, E.N., and 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., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* *36*, 1073–1078.
- Miller, J.N., and Pearce, D.A. (2014). Nonsense-Mediated Decay in Genetic Disease: Friend or Foe? *Mutat. Res. Rev. Mutat. Res.* *0*, 52–64.
- Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol. Cell* *61*, 874–885.
- Mitkevich, V.A., Kononenko, A.V., Petrushanko, I.Y., Yanvarev, D.V., Makarov, A.A., and 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.
- Mocquet, V., Durand, S., and Jalinot, P. (2015). How Retroviruses Escape the Nonsense-Mediated mRNA Decay. *AIDS Res. Hum. Retroviruses* *31*, 948–958.
- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell* *136*, 688–700.
- Mort, M., Ivanov, D., Cooper, D.N., and Chuzhanova, N.A. (2008). A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.* *29*, 1037–1047.
- Mugridge, J.S., Ziemniak, M., Jemielity, J., and Gross, J.D. (2016). Structural basis of mRNA-cap recognition by Dcp1-Dcp2. *Nat. Struct. Mol. Biol.* *23*, 987–994.
- Mugridge, J.S., Tibble, R.W., Ziemniak, M., Jemielity, J., and Gross, J.D. (2018). Structure of the activated Edc1-Dcp1-Dcp2-Edc3 mRNA decapping complex with substrate analog poised for catalysis. *Nat. Commun.* *9*.
- Muhlrad, D., and Parker, R. (1992). Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev.* *6*, 2100–2111.
- Muhlrad, D., and Parker, R. (1994). Premature translational termination triggers mRNA decapping. *Nature* *370*, 578–581.
- Muhlrad, D., and Parker, R. (1999a). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA N. Y. N* *5*, 1299–1307.
- Muhlrad, D., and Parker, R. (1999b). Recognition of yeast mRNAs as “nonsense containing” leads to both inhibition of mRNA translation and mRNA degradation: implications for the control of mRNA decapping. *Mol. Biol. Cell* *10*, 3971–3978.
- Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'→3' digestion of the transcript. *Genes Dev.* *8*, 855–866.

Muhlrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol. Cell. Biol.* *15*, 2145–2156.

Nadler-Holly, M., Breker, M., Gruber, R., Azia, A., Gymrek, M., Eisenstein, M., Willison, K.R., Schuldiner, M., and Horovitz, A. (2012). Interactions of subunit CCT3 in the yeast chaperonin CCT/TRiC with Q/N-rich proteins revealed by high-throughput microscopy analysis. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 18833–18838.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* *23*, 198–199.

Nakano, K., Ando, T., Yamagishi, M., Yokoyama, K., Ishida, T., Ohsugi, T., Tanaka, Y., Brighty, D.W., and Watanabe, T. (2013). Viral interference with host mRNA surveillance, the nonsense-mediated mRNA decay (NMD) pathway, through a new function of HTLV-1 Rex: implications for retroviral replication. *Microbes Infect.* *15*, 491–505.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* *457*, 1038–1042.

Nelson, J.O., Moore, K.A., Chapin, A., Hollien, J., and Metzstein, M.M. (2016). Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. *ELife* *5*.

Neu-Yilik, G., Raimondeau, E., Eliseev, B., Yeramala, L., Amthor, B., Deniaud, A., Huard, K., Kerschgens, K., Hentze, M.W., Schaffitzel, C., et al. (2017). Dual function of UPF3B in early and late translation termination. *EMBO J.* e201797079.

Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A., and Mühlemann, O. (2014). A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res.* *42*, 9217–9235.

Nicholson, P., Gkratsou, A., Josi, C., Colombo, M., and Mühlemann, O. (2018). Dissecting the functions of SMG5, SMG7 and PNRC2 in nonsense-mediated mRNA decay of human cells. *RNA* rna.063719.117.

Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* *47*, 1588–1602.

Nissan, T., Rajyaguru, P., She, M., Song, H., and Parker, R. (2010). Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. *Mol. Cell* *39*, 773–783.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 Induces Formation of mRNA Surveillance Complexes Containing hSMG-5 and hSMG-7. *Mol. Cell* *12*, 1187–1200.

- Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res.* *40*, 1251–1266.
- Olivas, W., and Parker, R. (2000). The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J.* *19*, 6602–6611.
- Oliveira, C.C., and McCarthy, J.E. (1995). The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* *270*, 8936–8943.
- Ottens, F., Boehm, V., Sibley, C.R., Ule, J., and Gehring, N.H. (2017). Transcript-specific characteristics determine the contribution of endo- and exonucleolytic decay pathways during the degradation of nonsense-mediated decay substrates. *RNA* *rna.059659.116*.
- Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., and Milos, P.M. (2010). Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. *Cell* *143*, 1018–1029.
- Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 Is a Phosphorylated Protein Required for mRNA Surveillance in *Caenorhabditis elegans* and Related to Upf1p of Yeast. *Mol. Cell. Biol.* *19*, 5943–5951.
- Parker, R., and Sheth, U. (2007). P Bodies and the Control of mRNA Translation and Degradation. *Mol. Cell* *25*, 635–646.
- Peccarelli, M., and Kebaara, B.W. (2014). Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot. Cell* *13*, 1126–1135.
- Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* *497*, 127–131.
- Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* *161*, 1400–1412.
- Perlick, H.A., Medghalchi, S.M., Spencer, F.A., Kendzior, R.J., and Dietz, H.C. (1996). Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 10928–10932.
- Petit, A.-P., Wohlbold, L., Bawankar, P., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2012). The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res.* *40*, 11058–11072.
- Pisarev, A.V., Hellen, C.U.T., and Pestova, T.V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. *Cell* *131*, 286–299.

- Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U.T., and Pestova, T.V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol. Cell* 37, 196–210.
- Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014). Cryoelectron Microscopic Structures of Eukaryotic Translation Termination Complexes Containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* 8, 59–65.
- Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124.
- Pulak, R., and Anderson, P. (1993). mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes Dev.* 7, 1885–1897.
- Radhakrishnan, A., and Green, R. (2016). Connections Underlying Translation and mRNA Stability. *J. Mol. Biol.* 428, 3558–3564.
- Ramanathan, A., Robb, G.B., and Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res.* 44, 7511–7526.
- Ramani, A.K., Nelson, A.C., Kapranov, P., Bell, I., Gingeras, T.R., and Fraser, A.G. (2009). High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective *Caenorhabditis elegans*. *Genome Biol.* 10, R101.
- Ray, P., Basu, U., Ray, A., Majumdar, R., Deng, H., and Maitra, U. (2008). The *Saccharomyces cerevisiae* 60 S Ribosome Biogenesis Factor Tif6p Is Regulated by Hrr25p-mediated Phosphorylation. *J. Biol. Chem.* 283, 9681–9691.
- Rees, D.C., Johnson, E., and Lewinson, O. (2009). ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* 10, 218–227.
- Reichenbach, P., Höss, M., Azzalin, C.M., Nabholz, M., Bucher, P., and Lingner, J. (2003). A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. *Curr. Biol. CB* 13, 568–574.
- Rieser, E., Cordier, S.M., and Walczak, H. (2013). Linear ubiquitination: a newly discovered regulator of cell signalling. *Trends Biochem. Sci.* 38, 94–102.
- Roque, S., Cerciat, M., Gaugué, I., Mora, L., Floch, A.G., de Zamaroczy, M., Heurgué-Hamard, V., and Kervestin, S. (2015). Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in *Saccharomyces cerevisiae*. *RNA N. Y. N* 21, 124–134.
- Ruiz-Echevarría, M.J., and Peltz, S.W. (2000). The RNA Binding Protein Pub1 Modulates the Stability of Transcripts Containing Upstream Open Reading Frames. *Cell* 101, 741–751.

- Ruiz-Echevarria, M.J., Czaplinski, K., and Peltz, S.W. (1996). Making sense of nonsense in yeast. *Trends Biochem. Sci.* *21*, 433–438.
- Ruth, A., Dodds, E., Li, W., and Wysocki, V. (2010). Understanding and Exploiting Peptide Fragment Ion Intensities Using Experimental and Informatic Approaches. *Methods Mol. Biol.* *604*, 73–94.
- Sachs, A.B., and Davis, R.W. (1989). The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. *Cell* *58*, 857–867.
- Sachs, A.B., Bond, M.W., and Kornberg, R.D. (1986). A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. *Cell* *45*, 827–835.
- Saito, S., Hosoda, N., and Hoshino, S. (2013). The Hbs1-Dom34 protein complex functions in non-stop mRNA decay in mammalian cells. *J. Biol. Chem.* *288*, 17832–17843.
- Salas-Marco, J., and Bedwell, D.M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol. Cell. Biol.* *24*, 7769–7778.
- Sayani, S., Janis, M., Lee, C.Y., Toesca, I., and Chanfreau, G.F. (2008). Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Mol. Cell* *31*, 360–370.
- Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat. Struct. Mol. Biol.* *16*, 56–62.
- Schmidt, C., Kowalinski, E., Shanmuganathan, V., Defenouillère, Q., Braunger, K., Heuer, A., Pech, M., Namane, A., Berninghausen, O., Fromont-Racine, M., et al. (2016). The cryo-EM structure of a ribosome–Ski2-Ski3-Ski8 helicase complex. *Science* *354*, 1431–1433.
- Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res.* *37*, 1127–1140.
- Schoenberg, D.R., and Maquat, L.E. (2012). Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* *13*, 246–259.
- Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay — Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* *1829*, 612–623.
- Serdar, L.D., Whiteside, D.L., and Baker, K.E. (2016). ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nat. Commun.* *7*, 14021.

