Fidélité de la traduction chez les eucaryotes. De la molécule au génome

Hélène Chommy

▶ **To cite this version:**

Hélène Chommy. Fidélité de la traduction chez les eucaryotes. De la molécule au génome. Sciences agricoles. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA112193 . tel-00749760

HAL Id: tel-00749760

https://theses.hal.science/tel-00749760

Submitted on 8 Nov 2012

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

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

UNIVERSITÉ PARIS SUD XI
UFR SCIENTIFIQUE ORSAY

Ecole doctorale : Gènes, Génomes, Cellules
Spécialité : Génétique

THÈSE

En vue de l'obtention du grade de
DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS XI ORSAY

Fidélité de la traduction chez les eucaryotes : de la molécule au génome.

Présentée par
Hélène CHOMMY

Soutenance le 21 septembre 2012 devant la commission d'examen composée de :

Rapportrices :	Dr. Micheline FROMONT-RACINE
	Dr. Emmanuelle SCHMITT
Examinateurs :	Dr. Valérie HEURGUE-HAMARD
	Pr. Alain DENISE
Directeur de thèse :	Dr. Olivier NAMY

"... This affair must be unravelled from within. " He tapped his forehead. "These little grey cells. It is « up to them »- as you say over here."

Hercule Poirot

Chapter 10, *The mysterious affair at Styles.*
Agatha Christie

Cette aventure a été rendue possible grâce à Olivier Namy qui m'a offert l'opportunité de travailler dans son laboratoire alors qu'il ne me connaissait pas. *Olivier je te remercie de m'avoir fourni l'occasion de développer ces 3 projets aussi variés que formateurs, et de m'avoir fait confiance pour présenter régulièrement ces travaux en congrès. J'ai appris beaucoup à ton contact.*

Une des craintes que l'on peut avoir en commençant une thèse en biologie est d'étudier la vie en passant à côté de la sienne. Avec vous les GMT/GST, je peux affirmer que cette tranche de ma vie en compagnie des « fous du bout » a été riche à bien des égards et je ne saurai jamais assez vous en remercier. Il faut bien l'avouer, comme labo il est difficile de faire plus barré !! Grâce à vous, même dans les moments difficiles, je n'ai jamais regretté d'avoir signé chaque année pour ce drôle de métier.

Je tiens à témoigner ma reconnaissance aux personnes suivantes : Céline Fabret, ma chère voisine de bureau, pour sa disponibilité et ses conseils. Damien Brégeon dont la douce voix grave « tinte » parfois aux oreilles… :-p *Je te souhaite de réussir dans ta nouvelle équipe.* Henri Grosjean pour sa passion des ARNt et pour le cours accéléré de biochimie. Stéphane Demais qui a des mains en or pour les ARN et toujours de bons conseils filmographiques. Matthieu Saguy à l'organisation horizontale et l'humour dévastateur. *Je vous remercie tous les deux pour avoir repris le flambeau sur la génomique et VOA1.* Sandra Blanchet pour sa gentillesse. *Je te souhaite une bonne « terminaison » !* Olivier Bugaud avec son enthousiasme. *Je te souhaite un bon marathon scientifique.* Tous trois nous avons su affronter les péripéties parfois cocasses du labo en août ! Au passage, merci à Murielle Dutertre de nous avoir sauvés sur certains coups durs ! Laure Bidou avec ses chansons fredonnées et sa ténacité face aux représentants. Célia Floquet pour sa persévérance et avec qui j'ai eu plaisir à personnaliser le bain-marie. Vive les baleines ! Agnès Baudin-Baillieu pour son soutien, ses conseils de lectures, et les bières parties ! Jean-Pierre pour les tartines de cantal ainsi que la tasse de thé et le carré de chocolat. Nelly Sainsard (« super Nelly ») qui en se chargeant de l'intendance au quotidien nous permet de réaliser nos expériences dans les meilleures conditions. Magali Prigent et Marie-Hélène Cuif pour leur aide sur certaines techniques, et pour avoir partagé nos réunions de labo. Je tiens aussi à remercier Catherine Drouet et Muriel Decraene (« Mumu !!! ») qui sont de super secrétaires.

Enfin, last but not least : Isabelle !! Le pilier du labo. Evidemment, avec tes étudiantes en BTS (Noémie Brisemeur, Coline Chabrier et Mathilde Michelet), je vous remercie pour l'aide apportée pour le projet eEF2 et le criblage phénotypique. Mais c'est surtout pour tout le reste que je te suis reconnaissante : les innombrables discussions en voiture, les nombreux fous-rires, les conseils, les surnoms improbables (« tartine » ??), la course en sac, les sorties (1h sur ce carrefour à Londres…). *MERCI Isa d'avoir préservé en moi la part du lapin de 6 semaines.*

Lors de mes projets, j'ai eu l'occasion de collaborer avec d'autres équipes. Je remercie donc Alain Denise, Claire Kuchli et surtout Christine Drevet pour la partie bioinfo. *Merci Christine de t'être laissée embringuer dans l'analyse des ribosome profiling et surtout d'avoir été là lors de la rédaction du papier.*

Pour la partie particule unique, je remercie Karen Perronet, Nicolas Fiszman, Antoine Le Gall, David Dulin et Nathalie Westbrook. *Merci pour votre enthousiasme et vos explications et surtout votre*

patience face aux aléas de la biologie… Ca a été un véritable plaisir de découvrir votre univers et de travailler en nocturne avec vous. Je vous remercie aussi de m'avoir adoptée et soutenue le temps de l'écriture de ce manuscrit. J'espère que l'excursion chez Puglisi va apporter des solutions à certains de nos problèmes.

Je tiens à remercier les personnes de l'institut qui assurent le fonctionnement au quotidien. Ainsi que Barry Holland et Simone Seror pour les groupes de discussion en anglais, Anna Bizard pour avoir organisé ensemble les journées de l'IGM et le BBQ. Stéphanie Beaunay pour l'organisation de la fête de la science.

Je pense aussi à Alain Raynal, Thierry Locatelli, Catherine Esnault, Thierry Hoyau, Claire Toffano, Stéphane Descorps-Declere, Vladimir Daric, Jean Lehmann, Miche Guérineau, Jean-Luc Pernodet et son équipe pour avoir fait vivre la cafet' au fil des saisons. Sans oublier les jeunes : Jules, Cyrielle, Jennifer, Mohea, Audrey, Alban, Cécile, Marc, Rémy, Zohra, Antonin et Magali N.

Je tiens aussi à remercier mes amis, mes parents et mon frère pour leur soutien indéfectible malgré la distance de certain(e)s.

Quant à l'orthographe de ce manuscrit, je tiens à préciser qu'il doit beaucoup aux qualités de « proofreading » de Karen et surtout de ma mère.

Enfin, je tiens à remercier Micheline Fromont-Racine et Emmanuelle Schmitt pour avoir accepté de rapporter mes travaux de thèse, ainsi que Valérie Heurgué-Hamard et Alain Denise pour avoir accepté d'être examinateurs. D'autre part, je tiens aussi à remercier Valérie et Daniel Gautheret pour avoir été les parrains de ma thèse.

Ces travaux ont été financés par l'Agence Nationale pour la Recherche, l'Association pour la Recherche contre le Cancer, Photonics 4 Life, C'NANO Ile-de-France et le Pôle de Recherche et d'Enseignement Supérieur UniverSud Paris .

Mes trois premières années de thèse ont été financées par le Ministère de l'éducation et de la recherche et la quatrième année a été financée par la Fondation pour la Recherche Médicale.

Table des matières

Liste des abréviations

Introduction..... 1

Chapitre 1 : Etude du rôle d'eEF2 dans le maintien du cadre de lecture..... 55

Chapitre 2 : Etude de l'élongation en molécule unique..... 72

Chapitre 3 : Recherche d'événements de recodage dans le génome de *S. cerevisiae* et étude de *VOA1*. 92

Conclusions générales et perspectives..... 140

Annexe..... 143

Matériels et méthodes..... i

Liste des abréviations

-1CEV	**-1** frameshifting **C**ompatible **EV**ent
+1CEV	**+1** frameshifting **C**ompatible **EV**ent
A, C, G, T, U	Bases azotées : **A**dénine, **C**ytosine, **G**uanine, **T**hymine et **U**racile
ADN	**A**cide **D**ésoxyribo**N**ucléique
ADNc	**ADN c**omplémentaire
ADNr	**ADN r**ibosomique
AMPc	**A**dénosine **M**ono**P**hosphate **c**yclique
ARNm	**A**cide **R**ibo**N**ucléique **m**essager
ARNr	**A**cide **R**ibo**N**ucléique **r**ibosomique
ARNt	**A**cide **R**ibo**N**ucléique de **t**ransfert
ATP	**A**dénosine **T**ri**P**hosphate
CDS	**C**oding **D**NA **S**equence
CEV	**C**ompatible **EV**ent
CrPV	**Cr**icket **P**aralysis **V**irus
cryo-EM	**cryo-E**lectron **M**icroscopy
E. coli	*Escherichia coli*
eEF	**e**ukaryotic **E**longation **F**actor
GAC	**G**-ATPase **A**ctivating **C**enter
GTP	**G**uanosine **T**ri**P**hosphate
IBV	**I**nfectious **B**ronchitis **V**irus
IRES	**I**nternal **R**ibosome **E**ntry **S**ite
NGD	**N**o-**G**o **D**ecay
NMD	**N**on-sense-**M**ediated **D**ecay
NSD	**N**o-**S**top **D**ecay
ORF	**O**pen **R**eading **F**rame
RCEV	**R**eadthrough **C**ompatible **EV**ent
RRL	**R**abbit **R**eticulocyte **L**ysat
S. cerevisiae	*Saccharomyces cerevisiae*
smFRET	**s**ingle **m**olecule **F**orster **R**esonance **E**nergy **T**ransfer
SRL	**S**arcin-**R**icin **L**oop
TIRFM	**T**otal **I**nternal **R**eflexion **F**luorescence **M**icroscopy
UTR	**UnT**ranslated **R**egion
ZMW	**Z**ero **M**ode **W**aveguide

Introduction

Sommaire

Préambule : Rôle de la traduction 1

I. Le code génétique 2

1. Le code génétique universel 2

2. Les évolutions du code génétique. 2

II. Les acteurs de la traduction 3

1. L'ARNm 3

A. La coiffe 3

B. La région 5'UTR 3

C. La phase codante 4

D. La région 3' UTR 4

E. La queue polyA 4

2. Les ARNt 5

A. Structure des ARNt 5

B. Modifications des ARNt 5

C. Chargement des ARNt 6

D. L'ARNt iniateur $ARNt^{MET}_{i}$ 6

3. Le ribosome 7

A. Biogenèse du ribosome 7

B. Composition et structure du ribosome 7

C. Aparté historique : le ribosome et le prix Nobel 8

III. Les étapes de la traduction 10

1. L'initiation 10

A. Les sous-unités dans le cytoplasme 10

B. Le complexe ternaire et le pré-complexe d'initiation (PIC) 10

C. L'initiation dépendante de la coiffe 10

D. Initiation interne 11

2. L'élongation 11

A. Préambule à l'élongation 12

B. Reconnaissance initiale du codon 12

C. Accommodation de l'ARNt au site A : une étape de correction. 14

D. Formation de la liaison peptidique ou transpeptidation. 14

E. Translocation de l'ARNm et des ARNt 15

F. Maintien du cadre de lecture 17
G. Passage des structures secondaires 19
H. L'élongation chez les eucaryotes 19
3. La terminaison 21
A. Préambule à la terminaison 21
B. Le complexe de terminaison 21
C. Recrutement du complexe de terminaison par le ribosome 22
D. Reconnaissance du codon stop 23
E. Recrutement d'ABCE1/Rli1p 24
F. Libération de la protéine 25
4. Le recyclage des sous-unités ribosomiques 25
A. Dissociation des sous-unités ribosomiques 25
B. Recyclage de l'ARNm, de l'ARNt et des sous-unités ribosomiques 25
IV. Traduction et régulation de l'expression génique. 27
1. NGD (No-Go Decay) 27
A. Dom34p-Hbs1p 28
B. Dom34p-Rli1p 29
2. NSD (No-Stop Decay) 29
3. NMD (Non-sense-Mediated Decay) 30
V. Le recodage 31
1. La translecture 31
2. Le décalage du cadre de lecture en -1 34
A. Les signaux de décalage du cadre de lecture en -1. 34
B. Un mécanisme encore à préciser. 35
C. Régulation physiologique : exemple avec le virus du VIH 37
3. Le décalage du cadre de lecture en +1 37
A. Le rôle de l'ARNt au site P. 37
B. Les éléments stimulateurs 38
C. Abondance de décalages en +1 chez les Euplotes sur des stops « glissants ». 39
4. Saut de ribosome ou « By-passing » 40
VI. Travaux de thèse 41
Références bibliographiques 42
Références bibliographiques des figures 53

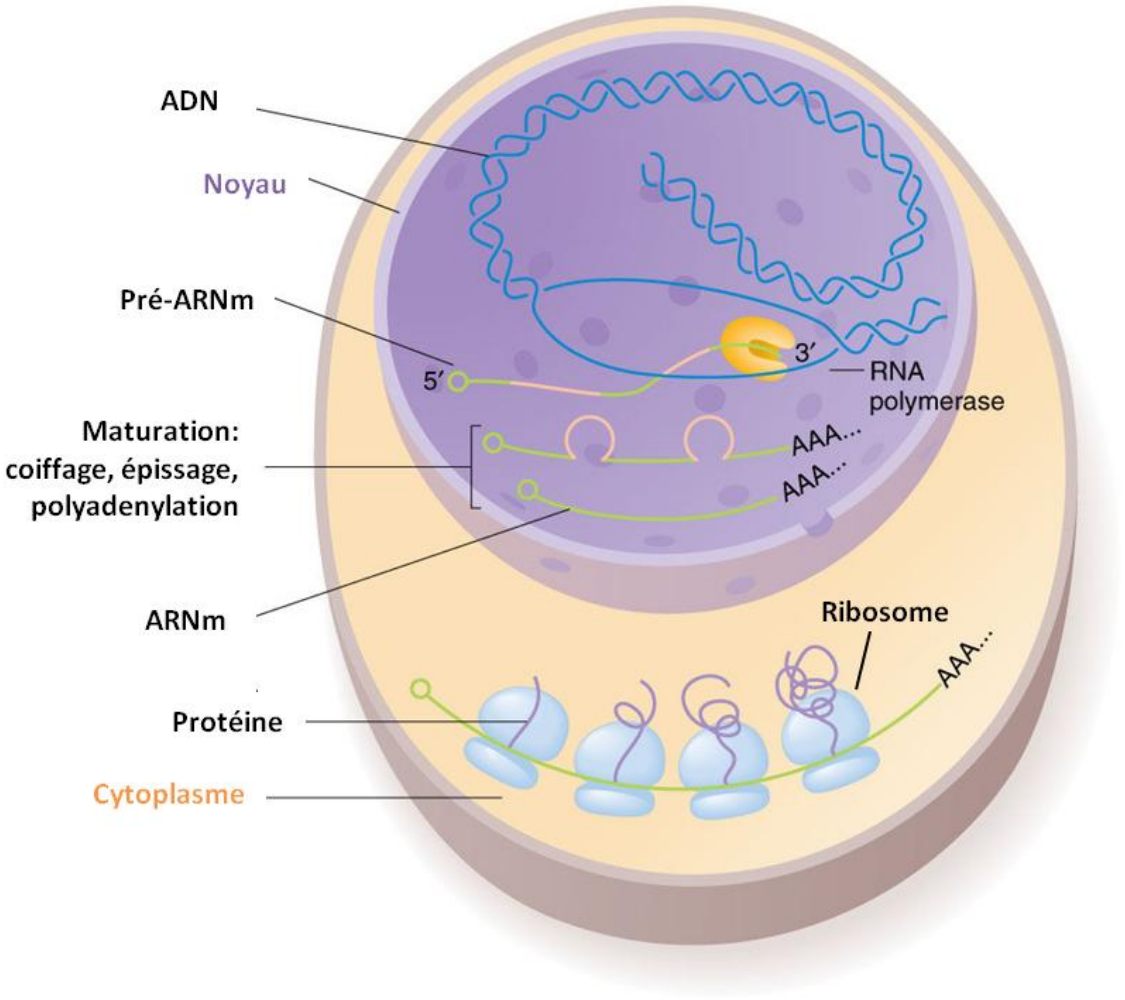


Figure 1 : Transcription et traduction dans une cellule eucaryote

Chez les eucaryotes, l'ADN est transcrit par l'ARN polymérase dans le noyau. Une fois le pré-ARNm maturé, l'ARNm est exporté dans le cytoplasme pour y être traduit par les ribosomes. Figure adaptée de Pearson education Inc 2010

Préambule : Rôle de la traduction

Dans l'ensemble du vivant, la production des protéines est un processus essentiel qui permet à chaque cellule de vivre et se reproduire. Cette activité consomme plus de la moitié de l'énergie produite par une cellule, sa mise en œuvre est donc régulée en fonction des ressources en énergies disponibles.

L'information permettant la production de chaque protéine est portée par des portions d'ADN (Acide DésoxyriboNucléique) appelées gènes. Les différentes étapes permettant de passer de l'information portée par un gène à une protéine fonctionnelle sont universellement conservées.

Tout d'abord, il y a une étape dite de transcription où la séquence ADN du gène est recopiée sous forme d'un ARN (Acide RiboNucléique) dit messager (ARNm). La molécule d'ARNm simple brin est ensuite lue pour synthétiser une protéine lors d'une étape dite de traduction. Ce terme découle du fait qu'à cette étape, l'information change de support moléculaire. Il y a une transition des acides nucléiques vers des acides aminés (figure 1).

Chez les eucaryotes, les processus de transcription et de traduction sont découplés. Dans le noyau, le pré-ARNm est transcrit et maturé en ARNm. Ce dernier est ensuite exporté dans le cytoplasme pour y être traduit (figure 1). Le processus de traduction est commun à tous les ARNm mais la traduction de chaque ARNm est régulée dans le temps et dans l'espace de la cellule. Cette régulation spatio-temporelle permet en partie le maintien de la physiologie cellulaire.

Au cours de ma thèse, je me suis intéressée à la fidélité de la traduction chez les eucaryotes, et en particulier chez la levure *Saccharomyces cerevisiae*. Au cours de cette introduction, j'utiliserai cet organisme comme un fil conducteur pour décrire différents aspects de la traduction. Tout d'abord, je décrirai les règles de la traduction, ses acteurs et les mécanismes moléculaires régissant les interactions entre ces différents acteurs. Je poursuivrai ensuite sur le rôle de la traduction dans la surveillance et la régulation de l'expression génique. Pour finir, je présenterai des événements dits de recodage dans lesquels les règles de la traduction sont transitoirement contournées afin de réguler l'expression de certains gènes.

Base 1	Base 2 T		Base 2 C		Base 2 A		Base 2 G		Base 3
T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T
	TTC		TCC		TAC		TGC		C
	TTA	Leu	TCA		TAA	**Stop**	TGA	**Stop**	A
	TTG		TCG		TAG	**Stop**	TGG	Trp	G
C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	Gln	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	Lys	AGA	Arg	A
	ATG	**Met**	ACG		AAG		AGG		G
G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	Glu	GGA		A
	GTG		GCG		GAG		GGG		G

Figure 2 : Le code génétique

Tableau regroupant les correspondances entre les 20 acides aminés naturels et les 61 codons sens. Les trois codons stop sont aussi indiqués.

I. Le code génétique

1. Le code génétique universel

Lors de la traduction, l'ARNm est décodé par unité de 3 bases. Chaque triplet, appelé codon, correspond à un acide aminé. Le code génétique regroupe l'ensemble des correspondances entre un codon et un acide aminé (figure 2). Parmi les 64 codons possibles, 61 sont assignés à un acide aminé et sont donc appelés codon « sens ». Les 3 codons restant : ambre (UAG), ocre (UAA) et opale (UGA) sont des codons « stop », ils servent à arrêter la traduction.

Les 61 codons sens sont associés à un des 20 acides aminés les plus communément trouvés dans les protéines. En général, plusieurs codons correspondent au même acide aminé c'est pourquoi on parle de redondance du code génétique. Seuls les acides aminés méthionine (AUG) et tryptophane (UGG) dérogent à cette règle. Dans la vaste majorité des cas, les protéines commencent par l'acide aminé méthionine.

Les différents codons associés à un même acide aminé diffèrent souvent au niveau de la dernière base. Cette observation a mené à parler de dégénérescence du code génétique. Les codons correspondant aux acides aminés leucine, sérine et arginine suivent cette règle mais pour chaque acide aminé les codons se répartissent en 2 groupes en fonction des 2 premières bases. Pour la leucine, les 2 groupes sont (CU-) et (UU-).

En plus des 20 acides aminés pris en charge par le code génétique, il existe 2 acides aminés naturels rares qui sont la sélénocystéine et la pyrrolysine. Elles sont incorporées dans certaines protéines par des mécanismes de recodage évoqués dans la partie V.3.c .

Le code génétique est commun à tous les organismes vivants. Néanmoins, la perception d'un code génétique universel est à moduler car il a évolué chez certains organismes.

2. Les évolutions du code génétique.

Le code génétique a longtemps été évoqué comme un ensemble de règles qui ont été fixées par accident à un temps donné de l'évolution et qui depuis sont immuables (Crick 1968). Cet axiome a été écorné par l'observation de petites différences entre le code génétique utilisé pour les gènes nucléaires de vertébrés et celui utilisé pour les gènes de leurs mitochondries. Chez les mitochondries, les codons AGA et AGG sont réassignés en codons stop et le codon UGA correspond au tryptophane. Depuis, la découverte de nouveaux génomes a enrichi la liste des codons réassignés et a montré que le code génétique évolue encore dans certaines lignées. Etonnement, certains changements se sont produits de manière répétée et indépendante dans différents taxons (figure 3). De nombreuses études essayent de comprendre comment le code génétique évolue et favorise la réassignation de certains codons en particulier (Knight, Freeland and Landweber 2001).

Dans la partie V.3.c, nous verrons comment chez les Euplotes la réassignation du codon stop UGA en codon sens affecte l'expression cellulaire d'une sous-population de gènes.

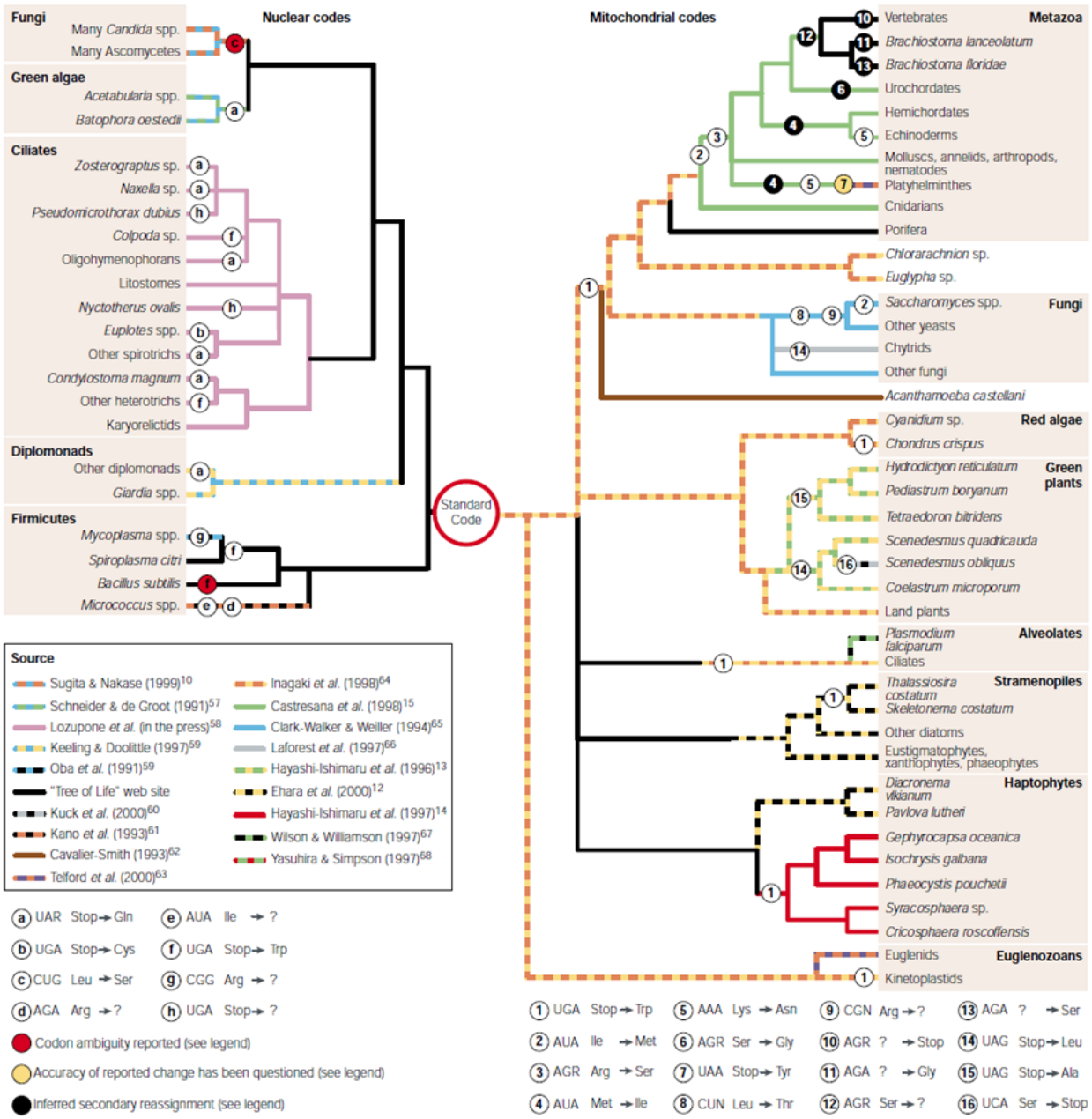


Figure 3 : Réassignation de codon au cours de l'évolution

L'arbre phylogénétique replace l'évolution de la réassignation de différents codons dans les génomes nucléaires (à gauche) et mitochondriaux (à droite). Figure issue de (Knight, Freeland and Landweber 2001)

II. Les acteurs de la traduction

Le processus de traduction est réalisé par l'action conjointe et coordonnée de plusieurs acteurs qui sont :

(1) l'ARNm qui porte la séquence codant la protéine.
(2) les ARN de transfert (ARNt) qui permettent de décoder chaque codon.
(3) le ribosome qui valide le décodage et synthétise la protéine
(4) les facteurs protéiques qui vont permettre la réalisation des différentes étapes de la traduction

Avant de décrire les différents mécanismes impliqués dans la traduction, je vais tout d'abord présenter les 3 acteurs principaux : l'ARNm, les ARNt et le ribosome.

1. L'ARNm

Chez les eucaryotes, les pré-ARN messagers sont transcrits par l'ARN polymérase II. Ils sont maturés co-transcriptionnellement par plusieurs enzymes pour donner des ARN messager matures (ARNm). L'ARNm est une molécule orientée qui se lit de l'extrémité 5' vers l'extrémité 3'. Lors de la maturation, une coiffe est ajoutée à l'extrémité 5' de l'ARNm, les introns sont épissés et l'extrémité 3' est polyadénylée (figure 4). Chez *S. cerevisiae*, les ARNm représentent 5% des ARN totaux. Bien que très peu de gènes soient épissés, les ARNm épissés font partis des ARNm les plus fortement exprimés et représentent 40% des ARNm d'une cellule. (Warner 1999).

A. La coiffe

Le coiffage consiste à ajouter co-transcriptionnellement un 7-méthylguanosine triphosphate (m7G) à l'extrémité 5' de l'ARNm. Cette modification protège l'ARNm de la dégradation par l'exonucléase Xrn1p qui a une activité de dégradation 5'-3'. L'interaction de la coiffe avec différents complexes protéiques joue un rôle dans le recrutement de la machinerie d'épissage, la polyadénylation et le transport de l'ARNm. De plus, cette modification permet le recrutement du ribosome lors de l'initiation traductionnelle. Ce point est abordé dans le paragraphe III.1 (Furuichi and Shatkin 2007).

B. La région 5'UTR

L'ARNm ne commence pas directement par la séquence codant la protéine mais par une région non traduite appelée 5'UTR (5' UnTranslated Region). La taille de cette région est variable, elle dépend de la position du promoteur qui est utilisé par l'ARN polymérase II pour initier la transcription. En moyenne, elle est de ~ 80 nucléotides chez *S. cerevisiae* (Tuller, Ruppin and Kupiec 2009). Le 5'UTR joue un rôle dans l'initiation de la traduction en permettant l'accroche du ribosome et la reconnaissance du codon d'initiation (cf partie III.1).

Malgré les termes de « région non traduite », certains 5'UTR incluent de petites régions uORF (upstream Open Reading Frame) dont la traduction sert à réguler la traduction de la phase codante en aval. Chez *S. cerevisiae*, une étude récente a montré que les uORF étaient assez abondantes dans les 5'UTR des gènes. Parmi les 4648 gènes exprimés dans les conditions de l'expérience (sur 5295

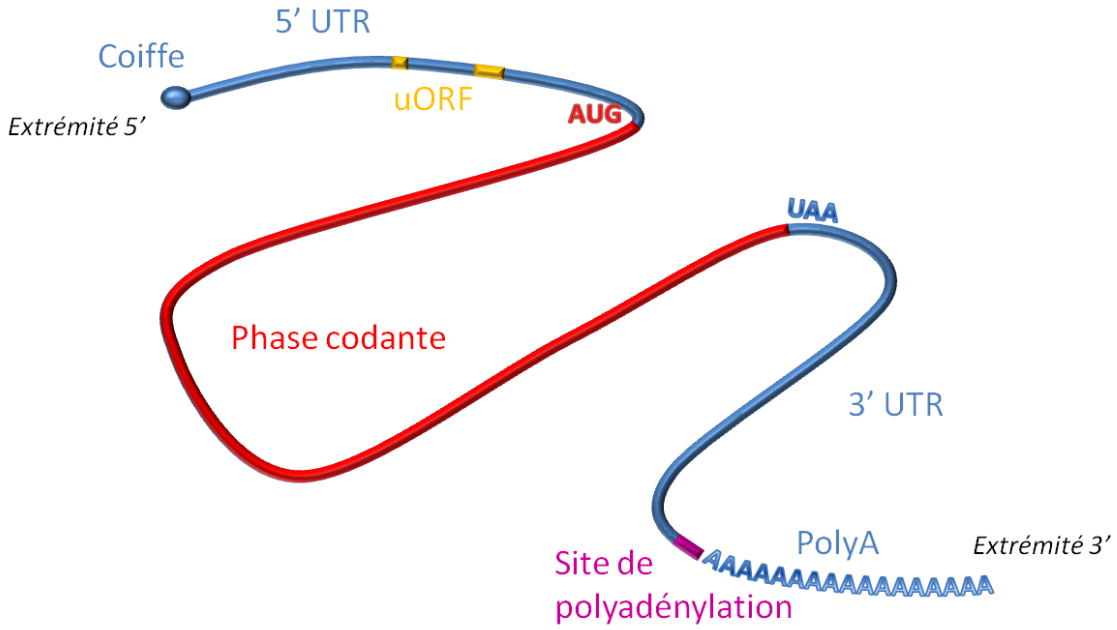


Figure 4 : Structure schématique d'un ARNm

Chez les eucaryotes, l’ARNm mature comprend : une coiffe à l’extrémité 5’, une région non-codante en 5’ (5’UTR) pouvant contenir de petites phases codantes (uORF), une région codante commençant par un codon de démarrage et se finissant au premier codon stop rencontré, une région non-codante en 3’ (3’UTR) et en enfin une queue poly A.

gènes au total), 706 présentent des uORF dans leur 5' UTR. Il semblerait que les uORF ne commençant pas par un AUG soient les plus abondantes, cependant une quantification précise n'est pas mentionnée dans l'article (Ingolia et al. 2009). Le cas le plus connu de régulation traductionnelle par des uORF est celui du gène *GCN4*. (Miller and Hinnebusch 1990).

C. La phase codante

La phase codante correspond à la région de l'ARNm qui est traduite par le ribosome. Elle commence par un codon de démarrage et se termine par le premier codon stop rencontré dans la même phase de lecture. Le codon de démarrage est généralement un AUG mais les codons GUG et UUG sont parfois utilisés. Le codon stop le plus utilisé chez *S. cerevisiae* est le codon UAA (47%), viennent ensuite les codons UGA (30%) et UAG (23%) (Saccaromyces Genome Database).

La traduction de la phase codante peut être modulée par des structures secondaires présentes dans l'ARNm. Nous verrons au cours de cette introduction que certaines structures peuvent influencer la traduction à différents moments : initiation, vitesse et fidélité de l'élongation et de la terminaison.

D. La région 3' UTR

La fin d'un ARNm n'est pas définie par le codon stop. Il existe après la phase codante une région non traduite appelée 3'UTR (3' UnTranslated Region). La taille de cette région est déterminée par la position du site de polyadénylation dans le pré-ARNm. Chez *S. cerevisiae*, sa taille est en moyenne de ~ 140 nucléotides (Tuller et al. 2009). Chez *S. cerevisiae*, le site de polyadénylation est composé d'une séquence de reconnaissance dont le consensus est AAUAAA. Ce motif est reconnu par le complexe CPSF (Cleavage and Polyadenylation Stimulation Factor). A 20-30 nucléotides en aval du motif AAUAAA, se trouve une courte séquence G/U riche qui est reconnue par le complexe CstF (Cleavage Stimulation factor). Les 2 complexes CPSF et CstF recrutent les protéines CF1 et CF2 (Cleavage Factor). Ces dernières coupent le pré-ARNm juste en amont de la séquence G/U riche (Proudfoot and O'Sullivan 2002).

Chez *S. cerevisiae*, 10 à 15% des transcrits possèdent des sites alternatifs de polyadénylation (Nagalakshmi et al. 2008). Dans des cellules mammifères, il existe une corrélation entre la taille de la région 3'UTR et l'efficacité de traduction de certaines protéines, plus le 3' UTR est petit plus la synthèse est abondante (Di Giammartino, Nishida and Manley 2011).

E. La queue polyA

La queue poly A est une extension composée de ~ 250 adénines chez les mammifères et de 50 à 80 adénines chez *S. cerevisiae* (Viphakone, Voisinet-Hakil and Minvielle-Sebastia 2008). Elle est ajoutée post-transcriptionnellement en 3' de l'ARNm par la polyA polymérase (PAP). Lors de la polyadénylation, des protéines PABP (Poly A Binding Protein) nucléaires Nab2p se fixent sur l'extension en cours de synthèse. Ces PABP jouent un rôle dans la régulation de la polyadénylation, la stabilisation de l'ARNm et l'export de celui-ci dans le cytoplasme.