Shalem, O., Groisman, B., Choder, M., Dahan, O., and Pilpel, Y. (2011). Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS Genet.* 7, e1002273.

Shao, S., Murray, J., Brown, A., Taunton, J., Ramakrishnan, V., and Hegde, R.S. (2016). Decoding Mammalian Ribosome-mRNA States by Translational GTPase Complexes. *Cell* 167, 1229-1240.e15.

She, M., Decker, C.J., Sundramurthy, K., Liu, Y., Chen, N., Parker, R., and Song, H. (2004). Crystal structure of Dcp1p and its functional implications in mRNA decapping. *Nat. Struct. Mol. Biol.* 11, 249–256.

She, M., Decker, C.J., Chen, N., Tumati, S., Parker, R., and Song, H. (2006). Crystal structure and functional analysis of Dcp2p from *Schizosaccharomyces pombe*. *Nat. Struct. Mol. Biol.* 13, 63–70.

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrud, D., Parker, R., and Song, H. (2008). Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell* 29, 337–349.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300, 805–808.

Sheth, U., and Parker, R. (2006). Targeting of aberrant mRNAs to cytoplasmic processing bodies. *Cell* 125, 1095–1109.

Shigeoka, T., Kato, S., Kawaichi, M., and Ishida, Y. (2012). Evidence that the Upf1-related molecular motor scans the 3'-UTR to ensure mRNA integrity. *Nucleic Acids Res.* 40, 6887–6897.

Shoemaker, C.J., Eyler, D.E., and Green, R. (2010). Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop off to initiate no-go decay. *Science* 330, 369–372.

Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.-K., Lou, C.-H., Espinoza, J.L., Song, H.-W., Phan, M.H., et al. (2016). The Antagonistic Gene Paralogues Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell*.

Siddiqui, N., Mangus, D.A., Chang, T.-C., Palermino, J.-M., Shyu, A.-B., and Gehring, K. (2007). Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *J. Biol. Chem.* 282, 25067–25075.

Silva, A.L., Ribeiro, P., Inácio, A., Liebhaber, S.A., and Romão, L. (2008). Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA N. Y. N* 14, 563–576.

Silva, J.C., Gorenstein, M.V., Li, G.-Z., Vissers, J.P.C., and Geromanos, S.J. (2006). Absolute Quantification of Proteins by LCMSE A Virtue of Parallel ms Acquisition. *Mol. Cell. Proteomics* 5, 144–156.

- Simms, C.L., Yan, L.L., and Zaher, H.S. (2017). Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Mol. Cell* 68, 361-373.e5.
- Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* 6.
- Sparks, K.A., and Dieckmann, C.L. (1998). Regulation of poly(A) site choice of several yeast mRNAs. *Nucleic Acids Res.* 26, 4676–4687.
- Spellman, R., Rideau, A., Matlin, A., Gooding, C., Robinson, F., McGlincy, N., Grellscheid, S.N., Southby, J., Wollerton, M., and Smith, C.W.J. (2005). Regulation of alternative splicing by PTB and associated factors. *Biochem. Soc. Trans.* 33, 457–460.
- Standart, N., and Weil, D. (2018). P-Bodies: Cytosolic Droplets for Coordinated mRNA Storage. *Trends Genet.*
- Stark, C., Breikreutz, B.-J., Reguly, T., Boucher, L., Breikreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535-539.
- Steiger, M., Carr-Schmid, A., Schwartz, D.C., Kiledjian, M., and Parker, R. (2003). Analysis of recombinant yeast decapping enzyme. *RNA N. Y.* N 9, 231–238.
- Stevens, A. (1978). An exoribonuclease from *Saccharomyces cerevisiae*: effect of modifications of 5' end groups on the hydrolysis of substrates to 5' mononucleotides. *Biochem. Biophys. Res. Commun.* 81, 656–661.
- Stevens, A. (1980). An mRNA decapping enzyme from ribosomes of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 96, 1150–1155.
- Stolz, A., Besser, S., Hottmann, H., and Wolf, D.H. (2013). Previously unknown role for the ubiquitin ligase Ubr1 in endoplasmic reticulum-associated protein degradation. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15271–15276.
- Stowell, J.A.W., Webster, M.W., Kögel, A., Wolf, J., Shelley, K.L., and Passmore, L.A. (2016). Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1. *Cell Rep.* 17, 1978–1989.
- Stowell, J.A.W., Wagstaff, J.L., Hill, C.H., Yu, M., McLaughlin, S.H., Freund, S.M.V., and Passmore, L.A. (2018). A low-complexity region in the YTH domain protein Mmi1 enhances RNA binding. *J. Biol. Chem.* 293, 9210–9222.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71.
- Sun, X., and Maquat, L.E. (2000). mRNA surveillance in mammalian cells: the relationship between introns and translation termination. *RNA* 6, 1–8.