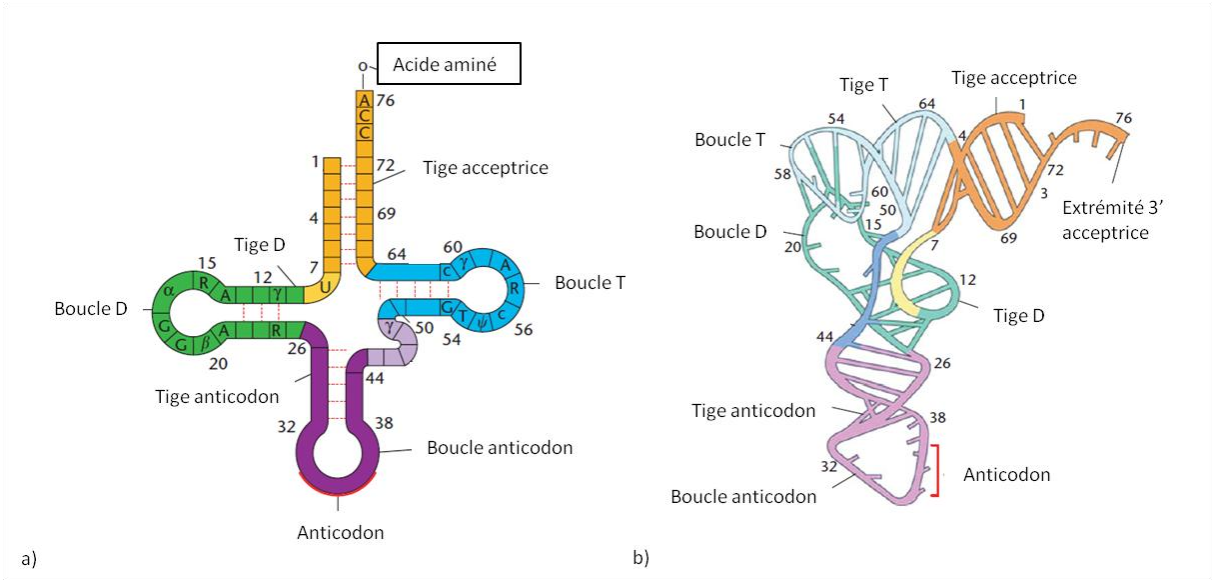


Figure 5 : Structure secondaire et tertiaire d'un ARNt

A gauche, structure secondaire en forme de trèfle d'un ARNt aminoacylé. A doite, structure tertiaire en forme de L d'un ARNt déacylé. Les codes couleurs des domaines sont conservés pour les 2 schémas. Figure adaptée de (Goldman 2008)

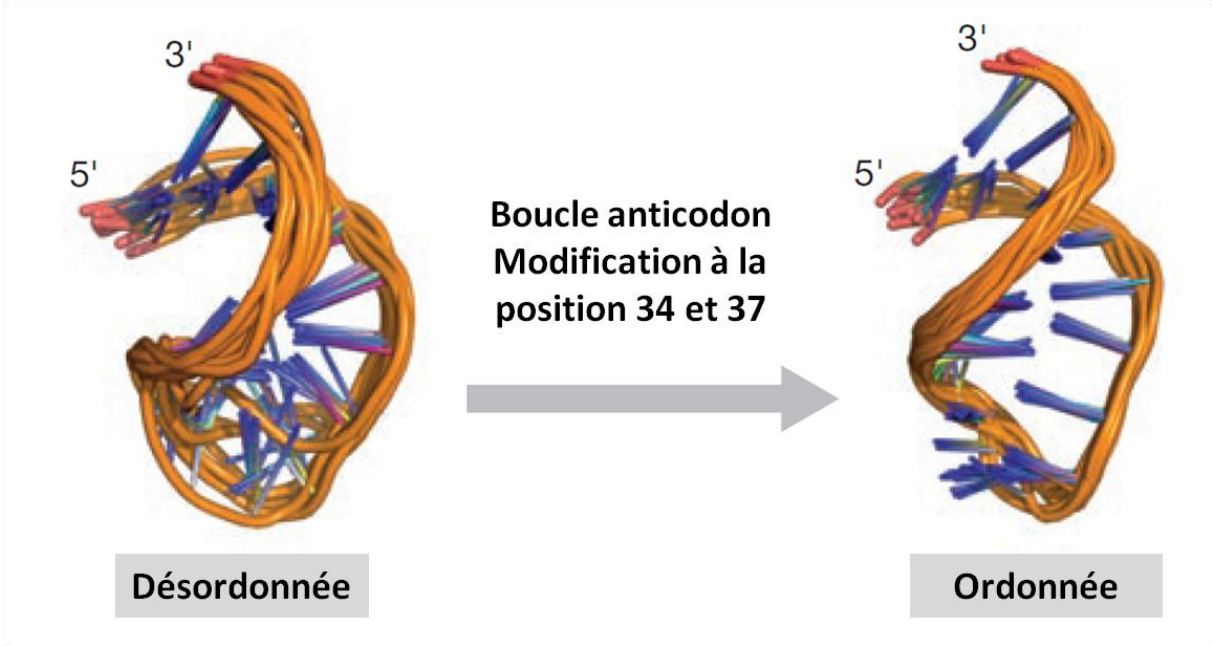


Figure 6 : Modifications de la tige et de la boucle de l'anticodon : effet sur la stabilité de la structure.

Sur la gauche, la tige-boucle de l'anticodon n'est pas modifiée et est désordonnée. Sur la droite, les modifications aux bases 34 et 37 restreignent la dynamique de la boucle anticodon et contraignent les bases à s'orienter vers l'extérieur de la boucle. Figure adaptée de (Agris 2008)

Les PABP nucléaires sont ensuite remplacées par des PABP cytoplasmiques Pab1p qui aideront à l'initiation traductionnelle et au recyclage des ribosomes (Mangus, Evans and Jacobson 2003). Le remplacement des PABP nucléaires par des PABP cytoplasmiques n'est pas encore bien compris. Par exemple, le mécanisme et la localisation cellulaire de l'échange des protéines Nab2p par des protéines Pab1p demeurent inconnus.

Pour finir, la queue polyA permet de réguler la dégradation de l'ARNm par l'exosome. L'exosome est un complexe dont certaines protéines sont des exonucléases avec des activités de dégradation 3'-5'.

2. Les ARNt

Les ARNt sont de petites molécules d'ARN de 75 à 95 nucléotides. Ils servent d'adaptateurs entre l'ARNm et les acides aminés (Goldman 2008). Chez *S. cerevisiae*, ils comptent pour 15% des ARN totaux (Warner 1999).

A. Structure des ARNt

Ces molécules sont hautement structurées. La structure secondaire forme un trèfle à 4 branches dont 3 des branches sont des tiges-boucles (figure 5).

La première tige correspond au bras accepteur. Elle est formée par l'appariement des extrémités 5' et 3'. Les 3 dernières bases de l'extrémité 3' ne sont pas appariées et sont toujours constituées du motif CCA. Lorsque ce triplet n'est pas présent dans le gène de l'ARNt, une ATP(CTP):tRNA nucleotidyltransferase l'ajoute post-transcriptionnellement (Toh et al. 2009). L'extrémité 3'OH de l'ARNt lie l'acide aminé via une liaison ester avec le ribose du dernier nucléotide.

La tige-boucle D tire son nom de la présence presque universelle de la base modifiée dihydrouridine (D) dans la boucle. De la même manière, la tige-boucle T ou TψC est nommée ainsi à cause de la présence d'une ribothymidine (T) et d'une pseudouridine (ψ).

La boucle A contient un triplet de bases appelé anticodon permettant la reconnaissance du codon. Cette reconnaissance passe par l'appariement respectif des bases 34, 35 et 36 avec les bases 3, 2 et 1 du codon (Goldman 2008).

Dans la structure tertiaire, les 4 tiges se replient pour former une molécule en forme de L (figure 5). Chaque côté est composé d'un empilement coaxial de 2 tiges : le bras accepteur avec la tige T ; la tige A avec la tige D (Goldman 2008).

B. Modifications des ARNt

Les ARNt sont transcrits par l'ARN polymérase III. L'épissage et les modifications post-transcriptionnelles sont effectués à différentes étapes de leur synthèse (Motorin and Helm 2010).

Les bases modifiées participent à la stabilisation de la structure tertiaire. Comme dans un ressort, elles rigidifient la structure globale de l'ARNt tout en laissant une flexibilité locale. Comme nous le verrons par la suite, la rigidité de la tige A facilite une entrée rapide de l'ARNt dans le ribosome mais la souplesse de certaines régions est nécessaire à son incorporation. De la même manière, les bases

modifiées de la boucle anticodon ont un rôle structural. Elles contraignent les bases à s'orienter vers l'extérieur de la boucle afin de faciliter la reconnaissance codon:anticodon (Agris 2008) (figure 6). Ces modifications contribuent aussi à l'efficacité de la traduction en augmentant la surface de contact accessible pour différents partenaires protéiques sans avoir à changer la séquence même de l'ARNt (Goldman 2008).

Les modifications post-transcriptionnelles jouent un rôle dans le décodage des codons. La base 34 de l'anticodon, qui est impliquée dans la reconnaissance de la 3ème base du codon, est souvent modifiée. Ces différentes modifications autorisent l'appariement avec d'autres bases que celles établies dans les interactions Watson-Crick (A:U ; C:G ; G:C et U:A) (Söll et al. 1966, Crick 1966, Agris 2004). Ces appariements non-canoniques ont donné le nom de « base wobble » (« bancal » en anglais) à la 3ème base du codon. Comme la modification de la base 34 permet à un ARNt de lire des codons qui diffèrent au niveau de la base wobble, une cellule n'a pas besoin de produire 61 ARNt pour lire les 61 codons sens. En pratique, on constate que chaque organisme utilise préférentiellement certains codons pour l'incorporation d'un acide aminé donné.

C. Chargement des ARNt

L'ARNt existe sous 2 formes : déacylé quand il est seul ou aminoacylé quand il porte un acide aminé. Une convention d'écriture a été adoptée pour savoir quelle forme est évoquée. Lorsqu'un ARNt par exemple liant la Leucine (Leu) est déacylé, il s'écrit $ARNt^{LEU}$. Quand il est aminoacylé le couple est écrit $Leu\text{-}ARNt^{LEU}$ (Goldman 2008).

Tous les ARNt fixant un même acide aminé sont des isoaccepteurs. Ils sont reconnus pas la même aminoacyl ARNt synthétase. Il en existe une pour chacun des 20 acides aminés. Chaque aminoacyl ARNt synthétase reconnait l'acide aminé à charger sur des critères de taille et de groupement chimique particulier. Elles discernent les ARNt à aminoacyler essentiellement grâce à la séquence de l'anticodon. Cependant pour certains ARNt l'élément de reconnaissance est situé dans la tige du bras accepteur. La fidélité de reconnaissance de l'acide aminé et des ARNt ainsi que l'activité d'édition de certaines de ces enzymes permettent un taux d'erreur inferieur à 1 mauvais acide aminé pour 1 000 000 chargés (Goldman 2008). La fidélité du chargement de l'acide aminé sur l'ARNt est cruciale car en cas d'erreur la cellule n'a pas de moyen pour la corriger.

D. L'ARNt iniateur $ARNt^{MET}_i$

L'ARNt initiateur (ARNti) $Met\text{-}ARNt^{MET}_i$ est différent des autres ARNt car il participe à la reconnaissance du codon d'initiation mais il n'est jamais utilisé par la suite lors de la synthèse des protéines. Chez les bactéries, l'acide aminé Met porté par l'ARNti est formylé, ce n'est pas le cas chez les eucaryotes. Les $ARNt^{MET}_i$ et l'$ARNt^{MET}$ sont aminoacylés par la même ARNt synthétase mais leurs séquences et modifications post-transcriptionnelles divergent. Ces caractéristiques leur permettent d'être reconnus exclusivement par des protéines différentes en fonction des étapes de la traduction (Rasmussen et al. 2009).

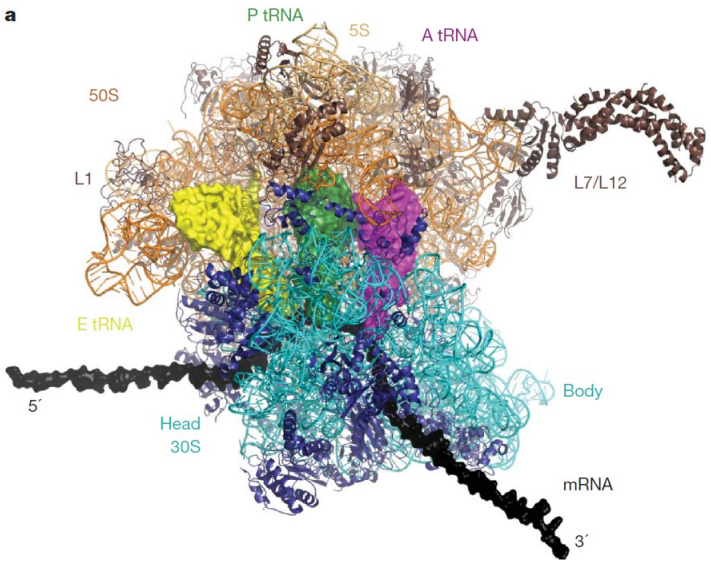


Figure 7 : Le ribosome avec les sites A, P et E

Ribosome bactérien 70S avec la petite sous-unité en bleue et la grande sous-unité en jaune. L'ARNm est en noir et les ARNt au site A, P et E sont respectivement en rose, vert et bleu. Figure adaptée de (Schmeing and Ramakrishnan 2009)

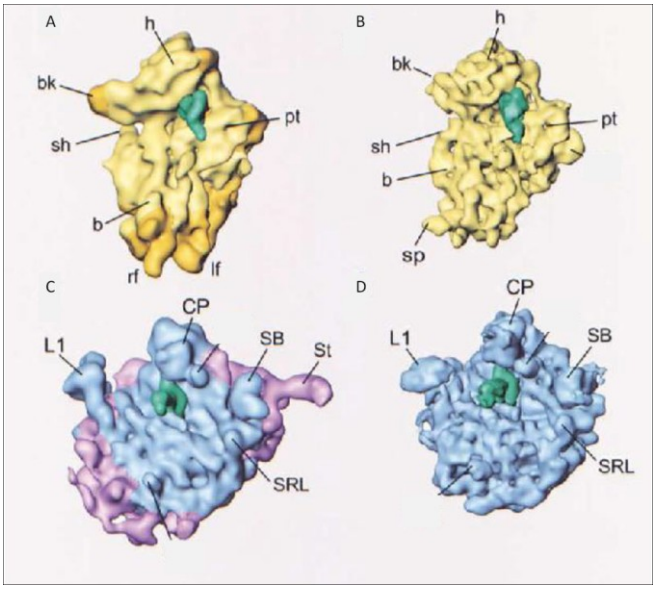


Figure 8 : Comparaison des petites et grandes sous-unités des ribosomes eucaryote et bactérien

A) et **B)** correspondent respectivement aux petites sous-unités eucaryote (40S) et bactérienne (30S). La face présentée est celle à l'interface avec la grande sous-unité. h : tête ; bk : becque ; sh : épaule ; b : corps ; rf : pied droit ; lf : pied gauche ; sp : aiguillon et pt : plateforme. La densité verte correspond à l'ARNt au site P

C) et **D)** correspondent respectivement aux grandes sous-unités eucaryote (60S) et bactérienne (50S). La face présentée est celle à l'interface avec la petite sous-unité. L1 : région L1 ; CP : protubérance centrale ; SB : « stalk base » ; St : complexe L7/L12 ; SRL : boucle Sarcine-Ricine.

Ribosome eucaryote : *Saccharomyces cerevisiae*, ribosome bactérien : *Escherichia coli*. Figure adaptée de (Spahn et al. 2001)

3. Le ribosome

Le ribosome est une macromolécule composée d'ARN ribosomique (ARNr) et de protéines, d'où son nom de complexe ribonucléoprotéique.

A. Biogenèse du ribosome

En phase de croissance, une cellule de *S. cerevisiae* compte ~ 200 000 ribosomes. La production est de ~ 2000 ribosomes par minute. (Warner 1999). Pour fabriquer cette quantité de ribosomes, il y a ~ 150 copies des gènes ADNr (10% du génome) et 137 gènes codant les protéines ribosomiques par génome haploïde.

Seule la moitié des gènes ADNr sont transcrits en ARNr (Ide et al. 2010), ces derniers représentent 80% des ARN totaux. Les ARNr 25S, 18S et 5,8S sont transcrits par l'ARN polymérase I et l'ARNr 5S est transcrit par l'ARN polymérase III. La plupart des 79 protéines ribosomiques sont codées par 2 copies du même gène. Bien que les gènes codant les protéines ribosomiques représentent seulement 2% des gènes totaux, ils représentent à eux seuls 90% des gènes épissés (Warner 1999).

La formation d'un ribosome est un processus long et complexe que je n'évoquerai pas dans cette introduction.

B. Composition et structure du ribosome

La comparaison des structures des ribosomes bactériens (Yusupov et al. 2001, Schuwirth et al. 2005), archéens (Ban et al. 1999) et eucaryotes (Chandramouli et al. 2008, Armache et al. 2010, Ben-Shem et al. 2011) montre qu'ils partagent un cœur commun : 2 sous-unités à l'interface desquelles se rencontrent l'ARNm et les ARNt dans les sites A, P et E (figure 7).

Le ribosome eucaryote 80S (densité 80 svedberg) est composé d'une petite sous-unité 40S et d'une grande sous-unité 60S. Les sites A, P et E accueillent les ARNt selon leur état d'acylation. L'ARNt aminoacylé entre au site A (Aminoacyl-ARNt). Le site P (Peptidyl-ARNt) contient l'ARNt lié à la protéine en cours de synthèse. Le site E (Exit) admet l'ARNt déacylé avant sa sortie du ribosome.

La sous-unité 40S est composée de 33 protéines et d'un ARNr 18S (Ben-Shem et al. 2011). Comme la petite sous-unité bactérienne, la structure générale de la sous-unité 40S comprend une tête, un corps et une plateforme (Spahn et al. 2001) (figure 8 A et B). La tête possède un bec et le corps a une épaule, tous deux sont du côté du site A. En plus des caractéristiques communes, la sous-unité 40S a en bas du corps un pied droit et un pied gauche (Spahn et al. 2001). Les bords de la tête et du corps forment un sillon dans lequel l'ARNm va venir se loger et être décodé par la boucle anticodon des ARNt. Le tunnel d'entrée de l'ARNm est formé par la tête et l'épaule alors que le tunnel de sortie est délimité par la tête et la plateforme (Kitzan et al. 1999). Sur la face contactant la grande sous-unité 60S, l'hélice 44 de l'ARNr 18S traverse la sous-unité du haut du corps jusqu'au pied droit (Spahn et al. 2001). Cette hélice porte dans sa boucle des bases conservées universellement et qui forment le site de décodage au site A (Ogle et al. 2001, Demeshkina et al. 2012).

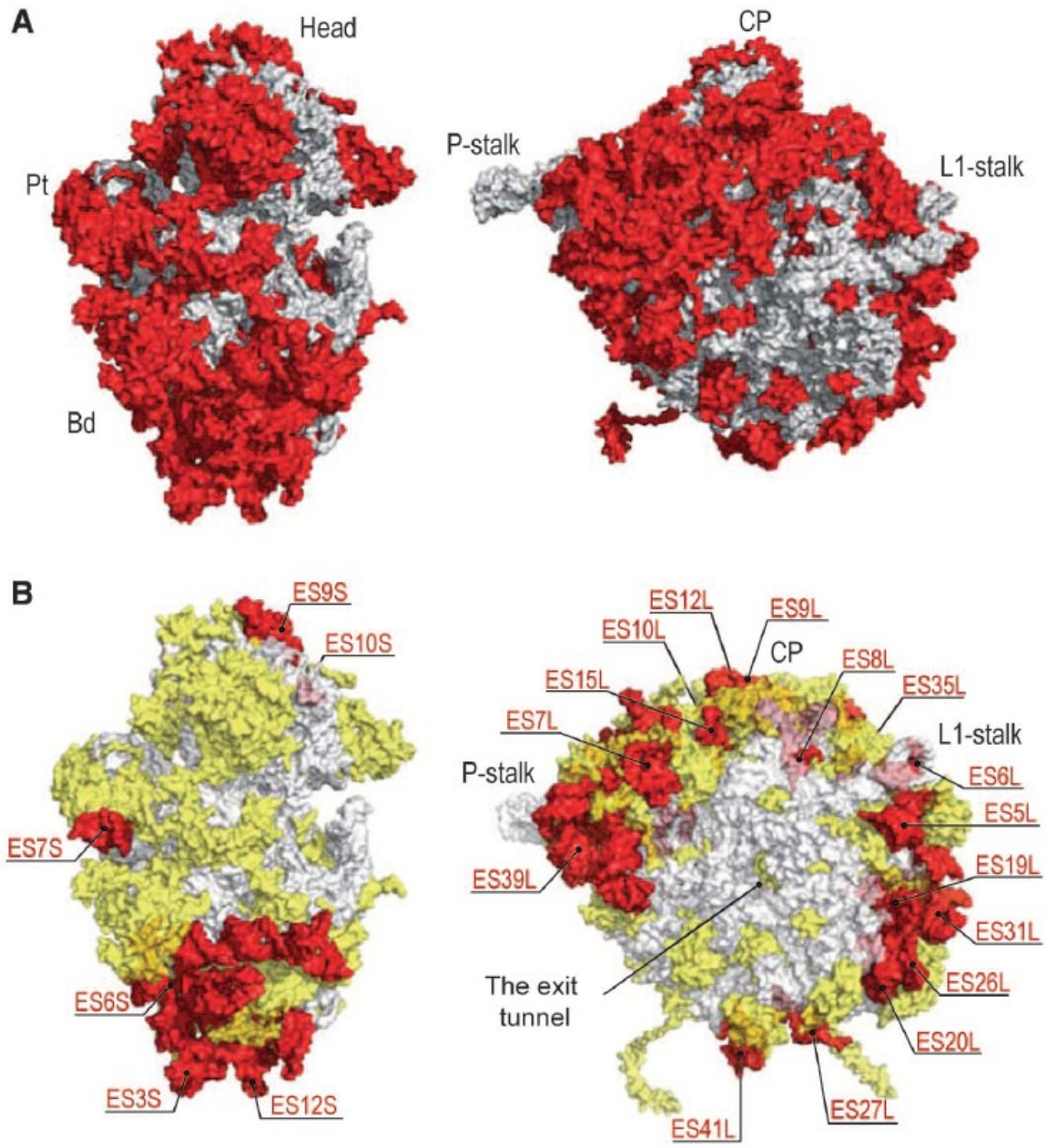


Figure 9 : Répartition des protéines et segments d'ARNr spécifiques au ribosome eucaryote

A) Répartition des protéines spécifiques au ribosome eucaryote (rouge) à la surface externe de la sous-unité 40S à gauche et 60S à droite. Head : tête ; Pt : plateforme ; Bd : corps ; P-stalk : région « stalk-base »; CP : protubérence centrale ; L1-stalk : région L1.

B) Répartition des protéines (jaune) et segments d'ARNr (rouge) spécifiques au ribosome eucaryote. A la surface de la sous-unité 40S (à gauche), protéines et ARNr spécifiques ne sont pas fortement colocalisés. A l'inverse, sur la sous-unité 60S, ils sont fortement associés.

Ribosome de *Saccharomyces cerevisiae*. Figure issue de (Ben-Shem et al. 2011).

La sous-unité 60S est composée de 46 protéines et des ARNr 25S, 5,8S et 5S (Ben-Shem et al. 2011). Comme la grande sous-unité bactérienne, la structure générale de la sous-unité 60S comprend une partie centrale couronnée de la « stalk base », d'une protubérance centrale et d'une région L1 (Spahn et al. 2001) (figure 8 C et D). La « stalk base » sert de support à une tige mobile composée de phosphoprotéines acides (protéines P). Chez les bactéries, cette tige est constituée des protéines L7 et L12 (Rich and Steitz 1987). La région L1, régule la sortie de l'ARNt déacylé du site E (Trabuco et al. 2010). La sous-unité 60S interagit avec les ARNt en fixant leurs tiges acceptrices. La liaison peptidique entre le nouvel acide aminé et la chaîne peptidique est catalysée par le centre peptidyl-transférase (PTC) situé entre les sites A et P. Un tunnel de sortie part de la cavité PTC et traverse la sous-unité 60S. Il permet à la protéine de sortir du ribosome durant sa synthèse. L'activité enzymatique du ribosome est portée par l'ARNr 25S, le ribosome est donc un ribozyme.

Les ribosomes eucaryotes sont environs 40% plus gros que les ribosomes bactériens. Leurs ARNr sont plus longs car ils contiennent des segments additionnels, et il y a plus protéines ribosomiques. Au sein des ribosomes eucaryotes, le nombre de protéines est presque constant (79 chez *S. cerevisiae*, 80 chez l'Homme) mais la taille des ARNr varie. Le ribosome du chien (Chandramouli et al. 2008) est plus gros que celui du blé *Triticum aestivum* (Armache et al. 2010) et des levures *Saccharomyces cerevisiae* et *Thermomyces lanuginosus* (Spahn et al. 2001, Taylor et al. 2009).

En 2011, l'équipe de Yusupov a publié la première carte du ribosome de *S. cerevisiae* à une résolution atomique de 3Å (Ben-Shem et al. 2011). Parmi les 79 protéines de *S. cerevisiae*, 32 protéines ont des homologues chez les bactéries et les archées. Ces protéines sont plus grosses que leurs homologues car elles possèdent des domaines spécifiques supplémentaires. Les 47 protéines spécifiques aux eucaryotes sont situées majoritairement à la périphérie du ribosome autour du cœur commun (figure 9 A). De même, les segments d'ARNr spécifiques aux eucaryotes se retrouvent essentiellement à la surface. Pour la sous-unité 40S, il y a peu de corrélation entre la position des protéines et des segments d'ARNr spécifiques. A l'inverse, pour la sous-unité 60S, les protéines et segments d'ARNr spécifiques sont fortement associés. Par exemple, dans la région « P-stalk », ils forment un site de contact pour le complexe qui va adresser le ribosome au réticulum endoplasmique (RE) (figure 9 B).

Le fait que les protéines ribosomiques eucaryotes soient retrouvées en surface conforte l'hypothèse d'un rôle dans la régulation spécifique de la traduction de certains ARNm. En 2011, l'équipe de Barna à montrer que la protéine Rpl38 est impliquée dans la régulation traductionnelle de certains gènes Hox chez la souris (Kondrashov et al. 2011). Les gènes Hox sont responsables de la mise en plan d'organisation du corps chez les mammifères.

C. Aparté historique : le ribosome et le prix Nobel

L'établissement de la structure du ribosome a longtemps été un challenge pour les biochimistes. Les premières études de cristallographie se sont concentrées sur les ribosomes d'archée et de bactérie résistants aux fortes concentrations salines utilisées pour la formation des cristaux. En 1983, Ada Yonath *et al.* ont montré pour la première fois que des cristaux pouvaient être produits à partir de sous-unités 50S de *Bacillus stearothermophilus* (Yonath et al. 1983). Durant les 2 décennies suivantes, des structures des sous-unités 30S et 50S ont été proposées mais avec de faibles résolutions. Ce n'est

qu'en 1998 que le laboratoire de Thomas Steitz présente une structure de la sous-unité 50S d'*Haloarcula marismortui* (archae) à une résolution atomique de 9 Å (Ban et al. 1998), puis de 5 Å en 1999 (Ban et al. 1999). La même année, le laboratoire de Venkatraman Ramakrishnan publie la structure de la sous-unité 30S de *Thermus thermophylus* (Clemons et al. 1999) à une résolution de 5 Å. En 2000, le laboratoire d'Ada Yonath descend la résolution de cette sous-unité à 3,3 Å (Schluenzen et al. 2000). La première structure complète d'un ribosome 70S de *Thermus thermophylus* avec des ARNt liés aux sites A, P et E à une résolution de 7,8 Å est présentée par l'équipe d'Harry Noller en 1999 (Cate et al. 1999). Une structure à 5,5 Å est rendu accessible en 2001 (Yusupov et al. 2001).

De nombreuses autres structures continuent à être publiées depuis. Elles permettent de saisir les interactions au sein du ribosome et ses changements de conformation à chaque stade de la traduction. Les changements de conformation observés en présence de certains antibiotiques ou inhibiteurs ont permis d'élucider leurs modes d'action sur la traduction ou sa fidélité.

En 2009, ce travail de longue haleine est récompensé par l'attribution du prix Nobel de Chimie à Ada Yonath, Thomas Steitz et Venkatraman Ramakrishnan. « *Ces 3 lauréats ont créé des modèles 3D montrant comment différents antibiotiques se fixent sur le ribosome. Ces modèles sont à présent utilisés par les chercheurs pour créer de nouveaux antibiotiques, aidant ainsi à sauver des vies et à diminuer les souffrances de l'humanité* » (communiqué publié par l'organisation Nobel).

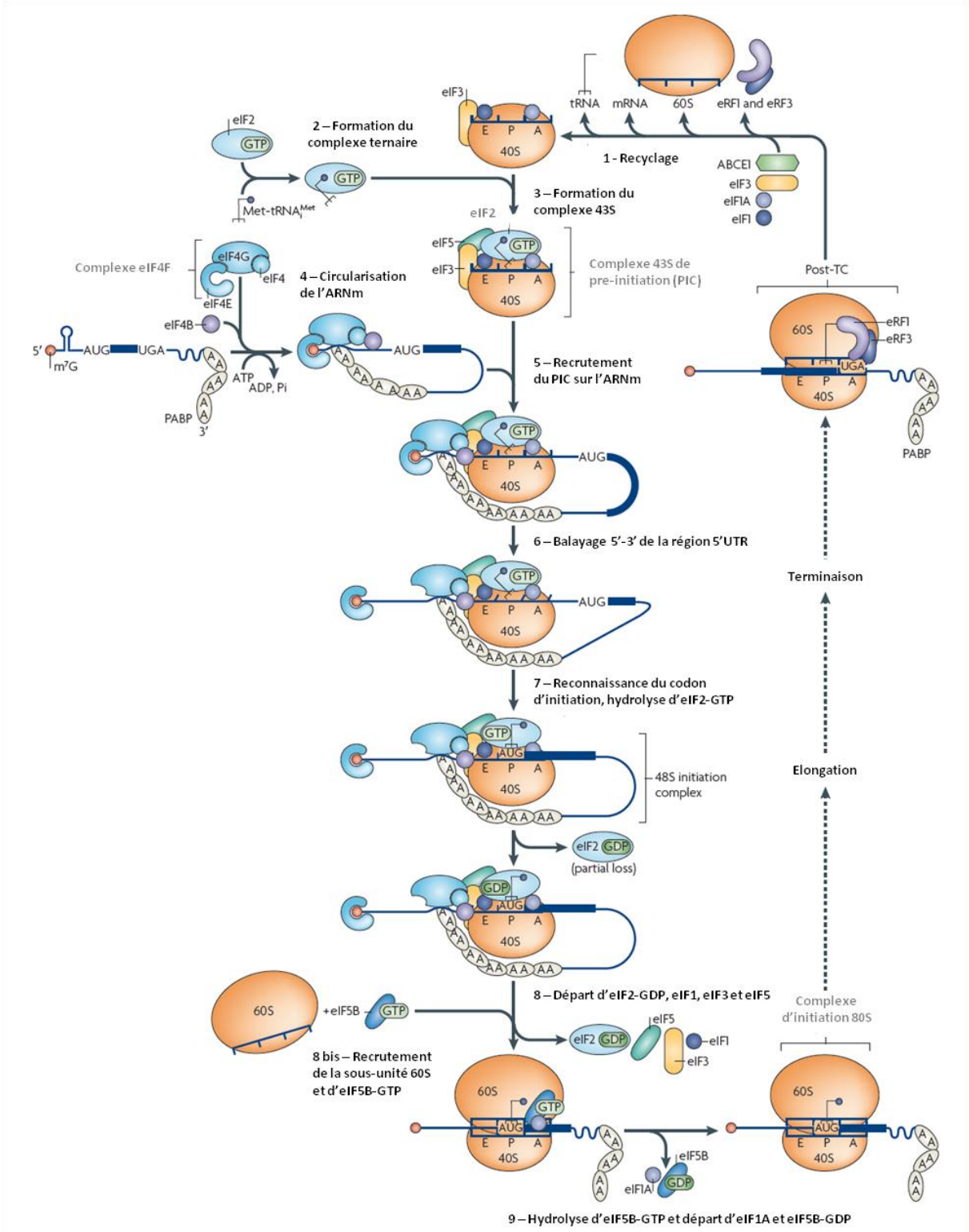


Figure 10 : Schéma des différentes étapes de l'initiation de la traduction dépendante de la coiffe chez les eucaryotes.

Figure adaptée de (Jackson, Hellen and Pestova 2010)

III. Les étapes de la traduction

La traduction est un processus d'amplification où une molécule d'ARNm est à l'origine de la synthèse de plusieurs protéines. L'ARNm est traduit simultanément par plusieurs ribosomes, on parle alors de polysomes. Chaque ribosome accomplit un cycle de traduction qui se décompose en 4 étapes :

(1) l'initiation durant laquelle les sous-unités ribosomiques s'assemblent sur l'ARNm et se positionnent sur le codon d'initiation.
(2) l'élongation correspond à la lecture de l'ARNm et à la synthèse de la protéine.
(3) la terminaison où la protéine est relâchée.
(4) la dissociation et le recyclage des sous-unités ribosomiques.

1. L'initiation

Chez les eucaryotes, l'initiation est complexe car elle requiert l'action d'un grand nombre de facteurs. Comme elle ne constitue pas le sujet central de ce manuscrit, je la décrirai dans ses grandes lignes (Sonenberg and Hinnebusch 2009, Jackson, Hellen and Pestova 2010, Hinnebusch 2011) (figure 10). Les facteurs d'initiation (eIF) sont au nombre de 13 et comprennent 33 protéines dont certaines forment des complexes. Il y existe aussi 5 facteurs auxiliaires dont seule la PABP sera mentionnée.

A. Les sous-unités dans le cytoplasme

Pour que les sous-unités restent dissociées dans le cytoplasme, elles sont prises en charge par des facteurs d'initiation. La sous-unité 40S est associée aux facteurs eIF1, eIF1A et eIF3 et la sous-unité 60S est liée à eIF6.

B. Le complexe ternaire et le pré-complexe d'initiation (PIC)

Le complexe ternaire est formé de l'ARNt Met-ARNt$^{Met}_{i}$ et d'eIF2-GTP. Il s'associe à la sous-unité 40S grâce à la GTPase eIF5 pour former le pré-complexe d'initiation 43S (PIC). L'ARN hélicase eIF1A est au dessus du site A, l'ARNt initiateur au site P et eIF1 interagit près du site E.