Sun, X., Moriarty, P.M., and Maquat, L.E. (2000). Nonsense-mediated decay of glutathione peroxidase 1 mRNA in the cytoplasm depends on intron position. *EMBO J.* *19*, 4734–4744.

Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., and Villén, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* *10*, 676–682.

Swisher, K.D., and Parker, R. (2011). Interactions between Upf1 and the Decapping Factors Edc3 and Pat1 in *Saccharomyces cerevisiae*. *PLoS ONE* *6*.

Szoradi, T., Schaeff, K., Garcia-Rivera, E.M., Itzhak, D.N., Schmidt, R.M., Bircham, P.W., Leiss, K., Diaz-Miyar, J., Chen, V.K., Muzzey, D., et al. (2018). SHRED Is a Regulatory Cascade that Reprograms Ubr1 Substrate Specificity for Enhanced Protein Quality Control during Stress. *Mol. Cell* *70*, 1025-1037.e5.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* *22*, 947–956.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Molina, M.M.S., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in Vivo Map of the Yeast Protein Interactome. *Science* *320*, 1465–1470.

Tarun, S.Z., Jr, and Sachs, A.B. (1997). Binding of eukaryotic translation initiation factor 4E (eIF4E) to eIF4G represses translation of uncapped mRNA. *Mol. Cell. Biol.* *17*, 6876–6886.

Tarun, S.Z., Jr, Wells, S.E., Deardorff, J.A., and Sachs, A.B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 9046–9051.

Tharun, S., and Parker, R. (1999). Analysis of mutations in the yeast mRNA decapping enzyme. *Genetics* *151*, 1273–1285.

Tharun, S., and Parker, R. (2001). Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol. Cell* *8*, 1075–1083.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeyer, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* *17*, 3484–3494.

Thore, S., Mauxion, F., Séraphin, B., and Suck, D. (2003). X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep.* *4*, 1150–1155.

Tian, M., Yang, W., Zhang, J., Dang, H., Lu, X., Fu, C., and Miao, W. (2017). Nonsense-mediated mRNA decay in *Tetrahymena* is EJC independent and requires a protozoa-specific nuclease. *Nucleic Acids Res.* *45*, 6848–6863.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. *Science* *303*, 808–813.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* *104*, 377–386.

Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhlrud, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *EMBO J.* *21*, 1427–1436.

Tuller, T., Ruppin, E., and Kupiec, M. (2009). Properties of untranslated regions of the *S. cerevisiae* genome. *BMC Genomics* *10*, 391.

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

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* *302*, 1212–1215.

Unterholzner, L., and Izaurralde, E. (2004). SMG7 Acts as a Molecular Link between mRNA Surveillance and mRNA Decay. *Mol. Cell* *16*, 587–596.

Valkov, E., Jonas, S., and Weichenrieder, O. (2017). Mille viae in eukaryotic mRNA decapping. *Curr. Opin. Struct. Biol.* *47*, 40–51.

Vermulst, M., Denney, A.S., Lang, M.J., Hung, C.-W., Moore, S., Mosely, M.A., Thompson, J.W., Madden, V., Gauer, J., Wolfe, K.J., et al. (2015). Transcription Errors Induce Proteotoxic Stress and Shorten Cellular Lifespan. *Nat. Commun.* *6*, 8065.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* *1829*, 561–570.

Wang, W., Czaplinski, K., Rao, Y., and Peltz, S.W. (2001). The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J.* *20*, 880–890.

Wang, W., Cajigas, I.J., Peltz, S.W., Wilkinson, M.F., and González, C.I. (2006). Role for Upf2p phosphorylation in *Saccharomyces cerevisiae* nonsense-mediated mRNA decay. *Mol. Cell. Biol.* *26*, 3390–3400.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 5860–5865.

- Webster, M.W., Chen, Y.-H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Collier, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol. Cell* 70, 1089-1100.e8.
- Weil, J.E., and Beemon, K.L. (2006). A 3' UTR sequence stabilizes termination codons in the unspliced RNA of Rous sarcoma virus. *RNA N. Y. N* 12, 102–110.
- Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Mönch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E.W., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* 22, 1381–1396.
- Welch, E.M., and Jacobson, A. (1999). An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. *EMBO J.* 18, 6134–6145.
- Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by Eukaryotic Translation Initiation Factors. *Mol. Cell* 2, 135–140.
- Wen, J., and Brogna, S. (2010). Splicing-dependent NMD does not require the EJC in *Schizosaccharomyces pombe*. *EMBO J.* 29, 1537–1551.
- Weng, Y., Czaplinski, K., and Peltz, S.W. (1996a). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol. Cell. Biol.* 16, 5477–5490.
- Weng, Y., Czaplinski, K., and Peltz, S.W. (1996b). Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Mol. Cell. Biol.* 16, 5491–5506.
- Weng, Y., Czaplinski, K., and Peltz, S.W. (1998). ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA N. Y. N* 4, 205–214.
- Wilkening, S., Pelechano, V., Järvelin, A.I., Tekkedil, M.M., Anders, S., Benes, V., and Steinmetz, L.M. (2013). An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* 41, e65–e65.
- Withers, J.B., and Beemon, K.L. (2010). Structural features in the Rous sarcoma virus RNA stability element are necessary for sensing the correct termination codon. *Retrovirology* 7, 65.
- Wittkopp, N., Huntzinger, E., Weiler, C., Saulière, J., Schmidt, S., Sonawane, M., and Izaurralde, E. (2009). Nonsense-Mediated mRNA Decay Effectors Are Essential for Zebrafish Embryonic Development and Survival. *Mol. Cell. Biol.* 29, 3517–3528.
- Wittmann, J., Hol, E.M., and Jäck, H.-M. (2006). hUPF2 Silencing Identifies Physiologic Substrates of Mammalian Nonsense-Mediated mRNA Decay. *Mol. Cell. Biol.* 26, 1272–1287.

- Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W.J. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* *13*, 91–100.
- Wurm, J.P., Holdermann, I., Overbeck, J.H., Mayer, P.H.O., and Sprangers, R. (2017). Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* *114*, 6034–6039.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* *121*, 725–737.
- Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M., and Varshavsky, A. (2008). Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. *J. Biol. Chem.* *283*, 24011–24028.
- Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* *457*, 1033–1037.
- Yamashita, A. (2013). Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. *Genes Cells* *18*, 161–175.
- Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and 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.
- Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T., et al. (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.
- Yepiskoposyan, H., Aeschmann, F., Nilsson, D., Okoniewski, M., and Mühlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* *N. Y. N* *17*, 2108–2118.
- Yi, H., Park, J., Ha, M., Lim, J., Chang, H., and Kim, V.N. (2018). PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol. Cell* *70*, 1081-1088.e5.
- Yoon, J.-H., Choi, E.-J., and Parker, R. (2010). Dcp2 phosphorylation by Ste20 modulates stress granule assembly and mRNA decay in *Saccharomyces cerevisiae*. *J. Cell Biol.* *189*, 813–827.
- Zaher, H.S., and Green, R. (2009). Fidelity at the molecular level: lessons from protein synthesis. *Cell* *136*, 746–762.

- Zhang, Z., and Dietrich, F.S. (2005). Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in *Saccharomyces cerevisiae*. *Curr. Genet.* *48*, 77–87.
- Zhang, B., Shi, Q., Varia, S.N., Xing, S., Klett, B.M., Cook, L.A., and Herman, P.K. (2016). The Activity-Dependent Regulation of Protein Kinase Stability by the Localization to P-Bodies. *Genetics* *203*, 1191–1202.
- Zhang, B., Butler, A.M., Shi, Q., Xing, S., and Herman, P.K. (2018). P-body localization of the Hrr25/CK1 protein kinase is required for the completion of meiosis. *Mol. Cell. Biol.*
- Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Mol. Cell. Biol.* *18*, 5272–5283.
- Zhang, S., Ruiz-Echevarria, M.J., Quan, Y., and Peltz, S.W. (1995). Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. *Mol. Cell. Biol.* *15*, 2231–2244.
- Zhang, S., Welch, E.M., Hogan, K., Brown, A.H., Peltz, S.W., and Jacobson, A. (1997). Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in *Saccharomyces cerevisiae*. *RNA* *3*, 234–244.
- Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* *31*, 88–100.
- Zünd, D., Gruber, A.R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat. Struct. Mol. Biol.* *20*, 936–943.