C. L'initiation dépendante de la coiffe

Chez les eucaryotes, l'initiation de la traduction est majoritairement dépendante de la coiffe. Dans le cytoplasme, le complexe eIF4F (formé par les facteurs d'initiation eIF4A, eIF4E et eIF4G) et la protéine eIF4B se fixent sur la coiffe de l'ARNm via eI4E. Le facteur eIF4B stimule l'hélicase eIF4A pour qu'elle ouvre les structures secondaires présentes au début de l'ARNm. Ce dernier est circularisé par une interaction entre eIF4G et les protéines PABP liées à la queue poly A. Cette interaction stimule le recrutement du PIC.

Le PIC balaye ensuite la région 5'UTR jusqu'à trouver le codon d'initiation. L'activité hélicase d'eIF4A permet de fondre les faibles structures secondaires. Le codon initiateur est en général le premier AUG rencontré dans un contexte favorable appelé séquence Kozak (Kozak 1986). La séquence consensus chez les mammifère est GCC(A/G)CC**AUG**G (le codon d'initiation est indiqué en gras). L'ARNt initiateur mais aussi les facteurs eIF1, eIF2 et eIF5 participent à la reconnaissance du

contexte. Lorsque l'appariement codon:anticodon est validé, eIF5 hydrolyse le GTP d'eIF2 entraînant la dissociation des facteurs eIF1, eIF3 et eIF5. Le devenir d'eIF4B et eIF4F n'est pas encore éclairci.

La fixation de la sous-unité 60S (sans eIF6) et d'eIF5B-GTP sur la sous-unité 40S déclenche la dissociation d'eIF2-GDP. L'hydrolyse d'eIF5B permet de dissocier eIF1A du site A. A cette étape, le ribosome est complètement assemblé et prêt à démarrer l'élongation.

D. Initiation interne

Découverte chez les virus picoranavidae, l'initiation interne implique un recrutement direct du ribosome par l'ARNm sur le codon d'initiation. Ces ARNm possèdent des structures secondaires IRES (Internal Ribosome Entry Site) qui miment l'ARNt initiateur et l'action de certains facteurs d'initiation. La dépendance aux différents eIF varie pour chaque IRES (Hellen 2009). Le virus CrPV (Cricket Paralysis Virus) possède une IRES remarquable car elle est indépendante de tous facteurs d'initiation (Jan and Sarnow 2002). Sa structure et son mode d'action sont détaillés au chapitre III.

Il existe aussi des IRES cellulaires qui permettent de traduire certains gènes lorsque la traduction est inhibée, par exemple lors d'une phase du cycle cellulaire, d'une carence en acide aminé ou encore lors d'une infection virale. En 2010, W. Gilbert a livré une réflexion intéressante sur le rôle de ces IRES (Gilbert 2010). La vision traditionnelle oppose l'initiation par la coiffe et l'initiation par une IRES. Pour W. Gilbert, les IRES cellulaires pourraient compenser la carence en certains facteurs d'initiation et stimuler ainsi une initiation dépendante de la coiffe.

2. L'élongation

La protéine est synthétisée lors de l'élongation par ajout successif d'acides aminés au peptide en cours de synthèse. Chaque ajout peut être décrit comme un cycle de 4 étapes :

(1) Reconnaissance du codon par un ARNt aminoacylé.
(2) Accommodation de l'ARNt aminoacylé au site A.
(3) Transfert de la chaine peptidique présente au site P sur l'acide aminé au site A.
(4) Translocation :
    de l'ARNm
    de l'ARNt déacylé du site P au site E
    de l'ARNt peptidyl du site A au site P

Les mécanismes moléculaires sont très conservés parmi les archées, les bactéries et les eucaryotes. Les connaissances sur l'étape d'élongation ont été majoritairement apportées par l'étude du ribosome bactérien. Je présenterai donc l'élongation chez les bactéries et dans un second temps je décrirai les caractéristiques propres à l'élongation chez les eucaryotes.
Pour plus de clarté, voici un tableau présentant la correspondance entre les composants bactériens et eucaryotes.

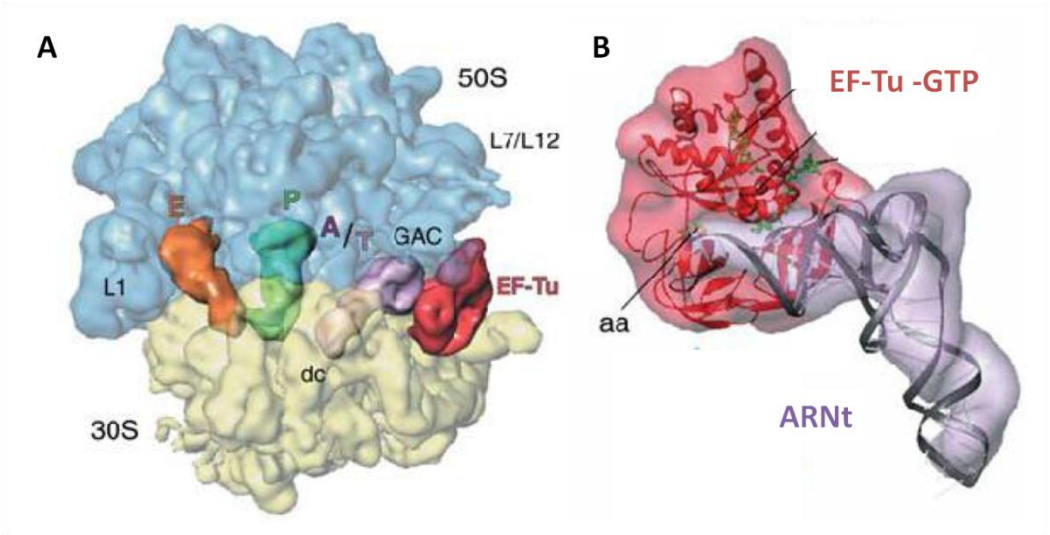


Figure 11 : Recrutement du complexe ternaire au site A du ribosome

A) Carte d'un ribosome à 9Å montrant le recrutement du complexe ternaire au site A par les protéines L7/L12. EF-Tu est en contact avec le centre d'activation des GTPase (GAC). L'anticodon de l'ARNt aminoacylé entre dans le site de décodage (dc) du site A. L'ARNt T est dans l'état A/T. **B)** Détail du complexe ternaire au sein du ribosome. EF-Tu-GTP est rouge et l'ARNt aminoacylé en violet. Figure adaptée de (Valle et al. 2003)

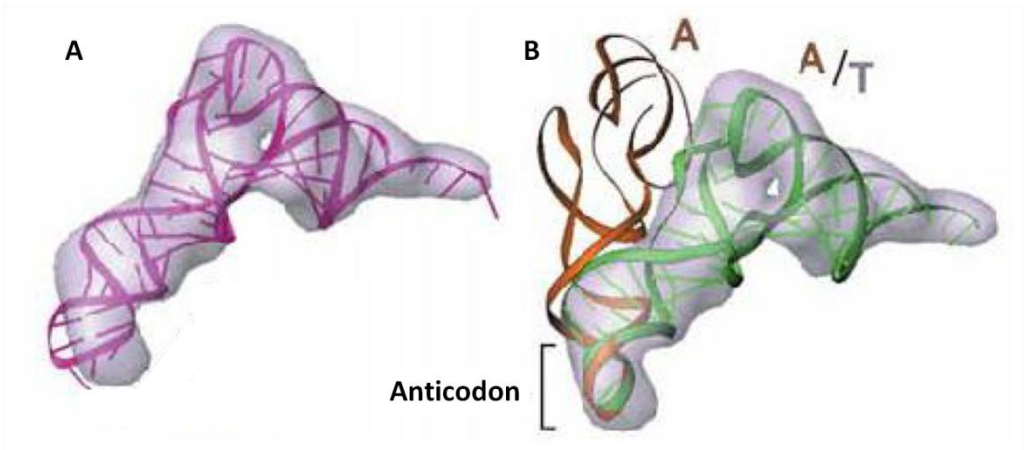


Figure 12 : Tension dans le bras de l'anticodon de l'ARNt lors de l'état A/T

A) La densité (parme) correspond à la carte d'un ARNt au site A lors de l'état A/T. La structure en ruban (magenta) représente la structure de l'ARNt hors du ribosome. La superposition des 2 structures montre, que dans l'état A/T, l'ARNt est déformé. Le bras de l'anticodon est tordu pour permettre l'appariement codon:anticodon au site A. **B)** La figure montre l'évolution de la position de l'ARNt aminoacylé dans le ribosome entre l'état A/T et l'état A/A après accommodation. Figure adaptée de (Valle et al. 2003)

Bactérien	Eucaryote
Ribosome 70S	**Ribosome 80S**
Sous-unité 50S	**Sous-unité 40S**
ARNr 16S (G530, A1492, A1493)	ARNr 18S (G577 A1755 A1756 chez *S. cerevisiae*)
Sous-unité 30S	**Sous-unité 60S**
ARNr 23S et 5S	ARNr 25S ; 5,8S et 5S
EF-Tu **(Elongation Factor-Thermo unstable)**	**eEF1-A**
EF-G	**eEF2**

Comme pour l'initiation, chaque étape est dynamique et induit des changements de conformation de chaque partenaire.

A. Préambule à l'élongation

En sortie d'initiation, le ribosome est positionné sur l'ARNm avec l'ARNt initiateur au site P et un site A vide qui contient le codon à décoder. Pour chaque codon, le pool d'ARNt se répartit en 3 classes : les ARNt cognats qui s'apparient complètement au codon, les ARNt proches-cognats qui s'apparient au codon avec un mésappariement d'une base et les ARNt non-cognats qui diffèrent totalement du codon. Alors que les ARNt cognats constituent 1 à 10% du pool total d'ARNt (Rodnina et al. 2005), le ribosome doit les sélectionner pour maintenir un niveau de fidélité optimale compatible avec le fonctionnement de la cellule.

B. Reconnaissance initiale du codon

Le complexe ternaire est composé du facteur d'élongation Tu (EF-Tu) qui est une GTPase, d'un GTP et d'un ARNt aminoacylé (figure 11 B). Le complexe est recruté via EF-Tu par le complexe protéique L7/L12 situé à la surface de la sous-unité 50S près du site A (figure 11 A). Chez *Escherichia coli*, les complexes L7/L12 sont au nombre de 4 (Rodnina and Wintermeyer 2011). Ils agissent comme des bras en apportant les complexes ternaires dans le ribosome (Wieden, Wintermeyer and Rodnina 2001, Valle et al. 2003a). Les complexes ternaires avec un ARNt non-cognat sont rejetés lors d'interactions labiles advenant à cette étape (Blanchard et al. 2004).

Une fois le complexe ternaire fixé au ribosome, l'ARNt est dans une configuration dite A/T (A pour la site A et T pour EF-Tu) (Noller et al. 2002). La boucle anticodon de l'ARNt est dans le site de décodage du site A et la tige acceptrice est en dehors du site A. La déformation que subit l'ARNt produit une conformation hautement énergétique (Valle et al. 2003a)(figure 12).

Lors de la fixation d'un ARNt au site A, le ribosome subit des changements globaux de configuration. La sous-unité 30S passe d'une configuration « ouverte » à une configuration « fermée ». En présence d'un ARNt cognat ou proche-cognat au site A, la tête et l'épaule de la sous-unité 30S se referment autour du centre de décodage (Ogle et al. 2001, Demeshkina et al. 2012) et rétrécissent le tunnel d'entrée de l'ARNm (Frank and Agrawal 2000). La fermeture de la sous-unité 30S semble être un pré-requis pour

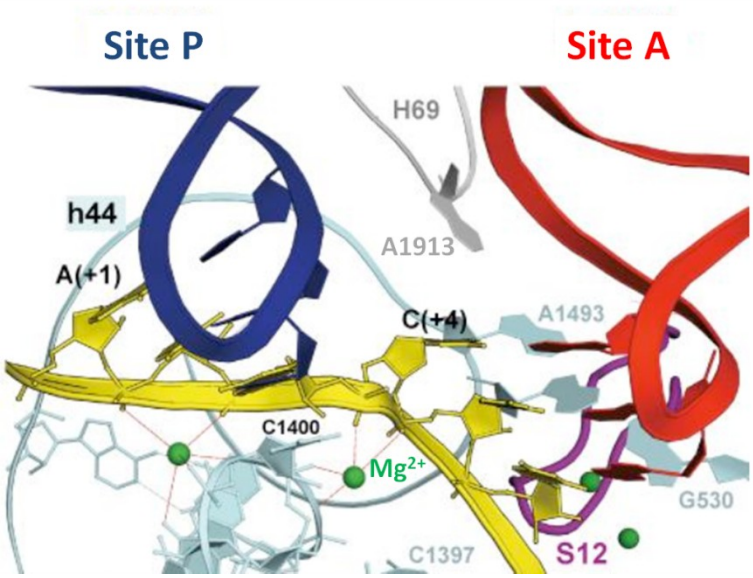


Figure 13 : Le centre de décodage

Les codons aux sites P et A sont physiquement délimités par une courbure de l'ARNm (en jaune). Le coude dans l'ARNm est provoqué par un cation Mg^{2+}(boule verte). La boucle anticodon de l'ARNt peptidyl au site P est représentée en bleu et celle de l'ARNt aminoacylé au site A en rouge.

Le centre de décodage est situé dans le site A de la petite sous-unité ribosomique (ici 30S). Il comprend les bases A1492 et A1493 dans la boucle h44, G530 dans l'hélice h18 chez les bactéries. La base A1913 dans la boucle H69 de la grande sous-unité intervient aussi dans l'activité du centre de décodage.

Figure adaptée de (Demeshkina et al. 2012)

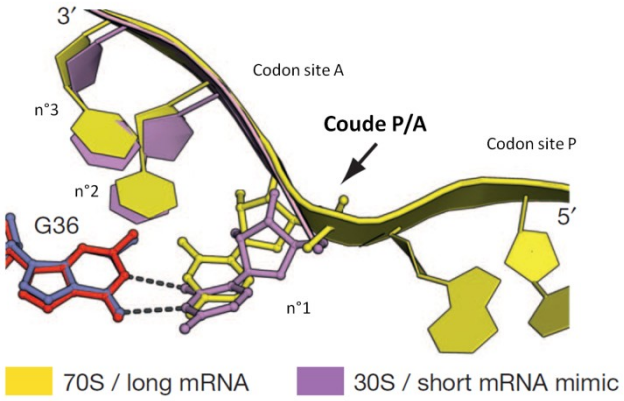


Figure 14 : Le coude dans l'ARNm entre le site A et P contraint l'orientation de la 1ère base du codon au site A

Selon la taille de l'ARNm utilisé lors des études cristallographiques (court en violet et long en jaune), l'orientation de la première base du codon au site A varie. Dans l'ARNm long, le coude entre les sites A et P est présent et contraint l'orientation de la première base du codon qui forme alors une interaction de type Watson-Crick avec la base 36 de l'anticodon (bleu et rouge). Figure adaptée de (Demeshkina et al. 2012)

former le centre de décodage et y induire des changements de configuration locaux (Demeshkina et al. 2012).

Au centre de décodage, la boucle anticodon de l'ARNt (cognat ou proche-cognat) s'apparie au codon du site A en formant une hélice codon:anticodon. Cet appariement entraîne un changement de configuration locale des bases A1492, A1493 (hélice 44) et G530 (hélice 18) de l'ARNr 16S (figure 13). Les bases A1493 et A1492 s'orientent vers l'extérieur de leur boucle et contactent le petit sillon de l'hélice codon:anticodon respectivement au niveau la 1ère et 2ème base du codon (Ogle et al. 2001). Quant à la base G530, elle change de configuration pour interagir avec la 3ème base wobble du codon et la 2ème base de l'anticodon (Ogle et al. 2001). En plus de ces 3 bases universellement conservées, la base C518 de l'ARNr 16S et une partie de la protéine ribosomique S12 vérifient l'appariement de la base wobble (Zaher and Green 2009). La stabilité du complexe ARNm:ARNt est favorisée par la purine en position 37 de la boucle anticodon de l'ARNt (Rodnina et al. 2005, Jenner et al. 2010). A cette étape, les ARNt cognats et proches-cognats ont la même cinétique de reconnaissance du codon. Néanmoins, l'interaction avec un ARNt proche-cognat déstabilise fortement l'hélice codon:anticodon ce qui conduit à un rejet d'un certain nombre d'entre eux (Blanchard et al. 2004, Rodnina et al. 2005).

Jusqu'à cette année, les structures partielles du ribosome proposaient un modèle où les bases G530, A1492 et A1493 vérifiaient la spécificité de l'interaction codon:anticodon. Ces bases examineraient la géométrie des interactions présentes au niveau des bases 1 et 2 du codon. La géométrie à la position wobble ne serait pas prise en compte ce qui expliquerait la dégénérescence du code génétique (Yoshizawa, Fourmy and Puglisi 1999, Cochella and Green 2004, Schmeing and Ramakrishnan 2009). Cette année, des structures du ribosome 70S avec un ARNt au site P et un ARNt cognat ou proche-cognat au site A ont révolutionné la compréhension du rôle des bases du centre de décodage dans la sélection de l'ARNt (Demeshkina et al. 2012).

L'ARNm forme un coude entre le site A et le site P. Ce coude est stabilisé par la présence d'un ARNt au site P, par l'ARNr 16S (bases 1401 et 1402) et la présence d'un cation Mg^{2+} (Selmer et al. 2006, Jenner et al. 2010). Cette courbure contraint l'orientation de la 1ère base du codon au site A et restreint le mouvement de la 2nde base du codon (figure 14). La conséquence de cette contrainte est qu'au site A, l'anticodon d'un ARNt proche-cognat ayant un mésappariement de bases avec la 1ère ou la 2nde base du codon interagit avec les bases du codon de la même manière qu'un ARNt cognat. Les bases mésappariées adoptent, de façon étonnante, une géométrie de type Watson-Crick et non pas de type wobble (figure 15). Selon l'emplacement du mésappariement dans l'hélice codon:anticodon, les plateaux de bases sont plus ou moins courbés ce qui déstabiliserait la boucle anticodon d'un ARNt proche-cognat. La géométrie des plateaux de bases pourrait influencer la conformation de l'hélice 69 dont la base A1913 saille dans le centre de décodage (Demeshkina et al. 2012) (figure 13). D'autre part, les bases A1492, A1493 et G530 se réorientent de la même manière pour contacter l'hélice codon:anticodon avec un ARNt cognat ou proche-cognat (figure 15). Le mouvement de ces bases n'est donc pas une étape de validation de l'interaction codon:anticodon mais une propriété inhérente au ribosome. Le modèle proposé est qu'en contraignant les bases 1 et 2 du codon à adopter des interactions Watson-Crick avec l'anticodon, les répulsions électrostatiques fourniraient l'énergie pour rejeter les ARNt proches-cognats (Demeshkina et al. 2012). La déformation de la boucle anticodon se propagerait dans tout l'ARNt et le déstabiliserait, facilitant ainsi son rejet hors du

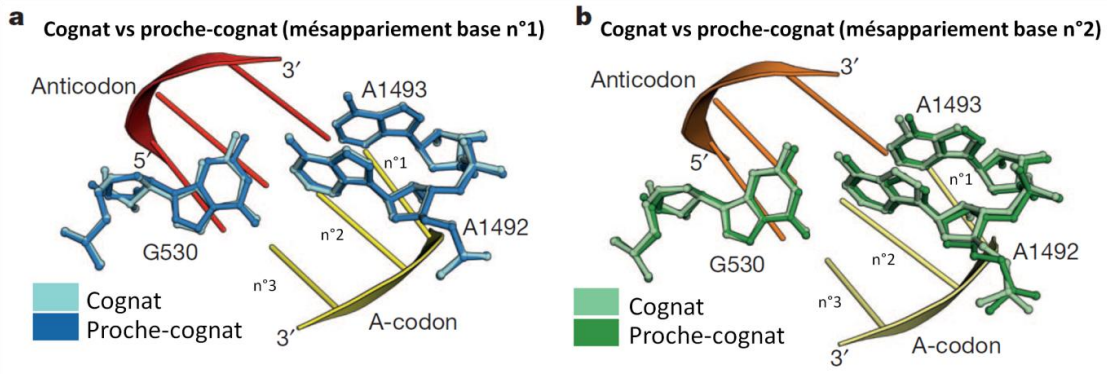


Figure 15 : L'orientation des bases universellement conservées A1492, A1493 et G530 ne discrimine pas les ARNt cognats des ARNt proches-cognats.

Comparaison de l'orientation des bases A1492, A1493 et G530 dans différentes structures cristallographiques. Le codon au site A est en jaune et l'anticodon en rouge. **A)** Comparaison avec (bleu foncé) ou sans (bleu clair) mésappariement au niveau de la première base du codon au site A. **B)** Comparaison avec (vert foncé) ou sans (vert clair) mésappariement au niveau de la seconde base du codon au site A. Dans tous les cas, l'orientation des bases A1492, A1493 et G530 est identique. Figure adaptée de (Demeshkina et al. 2012)

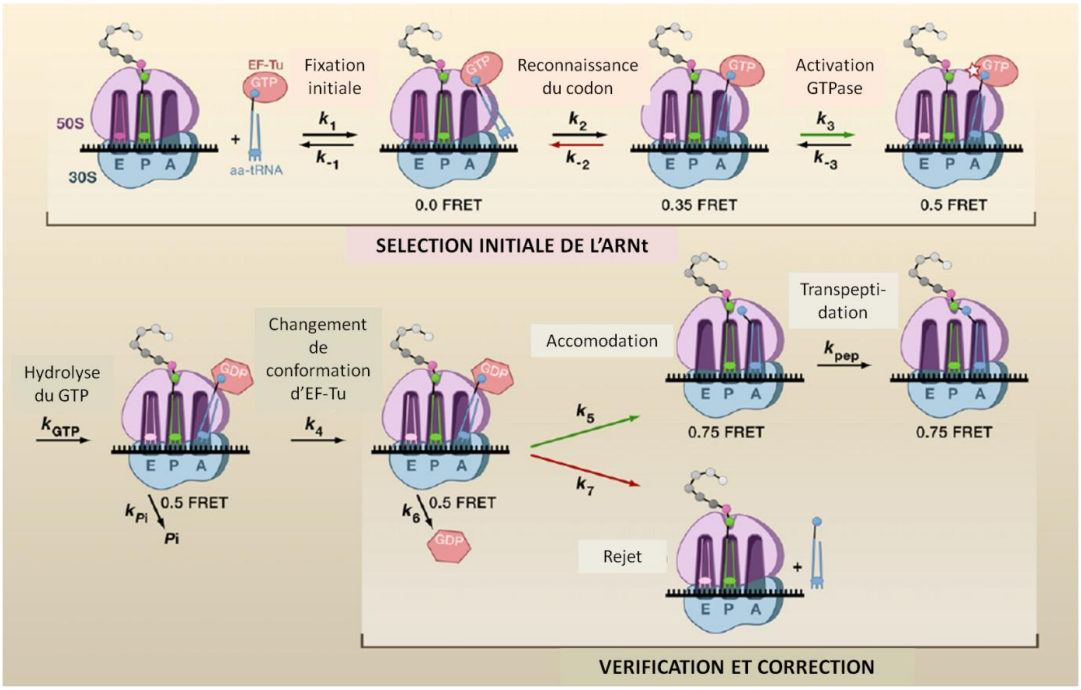


Figure 16 : Schéma des étapes de sélection de l'ARNt aminoacylé au site A

La sélection de l'ARNt aminoacylé au site A comprend 2 grandes étapes : une sélection initiale et une étape de vérification et correction. Ces 2 étapes sont séparées par l'hydrolyse du GTP par le facteur d'élongation EF-Tu. Lors de la sélection initiale, les ARNt non-cognats sont rejetés (constante cinétique k_{-2}) et les ARNt cognats et proches-cognats forment l'hélice codon:anticodon. L'activation d'EF-Tu est accélérée pour un ARNt cognat (k_3) L'ARNt est alors dans une configuration A/T. Suite à l'hydrolyse du GTP et au départ d'EF-Tu, l'ARNt cognat est accommodé au site A (k_5). Les ARNt proches-cognats sont rejetés lors de l'accommodation (k_7). Une fois accommodé au site A dans une configuration A/A, le transfert de la chaine peptidique sur l'ARNt aminoacylé est quasi instantané.

Figure adaptée de (Zaher and Green 2009)

ribosome. Ce modèle appuie l'idée que la structure l'ARNt a un rôle actif dans le processus de sélection (Cochella and Green 2005, Schmeing et al. 2011).

Une fois fixée, l'orientation relative de l'ARNt par rapport à EF-Tu change un peu, contribuant ainsi à l'activation d'EF-Tu (Schuette et al. 2009). Le signal est aussi transmis au centre d'association des GTPase (GAC) et à la région Sarcine-Ricine (SRL) de la sous-unité 50S. Le complexe L7/L12 et les protéines L10, L11 du CAG ainsi que la boucle Sarcine-Ricine du domaine IV de l'ARNr 23S, sont impliqués dans le changement de configuration du site GTP-ase d'EF-Tu (régions switch 1 et switch 2) et dans l'hydrolyse du GTP (Valle et al. 2003a, Blanchard et al. 2004, Villa et al. 2009) L'ensemble des changements de conformation et la stabilité des interactions ARNm:ARNt:ribosome favorisent une hydrolyse rapide du GTP dans le cas d'un ARNt cognat.

C. Accommodation de l'ARNt au site A : une étape de correction.

La dissociation du phosphate inorganique (Pi) induit un changement de configuration d'EF-Tu qui libère l'ARNt et se détache du ribosome (Rodnina et al. 2005, Schmeing and Ramakrishnan 2009). L'ARNt est alors accommodé au site A, son extrémité 3' parcourt 70Å dans le ribosome afin de se placer dans le centre peptidyl transférase (PTC). L'accommodation est une étape de correction qui discrimine les ARNt cognats des ARNt proches-cognats grâce à leur stabilité dans le ribosome et leurs vitesses d'accommodation différentes. Les ARNt cognats, plus stables dans la conformation fermée de la sous-unité 30S, sont accommodés rapidement et efficacement alors que les ARNt proches-cognats sont rejetés (Rodnina et al. 2005). A la fin de cette étape, les ARNt sont dans un état « classique » (A/A ; P/P) avec l'ARNt peptidyl au site P et l'ARNt aminoacylé au site A.

- Considération énergétique I

La sélection des ARNt utilise un mécanisme de fidélité basé sur une cinétique de « vérification et correction » (« proofreading » en anglais). Cette cinétique repose sur l'existence de 2 étapes séparées par une réaction chimique irréversible. A chaque étape, l'exploitation des différences d'énergie libre (ΔΔG) entre une enzyme et un substrat permettent de vérifier et sélectionner le substrat le plus affin (Zaher and Green 2009) (figure 16).

Dans le cas de la sélection d'un ARNt, les 2 étapes sont la reconnaissance du codon et l'accommodation au site A. La réaction chimique irréversible est apportée par l'hydrolyse d'un GTP d'EF-Tu. En plus de ce mécanisme, la sélection de l'ARNt cognat utilise une série de mécanismes « clé-serrure » (« induced fit » en anglais). A chaque étape, les l'ARNt dont la conformation s'adaptent le mieux aux interactions avec l'ARNm et le ribosome sont favorisés. Leur probabilité d'incorporation est augmentée par des constantes d'incorporation rapides (Rodnina et al. 2005, Zaher and Green 2009). Les constantes d'activation d'EF-Tu et de l'accommodation d'un ARNt cognat sont les 2 principaux facteurs responsables de la sélectivité du ribosome (Zaher and Green 2009).

D. Formation de la liaison peptidique ou transpeptidation.

Une fois l'accommodation réalisée, le transfert de la chaine peptidique du site P sur l'acide aminé au site A est immédiate (> $300s^{-1}$). Le taux d'incorporation varie néanmoins avec la réactivité et les contraintes stériques de l'acide aminé présent au site A (Beringer 2008).

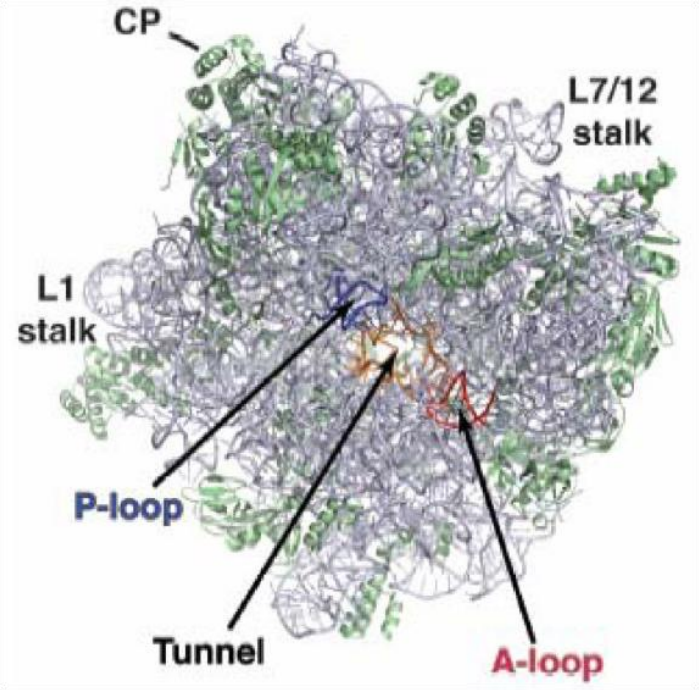


Figure 17 : Localisation du centre peptidyl transférase (PTC) dans la sous-unité 60S

Représentation cristallographique de la grande sous-unité ribosomique. Le tunnel de sortie du peptide naissant est indiqué en jaune. Les boucles d'ARNr maintenant les extrémités des ARNt CCA des ARNt aux sites A et P sont respectivement indiquées en rouge et bleu. Ces boucles se situent de part et d'autre de la cavité du centre peptidyl transférase (PTC). Figure issue de (Cruz-Vera et al. 2011)

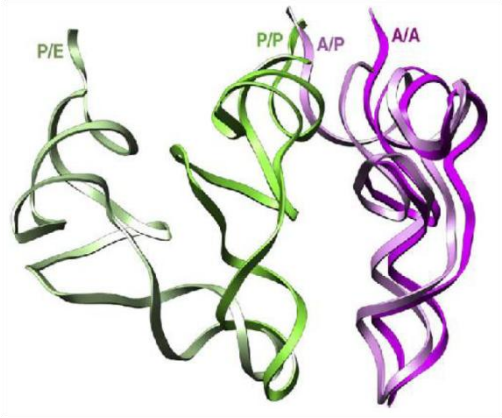


Figure 18 : Fluctuation des ARNt entre l'état "classique" (A/A ; P/P) et l'état "hybride" (A/P ; P/E)

Suite au transfert de la chaine peptidique, les ARNt aux sites A et P fluctuent entre un état classique (A/A ; P/P) et un état hybride (A/P ; P/E). La distance parcourue par l'ARNt au site P entre les 2 états est plus grande que pour l'ARNt au site A, elle implique la rupture d'un plus grand nombre de contacte avec le ribosome. Figure issue de (Agirrezabala et al. 2008)

Le site catalytique du centre peptidyl transférase (PTC) est situé dans le domaine V de l'ARNr 23S et implique 5 bases conservées : A2451, U2506, U2585, C2452 et A2602 (Beringer 2008). Deux boucles A et P, situées de part et d'autre de la cavité PTC, fixent les extrémités CCA des ARNt situés respectivement aux sites A et P (Simonović and Steitz 2009) (figure 17). La cavité du PTC est aussi en contact avec le tunnel de sortie du ribosome. Ce tunnel contient la chaine naissante de la protéine dont les acides aminés peuvent réguler l'activité du ribosome. Certaines séquences sont capables d'arrêter la traduction en induisant une pause du ribosome (Cruz-Vera et al. 2011).

E. Translocation de l'ARNm et des ARNt

La translocation est une étape délicate qui implique un déplacement coordonné et unidirectionnel de l'ARNm et des ARNt dans le ribosome. Pour éviter un changement du cadre de lecture, cette translation doit maintenir l'appariement codon:anticodon et faire progresser le ribosome d'un seul codon sur l'ARNm. La translocation repose sur un mécanisme à cliquet qui permet de passer d'un état de pré-translocation à un état de post-translocation (Spirin 1969, Frank and Agrawal 2000). La réaction est catalysée par le facteur d'élongation EF-G. Comme EF-Tu, EF-G est une GTPase.

❖ *Etat hybride (A/P ; P/E) et encliquetage des sous-unités 30S et 50S*

Suite au transfert de la chaine peptidique sur l'ARNt au site A, le statut chimique des bras accepteurs des ARNt a changé. L'ARNt au site A est à présent un ARNt peptidyl et l'ARNt au site P est déacylé. Ces nouvelles propriétés changent les affinités entre l'extrémité des bras accepteurs des ARNt et leurs environnements. Alors que les boucles anticodons des ARNt sont maintenues aux sites A et P de la sous-unité 30S, les bras accepteurs se logent respectivement dans les sites P et E de la sous-unité 50S (Moazed and Noller 1989, Dorner et al. 2006) (figure 18). C'est l'état « hybride » (A/P ; P/E) (Bretscher 1968, Squires and Berry 2008). A cause de la structure interne du ribosome, la répercussion spatiale sur les 2 ARNt est différente. Alors que le coude de l'ARNt au site P bouge de ~ 40Å entre le site P et E, l'ARNt au site A pivote légèrement sur lui même (Agirrezabala et al. 2008). Le mouvement de l'ARNt déacylé implique donc le remodelage d'un plus grand nombre de contacts avec le ribosome que pour le mouvement de l'ARNt peptidyl. En plus de ces 2 états, il existerait un état intermédiaire (A/A ; P/E) plus rare et plus instable (Munro et al. 2007, Pan, Kirillov and Cooperman 2007). Cette observation suggère que les mouvements des 2 ARNt seraient indépendants l'un de l'autre.

Les ARNt fluctuent dans la sous-unité 50S entre les états « classique » et « hybride ». Cette fluctuation est spontanée et indépendante du facteur d'élongation EF-G (Moazed and Noller 1989, Dorner et al. 2006, Frank et al. 2007, Cornish et al. 2008, Agirrezabala et al. 2008, Julián et al. 2008).

L'état « hybride » (A/P ; P/E) engendre une réorganisation globale des deux sous-unités. Les sous-unités 30S et 50S pivotent l'une par rapport à l'autre avec un angle de 6 degrés. Le ribosome est alors en position « encliquetée » (« ratcheted » en anglais) (Frank and Agrawal 2000, Julián et al. 2008, Agirrezabala et al. 2008) (figure 19 et 20). Comme pour les ARNt, en absence d'EF-G il existe une fluctuation spontanée entre les états « décliqueté » et « encliqueté » des sous-unités ribosomiques (Cornish et al. 2008).