Annexes

Liste des abréviations

CRAC	<i>Cross Link and Analysis of cDNA</i>
3'-UTR	<i>3' UnTranslated Region</i> / Région non traduite située en 3' de l'ARN
5'UTR	<i>5' UnTranslated Region</i> / Région non traduite située en 5' de l'ARN
ADN	<u>A</u> cide <u>D</u> éoxyribo- <u>N</u> ucléique
ADNc	ADN complémentaire
ARN	<u>A</u> cide <u>R</u> ibo- <u>N</u> ucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ARNr	ARN ribosomal
AS-NMD	<i>Alternative-Splicing - NMD</i>
BSA	<i>Bovin Serum Albumin</i> / Sérum Albumine Bovin
CBC	<i>Cap Binding Complex</i>
CLIP	<u>C</u> ross <u>L</u> ink et <u>I</u> mmuno- <u>P</u> réci <u>P</u> itation
CTD	Domaine C-terminal de l'ARN polymérase II
CUT	<i>Cryptic Unstable Transcript</i>
CytoQC	<i>Cytoplasmic protein Quality Control</i>
DECID	<i>Decay-Inducing</i>
DSE	<i>Downstream Sequence Element</i>
EBM	<i>EJC Binding Motif</i>
EJC	<i>Exon Junction Complex</i> / Complexe de jonction des exons
ERAD	<i>Endoplasmic-Reticulum Associated Degradation</i>
FAT	<u>F</u> RAP, <u>A</u> TM, and <u>T</u> RRAP
FATC	<u>F</u> RAP, <u>A</u> TM, and <u>T</u> RRAP <u>C</u> -terminal

GGQ	Motif Glutamine - Glutamine - Arginine
GMD	<i>Glucocorticoid receptor-Mediated mRNA Decay</i>
HEAT	Huntington, <u>E</u> F3A, <u>A</u> TM, <u>T</u> OR
HHR	<i>Helical Hairpin Repeat</i>
HPLC	<i>High-Performance (ou Pressure) Liquid Chromatography</i>
HTP	étiquette His6-TEV-protéine A
IgG	<u>I</u> mmuno-globuline <u>G</u>
Kda	Kilo Dalton
L	Litre
LTOP2	Log2 du score Top3 ou 2 si besoin
m/z	ratio de la masse du peptide sur sa charge
MS	<i>Mass spectrometry</i>
MS/MS	<i>Tandem mass spectrometry</i>
nm	nanomètre
NMD	<i>Nonsense-Mediated mRNA Decay</i>
p-bodies	<i>processing-bodies</i>
PABP	<i>Poly(A) Binding Protein</i>
PCR	Réaction de polymérisation en chaîne
PDB	<i>Protein DataBase</i>
PIKK	<i>Phosphatidyl Inositol (PI) 3-Kinase-related Kinase</i>
PIN	<u>P</u> ilt <u>N</u> -terminus
PTC	<i>Premature Termination Codon</i> / Codon Stop précoce
Puf	<i>Pumilio family protein</i>
qPCR	PCR quantitative
RIP	<i>RNA-ImmunoPrecipitation</i>

RIP-seq	RIP suivi du séquençage des ARNs purifiés
RNase	<u>R</u> ibo- <u>N</u> ucléase
RQC	<u>R</u> ibosome <u>Q</u> uality <u>C</u> ontrol
RRM	<i>RNA Recognition Motif</i> / Motif de reconnaissance de l'ARN
RT-qPCR	<i>Reverse Transcription des ARNs et PCR quantitative</i>
SHRED	<i>Stress-induced <u>H</u>omeostatically <u>R</u>egulated protein <u>D</u>egradation</i>
siRNA	<i>RNA silencing</i> / ARN silencieux
SMD	<i><u>S</u>taufen-<u>M</u>ediated mRNA <u>D</u>ecay</i>
SURF	Smg1-Upf1-eRF1 & 3
SUT	<i><u>S</u>table <u>U</u>nannotated <u>T</u>ranscript</i>
TAP	étiquette CBP-TEV-protéine A pour la purification d'affinité
TSS	<i><u>T</u>ranscription <u>S</u>tart <u>S</u>ite</i> / site de démarrage de la transcription
TTS	<i><u>T</u>ranscription <u>T</u>ermination <u>S</u>ite</i> / site de terminaison de la transcription
TUTase	<i><u>T</u>erminal <u>U</u>ridyl <u>T</u>ransferases</i>
Upf1-CH	domaines CH et N-terminal de Upf1, 1-208 chez la levure
Upf1-CH-HD	domaines CH et hélicase d'Upf1, 1-851 chez la levure
Upf1-Cter	domaine C-terminal d'Upf1, 854-971 chez la levure
Upf1-FL	protéine entière de levure, 1-971
Upf1-HD	domaine hélicase d'Upf1, 208-851 chez la levure, 295-914 chez l'Homme
Upf1-HD-Cter	domaines hélicase et C-terminal d'Upf1, 208-971 chez la levure
UPR	<i><u>U</u>nfolded <u>P</u>rotein <u>R</u>esponse</i> / Contrôle des protéines mal repliées
USR	<i><u>U</u>nstructured Region</i> / Région protéique non structurée
UV	Rayon <u>U</u> ltra- <u>V</u> iolet
XUT	<i><u>X</u>rn1-dependent <u>U</u>nannotated <u>T</u>ranscript</i>

Table des Figures

Figure 1 : Étapes de la vie d'un ARNm dans la cellule et mécanismes de contrôle qualité associés	9
Figure 2 : Voies générales de dégradation des ARNs cytoplasmiques	12
Figure 3 : Schéma des modèles de dé-adénylation des ARNs	14
Figure 4 : Processus de traduction des ARNm eucaryotes.....	20
Figure 5 : Voies cytoplasmiques de surveillance des ARNs.....	23
Figure 6 : Upf1 ; architecture, interactions, conservation et structures.....	29
Figure 7 : Upf2 – Upf3	31
Figure 8 : Architecture de la protéine Smg1 et ses interactions.....	33
Figure 9 : Rôle de Smg5-7 dans le cycle de phosphorylation de Upf1	35
Figure 10 : Modèle SURF-DECID.....	39
Figure 11: Schéma de l'analyse en spectrométrie de masse en tandem.....	47
Figure 12 : Graphique de l'intensité des peptides quantifiés pour Upf1 purifié dans différentes conditions et souches.	50
Figure 13 : Comparaison des LTOP2 des protéines quantifiées dans les extraits totaux avec l'abondance connue de ces protéines.....	52
Figure 14 : Coloration à l'argent du résultat de sept purifications différentes.....	52
Figure 15 : Principe du calcul de l'enrichissement	53
Figure 16 : Représentation des résultats de la purification de Dcp1	54
Figure 17 : Schéma de la méthode CRAC	58
Figure 18 : Exemples de la localisation d'Upf1 sur deux transcrits typiques des substrats de NMD	59
Figure 19 : Localisation d'Upf1 sur les transcrits de levure	61
Figure 20 : Caractérisation des mutants d'Upf1.....	62
Figure 21 : Les mutants d'Upf1 perdent la spécificité de liaison aux transcrits RPL28 immatures, cibles du NMD.....	64
Figure 22 : Upf1 muté au niveau de l'hélicase reste associé à ses partenaires	68
Figure 23 : Réseau des interactions des protéines Upf1, Upf2 et Upf3	70
Figure 24 : Stratégie pour la caractérisation des partenaires d'Upf1 par purification d'affinité et spectrométrie de masse quantitative	72
Figure 25 : Modèle proposé pour le mécanisme du NMD EJC-indépendant chez la levure <i>S. cerevisiae</i>	73
Fig. 1: Enrichment analysis accurately describes Upf1-associated proteins.....	80
Fig. 2: Purification of NMD factors reveals two distinct complexes containing Upf1.....	82
Fig. 3: Nmd4 and Ebs1 are the only Detector components that interact with Upf1 independent of the N-terminal CH domain.....	84
Fig. 4: Nmd4 interaction with the Effector complex is direct and mediated by Upf1	87
Fig. 5: Nmd4 and Ebs1 are essential for NMD elicited by the overexpression of the helicase domain of Upf1.....	88
Fig. 6: Upf2 and Upf3 function as a heterodimer in the formation of Detector.....	92
Fig. 7: Association of Nmd4 to NMD substrates depends on Detector.	93
Figure EV1: Tagged proteins are functional for NMD.	111
Figure EV2: N-terminal and C-terminal tagged Upf1 enrich similar sets of specific proteins.	112
Figure EV3: Deletion of <i>NMD4</i> and <i>EBS1</i> stabilize a set of transcripts that is also stabilized in the absence of <i>UPF1</i>	113