L'état « hybride » (A/P ; P/E) provoque aussi des réorganisations locales dans chaque sous-unités. Au niveau de la sous-unité 30S, l'entrée du tunnel de l'ARNm située entre la tête et l'épaule s'élargit

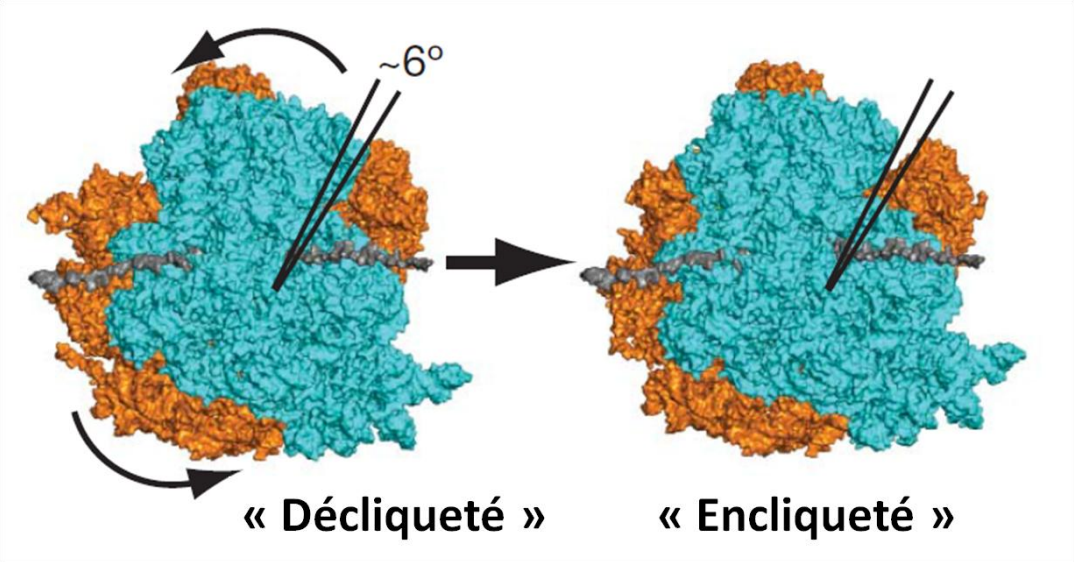


Figure 19 : Encliquetage des sous-unités ribosomiques.

L'encliquetage des sous-unités ribosomiques implique une rotation antihoraire de la sous-unité 30S par rapport à la sous-unité 50S. La sous-unité 30S est représentée en bleu et la sous-unité 50S en jaune. Figure adaptée de (Schmeing and Ramakrishnan 2009)

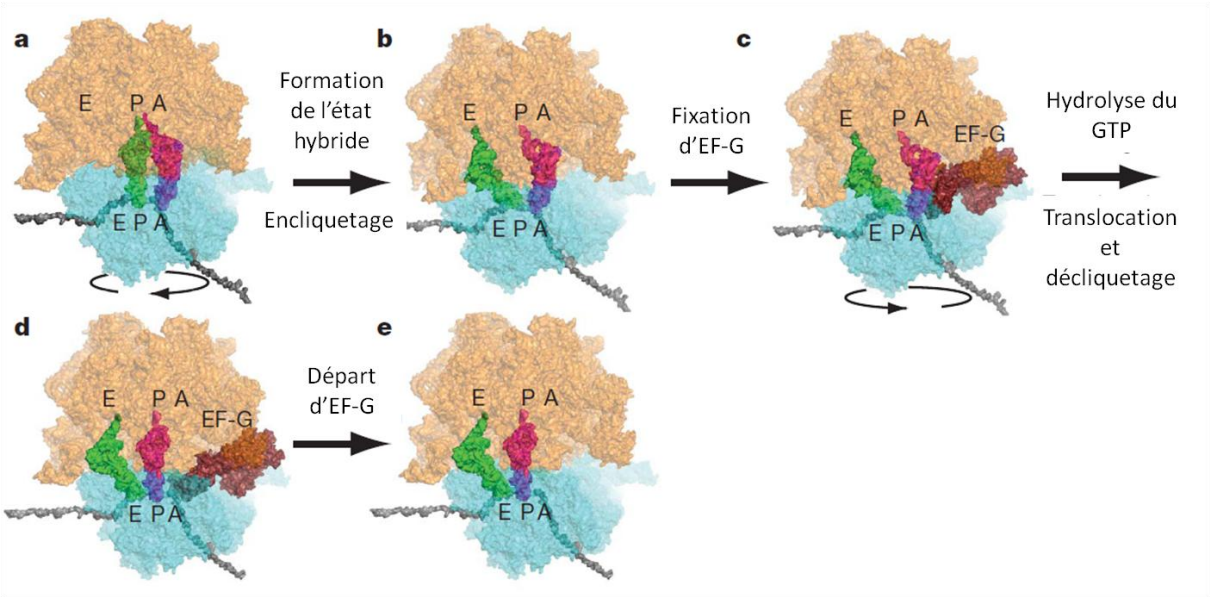


Figure 20 : Cycle de translocation.

A à B) Suite au transfert de la chaine peptidique, les ARNt aux sites A et P fluctuent entre un état classique (A/A ; P/P) et un état hybride (A/P et P/E). L'état hybride est associé à une conformation « encliquetée » des sous-unités du ribosome. **B à C)** La fixation d'EF-G stabilise l'état hybride. **C à D)** L'hydrolyse du GTP d'EF-G induit un changement de conformation d'EF-G et initie la translocation de l'ARNm et des ARNt. Les sous-unités ribosomiques retournent à l'état « décliqueté ». **E)** A la fin de la translocation, l'ARNt sont dans les sites P et E. Figure adaptée de (Schmeing and Ramakrishnan 2009)

(Frank and Agrawal 2000). Dans la sous-unité 50S, à la sortie du site E, la protubérance L1 pivote de 20 degrés à l'intérieur du site E et contacte la boucle T de l'ARNt P/E (Valle et al. 2003b). La protubérance L1 est alors dans une conformation « fermée » et obstrue la sortie du site E (Cornish et al. 2009, Fei et al. 2009) (figure 21).

Là encore, une fluctuation de la protubérance L1 entre les positions « ouverte » et « fermée » est observable en absence d'EF-G. Ces mouvements sont corrélés à ceux de la rotation des 2 sous-unités (Cornish et al. 2009, Fei et al. 2009). La synchronisation entre la conformation « encliquetée » des sous-unités et la fermeture de la protubérance L1 serait une propriété mécanique du ribosome. Elle ne dépendrait pas de la présence d'un ARNt au site P (Spahn et al. 2004, Fei et al. 2009).

- Considération énergétique II

En absence d'EF-G, un cycle de translocation peut se produire spontanément quoique lentement. L'énergie apportée par l'hydrolyse d'un GTP n'est donc pas nécessaire (Gavrilova et al. 1976).

Le déplacement des ARNt dans la sous-unité 50S utilise une partie de l'énergie libérée lors de la réaction de transpeptidation. L'affinité des sites A, P et E de la sous-unité 50S pour différents substrats contraint le mouvement unidirectionnel des ARNt dans le ribosome (Lill, Robertson and Wintermeyer 1989, Noller et al. 2002, Cornish et al. 2008).

Le mouvement de l'ARNm et des ARNt dans la sous-unité 30S nécessite le franchissement d'une barrière cinétique entre l'état « encliqueté » et « décliqueté ». La présence du facteur d'élongation EF-G facilite le passage de cette barrière (Noller et al. 2002).

- ***Recrutement d'EF-G–GTP et hydrolyse du GTP***

Le facteur d'élongation EF-G lié au GTP est recruté au site A du ribosome par le complexe L7/L12 (Diaconu et al. 2005) et se lie au CAG et au SRL de la sous-unité 50S (Clementi et al. 2010) (figure 20). Son recrutement par le complexe de pré-translocation est indépendant de la conformation des ARNt et des sous-unités ribosomiques (Chen et al. 2011). Sa fixation stabilise l'état hybride (A/P ; P/E) et la fermeture de la protubérance L1 (Fei et al. 2009) mais elle n'est pas suffisante pour déclencher la translocation (Spiegel, Ermolenko and Noller 2007).

Des changements de conformation au niveau des sites GAC et SRL permettent l'activation du domaine catalytique d'EF-G. L'hydrolyse du GTP induit un changement de conformation d'EF-G (Connell et al. 2007, Li et al. 2011). Le ribosome s'engage alors dans le processus de décliquetage des 2 sous-unités. De plus, la tête de la sous-unité 30S tourne sur elle-même du site A vers le site E. La base moléculaire liant le changement de conformation d'EF-G à celui du ribosome n'est pas connue.

- ***Rotation de la tête de la sous-unité 30S et décliquetage des 2 sous-unités***

Une large rotation de la tête (5° → 18°) modifie peu la fixation d'EF-G mais améliore beaucoup l'interaction entre le domaine IV et l'hélice 34 de l'ARNr 16S (Ratje et al. 2010, Matassova, Rodnina and Wintermeyer 2001) (figure 22 et 23). L'hélice 34 et le nucléotide 530 de l'ARNr 16S forment le verrou du tunnel d'entrée de l'ARNm. Lors de l'interaction avec le domaine IV d'EF-G, l'hélice 34 et la base 530 s'écartent l'une de l'autre ouvrant ainsi le tunnel d'entrée ce qui facilite la translocation de l'ARNm (Ratje et al. 2010).

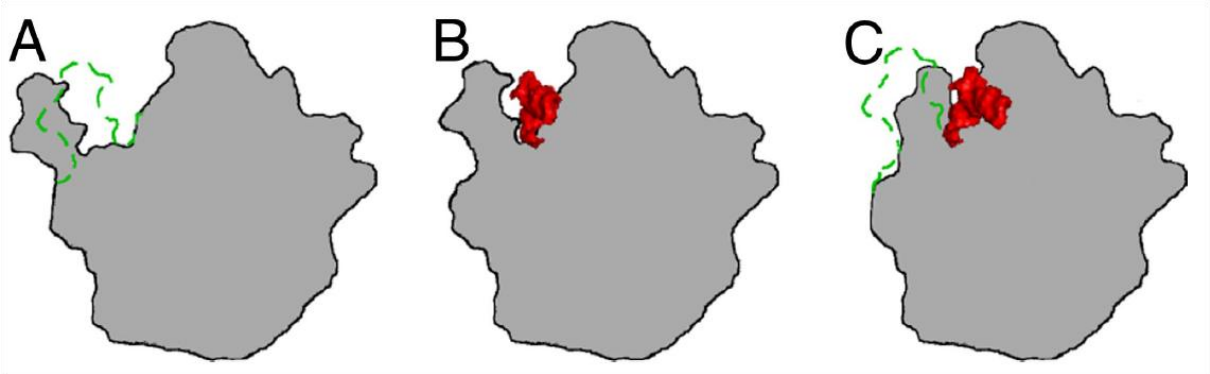


Figure 21 : Mouvement de la région L1 lors de l'élongation

Schéma de la grosse sous-unité ribosomique indiquant les différentes positions adoptées par la région L1 lors de la translocation. Le pointillé vert correspond à la position de la région L1 dans le cas **B**. **A)** En absence d'ARNt au site E, la région L1 est en position ouverte. **B)** Avec un ARNt déacylé (rouge) au site E dans un état classique E/E, la région L1 est en position à demi-fermée. **C)** Avec un ARNt déacylé dans un état hybride P/E, la région L1 est en position fermée. Figure issue de (Cornish et al. 2009)

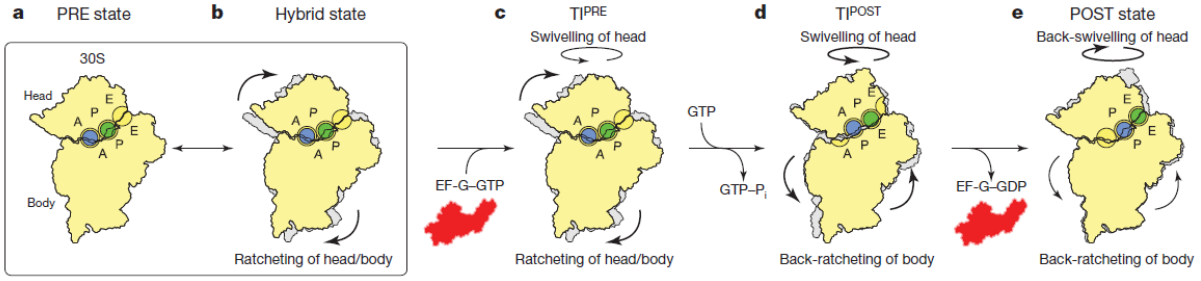


Figure 22 : Mouvements de la sous-unité 30S lors de la translocation

La translocation implique un pivotement des sous-unités l'une par rapport à l'autre (encliquetage : ratcheting ; décliquetage : back-ratcheting) mais aussi une rotation de la tête de la petite sous-unité par rapport au reste de la sous-unité (swivelling et back-swivelling). Figure issue de (Ratje et al. 2010)

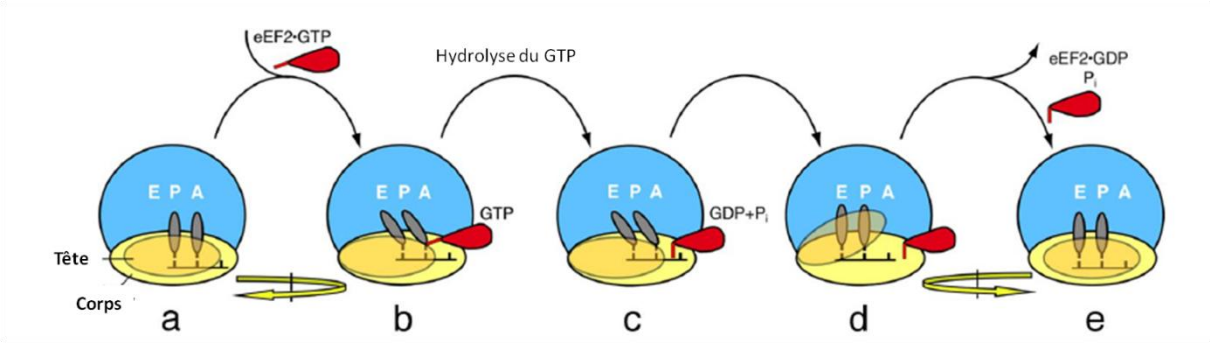


Figure 23 : Schéma des étapes de translocation impliquant un encliquetage des sous-unités ribosomiques et une rotation de la tête de la sous-unité 40S

Même légende que pour la figure 22. La grande sous-unité est représentée en bleu et la petite en jaune. Figure adaptée de (Taylor et al. 2007)

La rotation de la tête s'accompagne d'un pivotement des sous-unités 30S et 50S. Ce pivotement s'effectue dans le sens inverse de celui observé dans lors de l'état hybride (7° → 4°) (Weixel and Bradbury 2002, Ratje et al. 2010), il correspond au décliquetage des sous-unités. Le mouvement concerté de la tête et des sous-unités réduit la distance que l'ARNm et les ARNt ont à parcourir lors de la translocation (Ratje et al. 2010). Ce processus est essentiel à la translocation (Horan and Noller 2007).

❖ ***Retour de la tête, translocation de l'ARNm et départ d'EF-G-GDP***

A la fin de sa rotation, la tête de la sous-unité 30S revient en place (18° → 0°) et les sous-unités se décliquent totalement (4° → 0°) (figures 22 et 23). Pendant ce retour, l'ARNm se déplace d'un codon dans le ribosome et les ARNt se logent totalement aux sites P et E en adoptant un état « classique » (P/P ; E/E) (Ratje et al. 2010, Ermolenko and Noller 2011). La protubérance L1 contacte toujours l'ARNt au site E, elle est dans une conformation « semi-fermée » (Valle et al. 2003b, Cornish et al. 2009, Fei et al. 2009, Trabuco et al. 2010). Ce retour à l'état classique est indépendant de la présence d'EF-G (Ermolenko and Noller 2011, Valle et al. 2003b, Gao et al. 2009). Il est proposé que ce dernier ce détache après que le ribosome ait atteint l'état classique (Schmeing and Ramakrishnan 2009).

Les conformations complètement « encliqueté » et « décliqueté » du ribosome sont les 2 seuls intermédiaires stables de la translocation décrits à ce jour (Ermolenko and Noller 2011).

➢ Considération énergétique III

Le ribosome fonctionne comme un mécanisme à cliquet brownien (Spirin 2009). EF-G serait un cliquet dynamique qui permettrait de découpler le décliquetage des sous-unités et le retour à un état classique avec des ARNt aux sites P et E. Pour cela, EF-G facilite la transition entre les différentes conformations du ribosome et assure que le mouvement soit unidirectionnel (Ratje et al. 2010).

❖ ***Sortie de l'ARNt déacylé***

La sortie de l'ARNt déacylé du site E est couplée à l'ouverture de la protubérance L1. Les mécanismes moléculaires qui déclenchent l'ouverture du site E sont encore inconnus (Valle et al. 2003b, Cornish et al. 2009, Fei et al. 2009, Trabuco et al. 2010). L'ouverture de la protubérance L1 et le décliquetage des sous-unités ne sont pas corrélés. L'ouverture du site E est bien plus rapide que le décliquetage (Cornish et al. 2009) et sa vitesse d'ouverture n'est pas le facteur limitant la sortie de l'ARNt du site E (Fei et al. 2009).

Après le sortie de l'ARNt du site E, le ribosome entame un nouveau cycle d'élongation avec l'arrivé d'un nouveau complexe ternaire.

F. Maintien du cadre de lecture

Au cours de l'élongation 2 types d'erreurs peuvent se produire : la mauvaise incorporation d'un acide aminé ou un décalage du cadre de lecture. Dans le premier cas, une erreur ponctuelle d'incorporation n'est pas forcément délétère pour la protéine finale. Le taux d'erreur constaté est de 1 pour 200 à 1600 acides aminés incorporés (Edelmann and Gallant 1977, Bouadloun, Donner and Kurland 1983). A l'inverse du premier cas, le décalage du cadre de lecture conduit à la traduction d'une protéine dont la séquence est incorrecte et potentiellement délétère pour la cellule. Chez *E. coli*, le taux d'erreurs dues à un décalage du cadre de lecture est estimé à 1 pour 30000 acides aminés incorporés (Jørgensen and Kurland 1990). L'ARNm ne contient pas de signaux de ponctuation indiquant la bonne

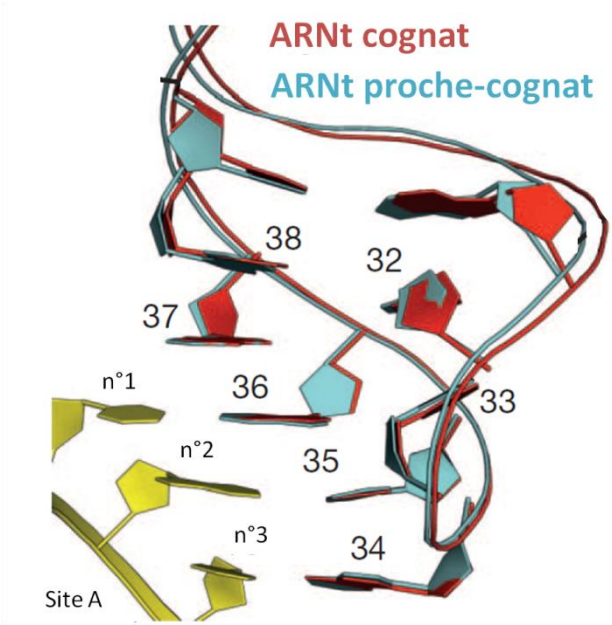


Figure 24 : Empilement de la base 37 de la boucle anticodon

Lors de la formation de l'hélice codon:anticodon au site A, la base 37 de l'anticodon s'empile sur l'appariement entre la première base du codon et la base 36 de l'anticodon. Cet empilement se produit dans le cas d'un ARNt cognat et proche-cognat. Le codon de l'ARNm est représenté en jaune et la boucle anticodon en bleu ou rouge. Figure adaptée de (Demeshkina et al. 2012)

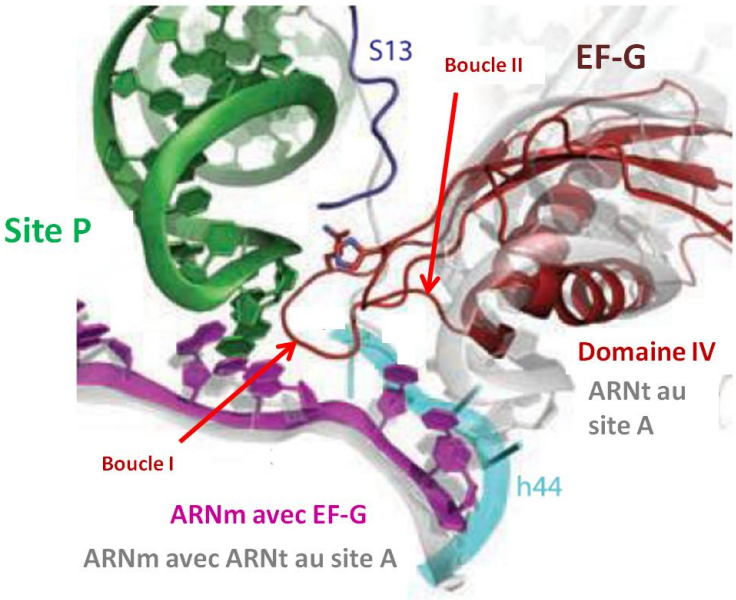


Figure 25 : Interaction EF-G avec le site de décodage

Suite à la translocation, l'ARNt peptidyl (vert) est au site P et EF-G (rouge) entre profondément au site A. e domaine IV est en contact avec l'hélice codon:anticodon au site P (hélice I du domaine IV) et avec le centre de décodage au site A (hélice II du domaine IV). La position d'un ARNt au site A est indiqué comme référence en gris. Figure adaptée de (Gao et al. 2009)

phase de lecture après l'initiation. Après un décalage, le ribosome ne peut pas savoir s'il est dans la bonne phase de lecture ni se corriger. Afin de limiter ce type d'erreurs, il est important que l'ARNm se décale seulement d'un codon dans le ribosome et que l'interaction entre l'ARNm et les ARNt soit maintenue lors de la translocation.

Concernant le déplacement de l'ARNm au sein du ribosome, les études structurales ont montré que l'ARNm est maintenu dans le ribosome par les protéines situées à l'entrée et à la sortie de son tunnel (Jenner et al. 2010). Au cours de la translocation, les diamètres des tunnels d'entrée et de sortie de l'ARNm varient. En présence d'EF-G-GTP, le tunnel d'entrée commence à s'élargir et atteint son ouverture maximale quand EF-G est sous forme GDP. Pour le tunnel de sortie, c'est l'inverse. Bien que son diamètre soit plus variable, il est plus large lorsqu'EF-G est sous sa forme GTP que GDP (Frank and Agrawal 2000). Cette opposition de fonctionnement laisse penser à un mécanisme dynamique qui maintiendrait l'ARNm en le pinçant à l'entrée ou à la sortie du ribosome selon l'état du cycle d'élongation (Frank and Agrawal 2000).

Concernant le maintien de l'appariement codon:anticodon, 2 points peuvent être mentionnés.

Le 1er point concerne la modification de la purine à la position 37 (ms^2i^6A37) de l'ARNt. Cette base modifiée augmente l'interaction codon:anticodon en s'empilant sur l'appariement entre la base 36 et la 1ère base du codon. Au site A et P, les anticodons deviennent alors virtuellement des anticodons à 4 bases stabilisant ainsi les interactions codons:anticodons (figure 24). L'interaction avec ms^2i^6A37 existe aussi au site E mais elle ne compense pas le fait que les bases #2 et #3 du codon soient trop éloignées de l'anticodon pour être correctement appariées. La stabilité codon:anticodon y est donc plus faible. L'ancrage des interactions ARNm:ARNt permettraient de diminuer le glissement des ARNt sur l'ARNm et éviterait ainsi un décalage du cadre de lecture (Jenner et al. 2010).

Le 2nd point concerne le rôle potentiel d'EF-G (ou eEF2) dans le maintien de l'interaction de l'ARNt peptidyl avec l'ARNm lors de la translocation. D'après la structure du ribosome en post-translocation avec des ARNt P/P et E/E, la pointe du domaine IV d'EF-G est en contact à la fois avec l'hélice codon:anticodon du site P et le centre de décodage au site A (figure 25). La pointe du domaine IV est composée de 2 boucles : la boucle I (acides aminés 496 à 509) et la boucle II (acides aminés 567 à 579). Au site P, les 2 glycines conservées (G502 et G503) dans la boucle I d'EF-G induisent une torsion de celle-ci qui permet son insertion dans le sillon mineur de l'hélice codon:anticodon. La glutamine Q500 contacte les bases 37 et 38 de l'ARNt. De plus, la base 37 est aussi contactée par l'histidine H573 de la boucle II. Au site de A, les résidus sérine S578 et glutamate E579 de la boucle II interagissent avec la base universellement conservée A1293. L'hypothèse avancée aujourd'hui est que ces interactions se mettent en place lorsque l'hélice codon:anticodon est encore au site A. Elles seraient ensuite maintenues pendant la translocation afin de prévenir tout glissement de l'ARNt peptidyl sur l'ARNm. Cette hypothèse implique que la boucle I d'EF-G aiderait au renversement des bases G530, A1292 et, A1293 dont les positionnements ont été établis lors de la reconnaissance du codon (Gao et al. 2009).

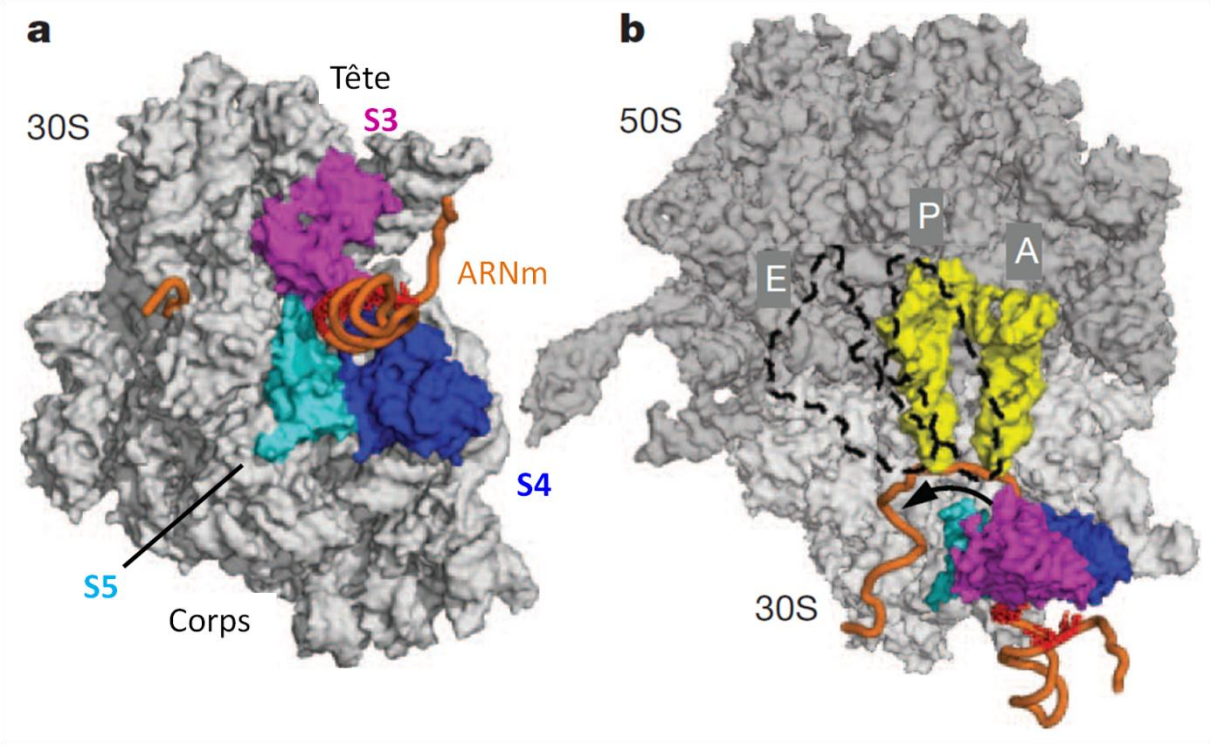


Figure 26 : Activité hélicase du ribosome : protéines S3, S4, et S5

L'activité hélicase du ribosome serait portée par les protéines S3 (violet), S4 (bleu) et S5 (cyan) situées à l'entrée du tunnel d'entrée de l'ARNm. **A)** Petite sous-unité 30S du ribosome vue de l'extérieur du côté du tunnel d'entrée de l'ARNm (orange). **B)** Coupe du ribosome 70S permettant de voir le trajet de l'ARNm et la position des protéines S3, S4 et S5. Figure adaptée de (Qu et al. 2011)

G. Passage des structures secondaires

Les ARNm sont des molécules structurées dans lesquelles se forment des hélices d'ARN plus ou moins stables. Le diamètre du tunnel d'entrée de l'ARNm dans le ribosome est trop petit pour accueillir un ARNm double brin (Yusupova et al. 2001). Les structures secondaires présentes dans l'ARNm doivent être débobinées lors de la traduction avant d'entrer dans le ribosome. Ce dernier possède une activité hélicase propre. Dans le tunnel d'entrée de l'ARNm, l'anneau composé des protéines S3, S4 et S5 seraient responsable de l'activité hélicase (Takyar, Hickerson and Noller 2005) (figure 26). Deux mécanismes sont utilisés pour ouvrir une jonction de 2 brins d'ARN : une action mécanique et une stabilisation de l'ouverture de la jonction entre les brins. Le 1er mécanisme permet d'initier l'ouverture d'une jonction. Le ribosome utilise l'énergie de l'hydrolyse du GTP, de la transpeptidation et de la translocation pour appliquer une force sur la jonction et l'ouvrir. Le 2nd mécanisme s'applique quand le ribosome est déjà confronté à une jonction ouverte. Comme le ribosome se lie et interagit préférentiellement avec une jonction ouverte, il favorise thermodynamiquement cette conformation ce qui facilite la translocation au codon suivant. Ce processus ne nécessite pas de source d'énergie externe. Les bases moléculaires de l'activité hélicase du ribosome restent encore peu connues (Qu et al. 2011).

H. L'élongation chez les eucaryotes

Les mécanismes moléculaires de l'élongation sont conservés entre les bactéries et les eucaryotes mais il existe quelques variations surtout concernant les facteurs d'élongation.

Les ribosomes eucaryotes adoptent aussi les conformations encliquetées et décliquetées mais la présence d'eEF2-GTP ne semble pas nécessaire pour stabiliser l'état « encliqueté » (Spahn et al. 2004). De plus la fixation d'eEF2 induit un rapprochement plus important du centre GAC vers la protubérance centrale (Spahn et al. 2004, Taylor et al. 2007).

❖ *Les facteurs d'élongation eEF1A et eEF2*

Les facteurs d'élongation eEF1A et eEF2 portent chacun une modification post-traductionnelle. Chacune de ces modifications est unique à ces protéines.

Le facteur eEF1A possède selon les espèces 0, 1 ou 2 phospoglycerol ethanolamine (EPG). Chez *S. cerevisiae* cette modification est absente bien que le site de modification soit présent. Le rôle de cette modification reste inconnu chez les autres espèces (Greganova, Altmann and Bütikofer 2011).

Le facteur eEF2 arbore un diphthamide à la pointe du domaine IV. Cette modification est aussi retrouvée chez les archées. Elle est la cible de plusieurs toxines bactériennes : la toxine diphtérique de *Corynebacterium diphtheriae*, l'endotoxine A de *Pseudomonas aeruginosa* et la toxine cholix de *Vibrio cholerae* (Greganova et al. 2011). Cette modification jouerait un rôle dans la fidélité de la traduction (Ohtsuka et al. 2009). Ce dernier point sera abordé au chapitre II.

❖ *Le facteur d'élongation eEF3 chez les Eumycètes*

Le modèle du ribosome avec les 3 sites allostériques A, P et E suggère que seulement 2 ARNt peuvent se trouver au même moment dans le ribosome. Le départ de l'ARNt déacylé du site E est nécessaire à l'entrée d'un ARNt aminoacylé au site A (Nierhaus 1990). Il a été proposé que la protéine eEF3 aiderait

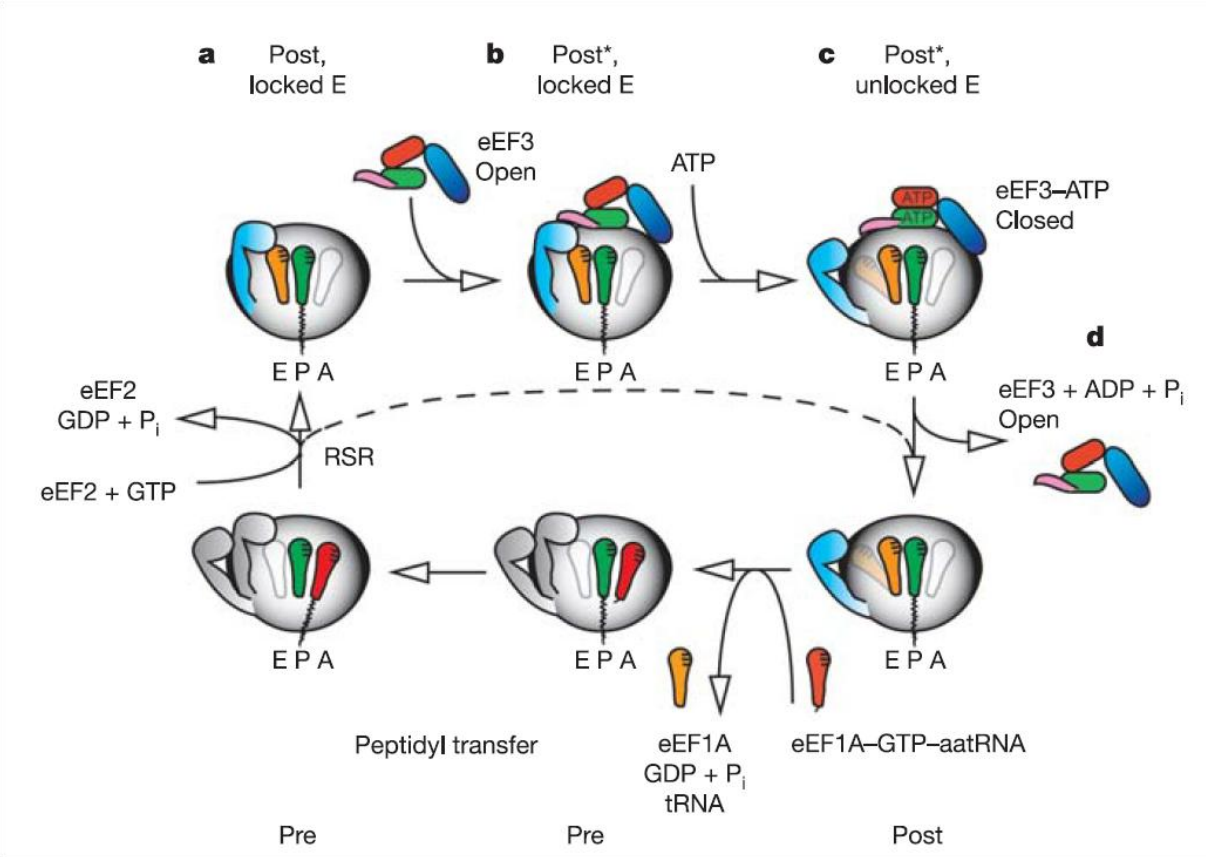


Figure 27 : Rôle d'eEF3 dans l'élongation chez *S. cerevisiae*

Le facteur d'élongation eEF3 se fixe sur la sous-unité 60S du coté du site E. Il aiderait à la sortie de l'ARNt déacylé du site E. Figure issue d' (Andersen et al. 2006)

ces 2 étapes en stimulant la fixation d'eEF1A avec un ARNt cognat au site A ainsi que l'éjection de l'ARNt déacylé du site E (Triana-Alonso, Chakraburtty and Nierhaus 1995).

Le facteur d'élongation eEF3 est propre au règne des Eumycètes. Chez *S. cerevisiae* il est codé par le gène essentiel *YEF3* (Qin, Moldave and McLaughlin 1987). La protéine eEF3 est une ATPase ribosome dépendante, membre de la famille des protéines ABC (ATP-Binding Cassette). Sous sa forme liée à l'ADP, eEF3 se fixe au domaine III d'eEF1A via son domaine C-terminal. Le rôle de cette interaction dans la fixation du complexe ternaire au ribosome est encore incompris (Kovalchuke et al. 1998, Anand et al. 2006).

D'autre part, eEF3 lié à l'ATP lie aussi directement le ribosome mais son site de fixation est différent de celui des 2 autres facteurs d'élongation eEF1a et eEF2. Il se fixe du côté du site E et forme un pont entre les 2 sous-unités reliant la protubérance centrale de la sous-unité 60S et la tête de la sous-unité 40S (Andersen et al. 2006).

A l'inverse du couple eIF1A eEF3, les facteurs eEF2 et eEF3 sont en compétition pour se fixer sur le ribosome (Kovalchuke et al. 1998). La raison tient au fait qu'eEF2 se fixe sur un ribosome en pré-translocation et induit une rotation de tête de la sous-unité 40S. Cette orientation de la tête diminue fortement l'accès d'eEF3 à son site de fixation. L'accroche d'eEF3 est rendue possible sur un ribosome en post-translocation (Andersen et al. 2006). Son rôle serait de déplacer la protubérance L1 vers l'extérieure et d'ouvrir ainsi le site E. Comme eEF3 chevauche en partie la sortie du site E, l'hydrolyse de l'ATP serait nécessaire pour son décrochage du ribosome. En facilitant la sortie de l'ARNt déacylé, eEF3 stimulerait l'incorporation d'un ARNt cognat comme suggéré dans le modèle des 3 sites allostériques (Andersen et al. 2006) (figure 27).

En plus de son rôle dans l'élongation, il a été rapporté qu'eEF3 aiderait au recyclage des sous unités ribosomiques (Kurata et al. 2010).

- ***Le facteur eIF5A***

Le facteur eIF5A est une protéine très conservée au sein des eucaryotes et des archées. Il est remarquable car il arbore une modification post-traductionnelle unique : une hypusine (Shiba et al. 1971). Cette modification est ajoutée sur la lysine conservée en position 50 (K50) chez les eucaryotes et certaines espèces d'archées (Park et al. 2010). La voie de modification est très conservée et implique 2 enzymes : la deoxyhypusine synthétase DHS et la deoxyhypusine hydroxylase DOHH. Le substrat de cette modification provient de la spermidine, une des trois polyamines naturelles de la cellule. Ces deux enzymes reconnaissent spécifiquement eIF5a grâce à la séquence entourant K50 et à la structure tertiaire de la protéine (Wolff et al. 2007, Park et al. 2010).

Lors de sa découverte, les premières études lui ont attribué un rôle dans l'initiation traductionnelle d'où son nom eIF5A (eucaryotic initiation factor 5A) (Kemper, Berry and Merrick 1976). Récemment cette vision a changé puisque le facteur eIF5A faciliterait l'élongation de la traduction (Gregio et al. 2009, Saini et al. 2009) mais son mécanisme d'action reste largement incompris. Chez *S. cerevisiae*, eIF5A serait un homo-dimère dont la formation nécessite la présence de l'hypusine et dépend de la fixation d'un ARNm (Gentz, Blatch and Dorrington 2009). Il a aussi été montré qu'eIF5A interagit physiquement avec le ribosome 80S (Jao and Chen 2006) ainsi qu'avec les facteurs d'élongation eEF1A et eEF2 (Zanelli et al. 2006). Ces observations renforcent l'hypothèse de son implication dans l'élongation de la traduction.

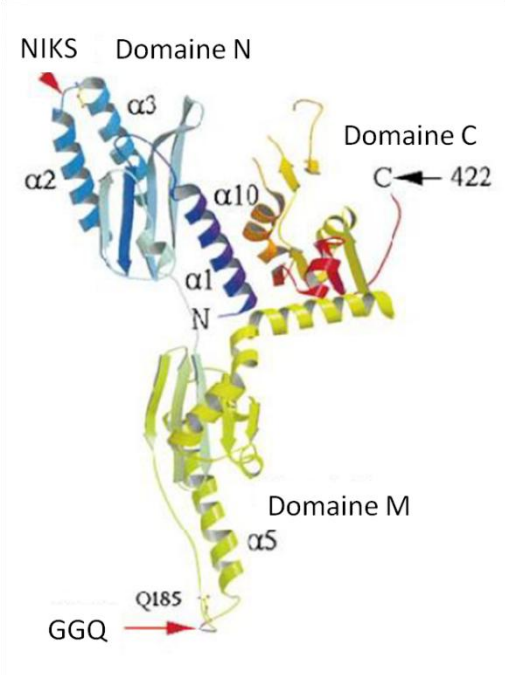


Figure 28 : Structure du facteur de terminaison eRF1

Le facteur de terminaison eRF1 est composé de 3 domaines : N, M et C. Le motif GGQ est à l'extrémité du domaine M et le motif NIKS à la pointe du domaine N. Figure adaptée de (Song et al. 2000)

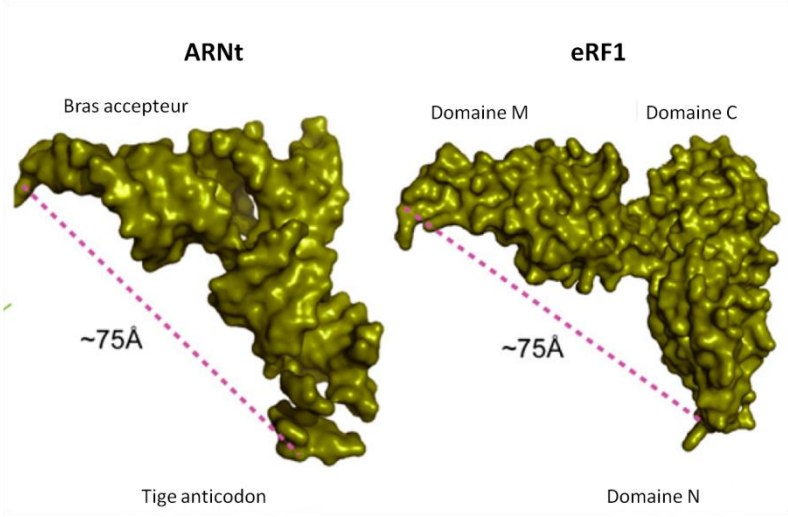


Figure 29 : Comparaison des structures d'un ARNt et d'eRF1

La structure d'eRF1 est très similaire à celle d'un ARNt. Figure adaptée de (Cheng et al. 2009)

Comme l'inhibition d'eIF5A induit un arrêt du cycle cellulaire il a été proposé qu'eIF5A soit impliqué dans la traduction d'ARNm du cycle cellulaire (Park, Lee and Joe 1997). Jusqu'à présent, aucune classe d'ARNm n'a été identifiée comme étant régulée par cette protéine.

3. La terminaison

La terminaison se produit lorsqu'un des 3 codons stop (UAA, UAG ou UGA) entre au site A.

A. Préambule à la terminaison

Le ribosome vient de finir un cycle d'élongation, il est positionné sur l'ARNm avec un site A vide et l'ARNt peptidyl au site P avec la protéine qui traverse le tunnel de sortie. Le ribosome est à ce stade un complexe de pré-terminaison (Alkalaeva et al. 2006).

Comme les 3 codons stops n'ont pas d'acide aminé assigné, ils ne sont pas reconnus par des ARNt. Pour eux le pool d'ARNt se répartit en 2 classes : les ARNt proches-cognats et les ARNt non-cognats. Ces ARNt sont en compétition avec un facteur de terminaison qui va reconnaître le codon stop. Chez les eucaryotes, il s'agit de la protéine eRF1 (eukaryotic Release Factor 1) (Kervestin et al. 2001, Seit-Nebi, Frolova and Kisselev 2002). Le ribosome doit sélectionner ce facteur de terminaison pour terminer la traduction de la protéine.

B. Le complexe de terminaison

Le complexe de terminaison est un hétéro-dimère composé des protéines eRF1 et eRF3 (eukaryotic Release Factor 3) (Cheng et al. 2009).

- ***eRF1***

La protéine eRF1 est un facteur de terminaison de classe I. Elle est responsable de la reconnaissance des 3 codons stop (Kervestin et al. 2001, Seit-Nebi et al. 2002) et de l'hydrolyse de la liaison ester reliant l'ARNt peptidyl à la protéine synthétisée (Loh and Song 2010). Cette hydrolyse est déclenchée par le motif universellement conservé Glycine-Glycine-Glutamine (GGQ) (Frolova et al. 1999). La protéine eRF1 comprend 3 domaines N, M et C qui s'organisent en Y (Frolova, Merkulova and Kisselev 2000, Salas-Marco and Bedwell 2004) (figure 28). La structure tertiaire d'eRF1 ressemble à celle d'un ARNt, indiquant une convergence fonctionnelle entre ces 2 molécules (Cheng et al. 2009) (figure 29). La région NIKS qui reconnait le codon stop est située à la pointe du domaine N, et la séquence GGQ est à la pointe du domaine M.

- ***eRF3***

La protéine eRF3 est un facteur de terminaison de classe II (Loh and Song 2010), c'est un paralogue d'eEF1 (Atkinson, Baldauf and Hauryliuk 2008). Elle possède une activité GTPase dépendante d'eRF1 et du ribosome (Pisareva et al. 2006). Les protéines eEF1A et eRF3 présentent des homologies de séquences (Inagaki and Ford Doolittle 2000) et de fonction (Salas-Marco and Bedwell 2004). La protéine eRF3 possède 4 domaines : N, G (GTPase), post-G et C. Les domaines G et post-G sont aussi appelés domaines M. Le domaine N contient chez les Eumycètes, des répétitions d'acides aminés glycine, glutamate, asparagine et tyrosine. Chez certaines levures, ces acides aminés sont impliqués dans la bascule sous forme prion d'eRF3. Ce point sera abordé au point V.1. Les domaines G (GTPase), post G et C sont

Figure 30 : Comparaison entre eRF1 et Dom34p

Superposition des structures des protéines eRF1 (beige) et Dom34p (bleu). Les structures sont superposables, Dom34p ne possède pas les extrémités NIKS et GGQ. Figure issue de (Graille, Chaillet and van Tilbeurgh 2008)

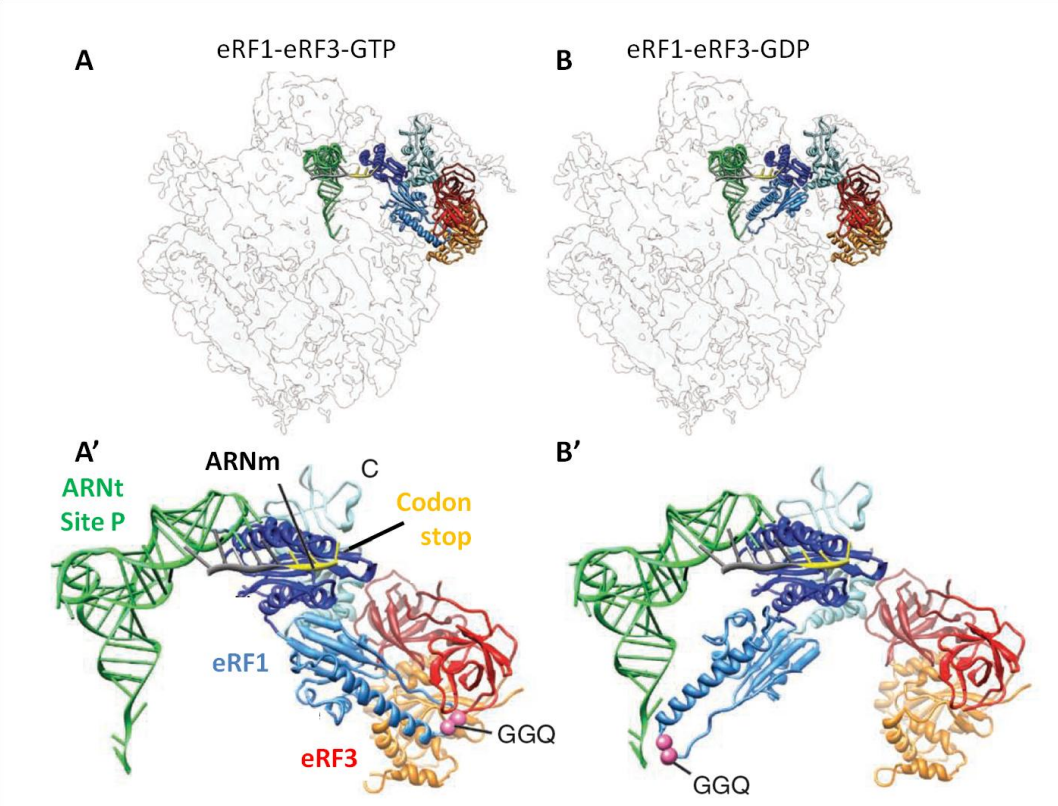


Figure 31 : Prédiction de la structure du complexe de terminaison dans le ribosome

Lors du recrutement du complexe de terminaison eRF1-eRF3-GTP au ribosome, seul le domaine N d'eRF1 entrerait au site A. Le domaine NIKS serait au site de décodage. Suite à l'hydrolyse du GTP par eRF3, eRF1 s'accommoderait au site A. Le domaine GGQ se retrouverait idéalement placé au site de transfert peptidique pour libérer la protéine. Figure adaptée de (Becker et al. 2011)

conservés au sein de la superfamille EF1 (Atkinson et al. 2008). Chez les mammifères, il existe 2 protéines eRF3 (Kisselev and Buckingham 2000, Kisselev, Ehrenberg and Frolova 2003).

- ***Le complexe ternaire eRF1-eRF3-GTP***

Le fonctionnement de chacune de ces 2 protéines est interdépendant. Elles coopèrent pour libérer la protéine rapidement et efficacement (Salas-Marco and Bedwell 2004, Alkalaeva et al. 2006). eRF1 et eRF3 interagissent fortement par leurs domaines C (Ito, Ebihara and Nakamura 1998, Merkulova et al. 1999, Ebihara and Nakamura 1999). La protéine eRF1 stabilise la fixation du GTP sur eRF3 (Hauryliuk et al. 2006) et stimule l'activité GTPase dépendante du ribosome d'eRF3 (Pisareva et al. 2006) grâce à une interaction entre son domaine M et les domaines M et C d'eRF3 (Kononenko et al. 2008, Cheng et al. 2009). La protéine eRF3 stimule la reconnaissance du codon stop (Salas-Marco and Bedwell 2004) et le relâchement du peptide par eRF1 (Alkalaeva et al. 2006).

C. Recrutement du complexe de terminaison par le ribosome

A ce jour, il n'y a pas de structure du complexe eRF1-eRF3 dans le ribosome. Cependant, il existe une structure avec des paralogues structuraux Dom34-Hbs1 (Inagaki and Ford Doolittle 2000) qui se fixent au ribosome lorsque celui-ci est bloqué sur l'ARNm en cours d'élongation. Ce mécanisme sera abordé au point IV.1. Dom34 est très similaire à eRF1, les différences majeures sont l'absence des régions portant le motif GGQ et de reconnaissance du codon stop (Graille, Chaillet and van Tilbeurgh 2008) (figure 30). La structure du complexe Dom34-Hbs1-GTP dans le ribosome 80S permet d'établir un modèle de fixation du complexe de terminaison dans le ribosome (Becker et al. 2012) (figure 31).

Le complexe ternaire est recruté par le ribosome. Le domaine GTPase d'eRF3 se fixerait sur la sous-unité 60S dans la même région qu'eEF2 (Alkalaeva et al. 2006), eEF1A et Hbs1 au niveau du centre CAG et de la boucle SRL. La région NIKS du domaine N d'eRF1 rentrerait dans le ribosome et serait à proximité du codon au site A (Becker et al. 2011). Quant au domaine M portant le motif GGQ, il ne serait pas près de la cavité PTC mais toujours en contact avec les domaines M et C d'eRF3 pour stimuler l'activité GTPase (Cheng et al. 2009, Becker et al. 2011).

La fixation du complexe ternaire eRF1-eRF3-GTP induit un changement de conformation du ribosome qui n'impliquerait pas un encliquetage des sous-unités. Dans ce complexe de post terminaison, l'extrémité CCA de l'ARNt peptidyl est toujours dans la cavité du PTC prête à être hydrolysée (Alkalaeva et al. 2006).

- ***Hydrolyse du GTP et départ d'eRF3.***

L'hydrolyse du GTP impliquerait un réarrangement des interactions entre eRF1 et eRF3 (Cheng et al. 2009, Becker et al. 2011) conduisant à un positionnement correct du motif GGQ au centre PTC (Alkalaeva et al. 2006) et à une dissociation d'eRF3 du ribosome (Shoemaker and Green 2011). L'énergie issue de l'hydrolyse du GTP permet un couplage efficace entre la reconnaissance du codon stop et la libération de la protéine (Alkalaeva et al. 2006, Cheng et al. 2009).

D. Reconnaissance du codon stop

Les mécanismes de reconnaissance de codon stop par eRF1 ne sont pas encore élucidés. L'hypothèse actuelle est que la reconnaissance ne se ferait pas sur une simple interaction codon:peptide mais par une reconnaissance d'un réseau tridimensionnel d'interactions (Song et al. 2000, Bertram et al. 2000, Inagaki et al. 2002, Seit-Nebi et al. 2002, Kolosov et al. 2005).

Deux régions conservées interviennent dans la reconnaissance du codon stop (Inagaki et al. 2002). Elles sont situées dans 2 boucles à la pointe du domaine N d'eRF1. Il s'agit du motif Asparagine-Isoleucine-Lysine-Sérine (NIKS) (Frolova, Seit-Nebi and Kisselev 2002) et de la séquence Tyrosine-X-Cystéine-XXX-Phénylalanine (YxCxxxF) (Seit-Nebi et al. 2002, Kolosov et al. 2005). Il a été montré que motif NIKS interagit directement avec la 1ère base du codon stop (Chavatte et al. 2002).

Deux autres cavités P1 et P2 (Poche 1 Poche 2) à l'extrémité du domaine N d'eRF1 sont impliquées dans la régulation de la fidélité de la terminaison. La cavité P1 est située près de la zone de décodage du codon stop, elle influencerait la reconnaissance de ce dernier. La cavité P2 aurait un rôle dans l'interaction d'eRF1 avec le site A (Hatin et al. 2009).

- Le contexte nucléotidique

Des analyses bioinformatiques et biochimiques ont montré qu'il existait un biais dans l'utilisation des nucléotides entourant le codon stop (Kohli and Grosjean 1981, Brown et al. 1990, Cridge et al. 2006). Cette influence du contexte nucléotidique est retrouvée chez tous les organismes mais il prend des formes différentes selon les espèces. Ce biais est influencé par le % global en bases G/C d'un génome (Brown et al. 1990). Afin de décrire facilement ces biais, j'utiliserai la nomenclature suivante : les 3 bases du codon stop sont numérotées de +1 à +3. Les bases en 5' (- x) et 3' (+ x) sont numérotées par rapport à la 1ère base du codon stop.

Le contexte en 5' du codon stop influence aussi l'efficacité de terminaison. Les génomes eucaryotes présentent un biais de bases aux positions -1, -2 et -3 (site P). En -3 le U est surreprésenté alors que le G est sous-représenté. En -2, le C est sous représenté alors qu'il est surreprésenté en position -1. De plus, lorsqu'il existe des ARNt isoaccepteurs pour les 2 derniers acides aminés ajoutés, il a été observé que se sont souvent les ARNt hyper-modifiés aux positions 32, 34 et 37 de la boucle anticodon qui sont incorporés. Cette observation est d'autant plus vraie pour les gènes fortement exprimés (Cridge et al. 2006). Etonnamment, les signaux de terminaisons (compris entre -6 et +3) les plus abondants chez les eucaryotes ne sont pas forcément les plus efficaces pour la terminaison. Ils pourraient être impliqués dans d'autres régulations que celle de la terminaison, comme le recyclage des sous-unités ribosomiques (Cridge et al. 2006).

Chez les eucaryotes, la base +4 est majoritairement une purine (A ou G) (Cridge et al. 2006). Chez *S. cerevisiae*, 65% des codons sont suivis par une purine (38% A, 27% G) alors que seulement 7% sont suivis par un C. Le tétranucléotide le plus utilisé est UAA(G/A). L'observation de ce biais dans plusieurs génomes a donné lieu à un modèle où les facteurs de terminaisons de classe I reconnaitraient un tétranucléotide au lieu d'un codon (Brown et al. 1990). Le biais (G/A) en +4 est surreprésenté dans la classe de gènes les plus exprimés ce qui indiquerait un rôle de cette base dans l'efficacité de terminaison (Brown et al. 1990, Cridge et al. 2006).

Chez *S. cerevisiae*, le biais dans l'utilisation des bases semble s'étendre jusqu'à la position +12 mais il est particulièrement important aux positions +5 et +8. La position +5 le G est favorisé et le U défavorisé. D'autre part, en +8 le A est favorisé et le U défavorisé (Brown et al. 1990).

Le rôle du contexte nucléotidique est aussi illustrée dans la partie « Translecture » V.1.

E. Recrutement d'ABCE1/Rli1p

Suite au départ d'eRF3, l'accommodation d'eRF1 au site A et la libération de la chaîne peptidique sont facilitées par la fixation de la protéine ABCE1/Rli1p sur le ribosome (Shoemaker and Green 2011).

❖ *Un nouveau partenaire : ABCE1 / Rli1p*

La protéine ABCE1 chez les mammifères, et son homologue Rli1p chez *S. cerevisiae*, stimulent l'hydrolyse de la liaison ester entre l'ARNt peptidyl et la chaine peptidique par eRF1 (Shoemaker and Green 2011) et la dissociation des sous-unités ribosomiques lors du recyclage du ribosome (Pisarev et al. 2010). ABCE1 appartient à la famille des protéines ABC (ATP-Binding Cassette) (Barthelme et al. 2011) mais elle n'a pas de préférence pour un nucléotide NTP en particulier (Loh and Song 2010). Elle possède 2 domaines de fixation identiques pour des NTP (NBD Nucleotide Binding Domaine). Ils sont orientés tête-bêche et sont reliés par une région charnière. Leur écartement dépend des nucléotides, la fixation de NTP induit une fermeture des domaines alors que l'hydrolyse les ouvre (Loh and Song 2010). Ce mécanisme de pince serait impliqué dans la dissociation des sous-unités ribosomiques (Loh and Song 2010, Barthelme et al. 2011, Becker et al. 2012). ABCE1 possède aussi un motif hélice-boucle-hélice (HLH Helix-Loop-Helix) et un domaine fer-souffre (FeS) essentiel pour dissocier le ribosome (Barthelme et al. 2011).

❖ *Recrutement de Rli1p par le ribosome*

Là encore, il n'existe pas de structure du complexe eRF1-Rli1p dans le ribosome. Cependant, celle avec Dom34-Rli1p existe et permet d'établir un modèle d'interactions entre eRF1, Rli1p et le ribosome (Becker et al. 2012).

Rli1p est recruté au complexe de post-terminaison sous sa forme GDP (Barthelme et al. 2011). Sous sa forme liée à un ATP, il se fixe sur la sous-unité 60S au même endroit que les autres GTPase impliquées dans la traduction mais ces interactions avec la boucle SRL sont plus distantes. Rli1p interagirait via son domaine FeS avec le domaine C d'eRF1. D'après la structure de Dom34p, eRF1 serait idéalement placé au site A avec le domaine GGQ en interaction avec le centre PTC et l'extrémité CCA du bras accepteur de l'ARNt peptidyl (Becker et al. 2012). La libération de la chaîne peptidique se produit avant l'hydrolyse d'ATP par Rli1p (Shoemaker and Green 2011).

La fixation d'ABCE1 sur le complexe de post-terminaison inverse le changement de conformation du ribosome observé lors de la fixation d'eRF1-eRF3-GTP. Dans cette conformation, l'affinité du ribosome pour eRF1 est diminuée mais l'interaction entre eRF1 et ABCE1 assure une fixation coopérative forte au ribosome (Pisarev et al. 2010).

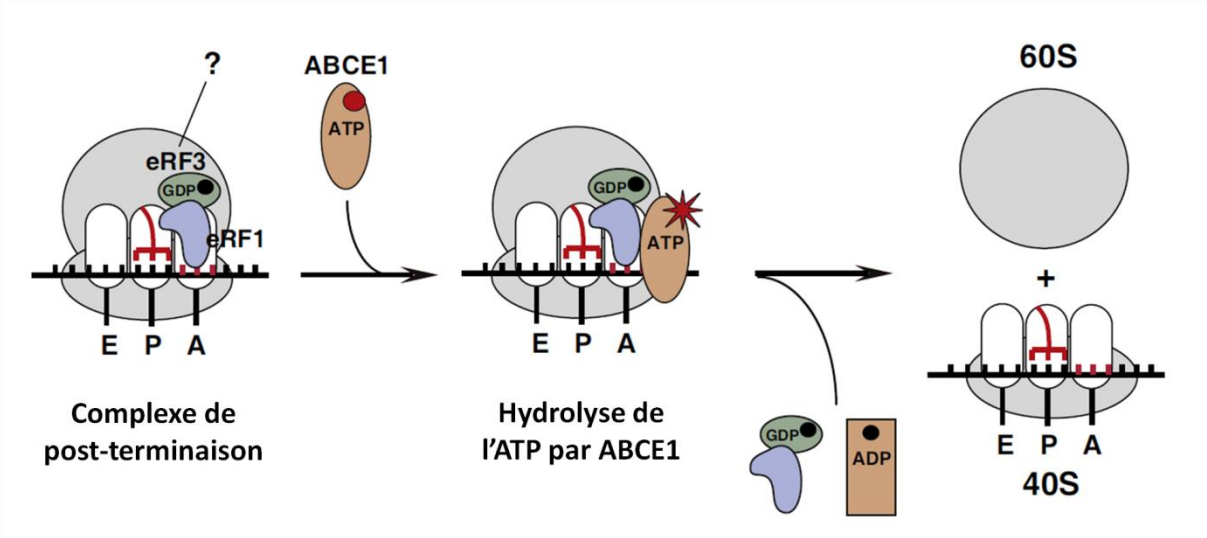


Figure 32 Dissociation des sous-unités ribosomiques

Suite au départ d'eRF3, ABCE1/Rli1p-GTP se fixe à eRF1 pour stimuler la libération de la protéine néo-synthétisée. ABCE1 hydrolyse alors le GTP ce qui induit la dissociation des sous-unités ribosomiques. Figure adaptée de (Pisarev et al. 2010)

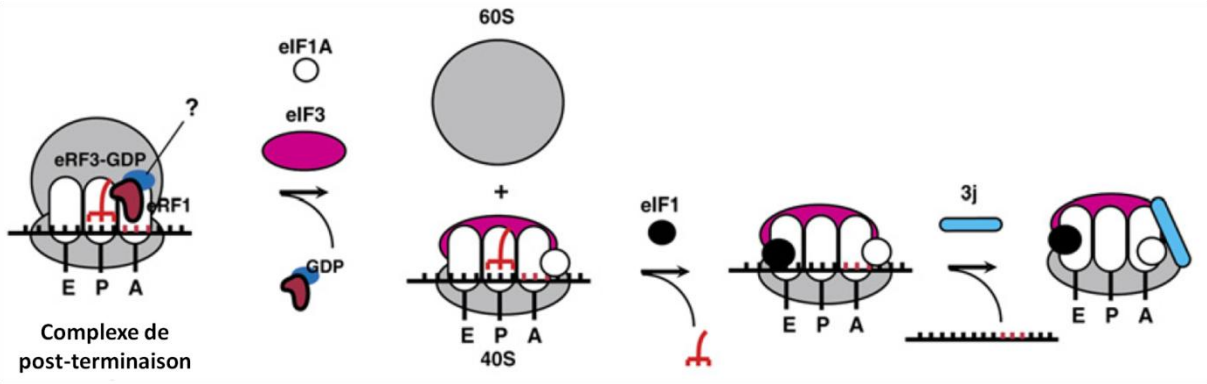


Figure 33 Recyclage des sous-unités ribosomiques

Suite à la dissociation du ribosome, la sous-unité 60S est prise en charge par eIF6. Les facteurs eIF1A et eIF3 se lient à la sous-unité 40S pour décrocher l'ARNt déacylé du site P. La fixation d'eIF1 et d'eIF3j permet le décrochage de l'ARNm. Figure adaptée de (Pisarev, Hellen and Pestova 2007)

F. Libération de la protéine

Le motif GGQ d'eRF1, la cavité PTC du ribosome et la liaison ester entre l'ARNtpeptidyl et la protéine sont universellement conservés. Bien qu'il n'y ait pas de modèle structural avec la protéine eRF1, les structures établies avec les facteurs de terminaisons RF1 et RF2 bactériens fournissent un éclairage sur le mécanisme d'hydrolyse de la liaison ester entre l'ARNt et la protéine (Laurberg et al. 2008, Weixlbaumer et al. 2008, Korostelev et al. 2008).

Le motif GGQ est en contact avec des bases du PTC et la dernière base A76 de l'ARNt. Il forme une boucle dont la conformation est rendue possible grâce à la présence des 2 acides aminés Glycine. Cette conformation ne peut être obtenue avec les autres acides aminés, ce qui explique l'universalité de ces 2 résidus. La boucle permet d'écarter la base U2585 de la liaison ester et de rendre cette liaison attaquable chimiquement par une molécule d'H_2O. Le résidu Glutamine (Q) est méthylé chez les bactéries et les eucaryotes. La méthylation stimule la terminaison chez les bactéries, mais ce n'est pas le cas chez *S. cerevisiae*(Graille et al. 2012). Il forme une liaison hydrogène avec la base A76 de l'ARNt déacylé. Deux modèles sont proposés pour expliquer l'action du motif GGQ. Dans le 1er, la glutamine catalyserait l'hydrolyse en coordonnant la molécule d'H_2O ; dans le 2nd, elle aiderait à la catalyse en stabilisant un des intermédiaire de réaction (Loh and Song 2010).

4. Le recyclage des sous-unités ribosomiques

Une fois la protéine libérée, la terminaison proprement dite est achevée. Le ribosome doit alors être dissocié et chaque élément recyclé pour entamer un nouveau cycle de traduction.

A. Dissociation des sous-unités ribosomiques

Chez les eucaryotes, la terminaison et la dissociation sont couplées car toutes deux font intervenir le complexe eRF1-ABCE1 (figure 32). Cependant ces 2 étapes sont chimiquement distinctes, la terminaison est liée à la fixation d'ABCE1 alors que la dissociation du ribosome nécessite l'hydrolyse d'NTP (Shoemaker and Green 2011).

Un 1er modèle propose une action directe d'ABCE1 dans la dissociation des sous-unités ribosomiques. Elles suivraient le mouvement de fermeture des domaines NBD lors de la fixation d'ATP. Le pincement d'un côté des sous-unités romprait les interactions inter sous-unités de l'autre côté en déstabilisant ainsi le ribosome (Becker et al. 2012).

Le 2nd modèle propose que le changement de conformation d'ABCE1 soit transmis via le domaine FeS à eRF1 induisant un changement d'interaction entre les domaines N et M d'eRF1, l'ARNt déacylé et les ponts entre les sous-unités (Becker et al. 2012).

B. Recyclage de l'ARNm, de l'ARNt et des sous-unités ribosomiques

Après dissociation, il y a d'un côté la sous-unité 60S et de l'autre la sous-unité 40S avec l'ARNm et l'ARNt déacylé au site P (complexe 48S) (Pisarev et al. 2010) (figure 33). La sous-unité 60S est prise en charge par le facteur d'anti-association eIF6. La petite sous-unité est reconnue par les facteurs d'initiation eIF1 et eIF3. Le recrutement d'eIF1A stimule la libération de l'ARNt déacylé du site P par

eIF1. L'ARNm est ensuite décroché grâce au recrutement d'eIF3j (Pisarev, Hellen and Pestova 2007, Pisarev et al. 2010).

Dans le cas des uORF, la sous-unité 40S n'est pas détachée de l'ARNm et reforme un complexe d'initiation avec le complexe ternaire pour scanner l'ARNm jusqu'au prochain codon d'initiation (Zhang and Hinnebusch 2011).

Les sous-unités ribosomiques sont prêtes à entrer dans un nouveau cycle de traduction.

IV. Traduction et régulation de l'expression génique.

Le contrôle de l'expression génique passe par une régulation temporelle de la transcription, de la traduction et de la dégradation des ARNm. Dans cette partie, nous verrons comment la traduction et la dégradation peuvent être liées afin de réguler l'expression génique.