Figure EV4: The helicase domain of Upf1 alone can destabilize RPL28 pre-mRNA, an NMD substrate.....	114
Appendix Fig. S1: Workflow for quantitative analysis of MS/MS results from affinity purified complexes	123
Appendix Fig. S2: Controls of total tagged protein levels in the presence or absence of other NMD components.....	125
Appendix Fig. S3: Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences	126
Appendix Fig. S4: Effect of NMD4 and EBS1 in the destabilization of an NMD substrate by the Upf1-HD-Cter domain.....	127
Appendix Fig. S5: Comparison between the canonical SURF/DECID model features (A) and our extended Detector/Effector model (B) for NMD.	128
Appendix Fig. S6. Similarities and differences between yeast Nmd4 PIN and PIN domains of human SMG6 and SMG5	129
Figure 26 : Précision sur les interactions d'Upf2	138
Figure 27 : Modèle de l'adressage au protéasome des peptides néo-synthétisés, par Ubr1 ..	140
Figure 28 : Comparaison des partenaires enrichis dans les purifications de Dcp1 et Hrr25..	142
Figure 29 : Impact de l'extrémité C-terminale d'Upf1 sur les interactions protéiques.....	145

Table des tableaux

Table 1: Caractéristiques des souches Upf1 sauvage et mutantes purifiées	65
Appendix Table S1: Summary of the number of replicates, number of proteins robustly quantified and number of proteins lost after RNase treatment.....	130
Appendix Table S2: RNA sequencing raw data analysis summary.....	131
Appendix Table S3 (strains).....	132
Appendix Table S4 (plasmids).....	133
Appendix Table S5 (oligonucleotides).....	134

Résumé

Le *Nonsense-Mediated mRNA Decay* (NMD) détecte et dégrade les ARNs qui ont une terminaison de la traduction précoce. Il affecte une grande diversité d'ARNs cytoplasmiques, c'est la voie majeure de dégradation des ARNs aberrants.

Les ARNs ciblés par le NMD chez la levure ont une phase ouverte de lecture courte et un long 3'-UTR. Comment ces caractéristiques permettent une dégradation efficace par le NMD et quelles sont les étapes du mécanisme reste peu clair.

À partir de 112 purifications d'affinités suivies d'une analyse en spectrométrie de masse quantitative, nous avons identifié deux complexes distincts impliqués dans le NMD. Ces deux complexes, mutuellement exclusifs, s'articulent autour d'Upf1, la protéine majeure du NMD. Le premier complexe, nommé Détecteur, aurait un rôle dans la reconnaissance des cibles alors que le deuxième complexe, nommé Effecteur, initierait la dégradation des ARNs par une interaction directe avec la machinerie de décapping.

Les facteurs impliqués dans ce modèle sont tous conservés à travers les eucaryotes et les étapes décrites sont transposables d'après la littérature. Cela semble indiquer un nouveau paradigme pour le mécanisme du NMD qui s'articulerait autour d'une base universelle commune à laquelle pourrait s'ajouter des étapes spécifiques des organismes ou des types d'ARNs.

Mots clés : [NMD – dégradation des ARNs – Spectrométrie de masse quantitative – Upf1 – levure]

Abstract

Composition and dynamic of Nonsense-mediated mRNA decay complexes in yeast

Nonsense-Mediated mRNA Decay (NMD) detects and degrades RNA for which translation ends prematurely. It affects a large diversity of cytoplasmic RNAs; it is the major decay pathway for aberrant RNAs.

In yeast, NMD targeted RNAs have a short open reading frame and a long 3'-UTR. How these two features lead to an efficient degradation through NMD and what are the steps of this mechanism is still unclear.

From 112 affinity purifications of NMD factors followed by an analysis using quantitative mass spectrometry, we identified two distinct complexes. Those complexes were mutually exclusive and both contained Upf1, the major NMD protein. A first complex, named Detector, might have a role in the NMD substrates recognition whereas the second one, named Effector, would initiate the degradation through a direct interaction with the decapping machinery.

The factors involved in our new model are all conserved throughout eukaryotes and the steps we describe have potential equivalents in other species. Our data suggest a new paradigm for the NMD mechanism that would be organised around a shared universal base to which specific steps could be added in certain organisms or for certain types of RNA substrates.

Keywords : [NMD – RNA decay – Quantitative mass spectrometry – Upf1 – yeast]