La transcription peut entraîner la synthèse d'ARNm aberrants. Afin d'éliminer ces ARNm, il y existe un contrôle de qualité des ARNm. Dans les cellules eucaryotes, ce contrôle s'effectue à 2 niveaux :

- Dans le noyau, les ARNm mal maturés sont polyadénylés par une polyA polymérase spéciale associée au complexe TRAMP. Cette queue polyA est dans ce cas un motif de dégradation reconnu par l'exosome nucléaire (Camier and Séraphin 2007).
- Dans le cytoplasme, la traduction permet de reconnaître les ARNm aberrant conduisant à la synthèse de protéines potentiellement délétères pour la cellule. Selon le défaut associé à un ARNm, il sera pris en charge par différentes voies de dégradation.

Les ARNm sur lesquels des ribosomes sont bloqués en cours d'élongation sont identifiés et dégradés par la voie NGD (No-Go Decay). Les ARNm dépourvus de codon stop sont pris en charge par la voie NSD (Non-Stop Decay) alors que ceux possédant un codon stop précoce sont dégradés par la voie NMD (Non-sense -Mediated Decay).

Outre des mécanismes moléculaires différents, ces voies se caractérisent par des vitesses de dégradation des ARNm différentes. Pour le gène rapporteur PGK1, les temps de demi-vie des ARNm sont estimés à (Passos et al. 2009) :

- NGD : 8 min pour un gène « PGK1-SL » contenant une structure tige-boucle très stable dans la phase codante (Doma and Parker 2006). (ARNm control : 17 min)
- NSD : 2-3 min pour un gène PGK1 dépourvu de codon stop (Frischmeyer et al. 2002). (ARNm control : 19 min)
- NMD : 1 min pour un gène PGK1 avec un codon stop précoce près du codon d'initiation (Cao and Parker 2003). (ARNm control : 28 min)
-

Chez *S. cerevisiae*, la dégradation « classique » des ARNm est initiée par une dégradation de la queue poly(A) médiée par l'enzyme Ccr4p. Cette déadénylation déclenche un décoiffage par l'enzyme Dcp1p et une digestion 5'-3' de l'ARNm par l'enzyme Xrn1p (Frischmeyer et al. 2002). La dégradation 3'-5' a seulement un rôle mineur dans la dégradation des ARNm (Wilusz, Wormington and Peltz 2001). La dégradation des ARNm dans les voies NGD, NSD et NMD utilise des mécanismes partiellement différents.

1. NGD (No-Go Decay)

La voie NGD cible et dégrade les ARNm sur lesquels des ribosomes sont bloqués en cours d'élongation par une structure secondaire forte (Doma and Parker 2006) ou une base dépurinée (Gandhi, Manzoor and Hudak 2008). Dans une mesure largement moindre et à basse température, un

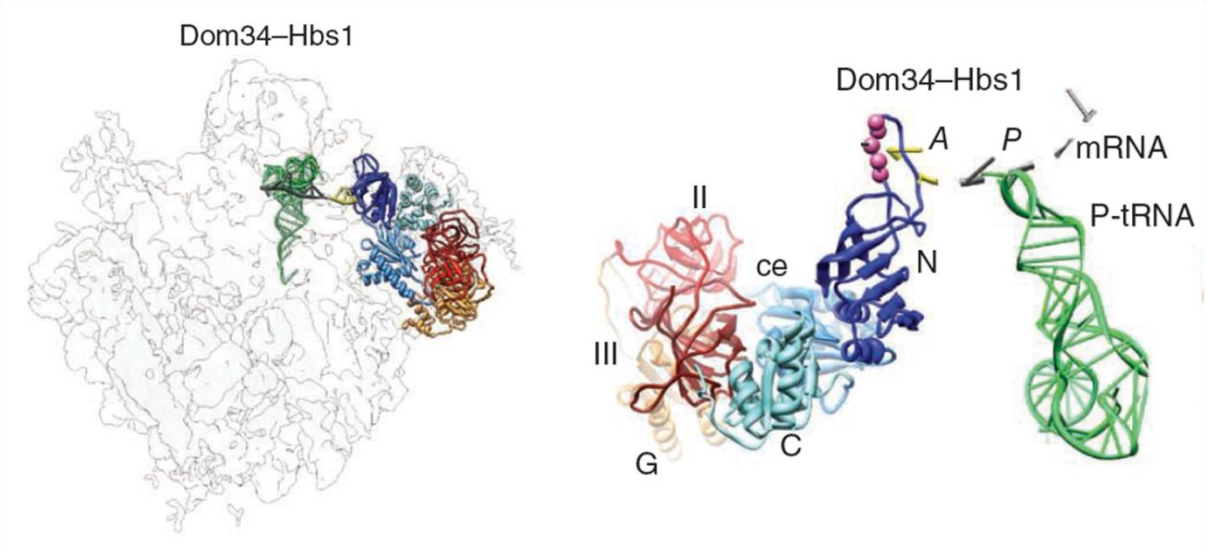


Figure 34 Structure du complexe Dom34p-Hbs1p dans le ribosome

Lors du recrutement du complexe Dom34p-Hbs1p-GTP sur un ribosome bloqué, Hbs1p (rouge) se fixe comme les autres GTPase sur la boucle SRL et au centre d'activation des GTPases (GAC). Le domaine N (ou I) d'Hbs1p contact le tunnel d'entrée de l'ARNm. Le domaine N de Dom34p (bleu) entre au site A à proximité du codon. Figure issue de (Becker et al. 2011)

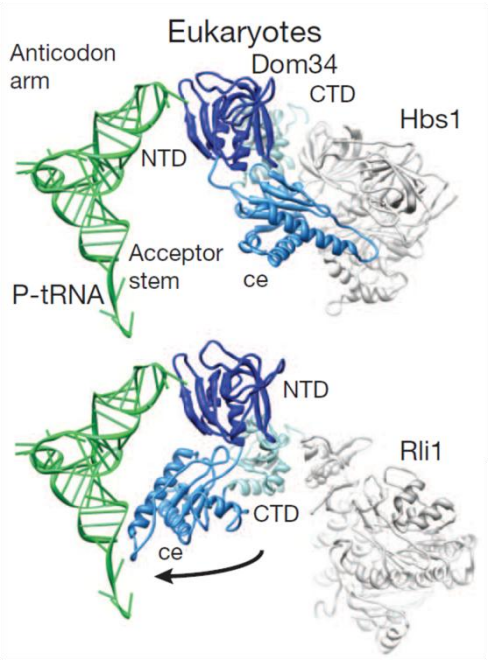


Figure 35 Comparaison des complexes Dom34p-Hbs1p et Dom34p-Rli1p dans le ribosome

Suite à l'hydrolyse du GTP et au départ d'Hsb1p, Dom34p est accommodé au site A avec son extrémité du domaine C en contact avec la tige acceptrice de l'ARNt peptidyl au site P. Cette conformation est stabilisée par la fixation de la protéine Rli1p. Figure issue de (Becker et al. 2012)

pseudonoeud, un codon rare et un codon stop prématuré peuvent induire la voie NGD (Doma and Parker 2006).

Les ribosomes bloqués sont reconnus par le dimère de protéines Dom34p-Hbs1p (Doma and Parker 2006). Ce complexe avec l'aide de Rli1p va dissocier le ribosome pour permettre le recyclage des sous-unités. Contrairement à la terminaison, la protéine synthétisée n'est pas libérée (Shoemaker, Eyler and Green 2010). De manière concomitante, la dégradation de l'ARNm est initiée par un clivage endonucléolytique près du ribosome bloqué (Doma and Parker 2006, Chen et al. 2010). Le fragment en 5' du clivage est dégradé par l'exosome alors que le fragment en 3' est dégradé par l'enzyme Xrn1p (Doma and Parker 2006).

A. Dom34p-Hbs1p

Dom34p et HBS1p sont respectivement des paralogues des protéines eRF1 et eRF3.

Hbs1p (Hsp70 subfamily B Suppressor1) appartient à la super famille EF1 et possède 4 domaines ; N, G (GTPase), post-G et C (Atkinson et al. 2008).

Dom34p (Duplication Of Multilocus region 34) et son homologue Pelota chez les mammifères possèdent 3 domaines : N, M et C. Le domaine N ne ressemble pas à celui d'eRF1, il adopte une structure « Sm-fold » qui est couramment retrouvée dans les protéines interagissant ou dégradant des ARN (Lee et al. 2007, Graille et al. 2008). La pointe du domaine N contient 2 boucles qui sont importantes pour l'activité de Dom34p dans les voies NGD et NSD (Passos et al. 2009, Kobayashi et al. 2010). Le domaine M ne contient pas le motif GGQ d'eRF1 ce qui explique qu'il n'y ait pas libération de la protéine dans la voie NGD. A la place, la boucle contient un motif conservé de résidus basiques Lysine-Lysine-Lysine-Arginine (KKKR) (Passos et al. 2009) nécessaire à la fonction de Dom34p (van den Elzen et al. 2010). Les domaines M et C sont impliqués dans l'interaction avec Hbs1p. Comme pour le complexe eRF1-eRF3, l'interaction de Dom34p avec Hbs1p augmente l'affinité de Hbs1p pour le GTP.

Une étude initiale de Dom34p proposait qu'il était responsable du clivage de l'ARNm (Lee et al. 2007) mais cette assertion a été largement contredite par la suite (Passos et al. 2009, Chen et al. 2010, Shoemaker et al. 2010). Le facteur clivant l'ARNm reste encore inconnu.

La structure de Dom34p/Hbs1p dans un ribosome 80S bloqué en cours de traduction par une structure secondaire forte montre que le ribosome est dans une conformation « décliquetée ». Comme pour eEF2, la fixation de Dom34p-Hbs1p provoque une fermeture de la « stalk base » vers l'intérieur du ribosome. Le domaine N de Dom34p entre profondément dans le site A où il contacte le centre de décodage (figure 34). Dans la structure, il semble que l'ARNm aux sites P et A soit désordonné. Dom34p induirait une déstabilisation de l'ARNm dans le ribosome. Le domaine M de Dom34p interagit peu avec le ribosome alors que le domaine C contacte la « stalk base » de la sous-unité 60S et est ramassé dans un contact avec Hbs1. Le domaine G d'Hbs1p contacte la sous-unité 40S et la boucle SRL dans la sous-unité 60S. De manière surprenante, l'extrémité du domaine N d'Hbs1p contacte la protéine rpS3 à l'orifice du tunnel d'entrée de l'ARNm (Becker et al. 2011).

La fixation de Pelota-Hbs1 n'induit pas de changement de conformation comme celui observé pour la fixation d'eRF1-eRF3. Chez l'Homme, le complexe Pelota/Hbs1 n'est pas suffisant pour dissocier le

ribosome, la dissociation requiert la présence d'ABCE1. En présence d'un court peptide sur l'ARNt peptidyl, la dissociation conduit à une sous-unité 60S d'un côté et d'une sous-unité 40S avec l'ARNm et l'ARNt peptidyl de l'autre (Pisareva et al. 2011). Chez *S. cerevisiae*, le complexe Dom34p-Hbs1p suffit à dissocier le ribosome (Shoemaker et al. 2010) indépendamment du codon présent au site A. La présence de Rli1p accélère la dissociation du ribosome d'un facteur 10 (Shoemaker and Green 2011).

Chez l'Homme, la dissociation du ribosome n'est observable que si la longueur de l'ARNm est inférieure à 9 nucléotides après le codon au site P (Pisareva et al. 2011). Cette dépendance est aussi observée chez *S. cerevisiae*, puisqu'au-delà de 23 nucléotides la dissociation est fortement réduite. Cette variation de longueur peut être le reflet de mécanismes de perception de la taille de l'ARNm qui diffèrent selon les espèces. La présence de Hbs1p est nécessaire pour une dissociation dépendante de la taille de l'ARNm (Shoemaker and Green 2011). Dans la structure décrite précédemment, le domaine N d'Hbs1p contacte l'entrée du tunnel et pourrait donc être impliqué dans la perception de la taille de l'ARNm en 3' du site P (Becker et al. 2011).

B. Dom34p-Rli1p

La structure de Dom34p/Rli1p dans le même ribosome montre qu'entre le départ d'Hbs1p et la présence de Rli1p, le domaine M de Dom34p a changé de conformation (figure 35). Dom34p est complètement accommodé au site A. Le domaine N n'a pas bougé, par contre la boucle KKKR du domaine M contacte à présent l'extrémité du bras accepteur de l'ARNt peptidyl. Le domaine C de Dom34p n'interagit plus qu'avec le domaine FeS de Rli1p. La « stalk base » et le domaine C sont resserrés autour de la boucle SRL. Il y a contact entre la boucle SRL et le CTD de Dom34p. La stalk base a la même conformation qu'en présence d'eEF2. Rli1p se fixe au ribosome dans la même région que les autres GTPase, cependant il ne contacte pas la boucle SRL (Becker et al. 2012).

Le recyclage de ces sous-unités ribosomiques permet qu'elles soient utilisées pour de nouveaux cycles de traduction.

2. NSD (No-Stop Decay)

La voie NSD reconnaît et dégrade les ARNm dépourvus de codon stop ce qui inclut les ARNm soumis à une polyadénylation prématurée (Frischmeyer et al. 2002). En l'absence de stop, le ribosome traduit la queue polyA. L'incorporation dans la protéine de 10 acides aminés lysines suffit à réprimer l'élongation en bloquant le ribosome dans la queue poly A (Ito-Harashima et al. 2007). Chez *S. cerevisiae*, les ribosomes bloqués dans la queue polyA seraient reconnus par la protéine cytoplasmique Ski7p (Super Killer 7) (van Hoof et al. 2002). La protéine Ski7p n'est présente que dans le règne des Eumycètes. Elle appartient à la superfamille des GTPase EF1, c'est un paralogue de la protéine HBS1p (Atkinson et al. 2008). Le domaine C-terminal de Ski7p est essentiel pour la dégradation des ARNm sans stop. Comme le domaine N-terminal de Ski7p s'associe de manière stable avec l'exosome, le modèle actuel suggère que l'exosome soit recruté au ribosome via le domaine C-terminal de Ski7p. L'intérêt de dégrader l'ARNm de 3' vers 5' serait d'arrêter rapidement la traduction des ribosomes en amont (van Hoof et al. 2002). Les protéines partiellement traduites sont marquées par l'ubiquitine-ligase Ltn1p (Bengtson and Joazeiro 2010) et dégradées par le protéasome (Ito-Harashima et al. 2007, Wilson, Meaux and van Hoof 2007).

Le complexe Dom34p-Hbs1p intervient aussi dans la voie NSD en dissociant les complexes ribosomiques et stimulant la dégradation de l'intermédiaire 5' de l'ARNm après clivage (Kobayashi et al. 2010, Tsuboi et al. 2012).

La voie NSD est aussi présente chez les mammifères (Frischmeyer et al. 2002). L'équivalent de la protéine Ski7p chez les autres eucaryotes reste inconnu.

La voie NSD ne servirait pas seulement à dégrader les ARNm aberrants. Elle pourrait réguler l'expression de gènes ayant des signaux alternatifs de polyadénylation dans leurs phases codantes (Frischmeyer et al. 2002).

3. NMD (Non-sense-Mediated Decay)

La voie NMD reconnait et dégrade spécifiquement les ARNm ayant un codon stop précoce. Au niveau de l'ADN, ces stops peuvent résulter d'une mutation ponctuelle dans le gène ainsi que d'une insertion ou délétion de bases qui changent le cadre de lecture. Au niveau de l'ARN, ils peuvent être la conséquence d'épissages alternatifs, d'erreurs d'épissage ou d'erreur de transcription (Nicholson and Mühlemann 2010).

L'étude de la voie NMD repose sur 2 questions centrales : comment est reconnu le codon stop prématuré ? et comment l'ARNm est-il dégradé ? Beaucoup d'éléments de réponses ont été apportés pour la première question. Il apparaît que les mécanismes moléculaires varient entre les organismes en fonctions des protéines impliquées. Il existe malgré tout des caractéristiques communes dans l'activation de la voie NMD. Pour être détecté, un ARNm aberrant doit être traduit jusqu'au codon stop précoce, ce dernier doit donc être en phase avec le codon d'initiation. La reconnaissance du caractère précoce du codon stop dépend de sa position sur l'ARNm. Lorsqu'il est près du 5'UTR, il est reconnu efficacement ce qui entraîne une dégradation rapide de l'ARNm. A l'inverse, plus il est proche du codon stop naturel moins il est détecté et moins l'ARNm est dégradé (Neu-Yilik and Kulozik 2008).

Le codon stop prématuré est reconnu par le complexe de terminaison qui recrute un complexe de 3 protéines : Upf1p, Upf2p et Upf3p. Selon les espèces, d'autres facteurs s'ajoutent à cette base. D'un point de vue moléculaire, l'activation de la voie NMD implique une interaction entre Upf1p et eRF3 lors de la terminaison (Brogna and Wen 2009).

La connaissance des mécanismes de dégradation de l'ARNm dans la voie NMD est encore incomplète. Selon les protéines partenaires activées, l'ARNm est dégradé soit directement par ses extrémités, ou soit via un clivage endonucléolytique (Nicholson and Mühlemann 2010).

Là encore, la voie NMD ne sert pas seulement à éliminer les ARNm aberrant, elle régule la traduction de certains gènes en affectant le temps de demi-vie de leurs ARNm. Cette régulation s'applique aux gènes régulés par des uORF, à des variants d'épissage alternatif, ou encore à des gènes soumis au recodage. Le recodage est abordé dans la partie suivante. Selon les espèces, il est estimé que 3 à 10% des transcrits sont régulés par la voie NMD (Nicholson and Mühlemann 2010).

V. Le recodage

Les règles de la traduction énoncées précédemment peuvent être altérées sur des séquences spécifiques de certains ARNm. Ces « erreurs programmées de la traduction », regroupées sous le terme de recodage (Gesteland, Weiss and Atkins 1992), sont en compétition avec les règles standards de la traduction. En conséquence, une seule partie des ribosomes traduisant ces séquences spécifiques subiront un recodage et produiront alors une protéine différente de celle prédite à partir des règles standard (Baranov, Gesteland and Atkins 2002).

Une liste des gènes répertoriés comme soumis à un événement de recodage est accessible sur la base de données RECODE : http://recode.genetics.utah.edu/ (Bekaert et al. 2010).

1. La translecture

La translecture est la reconnaissance d'un codon stop par un ARNt proche-cognat appelé, dans ce cas, ARNt suppresseur naturel. Par exemple, chez *S. cerevisiae* les $ARNt^{GLN}$ ayant comme anticodon GUC et UUG peuvent reconnaître respectivement les codons stops UAG et UAA (Edelman and Culbertson 1991, Pure et al. 1985). L'incorporation d'un acide aminé conduit la traduction à se poursuivre dans la même phase de lecture jusqu'au codon stop suivant. La translecture est en concurrence directe avec la terminaison pour le décodage du codon stop. Comme la terminaison et la translecture sont 2 faces d'une même pièce, leur étude permet un apport mutuel à chaque domaine.

In vivo, une surexpression des facteurs de terminaison eRF1 et eRF3 augmente l'efficacité de terminaison en concurrençant les ARNt suppresseurs naturels, ce phénomène est appelé anti-suppression. Chez les mammifères, la surexpression d'eRF1 seul suffit à augmenter l'anti-suppression (Le Goff, Philippe and Jean-Jean 1997) alors que chez *S. cerevisiae* la surexpression conjointe d'eRF1 et eRF3 est nécessaire (Stansfield et al. 1995).

La translecture est favorisée par des signaux situés de part et d'autre du codon stop. Ces signaux qui constituent le contexte nucléotidique, incluent parfois la présence d'une structure secondaire qui stimule la translecture. Les contextes nucléotidiques favorables à la terminaison de la traduction ont été abordés au point III.3.D. Je vais à présent décrire un contexte favorisant la translecture chez les eucaryotes (Cassan and Rousset 2001).

La séquence **CAA UAG CAA UAA** (le codon stop est souligné) a été identifiée dans le virus de la mosaïque du tabac TMV (Tobacco Mosaic Virus). La translecture de cette séquence est utilisée pour réguler le taux de traduction entre leur protéine d'enveloppe *gag* et une fusion protéique de la protéine d'enveloppe avec la polymérase *gag-pol* (Skuzeski et al. 1991).

Ce contexte nucléotidique provoque un taux de translecture de 4-5% chez les plantes (Skuzeski et al. 1991) et de 2% dans des cellules de souris (Cassan and Rousset 2001). Ce taux est plus élevé chez *S. cerevisiae*, il est compris entre 15 et 25% (Bonetti et al. 1995, Stahl et al. 1995). Il a été montré que le contexte en 5' et en 3' agissait en synergie pour stimuler la translecture (Cassan and Rousset 2001).

Un crible *in vivo* recherchant des contextes favorables de translecture a établi que la séquence consensus $^{+4}$CA(A/G) $^{+7}$N(U/C/G)A en 3' d'un stop suffisait à induire une translecture supérieure à

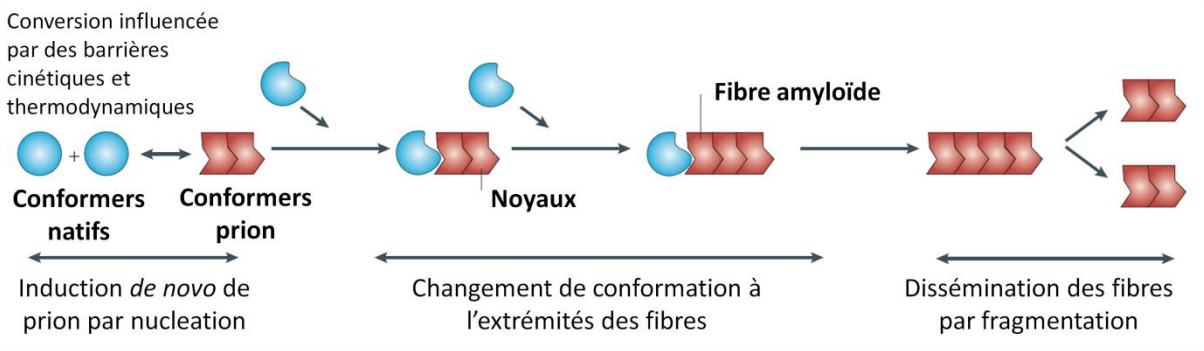


Figure 36 Du prion aux fibres amyloïdes

Un événement de nucléation stabilise une forme prion alternative. Les protéines prions s'agrègent pour former des noyaux qui sont à la base des fibres amyloïdes. La propagation des prions se fait par fragmentation des fibres et transfert aux cellules filles. Figure adaptée de (Shorter and Lindquist 2005)

5% chez *S. cerevisiae*. Etonnamment, la position +7 n'influence pas le taux de translecture (Namy, Hatin and Rousset 2001). Ces observations ont été confirmées par une analyse recherchant des contextes de codons stop favorable pour la translecture dans le génome de *S. cerevisiae* (Williams et al. 2004).

Chez *S. cerevisiae*, 2 gènes cellulaires possèdent des motifs compatibles avec un événement de translecture. Il s'agit des gènes *PDE2* (PhosphoDiEstherase 2) et *IMP3* (Interacting with Mpp10p). *PDE2* code une AMPc-phosphodiestherase à haute affinité pour l'AMPc (Adénosine MonoPhosphate cyclique). Le cas de *PDE2* est intéressant car le contexte nucléotidique du codon stop soumis à translecture est comparable à celui du virus TMV : **CAA UGA CAA GAA**, mais le taux de translecture est significativement plus faible (entre 1 et 2%) (Namy, Duchateau-Nguyen and Rousset 2002, Williams et al. 2004). La translecture du codon stop ajoute une courte séquence de 20 acides aminés qui déstabilise la protéine. Une variation du taux de translecture change la concentration intracellulaire d'AMPc ce qui module les capacités de la cellule à répondre à différents stress (Namy et al. 2002).

- Prévention de la translecture dans les gènes cellulaires

Une étude a analysé pour chaque région 3'UTR chez *S. cerevisiae* la position du prochain codon stop en aval de celui du gène. Comparée à une distribution attendue des codons stops, la distribution observée montre un enrichissement de codons stops aux positions +1, +2 et +3 juste après le stop du gène. Cet enrichissement traduit l'utilisation de garde-fous assurant une terminaison efficace. Dans la moitié des cas, le codon stop suivant est situé dans les 10 codons suivants (Williams et al. 2004).

❖ *Le prion [PSI+]*

Les prions (Proteinaceous Infectious particle (Prusiner 1982) sont des protéines pouvant adopter différentes structures tertiaires. Une des conformations dite prion a des capacités d'auto-agrégation et propage cette conformation aux autres molécules identiques. Les prions forment des particules protéiques infectieuses qui sont transmises aux cellules filles de manière non-mendélienne. Le changement de conformation d'une protéine prion modifie la relation entre le génotype[1] et le phénotype[2] sans qu'il y ait d'altération de la séquence nucléotidique, on parle alors de caractère épigénétique (Shorter and Lindquist 2005).

De façon occasionnelle, des protéines natives solubles adoptent une conformation prion et s'agrègent pour former des « conformers ». Ces conformers servent de bases pour propager la conformation prion aux autres protéines solubles de même séquence. Les nouvelles protéines prions s'ajoutent aux agrégats jusqu'à former des noyaux qui donneront des fibres amyloïdes très stables. La polymérisation des fibres amyloïdes est donc un phénomène d'auto-amplification. La fragmentation de ces fibres permet de disséminer des noyaux de nucléation dans le cytoplasme des cellules filles qui acquièrent alors un phénotype prion (Shorter and Lindquist 2005) (figure 36).

Chez *S. cerevisiae*, le prion [*PSI*$^{+}$] est un caractère phénotypique résultant d'un changement occasionnel de conformation du facteur de terminaison eRF3. Sous leur forme prion, les protéines eRF3 s'agrègent et propagent ce changement de conformation à d'autres protéines eRF3 solubles

[1] Le génotype est l'information portée par la séquence ADN d'une cellule ou d'un individu.

[2] Le Phénotype est l'état d'un caractère observable. L'observation peut être faite à l'échelle morphologique, physiologique ou moléculaire. Le phénotype peut être qualitatif (comme une couleur) ou quantitatif (comme le nombre de doigts). Le caractère observé est écrit entre crochets, par exemple : [*PSI+*].

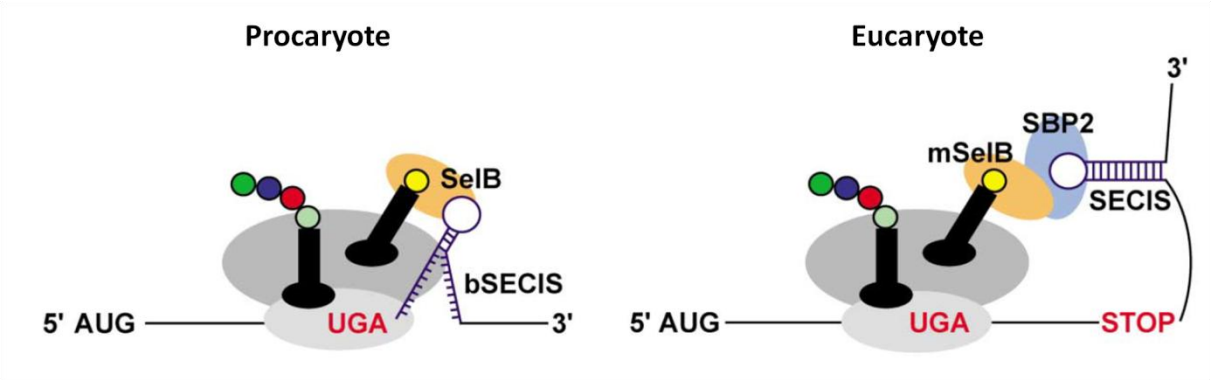


Figure 37 Incorporation de la sélénocystéine

L'incorporation de l'acide aminé sélénocystéine se fait au niveau du codon stop UGA. L'incorporation est rendue possible grâce à la structure secondaire SECIS qui recrute le Sec-ARNt$^{[Ser]Sec}$ via la protéine SelB. Chez les eucaryotes, la protéine SBP2 est aussi impliquée dans le recrutement de Sec-ARNt$^{[Ser]Sec}$ au site A. Figure issue de (Namy et al. 2004)

(King and Diaz-Avalos 2004, Tanaka et al. 2004). La bascule sous forme prion implique la région PrD (Prion determining domain) du domaine N ainsi que le domaine M d'eRF3 (Shorter and Lindquist 2005).

Dans une souche [*PSI*$^+$], la séquestration de protéines eRF3 solubles dans des fibres amyloïdes a pour conséquences de réduire l'efficacité de terminaison et d'augmenter la translecture des codons stops (Paushkin et al. 1996). Selon les souches de *S. cerevisiae* et la force du prion [*PSI*$^+$], la translecture d'un gène rapporteur pour les codons stops UAA, UAG et UGA peut varier respectivement entre 0,1% ; 0,2% ; 0,8% et 1,7% ; 4,5% ; 16% (True and Lindquist 2000). Dans le cas du contexte nucléotidique du virus TMV, les taux de translecture pour les 3 stops passent de 15-25% dans une souche [*psi*$^-$] à 55% dans une souche [*PSI*$^+$] (Cassan and Rousset 2001).

En laboratoire, l'apparition et la perte spontanée du caractère [*PSI*$^+$] est un événement rare qui se produit dans 1 cellule sur 10^5 à 10^7 (Shorter and Lindquist 2005). L'acquisition du caractère [*PSI*$^+$] influence de façon négative ou positive la croissance cellulaire dans différents milieux ainsi que la morphologie des colonies (True and Lindquist 2000). Certains de ces effets sont directement liés à la translecture induite par le prion [*PSI*$^+$] (True, Berlin and Lindquist 2004). Un exemple sera décrit au point V.3.B.

Suite à ces observations, il a été proposé que la translecture induite par le prion [*PSI*$^+$] conférerait un avantage adaptatif aux levures soumises à une pression de sélection occasionnelle. La traduction, au-delà des codons stops, de régions n'ayant pas été soumises à une pression de sélection permettrait à la cellule de tester l'effet de variations génétiques cryptiques. L'accroissement de la diversité phénotypique sans changement du génome serait un moyen d'adaptation rapide et transitoire à un nouvel environnement. Si un des caractères confère un avantage de survie à long terme pour la cellule, il pourrait être fixé de façon stable par une mutation dans le génome (Shorter and Lindquist 2005). Le rôle évolutif du prion [*PSI*$^+$] était très controversé car il n'était présent que dans des souches de laboratoires (Wickner et al. 2007, Halfmann, Alberti and Lindquist 2010). En 2012, il a été montré que ce prion pouvait être trouvé dans des souches sauvages (Halfmann et al. 2012).

❖ *Incorporation de sélénocystéine et pyrrolysine*

L'incorporation des acides aminés rares sélénocystéine et pyrrolysine n'est pas un événement de recodage utilisant la translecture comme mécanisme d'incorporation de ces acides aminés.

➢ Sélénocystéine

La sélénocystéine (Sec ou U) est le 21ème acide aminé naturel du code génétique, elle dérive de la sérine. Elle est incorporée dans les sélénoprotéines qui protègent les cellules contre les stress oxydatifs. Son incorporation se fait par translecture du codon stop opale (UGA) par un ARNt$^{[SER]SEC}$ spécial car la biosynthèse de la sélénocystéine est réalisée sur l'ARNt (Hatfield and Gladyshev 2002). Chez les bactéries, l'incorporation dépend du codon opale, d'une structure tige-boucle appelée SECIS (Sec insertion sequence) située immédiatement en 3' du codon opale, et du facteur SelB (Namy et al. 2004, Gupta and Copeland 2007). Chez les eucaryotes, plusieurs structures SECIS localisées dans le 3'UTR des ARNm peuvent incorporer des sélénocystéines sur plusieurs codons opale dans le même ARNm (Namy et al. 2004). Deux trans-facteurs SBP2 (SECIS Binding Brotein 2) et eEFSec (eukaryotic Elongation Eactor Sec) interviennent aussi dans l'incorporation (Gupta and Copeland 2007) (figure 37).

- Pyrrolysine

La pyrrolysine (Pyl ou O) est le 22ème acide aminé naturel, elle dérive de l'acide aminé lysine. Elle est retrouvée seulement chez les espèces d'archées *Methanosarcina*. Leur production de méthane est dépendante de 3 méthyl-transférases ayant dans leur séquence un acide aminé pyrrolysine (Gaston, Jiang and Krzycki 2011).

L'incorporation de la pyrrolysine lors de la traduction dépend d'un mécanisme de translecture. Les 3 gènes des méthyl-transférases sont interrompus par un codon stop Ambre (UAG). Ce codon est reconnu par un $ARNt^{PYL}$ spécifique à la pyrrolysine et dont l'anticodon est $^{34}CUA^{36}$. Cet ARNt se démarque des autres car il est plus court. Sa structure secondaire est un peu différente, par exemple la tige de l'anticodon est formée de 6 paires de bases au lieu de 5. D'autre part il lui manque quelques unes des modifications conservées universellement dans les autres ARNt. L'$ARNt^{PYL}$ est chargé par une lysyl-ARNt synthétase (Gaston et al. 2011).

L'incorporation de la pyrrolysine est très efficace car très peu de protéines terminant au site d'incorporation sont détectées. Une telle efficacité de translecture pourrait être expliquée par la présence d'une tige-boucle PYLIS (pyrolysine insertion elements) en aval du codon stop (Namy et al. 2004, Gaston et al. 2011).

2. Le décalage du cadre de lecture en -1.

Le décalage du cadre de lecture en -1 (« frameshift -1 » en anglais) est un événement durant lequel le ribosome et les ARNt se décalent d'une base vers le 5' de l'ARNm. La conséquence de ce décalage est que le ribosome termine la traduction dans une phase de lecture (phase -1) différente de celle dans laquelle il a commencé (phase 0). Le décalage en -1 a été retrouvé essentiellement dans des virus et éléments transposables. Cependant, quelques gènes cellulaires ont été identifiés comme soumis à un décalage en -1 chez les bactéries (Tsuchihashi and Kornberg 1990, Blinkowa and Walker 1990, Flower and McHenry 1990) et les mammifères Edr (Manktelow, Shigemoto and Brierley 2005) Ma3 (Wills et al. 2006).

A. Les signaux de décalage du cadre de lecture en -1.

Deux signaux au minimum sont nécessaires dans l'ARNm pour avoir une efficacité de décalage efficace : une séquence glissante et une structure secondaire en 3'.

- ***La séquence glissante***

La séquence glissante est la région sur laquelle le ribosome va pauser et se décaler d'un nucléotide en 5' (Somogyi et al. 1993). C'est un heptamère de séquence consensus **X XXY YYZ** (les codons sont dans la phase 0), dont les bases X et Y peuvent être identiques. Le codon XXY est au site P du ribosome et le codon YYZ au site A. Suite à un décalage en -1, les ARNt aux sites P et A rompent leurs interactions avec l'ARNm et se réapparient respectivement sur les séquences XXX et YYY. Une mutagénèse dirigée de cette séquence à montré que le codon YYY est celui qui affecte le plus l'efficacité de décalage -1 ce qui implique un rôle prépondérant de l'ARNt au site A (Jacks et al. 1988a). Seule l'interaction avec la base wobble change suite au décalage.

Le site de glissement seul suffit induire un décalage en -1 du ribosome. Le taux de décalage passe de 0.005% (Kurland 1992) à environ 0.5% par codon en fonction de la séquence considérée (Stahl, McCarty

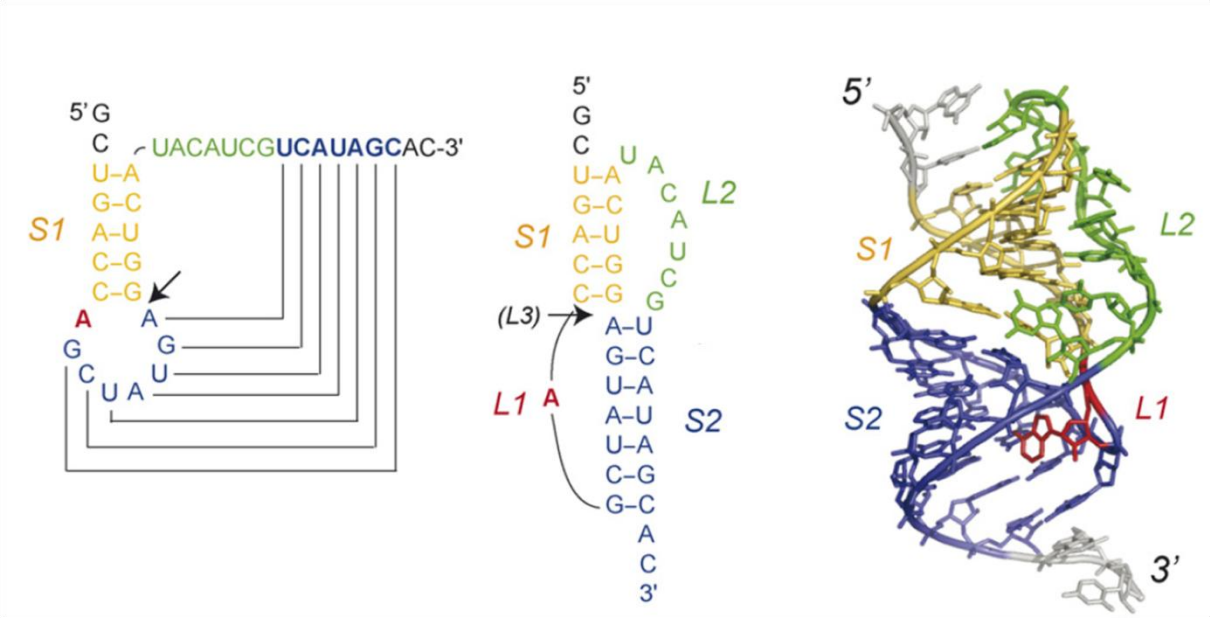


Figure 38 Structure secondaire et tertiaire d'un pseudonoeud

Dans un pseudonoeud, la boucle de la tige S1 (jaune) s'apparie avec une séquence en 3' (bleu) afin de former la tige S2 (structure secondaire à gauche et au centre). Dans la structure tertiaire (à droite), les 2 tiges s'empilent. Figure issue de (Giedroc and Cornish 2009)

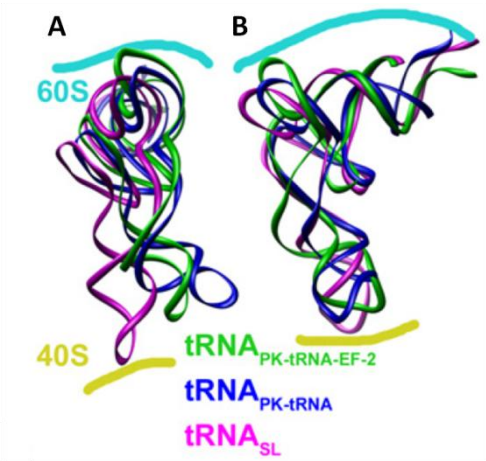


Figure 39 Position de l'ARNt au site A

Dans une population de ribosomes pausant à cause d'un pseudonoeud ($tRNA_{PK}$) provenant du site de décalage en -1 du virus IBV, l'ARNt au site A adopte une conformation hybride de type A/P. Contrairement à l'état A/P normal, l'ARNt est plaqué contre le plafond du site A. Selon la présence ou l'absence du facteur eEF2, l'inclinaison de l'ARNt varie et adopte soit l'état A/P' ou A/P''. **A)** et **B)** Comparaison de l'ARNt control ($tRNA_{SL}$) et les $ARNt_{PK}$ au site A. En absence d'eEF2, l'ARNt (bleu) est dans l'état A/P'. En présence d'eEF2, l'ARNt (vert) est dans l'état A/P''. Figure issue de (Flanagan et al. 2010)

and Farabaugh 2002). Cette efficacité de décalage est stimulée par la présence d'une structure secondaire en aval.

❖ ***La structure secondaire***

Le second élément est une structure secondaire stable située en aval de la séquence glissante (Brierley 1995). La distance entre la séquence glissante et la base de la structure secondaire est essentielle pour l'efficacité du décalage en -1 (Brierley, Digard and Inglis 1989). La distance optimale varie entre 6 et 8 nucléotides (Kollmus et al. 1994). D'après l'étude de l'activité hélicase du ribosome, cette distance placerait le début de la structure secondaire au site hélicase dans le tunnel d'entrée de l'ARNm (Giedroc, Theimer and Nixon 2000, Plant et al. 2003, Takyar et al. 2005).

La présence de cette structure secondaire augmente de 10 à 30 fois l'efficacité de décalage de la séquence glissante (Brierley 1995). Dans la plus part des cas étudiés en détail, la structure secondaire est un pseudonœud de type H (Hairpin). Ces pseudonoeuds sont composés d'une structure tige-boucle dans laquelle les bases de la boucle vont s'apparier à une séquence en 3' de la structure (figure 38). Il en résulte un pseudonoeud composé de 2 tiges et 2 boucles de tailles différentes. Selon les pseudonoeuds, les boucles peuvent se loger dans le sillon des hélices formées par les tiges. Il y a alors formation d'une triple hélice d'ARN (Giedroc et al. 2000) ce qui contribue à la stabilité du pseudonoeud (Chen et al. 2009). La forme de la structure tertiaire et la stabilité du pseudonoeud participent à l'efficacité de décalage en -1 (Somogyi et al. 1993, Staple and Butcher 2005).

La structure secondaire induit une pause du ribosome sur la séquence glissante (Somogyi et al. 1993, Lopinski, Dinman and Bruenn 2000) mais cette pause n'est pas suffisante en elle même pour induire un décalage en -1 sur n'importe quelle séquence (Tu, Tzeng and Bruenn 1992, Kontos, Napthine and Brierley 2001). De plus, tous les pseudonoeuds ne stimulent par le décalage en -1 sur la séquence glissante.

B. Un mécanisme encore à préciser.

Au fil des décennies, des modèles ont été proposés pour expliquer le mécanisme de décalage des ARNt en -1. Le modèle initial proposait un décalage simultané des 2 ARNt lors de l'accommodation de l'ARNt acétylé au site A (Jacks et al. 1988a, Plant et al. 2003). A présent, les derniers modèles proposent que le décalage en -1 se produise lors de la translocation du ribosome (Weiss et al. 1989, Namy et al. 2006, Flanagan et al. 2010). La structure secondaire servirait à abaisser la barrière énergétique en permettant un désappariement et rappariement des ARNt sur l'ARNm. Il existe une corrélation entre l'efficacité de décalage induite par un pseudonoeud et la force à appliquer pour ouvrir le pseudonoeud (Hansen et al. 2007).

Une structure en cryo-microscopie électronique d'un ribosome eucaryote pausant grâce au pseudonoeud issu du virus IBV montre que le pseudonoeud est encore largement structuré et au niveau de l'ouverture du tunnel d'entrée de l'ARNm où il contacte les protéines hélicases (Namy et al. 2006, Flanagan et al. 2010).

Les ribosomes sont dans un état « encliqueté » avec un ARNt au site A dans l'état hybride A/P. Cette conformation de l'ARNt laisse penser que la réaction de transpeptidation a eu lieu, mais la résolution de la structure n'est pas assez fine pour l'affirmer. En comparaison avec les autres états A/P observés

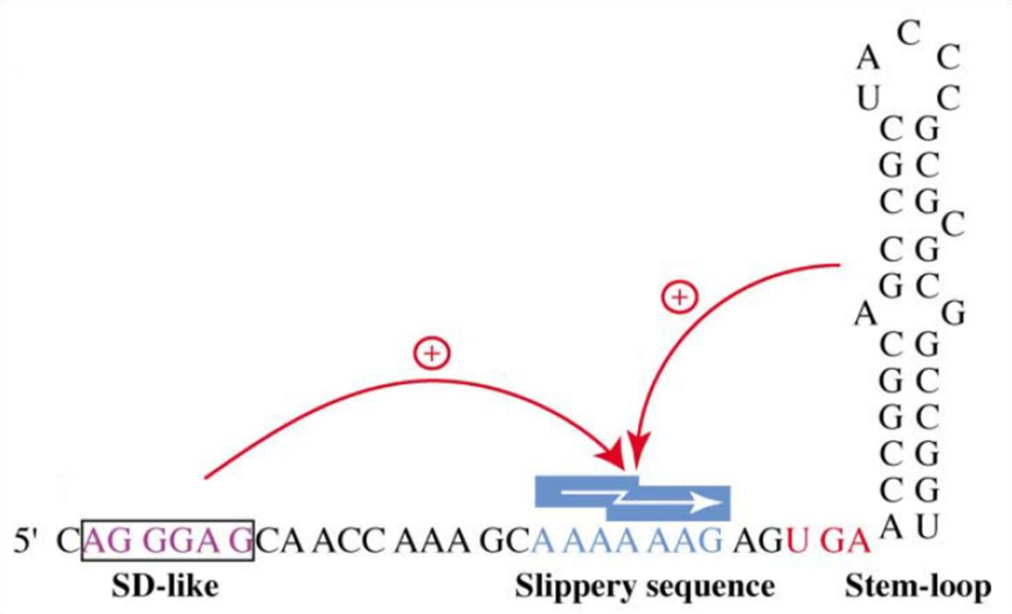


Figure 40 Site de recodage du gène dnaX

Le site de recodage du gène dnaX est constitué d'un site de glissement et d'une tige-boucle en aval. Une région comparable à la séquence Shine-Dalgarno (SD) stimule aussi le décalage en -1 en amont de la séquence de glissement. Figure issue de (Namy et al. 2004)

dans d'autres ribosomes, l'ARNt n'est pas en contact avec l'ARNm. Il est contre le plafond du site P de la sous-unité 60S (figure 39). Cet état est noté A/P''. Dans certains des complexes ribosomiques, le facteur eEF2 est présent au site A, l'ARNt est alors moins incliné mais toujours aussi haut dans le site P. Cette conformation est notée A/P' (Flanagan et al. 2010). Ces 2 populations de ribosomes varient aussi pour la position du pseudonoeud par rapport au tunnel d'entrée de l'ARNm. Dans le cas A/P' où eEF2 est présent, le pseudonoeud est orthogonal au tunnel alors que dans le cas A/P'' le pseudonoeud est incliné comme relâché. Ces différentes caractéristiques ne sont pas observées lorsque le ribosome pause à une tige-boucle de stabilité équivalente (V).

Il est difficile d'établir des liens de cause à effet entre l'orientation du pseudonoeud, la présence d'eEF2, l'inclinaison de l'ARNt et le devenir d'un décalage en -1 de ce dernier. Cependant ces structures supportent l'hypothèse que le décalage se produit lors de la translocation et implique l'ARNt au site A. Deux forces opposées s'exerceraient sur l'ARNm. La première entraîne l'ARNm dans le sens de la translocation alors que la résistance à l'ouverture du pseudonoeud s'opposerait à ce mouvement. La tension appliquée à l'ARNm figerait le système et entraverait la translocation des ARNt, piégeant un des ARNt dans une conformation déformée. Une solution pour résoudre cette situation serait de rompre les liaisons codons:anticodons. L'ARNt reprendrait alors sa conformation standard et se rapparierait dans la phase -1.

- ***Tension dans l'ARNr 16S : le cas dnaX.***

Chez les bactéries, le gène dnaX code les sous-unités γ et τ de l'ADN polymérase III. La synthèse de la sous-unité γ résulte d'un décalage du cadre de lecture en -1 qui amène le ribosome à rencontrer un codon stop prématuré. Avec un taux de décalage de 50%, cette séquence est la plus efficace des séquences décrites à ce jour (Tsuchihashi and Kornberg 1990, Blinkowa and Walker 1990, Flower and McHenry 1990). Cette efficacité s'explique par la convergence de 3 éléments favorables : une séquence glissante favorable entourée de 2 éléments qui vont coopérer pour stimuler le décalage en -1 (Larsen et al. 1994).

Le décalage se produit sur la séquence glissante **A AAA AAG**. L'ARNt^{LYS} a une affinité plus forte pour le codon AAA que le codon AAG ce qui favorise le décalage. A 5 nucléotides en 3' de la séquence glissante, il y a une structure secondaire qui, dans ce cas, est une tige-boucle. Son action est comparable à celle des autres structures secondaires précédemment décrites. Enfin, à 10 nucléotides en 5' de la séquence glissante, il y a une séquence de type Shine-Dalgarno (SD) qui est impliquée dans l'initiation traductionnelle chez les bactéries (Larsen et al. 1994) (figure 40).

L'initiation de la traduction chez les bactéries est différente de celle des eucaryotes. La petite sous-unité du ribosome se fixe sur une séquence d'ARNm appelée Shine-Dalgarno via une interaction avec l'ARNr 16S. Cette interaction permet de positionner le ribosome sur le codon d'initiation.

Dans dnaX, la séquence de type SD est plus éloignée du site P que lors de l'initiation de la traduction. En interagissant avec l'ARNr 16S, cette séquence de type SD va induire des tensions dans l'ARNr 16S et promouvoir un décalage du ribosome en 5' (Larsen et al. 1994). Ce cas illustre l'importance des contraintes imposées au ribosome lors du décalage en -1.

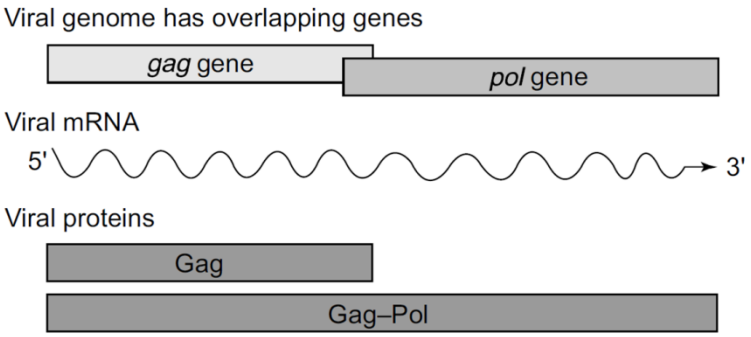


Figure 41 Synthèse des protéines virales Gag et Gag-Pol par recodage

Chez les virus, le site de recodage permet de réguler la production de la protéine d'enveloppe Gag et de la polymérase Gag-Pol. Figure issue de (Dinman, Ruiz-Echevarria and Peltz 1998)

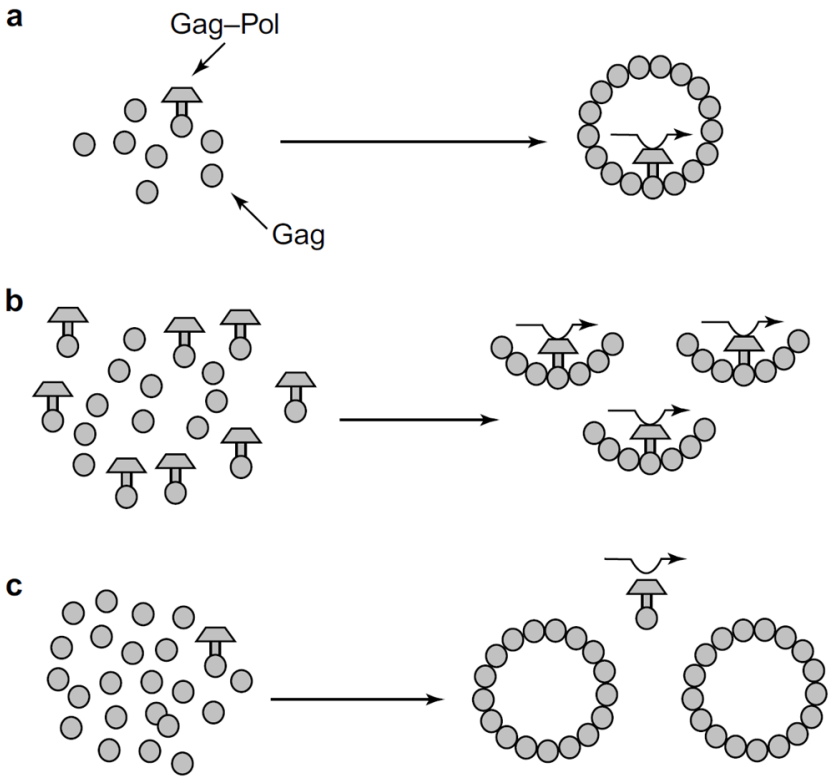


Figure 42 Rôle du recodage dans la production des particules virales

A) Le taux de recodage contrôlant la synthèse des protéines Gag et Gag-Pol permet une production optimale de particules virales. **B)** Augmenter le taux de recodage conduit à un encombrement stérique par les protéines Gag-Pol qui empêche la formation de particules virales. C) Diminuer le taux de recodage conduit à la formation de particules virales dépourvues de protéine Gag-Pol et de génome. Figure issue de (Dinman et al. 1998)

C. Régulation physiologique : exemple avec le virus du VIH

Le rétrovirus VIH-1 (Virus de l'Immunodéficience Humaine de type 1) régule l'expression de protéines Gag et Gag-Pol par un événement de décalage du cadre de lecture en -1. L'efficacité de décalage varie entre 5 et 10% (Jacks et al. 1988b). Sa séquence glissante **U UUU UUA** n'est pas suivi d'un pseudonoeud mais une tige-boucle en 2 parties (Dulude, Baril and Brakier-Gingras 2002, Gaudin et al. 2005, Staple and Butcher 2005). Cet élément est suffisant pour stimuler le décalage -1 du ribosome sur une séquence **U UUU UUA** particulièrement glissante (Brierley, Jenner and Inglis 1992).

Comme pour les autres virus, l'efficacité de décalage dicte le ratio de protéines structurales Gag et enzymatiques Gag-Pol disponible pour l'assemblage d'une particule virale (figure 41). Augmenter ce ratio conduit à la formation de particules virales incomplètes alors que le diminuer conduit à produire des particules sans génome ARN faute de réplicases Pol (figure 42). Une mutagénèse de la séquence glissante permettant de faire varier le taux de décalage en -1 a montré que le changement de ratio Gag:Gag-Pol affecte l'infectivité et la propagation du virus (Biswas et al. 2004).

3. Le décalage du cadre de lecture en +1

Le décalage du cadre de lecture en +1 (« frameshift +1 » en anglais) est un événement durant lequel le ribosome se décale d'une base vers le 3' de l'ARNm. Là encore, le ribosome termine la traduction dans une phase de lecture différente (phase +1) de celle dans laquelle il a commencé (phase 0).

A. Le rôle de l'ARNt au site P.

Le décalage en +1 se produit sur une séquence minimale tétramérique **XXX X**. Dans la plupart des cas, le codon adjacent stimule le décalage et la séquence forme alors un heptamère **XXX XXX X** (phase 0) dont les codons sont situés dans les sites A et P du ribosome. Contrairement au décalage en -1, il n'y a pas de séquence consensus (Sundararajan et al. 1999). Deux mécanismes peuvent expliquer un événement de décalage du cadre de lectures en +1 : soit l'ARNt au site P se décale d'un nucléotide en 3', soit l'ARNt rentrant au site A s'apparie directement dans la phase +1. La différence réside dans l'affinité de l'ARNt au site P pour son codon et dans sa capacité à glisser. Ce point sera illustré avec le gène *EST3* et le rétrotransposon Ty3.

Chez *S. cerevisiae*, le décalage en +1 nécessite qu'un ARNt proche-cognat soit présent au site P formant une interaction wobble faible pyrimidine:pyrimidine ou purine:purine (Sundararajan et al. 1999). Suite à cette étude, l'hypothèse est que l'interaction au site P perturberait la sélection de l'ARNt au site A, même dans le cas d'un ARNt cognat. Cette interférence semble éliminer la préférence pour l'incorporation d'un ARNt cognat en phase, au profit de l'acceptation d'un ARNt cognat en phase +1 (Stahl et al. 2002). La capacité de l'ARNt au site P de glisser + 1 peut augmenter l'efficacité de l'événement. Cette habilité peut se traduire par une stabilité plus grande entre l'appariement codon:anticodon dans la phase +1 (Belcourt and Farabaugh 1990, Sundararajan et al. 1999).

D'un point de vue enzymatique, l'efficacité de décalage est principalement influencée par la concentration de l'ARNt correspondant au codon du site A en phase 0 [A(0)], et par l'affinité de l'ARNt au site P pour le codon en phase +1. Une forte concentration d'ARNt pour le site A(0) favorise un décodage fidèle alors qu'une faible concentration favorise le décalage en +1. Une forte affinité de

l'ARNt au site P pour le codon +1 favorise un décalage en +1 alors qu'une faible affinité favorise un décodage fidèle du codon au site A (Baranov, Gesteland and Atkins 2004).

B. Les éléments stimulateurs

Pour un décalage en +1 efficace, la reconnaissance du codon au site A doit être lente afin d'induire une pause du ribosome. Cette pause peut être induite par un codon rare, un codon stop dans un contexte peu efficace ou une structure secondaire de type tige-boucle ou pseudonoeud.

❖ *Le codon rare : exemple EST3 et Ty3.*

Le gène *EST3* et le retrotransposon Ty3 sont tous deux trouvés chez *S. cerevisiae*. Dans ces 2 cas, le décalage en +1 est stimulé par la présence du codon **AGU U** au site A. Ce codon dit rare est décodé par l'ARNt^{SER} proche-cognat dont l'anticodon est GCU. Le décodage de ce codon rare induit une pause du ribosome.

EST3 (Ever Shorter Telomers 3) code une des sous-unités du complexe télomérase impliqué dans la maintenance des télomères *in vivo*. Le site de glissement est à la fin de la première phase ouverte de lecture. La traduction de la 1ère phase conduit à une protéine tronquée vraisemblablement dégradée. Suite au décalage en +1, la traduction se poursuit permettant la synthèse de la protéine EST3p fonctionnelle. Le décalage en +1 se fait avec une efficacité de 8% sur la séquence glissante **CUU AGU UGA**. Les codons sont dans la phase 0, la séquence glissante est en gras et le codon stop est souligné (Vimaladithan and Farabaugh 1994). Dans certains cas, la pause est assez longue pour laisser le temps à l'ARNt^{LEU} au site P (CUU) de se décaler sur le codon UUA. (Belcourt and Farabaugh 1990, Sundararajan et al. 1999, Morris and Lundblad 1997). Il semble qu'*in vivo*, l'efficacité de décalage en +1 soit plus importante que les 8% mesuré avec la séquence glissante (Morris and Lundblad 1997). Une mutagénèse des 30 nucléotides en aval de la séquence glissante a montré qu'il existait une région stimulatrice large et complexe dans cette zone. Son fonctionnement reste encore énigmatique (Taliaferro and Farabaugh 2007).

Ty3 est un élément transposable qui, à l'image des rétrovirus, utilise un événement de décalage en +1 pour réguler la production des protéines Gag3 et Gag3-Pol3. Le décalage en +1 se fait avec une efficacité de 11% (Farabaugh, Zhao and Vimaladithan 1993) sur la séquence glissante **GCG AGU U**. Dans ce cas, l'ARNt^{ALA} au site P ne glisse pas sur le codon CGA, c'est un ARNt^{VAL} entrant au site A qui s'apparie directement sur le codon GUU dans la phase +1 (Farabaugh et al. 1993). Comme pour *EST3*, une région de 14 nucléotides en aval du site de glissement de Ty3 stimule le décalage en +1 (Farabaugh et al. 1993). Dans les deux cas, il semblerait que cette stimulation passe par une interaction avec le tunnel d'entrée de l'ARNm (Guarraia et al. 2007).

❖ *Le stop « glissant » et la structure secondaire : ex* l'antizyme ornithine décarboxylase*.*

Chez *S. cerevisiae*, le gène *OAZ1* (Ornithine decarboxylase AntiZyme 1) code l'antizyme de l'ornithine décarboxylase OAZ1p. La traduction d'*OAZ1* nécessite un événement de décalage en +1 dépendant du codon stop de la première phase de lecture. Le site de glissement est **GCG UGA C**, le codon stop est souligné (Palanimurugan et al. 2004). GCG est un codon qui stimule le décalage en +1 (Vimaladithan and Farabaugh 1994) quant au contexte du codon stop il n'est pas favorable pour une terminaison efficace (Salas-Marco and Bedwell 2004). L'efficacité de glissement est en partie régulée par la disponibilité en

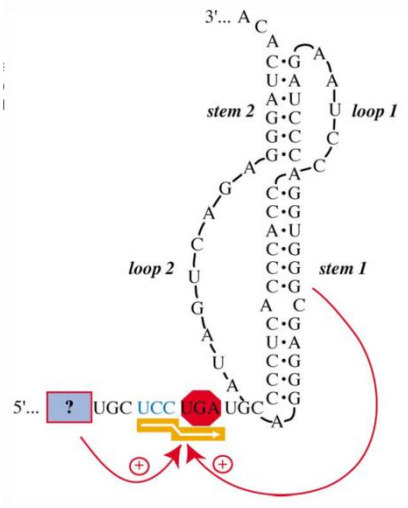


Figure 43 Site de recodage de l'antizyme AZ1 chez les mammifères

Le site de décalage en +1 du gène AZ1 est constitué d'une séquence glissante avec un pseudonoeud en aval. Une région en 5' du site de glissement stimule aussi le décalage. Figure issue de (Namy et al. 2004)

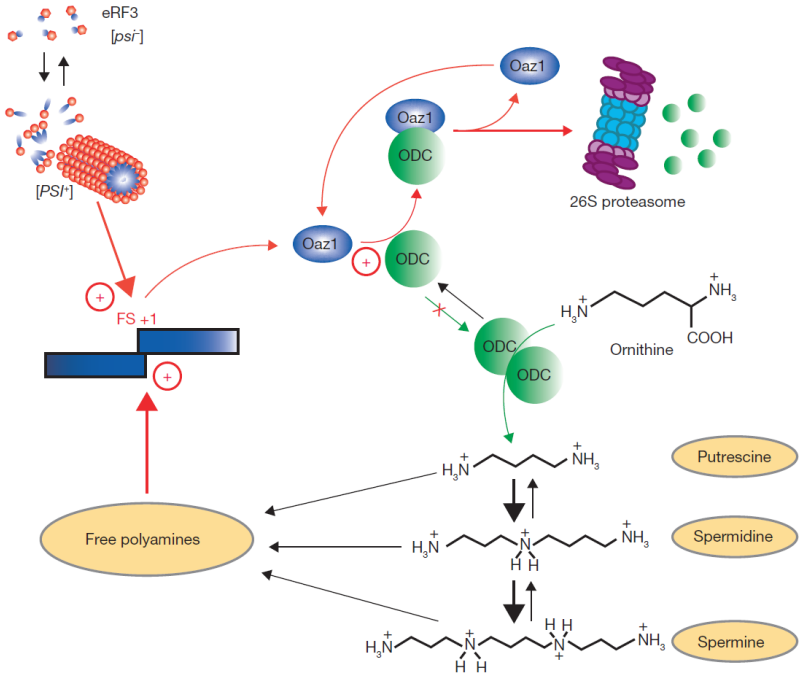


Figure 44 Régulation du taux de polyamines intracellulaires par le recodage de l'antizyme

Schéma illustrant la boucle d'autorégulation contrôlant la concentration intracellulaire en polyamines. Une forte concentration en polyamines stimule le décalage en +1 du gène de l'antizyme. La protéine Oaz1p stimule la dégradation de l'enzyme ODC et la synthèse de polyamine est alors inhibée. La baisse de la concentration en polyamines inhibe la synthèse alors d'Oaz1p. Le décalage en +1 est stimulé par un codon stop. En présence du prion [PSI$^+$], eRF3 est moins disponible ce qui stimule le décalage en +1 du gène de l'antizyme. Figure issue de (Namy et al. 2008)

facteurs de terminaison. Dans une souche [*PSI*$^{+}$], le facteur eRF3 est en quantité militante ce qui augmente l'efficacité de décalage en +1 et donc la traduction de OAZ1p (Namy et al. 2008).

Chez les mammifères, la synthèse de l'antizyme est aussi régulée par un événement de décalage en +1 stimulé par un stop (Matsufuji et al. 1995). De plus, des éléments de part et d'autre du site de glissement **UCC UGA U** stimulent le décalage en +1. Il s'agit d'une séquence conservée **CCG GGG CCU CGG** (Ivanov et al. 2000) 6 nucléotides en amont de la séquence glissante et d'un pseudonoeud à 3 nucléotides en aval (Matsufuji et al. 1995) (figure 43).

Le cas de l'antizyme est intéressant car la régulation du taux de recodage se fait par une boucle d'autorégulation. L'antizyme se fixe à l'ornithine décarboxylase (ODC) qui est alors adressée au protéasome pour être dégradée (Murakami et al. 1992). L'ODC est une enzyme qui catalyse la synthèse des polyamines (spermine, spermidine, putrescine). Cette étape est le point limitant dans la synthèse.

Les polyamines régulent l'efficacité de recodage de l'antizyme. En cas de fortes concentrations en polyamines, la traduction d'antizyme est stimulée. D'avantage de protéines ODC sont alors dégradées et la production de polyamines diminue. A l'inverse, si la concentration en polyamines diminue trop, l'efficacité de décalage en +1 décroît ce qui entraîne une diminution de la synthèse d'antizyme. Comme l'ODC est moins dégradée, la biosynthèse de polyamines n'est plus inhibée (Matsufuji et al. 1995) (figure 44).

Le lien entre la concentration intracellulaire en polyamines et l'efficacité de décalage en +1 de l'antizyme est indirect. La concentration en polyamines est perçue par les peptides naissants des antizymes en cours de traduction. Quand la concentration est trop faible, les peptides en cours d'élongation inhibent leurs synthèses en bloquant l'élongation de leurs ribosomes (Kurian et al. 2011).

C. Abondance de décalages en +1 chez les Euplotes sur des stops « glissants ».

Chez les Euplotes, qui sont des eucaryotes unicellulaires ciliés, une fraction significative des gènes utilise un événement de décalage en +1 pour synthétiser leurs protéines (Klobutcher and Farabaugh 2002). Un séquençage partiel du génome d'*Euplotes crassus* a dénombré 13% de gènes comme soumis à un décalage en +1. Une extrapolation au génome entier donne comme fourchette que 3,7 à 31,7% des gènes y serait soumis (Klobutcher 2005). Même en considérant la borne inférieure de cet intervalle, la valeur reste largement supérieure à celle observée pour les autres génomes. Chez *S. cerevisiae*, seuls 3 gènes cellulaires (ABP140, EST3 et OAZ1) parmi ~ 6000 ont été identifiés comme soumis à un décalage en +1.

Le décalage se produit sur les séquences **AAA UAA A** et parfois **AAA UAG A** avec un codon stop comme élément stimulateur (Klobutcher 2005). L'ARNt^{LYS} au site P glisserait sur le codon AAU et la traduction se poursuivrait dans la phase +1. La conservation du codon AAA implique une importance fonctionnelle. Il reste à déterminer si c'est le codon, l'ARNt ou l'acide aminé Lys qui stimule le décalage en +1 (Vallabhaneni et al. 2009).

Chez les Euplotes le codon UGA n'est pas reconnu comme un stop, il correspond à l'acide aminé cystéine (Meyer et al. 1991). Le facteur eRF1 des Euplotes est similaire à celui des autres eucaryotes. La

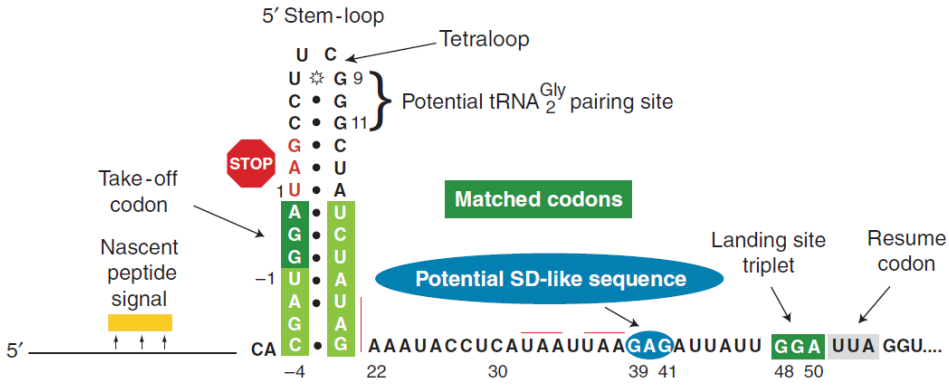


Figure 45 Site de recodage du gène 60 du bactériophage T4

Le site de recodage du gène 60 est composé d'un site de « décollage » (take-off) inclus dans une tige-boucle. Le « décollage » est stimulé par le codon stop au site A et le peptide naissant dans le tunnel de sortie du ribosome. Le site « d'atterrissage » en 3' est situé en aval d'une séquence de type Shine-Dalgarno (SD). Figure issue de (Wills et al. 2008)

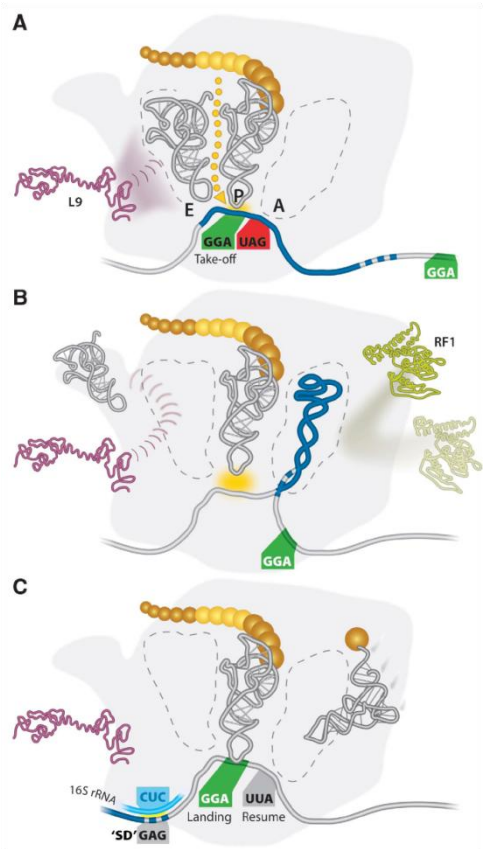


Figure 46 Schéma proposant un mécanisme moléculaire pour le « saut » de ribosome

A) Suite à l'entrée du codon stop au site A, le peptide naissant et l'absence du facteur de terminaison RF1 conduit à un décrochage de l'ARNt peptidyl au site P au niveau du codon de « décollage ». **B)** La tige-boucle aurait alors la possibilité de se reformer au site A et de leurrer le ribosome sur la présence d'un ARNt. **C)** Suite à un cycle de translocation, l'ARNt peptidyl se rapparie au codon « d'atterrissage ». Le ribosome continue ensuite la traduction jusqu'au prochain codon stop. Figure issue de (Wills et al. 2008)

présence d'une Alanine au lieu d'une sérine en position 70 (6 nucléotides en aval du NIKS) abolit la reconnaissance du codon UGA mais permet de garder une spécificité pour les codons UAA et UAG (Eliseev et al. 2011). La reconnaissance des codons stop par cet eRF1 bipotent est globalement moins efficace qu'avec un facteur eRF1 omnipotent ce qui favorise le décalage en +1 sur des séquences terminatrices faibles (Vallabhaneni et al. 2009).

4. Saut de ribosome ou « By-passing »

Le saut de ribosome est un événement très rare qui a été décrit initialement dans le gène 60 du bactériophage T4. Ce gène code une topoisomérase essentielle pour ce phage infectant la bactérie *Escherichia coli*. Contrairement à ce que laisse entendre le nom français de l'événement, le ribosome ne se décroche pas de l'ARNm. Durant l'élongation, des éléments en *cis* et *trans* de l'ARNm vont permettre au ribosome de suspendre l'élongation sur un codon de « décollage » GGA et de la reprendre 50 nucléotides en aval sur un codon d' « atterrissage » GGA. L'efficacité de l'événement est de 50 % (Herr, Gesteland and Atkins 2000b) (figure 45).

Lors du décollage, le codon GGA au site P est apparié avec l'ARNt peptidyl. Un codon stop UGA est présent de le site A mais n'est pas reconnu par RF1 le facteur de terminaison de classe I bactérien pour le codon UAG (Herr et al. 2000b). La séquence **GGA UAG** est comprise dans une tige boucle qui stimule le décollage, ce sont les éléments en *cis* (Weiss, Huang and Dunn 1990). L'élément en *trans* est une séquence peptidique de la protéine présente dans le tunnel de sortie du ribosome. Cette séquence est composée d'acides aminés hydrophiles et chargés (Weiss et al. 1990) qui vont favoriser la dissociation codon:anticodon. Cet élément augmente d'un facteur 9 l'efficacité de décollage. L'effet du peptide naissant est maintenu lorsque le complexe ribosomique scanne l'ARNm ce qui permet une sélection fidèle du site d'atterrissage (Herr, Atkins and Gesteland 2000a, Herr et al. 2000b, Herr et al. 2004). L'ARNt peptidyl au site P n'intervient pas directement dans le balayage de l'ARNm (Wills et al. 2008).

Le modèle actuel serait qu'en entrant au site A le codon stop induirait une pause du ribosome. Le peptide naissant déstabiliserait assez le ribosome et l'interaction codon:anticodon au site P pour permettre à la structure secondaire de se reformer dans le site A. Cela implique que les extrémités 5' et 3' de l'ARNm soient tirées à l'intérieur du ribosome. Une fois la structure formée, le ribosome « sentirait » un site A plein et engagerait un cycle de translocation permettant une reprise de la traduction au codon d'atterrissage. Cette reprise serait facilitée par la présence une région de type Shine-Dalgarno (Wills et al. 2008) (figure 46).

VI. Travaux de thèse

Au cours de ma thèse dans l'équipe Génomique Structure et Traduction (GST), J'ai développé 3 projets de recherche. Une introduction plus spécifique sera apportée pour chaque partie avant d'en présenter les résultats et de les discuter.

Les deux premiers chapitres porteront sur l'analyse de l'étape d'élongation à l'échelle moléculaire.

Le 1ère projet est une étude sur le rôle du facteur d'élongation eEF2 dans le maintien du cadre de lecture lors de l'élongation.

Le 2nd projet porte sur la mise au point d'un système de traduction eucaryote permettant l'étude du comportement de ribosomes individuels en cours d'élongation. Ce projet a été réalisé en étroite collaboration avec Nicolas Fiszman, Antoine Le Gall et Karen Perronet dans l'équipe de Nathalie Westbrook (Laboratoire Charles Fabry UMR8501, Institut d'Optique).

Le dernier projet est une recherche de nouveaux gènes cellulaires soumis au recodage chez *S. cerevisiae*. Le but étant de mieux comprendre l'influence du recodage dans la régulation de la physiologique cellulaire. Ce projet a été réalisé en collaboration avec Christine Drevet[1] et Alain Denise[123] ([1] Institut de Génétique et Microbiologie UMR8621, [2] Laboratoire de Recherche en Informatique UMR8623, [3] Institut National de Recherche en Informatique et en Automatique).

Références bibliographiques

Agirrezabala, X., J. Lei, J. L. Brunelle, R. F. Ortiz-Meoz, R. Green & J. Frank (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol Cell,* 32**,** 190-7.

Agris, P. F. (2004) Decoding the genome: a modified view. *Nucleic Acids Res,* 32**,** 223-38.

--- (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep,* 9**,** 629-35.

Alkalaeva, E. Z., A. V. Pisarev, L. Y. Frolova, L. L. Kisselev & T. V. Pestova (2006) In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell,* 125**,** 1125-36.

Anand, M., B. Balar, R. Ulloque, S. R. Gross & T. G. Kinzy (2006) Domain and nucleotide dependence of the interaction between Saccharomyces cerevisiae translation elongation factors 3 and 1A. *J Biol Chem,* 281**,** 32318-26.

Andersen, C. B., T. Becker, M. Blau, M. Anand, M. Halic, B. Balar, T. Mielke, T. Boesen, J. S. Pedersen, C. M. Spahn, T. G. Kinzy, G. R. Andersen & R. Beckmann (2006) Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature,* 443**,** 663-8.

Armache, J. P., A. Jarasch, A. M. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. N. Wilson & R. Beckmann (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Proc Natl Acad Sci U S A,* 107**,** 19748-53.

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

Ban, N., B. Freeborn, P. Nissen, P. Penczek, R. A. Grassucci, R. Sweet, J. Frank, P. B. Moore & T. A. Steitz (1998) A 9 A resolution X-ray crystallographic map of the large ribosomal subunit. *Cell,* 93**,** 1105-15.

Ban, N., P. Nissen, J. Hansen, M. Capel, P. B. Moore & T. A. Steitz (1999) Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature,* 400**,** 841-7.

Baranov, P. V., R. F. Gesteland & J. F. Atkins (2002) Recoding: translational bifurcations in gene expression. *Gene,* 286**,** 187-201.

--- (2004) P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA,* 10**,** 221-30.

Barthelme, D., S. Dinkelaker, S. V. Albers, P. Londei, U. Ermler & R. Tampé (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**,** 3228-33.

Becker, T., J. P. Armache, A. Jarasch, A. M. Anger, E. Villa, H. Sieber, B. A. Motaal, T. Mielke, O. Berninghausen & R. Beckmann (2011) Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol,* 18**,** 715-20.

Becker, T., S. Franckenberg, S. Wickles, C. J. Shoemaker, A. M. Anger, J. P. Armache, H. Sieber, C. Ungewickell, O. Berninghausen, I. Daberkow, A. Karcher, M. Thomm, K. P. Hopfner, R. Green & R. Beckmann (2012) Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature,* 482**,** 501-6.

Bekaert, M., A. E. Firth, Y. Zhang, V. N. Gladyshev, J. F. Atkins & P. V. Baranov (2010) Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res,* 38**,** D69-74.

Belcourt, M. F. & P. J. Farabaugh (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell,* 62**,** 339-52.

Ben-Shem, A., N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova & M. Yusupov (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science,* 334**,** 1524-9.

Bengtson, M. H. & C. A. Joazeiro (2010) Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature,* 467**,** 470-3.

Beringer, M. (2008) Modulating the activity of the peptidyl transferase center of the ribosome. *RNA,* 14**,** 795-801.

Bertram, G., H. A. Bell, D. W. Ritchie, G. Fullerton & I. Stansfield (2000) Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA,* 6**,** 1236-47.

Biswas, P., X. Jiang, A. L. Pacchia, J. P. Dougherty & S. W. Peltz (2004) The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol,* 78**,** 2082-7.

Blanchard, S. C., R. L. Gonzalez, H. D. Kim, S. Chu & J. D. Puglisi (2004) tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol,* 11**,** 1008-14.

Blinkowa, A. L. & J. R. Walker (1990) Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res,* 18**,** 1725-9.

Bonetti, B., L. Fu, J. Moon & D. M. Bedwell (1995) The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. *J Mol Biol,* 251**,** 334-45.

Bouadloun, F., D. Donner & C. G. Kurland (1983) Codon-specific missense errors in vivo. *EMBO J,* 2**,** 1351-6.

Bretscher, M. S. (1968) Translocation in protein synthesis: a hybrid structure model. *Nature,* 218**,** 675-7.

Brierley, I. (1995) Ribosomal frameshifting viral RNAs. *J Gen Virol,* 76 (Pt 8)**,** 1885-92.

Brierley, I., P. Digard & S. C. Inglis (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell,* 57**,** 537-47.

Brierley, I., A. J. Jenner & S. C. Inglis (1992) Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol,* 227**,** 463-79.

Brogna, S. & J. Wen (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol,* 16**,** 107-13.

Brown, C. M., P. A. Stockwell, C. N. Trotman & W. P. Tate (1990) Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res,* 18**,** 6339-45.

Camier, S. & B. Séraphin (2007) [Destroy this (RNA) message after reading it!]. *Med Sci (Paris),* 23**,** 850-6.

Cao, D. & R. Parker (2003) Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell,* 113**,** 533-45.

Cassan, M. & J. P. Rousset (2001) UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol Biol,* 2**,** 3.

Cate, J. H., M. M. Yusupov, G. Z. Yusupova, T. N. Earnest & H. F. Noller (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science,* 285**,** 2095-104.

Chandramouli, P., M. Topf, J. F. Ménétret, N. Eswar, J. J. Cannone, R. R. Gutell, A. Sali & C. W. Akey (2008) Structure of the mammalian 80S ribosome at 8.7 A resolution. *Structure,* 16**,** 535-48.

Chavatte, L., A. Seit-Nebi, V. Dubovaya & A. Favre (2002) The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J,* 21**,** 5302-11.

Chen, C., B. Stevens, J. Kaur, D. Cabral, H. Liu, Y. Wang, H. Zhang, G. Rosenblum, Z. Smilansky, Y. E. Goldman & B. S. Cooperman (2011) Single-molecule fluorescence measurements of ribosomal translocation dynamics. *Mol Cell,* 42**,** 367-77.

Chen, G., K. Y. Chang, M. Y. Chou, C. Bustamante & I. Tinoco (2009) Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A,* 106**,** 12706-11.

Chen, L., D. Muhlrad, V. Hauryliuk, Z. Cheng, M. K. Lim, V. Shyp, R. Parker & H. Song (2010) Structure of the Dom34-Hbs1 complex and implications for no-go decay. *Nat Struct Mol Biol,* 17**,** 1233-40.

Cheng, Z., K. Saito, A. V. Pisarev, M. Wada, V. P. Pisareva, T. V. Pestova, M. Gajda, A. Round, C. Kong, M. Lim, Y. Nakamura, D. I. Svergun, K. Ito & H. Song (2009) Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev,* 23**,** 1106-18.

Clementi, N., A. Chirkova, B. Puffer, R. Micura & N. Polacek (2010) Atomic mutagenesis reveals A2660 of 23S ribosomal RNA as key to EF-G GTPase activation. *Nat Chem Biol,* 6**,** 344-51.

Clemons, W. M., J. L. May, B. T. Wimberly, J. P. McCutcheon, M. S. Capel & V. Ramakrishnan (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. *Nature,* 400**,** 833-40.

Cochella, L. & R. Green (2004) Wobble during decoding: more than third-position promiscuity. *Nat Struct Mol Biol,* 11**,** 1160-2.

--- (2005) An active role for tRNA in decoding beyond codon:anticodon pairing. *Science,* 308**,** 1178-80.

Connell, S. R., C. Takemoto, D. N. Wilson, H. Wang, K. Murayama, T. Terada, M. Shirouzu, M. Rost, M. Schüler, J. Giesebrecht, M. Dabrowski, T. Mielke, P. Fucini, S. Yokoyama & C. M. Spahn (2007)

Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Mol Cell,* 25**,** 751-64.
Cornish, P. V., D. N. Ermolenko, H. F. Noller & T. Ha (2008) Spontaneous intersubunit rotation in single ribosomes. *Mol Cell,* 30**,** 578-88.
Cornish, P. V., D. N. Ermolenko, D. W. Staple, L. Hoang, R. P. Hickerson, H. F. Noller & T. Ha (2009) Following movement of the L1 stalk between three functional states in single ribosomes. *Proc Natl Acad Sci U S A,* 106**,** 2571-6.
Crick, F. H. (1966) Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol,* 19**,** 548-55.
--- (1968) The origin of the genetic code. *J Mol Biol,* 38**,** 367-79.
Cridge, A. G., L. L. Major, A. A. Mahagaonkar, E. S. Poole, L. A. Isaksson & W. P. Tate (2006) Comparison of characteristics and function of translation termination signals between and within prokaryotic and eukaryotic organisms. *Nucleic Acids Res,* 34**,** 1959-73.
Cruz-Vera, L. R., M. S. Sachs, C. L. Squires & C. Yanofsky (2011) Nascent polypeptide sequences that influence ribosome function. *Curr Opin Microbiol,* 14**,** 160-6.
Demeshkina, N., L. Jenner, E. Westhof, M. Yusupov & G. Yusupova (2012) A new understanding of the decoding principle on the ribosome. *Nature*.
Di Giammartino, D. C., K. Nishida & J. L. Manley (2011) Mechanisms and consequences of alternative polyadenylation. *Mol Cell,* 43**,** 853-66.
Diaconu, M., U. Kothe, F. Schlünzen, N. Fischer, J. M. Harms, A. G. Tonevitsky, H. Stark, M. V. Rodnina & M. C. Wahl (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell,* 121**,** 991-1004.
Doma, M. K. & R. Parker (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature,* 440**,** 561-4.
Dorner, S., J. L. Brunelle, D. Sharma & R. Green (2006) The hybrid state of tRNA binding is an authentic translation elongation intermediate. *Nat Struct Mol Biol,* 13**,** 234-41.
Dulude, D., M. Baril & L. Brakier-Gingras (2002) Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res,* 30**,** 5094-102.
Ebihara, K. & Y. Nakamura (1999) C-terminal interaction of translational release factors eRF1 and eRF3 of fission yeast: G-domain uncoupled binding and the role of conserved amino acids. *RNA,* 5**,** 739-50.
Edelman, I. & M. R. Culbertson (1991) Exceptional codon recognition by the glutamine tRNAs in Saccharomyces cerevisiae. *EMBO J,* 10**,** 1481-91.
Edelmann, P. & J. Gallant (1977) Mistranslation in E. coli. *Cell,* 10**,** 131-7.
Eliseev, B., P. Kryuchkova, E. Alkalaeva & L. Frolova (2011) A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. *Nucleic Acids Res,* 39**,** 599-608.
Ermolenko, D. N. & H. F. Noller (2011) mRNA translocation occurs during the second step of ribosomal intersubunit rotation. *Nat Struct Mol Biol,* 18**,** 457-62.
Farabaugh, P. J., H. Zhao & A. Vimaladithan (1993) A novel programed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: frameshifting without tRNA slippage. *Cell,* 74**,** 93-103.
Fei, J., J. E. Bronson, J. M. Hofman, R. L. Srinivas, C. H. Wiggins & R. L. Gonzalez (2009) Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc Natl Acad Sci U S A,* 106**,** 15702-7.
Flanagan, J. F., O. Namy, I. Brierley & R. J. Gilbert (2010) Direct observation of distinct A/P hybrid-state tRNAs in translocating ribosomes. *Structure,* 18**,** 257-64.
Flower, A. M. & C. S. McHenry (1990) The gamma subunit of DNA polymerase III holoenzyme of Escherichia coli is produced by ribosomal frameshifting. *Proc Natl Acad Sci U S A,* 87**,** 3713-7.
Frank, J. & R. K. Agrawal (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature,* 406**,** 318-22.
Frank, J., H. Gao, J. Sengupta, N. Gao & D. J. Taylor (2007) The process of mRNA-tRNA translocation. *Proc Natl Acad Sci U S A,* 104**,** 19671-8.
Frischmeyer, P. A., A. van Hoof, K. O'Donnell, A. L. Guerrerio, R. Parker & H. C. Dietz (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science,* 295**,** 2258-61.
Frolova, L., A. Seit-Nebi & L. Kisselev (2002) Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA,* 8**,** 129-36.

Frolova, L. Y., T. I. Merkulova & L. L. Kisselev (2000) Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. *RNA,* 6**,** 381-90.

Frolova, L. Y., R. Y. Tsivkovskii, G. F. Sivolobova, N. Y. Oparina, O. I. Serpinsky, V. M. Blinov, S. I. Tatkov & L. L. Kisselev (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA,* 5**,** 1014-20.

Furuichi, Y. & A. J. Shatkin. 2007. Caps on Eukaryotic mRNAs. Encyclopedia of Life Sciences.

Gandhi, R., M. Manzoor & K. A. Hudak (2008) Depurination of Brome mosaic virus RNA3 in vivo results in translation-dependent accelerated degradation of the viral RNA. *J Biol Chem,* 283**,** 32218-28.

Gao, Y. G., M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley & V. Ramakrishnan (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science,* 326**,** 694-9.

Gaston, M. A., R. Jiang & J. A. Krzycki (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr Opin Microbiol,* 14**,** 342-9.

Gaudin, C., M. H. Mazauric, M. Traïkia, E. Guittet, S. Yoshizawa & D. Fourmy (2005) Structure of the RNA signal essential for translational frameshifting in HIV-1. *J Mol Biol,* 349**,** 1024-35.

Gavrilova, L. P., O. E. Kostiashkina, V. E. Koteliansky, N. M. Rutkevitch & A. S. Spirin (1976) Factor-free ("non-enzymic") and factor-dependent systems of translation of polyuridylic acid by Escherichia coli ribosomes. *J Mol Biol,* 101**,** 537-52.

Gentz, P. M., G. L. Blatch & R. A. Dorrington (2009) Dimerization of the yeast eukaryotic translation initiation factor 5A requires hypusine and is RNA dependent. *FEBS J,* 276**,** 695-706.

Gesteland, R. F., R. B. Weiss & J. F. Atkins (1992) Recoding: reprogrammed genetic decoding. *Science,* 257**,** 1640-1.

Giedroc, D. P., C. A. Theimer & P. L. Nixon (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol,* 298**,** 167-85.

Gilbert, W. V. (2010) Alternative ways to think about cellular internal ribosome entry. *J Biol Chem,* 285**,** 29033-8.

Goldman, E. 2008. Transfer RNA. Encyclopedia of Life Sciences.

Graille, M., M. Chaillet & H. van Tilbeurgh (2008) Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. *J Biol Chem,* 283**,** 7145-54.

Graille, M., S. Figaro, S. Kervestin, R. H. Buckingham, D. Liger & V. Heurgué-Hamard (2012) Methylation of class I translation termination factors: Structural and functional aspects. *Biochimie*.

Greganova, E., M. Altmann & P. Bütikofer (2011) Unique modifications of translation elongation factors. *FEBS J,* 278**,** 2613-24.

Gregio, A. P., V. P. Cano, J. S. Avaca, S. R. Valentini & C. F. Zanelli (2009) eIF5A has a function in the elongation step of translation in yeast. *Biochem Biophys Res Commun,* 380**,** 785-90.

Guarraia, C., L. Norris, A. Raman & P. J. Farabaugh (2007) Saturation mutagenesis of a +1 programmed frameshift-inducing mRNA sequence derived from a yeast retrotransposon. *RNA,* 13**,** 1940-7.

Gupta, M. & P. R. Copeland (2007) Functional analysis of the interplay between translation termination, selenocysteine codon context, and selenocysteine insertion sequence-binding protein 2. *J Biol Chem,* 282**,** 36797-807.

Halfmann, R., S. Alberti & S. Lindquist (2010) Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol,* 20**,** 125-33.

Halfmann, R., D. F. Jarosz, S. K. Jones, A. Chang, A. K. Lancaster & S. Lindquist (2012) Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature,* 482**,** 363-8.

Hansen, T. M., S. N. Reihani, L. B. Oddershede & M. A. Sørensen (2007) Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc Natl Acad Sci U S A,* 104**,** 5830-5.

Hatfield, D. L. & V. N. Gladyshev (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol,* 22**,** 3565-76.

Hatin, I., C. Fabret, J. P. Rousset & O. Namy (2009) Molecular dissection of translation termination mechanism identifies two new critical regions in eRF1. *Nucleic Acids Res,* 37**,** 1789-98.

Hauryliuk, V., A. Zavialov, L. Kisselev & M. Ehrenberg (2006) Class-1 release factor eRF1 promotes GTP binding by class-2 release factor eRF3. *Biochimie,* 88**,** 747-57.

Hellen, C. U. (2009) IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochim Biophys Acta,* 1789**,** 558-70.
Herr, A. J., J. F. Atkins & R. F. Gesteland (2000a) Coupling of open reading frames by translational bypassing. *Annu Rev Biochem,* 69**,** 343-72.
Herr, A. J., R. F. Gesteland & J. F. Atkins (2000b) One protein from two open reading frames: mechanism of a 50 nt translational bypass. *EMBO J,* 19**,** 2671-80.
Herr, A. J., N. M. Wills, C. C. Nelson, R. F. Gesteland & J. F. Atkins (2004) Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *J Biol Chem,* 279**,** 11081-7.
Hinnebusch, A. G. (2011) Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev,* 75**,** 434-67, first page of table of contents.
Horan, L. H. & H. F. Noller (2007) Intersubunit movement is required for ribosomal translocation. *Proc Natl Acad Sci U S A,* 104**,** 4881-5.
Ide, S., T. Miyazaki, H. Maki & T. Kobayashi (2010) Abundance of ribosomal RNA gene copies maintains genome integrity. *Science,* 327**,** 693-6.
Inagaki, Y., C. Blouin, W. F. Doolittle & A. J. Roger (2002) Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. *Nucleic Acids Res,* 30**,** 532-44.
Inagaki, Y. & W. Ford Doolittle (2000) Evolution of the eukaryotic translation termination system: origins of release factors. *Mol Biol Evol,* 17**,** 882-9.
Ingolia, N. T., S. Ghaemmaghami, J. R. Newman & J. S. Weissman (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science,* 324**,** 218-23.
Ito, K., K. Ebihara & Y. Nakamura (1998) The stretch of C-terminal acidic amino acids of translational release factor eRF1 is a primary binding site for eRF3 of fission yeast. *RNA,* 4**,** 958-72.
Ito-Harashima, S., K. Kuroha, T. Tatematsu & T. Inada (2007) Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. *Genes Dev,* 21**,** 519-24.
Ivanov, I. P., S. Matsufuji, Y. Murakami, R. F. Gesteland & J. F. Atkins (2000) Conservation of polyamine regulation by translational frameshifting from yeast to mammals. *EMBO J,* 19**,** 1907-17.
Jacks, T., H. D. Madhani, F. R. Masiarz & H. E. Varmus (1988a) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell,* 55**,** 447-58.
Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr & H. E. Varmus (1988b) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature,* 331**,** 280-3.
Jackson, R. J., C. U. Hellen & T. V. Pestova (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol,* 11**,** 113-27.
Jan, E. & P. Sarnow (2002) Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J Mol Biol,* 324**,** 889-902.
Jao, D. L. & K. Y. Chen (2006) Tandem affinity purification revealed the hypusine-dependent binding of eukaryotic initiation factor 5A to the translating 80S ribosomal complex. *J Cell Biochem,* 97**,** 583-98.
Jenner, L. B., N. Demeshkina, G. Yusupova & M. Yusupov (2010) Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol,* 17**,** 555-60.
Julián, P., A. L. Konevega, S. H. Scheres, M. Lázaro, D. Gil, W. Wintermeyer, M. V. Rodnina & M. Valle (2008) Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc Natl Acad Sci U S A,* 105**,** 16924-7.
Jørgensen, F. & C. G. Kurland (1990) Processivity errors of gene expression in Escherichia coli. *J Mol Biol,* 215**,** 511-21.
Kemper, W. M., K. W. Berry & W. C. Merrick (1976) Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2Balpha and M2Bbeta. *J Biol Chem,* 251**,** 5551-7.
Kervestin, S., L. Frolova, L. Kisselev & O. Jean-Jean (2001) Stop codon recognition in ciliates: Euplotes release factor does not respond to reassigned UGA codon. *EMBO Rep,* 2**,** 680-4.
King, C. Y. & R. Diaz-Avalos (2004) Protein-only transmission of three yeast prion strains. *Nature,* 428**,** 319-23.
Kisselev, L., M. Ehrenberg & L. Frolova (2003) Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J,* 22**,** 175-82.
Kisselev, L. L. & R. H. Buckingham (2000) Translational termination comes of age. *Trends Biochem Sci,* 25**,** 561-6.

Kitzan, L. J., F. R. Ferraro, T. V. Petros & M. Ludorf (1999) The role of vocabulary ability during visual word recognition in younger and older adults. *J Gen Psychol,* 126**,** 6-16.
Klobutcher, L. A. (2005) Sequencing of random Euplotes crassus macronuclear genes supports a high frequency of +1 translational frameshifting. *Eukaryot Cell,* 4**,** 2098-105.
Klobutcher, L. A. & P. J. Farabaugh (2002) Shifty ciliates: frequent programmed translational frameshifting in euplotids. *Cell,* 111**,** 763-6.
Knight, R. D., S. J. Freeland & L. F. Landweber (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet,* 2**,** 49-58.
Kobayashi, K., I. Kikuno, K. Kuroha, K. Saito, K. Ito, R. Ishitani, T. Inada & O. Nureki (2010) Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1α complex. *Proc Natl Acad Sci U S A,* 107**,** 17575-9.
Kohli, J. & H. Grosjean (1981) Usage of the three termination codons: compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. *Mol Gen Genet,* 182**,** 430-9.
Kollmus, H., A. Honigman, A. Panet & H. Hauser (1994) The sequences of and distance between two cis-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human T-cell leukemia virus type II in vivo. *J Virol,* 68**,** 6087-91.
Kolosov, P., L. Frolova, A. Seit-Nebi, V. Dubovaya, A. Kononenko, N. Oparina, J. Justesen, A. Efimov & L. Kisselev (2005) Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res,* 33**,** 6418-25.
Kondrashov, N., A. Pusic, C. R. Stumpf, K. Shimizu, A. C. Hsieh, S. Xue, J. Ishijima, T. Shiroishi & M. Barna (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell,* 145**,** 383-97.
Kononenko, A. V., V. A. Mitkevich, V. I. Dubovaya, P. M. Kolosov, A. A. Makarov & L. L. Kisselev (2008) Role of the individual domains of translation termination factor eRF1 in GTP binding to eRF3. *Proteins,* 70**,** 388-93.
Kontos, H., S. Napthine & I. Brierley (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol,* 21**,** 8657-70.
Korostelev, A., H. Asahara, L. Lancaster, M. Laurberg, A. Hirschi, J. Zhu, S. Trakhanov, W. G. Scott & H. F. Noller (2008) Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A,* 105**,** 19684-9.
Kovalchuke, O., R. Kambampati, E. Pladies & K. Chakraburtty (1998) Competition and cooperation amongst yeast elongation factors. *Eur J Biochem,* 258**,** 986-93.
Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell,* 44**,** 283-92.
Kurata, S., K. H. Nielsen, S. F. Mitchell, J. R. Lorsch, A. Kaji & H. Kaji (2010) Ribosome recycling step in yeast cytoplasmic protein synthesis is catalyzed by eEF3 and ATP. *Proc Natl Acad Sci U S A,* 107**,** 10854-9.
Kurian, L., R. Palanimurugan, D. Gödderz & R. J. Dohmen (2011) Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature,* 477**,** 490-4.
Kurland, C. G. (1992) Evolution of mitochondrial genomes and the genetic code. *Bioessays,* 14**,** 709-14.
Larsen, B., N. M. Wills, R. F. Gesteland & J. F. Atkins (1994) rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol,* 176**,** 6842-51.
Laurberg, M., H. Asahara, A. Korostelev, J. Zhu, S. Trakhanov & H. F. Noller (2008) Structural basis for translation termination on the 70S ribosome. *Nature,* 454**,** 852-7.
Le Goff, X., M. Philippe & O. Jean-Jean (1997) Overexpression of human release factor 1 alone has an antisuppressor effect in human cells. *Mol Cell Biol,* 17**,** 3164-72.
Lee, H. H., Y. S. Kim, K. H. Kim, I. Heo, S. K. Kim, O. Kim, H. K. Kim, J. Y. Yoon, H. S. Kim, d. J. Kim, S. J. Lee, H. J. Yoon, S. J. Kim, B. G. Lee, H. K. Song, V. N. Kim, C. M. Park & S. W. Suh (2007) Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol Cell,* 27**,** 938-50.
Li, W., L. G. Trabuco, K. Schulten & J. Frank (2011) Molecular dynamics of EF-G during translocation. *Proteins,* 79**,** 1478-86.
Lill, R., J. M. Robertson & W. Wintermeyer (1989) Binding of the 3' terminus of tRNA to 23S rRNA in the ribosomal exit site actively promotes translocation. *EMBO J,* 8**,** 3933-8.

Loh, P. G. & H. Song (2010) Structural and mechanistic insights into translation termination. *Curr Opin Struct Biol,* 20**,** 98-103.
Lopinski, J. D., J. D. Dinman & J. A. Bruenn (2000) Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol,* 20**,** 1095-103.
Mangus, D. A., M. C. Evans & A. Jacobson (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol,* 4**,** 223.
Manktelow, E., K. Shigemoto & I. Brierley (2005) Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Res,* 33**,** 1553-63.
Matassova, A. B., M. V. Rodnina & W. Wintermeyer (2001) Elongation factor G-induced structural change in helix 34 of 16S rRNA related to translocation on the ribosome. *RNA,* 7**,** 1879-85.
Matsufuji, S., T. Matsufuji, Y. Miyazaki, Y. Murakami, J. F. Atkins, R. F. Gesteland & S. Hayashi (1995) Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell,* 80**,** 51-60.
Merkulova, T. I., L. Y. Frolova, M. Lazar, J. Camonis & L. L. Kisselev (1999) C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their in vivo interaction. *FEBS Lett,* 443**,** 41-7.
Meyer, F., H. J. Schmidt, E. Plümper, A. Hasilik, G. Mersmann, H. E. Meyer, A. Engström & K. Heckmann (1991) UGA is translated as cysteine in pheromone 3 of Euplotes octocarinatus. *Proc Natl Acad Sci U S A,* 88**,** 3758-61.
Miller, P. F. & A. G. Hinnebusch (1990) cis-acting sequences involved in the translational control of GCN4 expression. *Biochim Biophys Acta,* 1050**,** 151-4.
Moazed, D. & H. F. Noller (1989) Intermediate states in the movement of transfer RNA in the ribosome. *Nature,* 342**,** 142-8.
Morris, D. K. & V. Lundblad (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol,* 7**,** 969-76.
Motorin, Y. & M. Helm (2010) tRNA stabilization by modified nucleotides. *Biochemistry,* 49**,** 4934-44.
Munro, J. B., R. B. Altman, N. O'Connor & S. C. Blanchard (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell,* 25**,** 505-17.
Murakami, Y., S. Matsufuji, T. Kameji, S. Hayashi, K. Igarashi, T. Tamura, K. Tanaka & A. Ichihara (1992) Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature,* 360**,** 597-9.
Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein & M. Snyder (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320**,** 1344-9.
Namy, O., G. Duchateau-Nguyen & J. P. Rousset (2002) Translational readthrough of the PDE2 stop codon modulates cAMP levels in Saccharomyces cerevisiae. *Mol Microbiol,* 43**,** 641-52.
Namy, O., A. Galopier, C. Martini, S. Matsufuji, C. Fabret & J. P. Rousset (2008) Epigenetic control of polyamines by the prion [PSI+]. *Nat Cell Biol,* 10**,** 1069-75.
Namy, O., I. Hatin & J. P. Rousset (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep,* 2**,** 787-93.
Namy, O., S. J. Moran, D. I. Stuart, R. J. Gilbert & I. Brierley (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature,* 441**,** 244-7.
Namy, O., J. P. Rousset, S. Napthine & I. Brierley (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell,* 13**,** 157-68.
Neu-Yilik, G. & A. E. Kulozik (2008) NMD: multitasking between mRNA surveillance and modulation of gene expression. *Adv Genet,* 62**,** 185-243.
Nicholson, P. & O. Mühlemann (2010) Cutting the nonsense: the degradation of PTC-containing mRNAs. *Biochem Soc Trans,* 38**,** 1615-20.
Nierhaus, K. H. (1990) The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry,* 29**,** 4997-5008.
Noller, H. F., M. M. Yusupov, G. Z. Yusupova, A. Baucom & J. H. Cate (2002) Translocation of tRNA during protein synthesis. *FEBS Lett,* 514**,** 11-6.
Ogle, J. M., D. E. Brodersen, W. M. Clemons, M. J. Tarry, A. P. Carter & V. Ramakrishnan (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science,* 292**,** 897-902.
Ohtsuka, T., Y. Kitajima, T. Takahashi, S. Sato, A. Miyoshi, N. Kohya, K. Kitahara, Y. Nakafusa & K. Miyazaki (2009) Infectious complications after gastric cancer surgery accelerate a rapid hepatic recurrence. *Hepatogastroenterology,* 56**,** 1277-80.

Palanimurugan, R., H. Scheel, K. Hofmann & R. J. Dohmen (2004) Polyamines regulate their synthesis by inducing expression and blocking degradation of ODC antizyme. *EMBO J,* 23**,** 4857-67.
Pan, D., S. V. Kirillov & B. S. Cooperman (2007) Kinetically competent intermediates in the translocation step of protein synthesis. *Mol Cell,* 25**,** 519-29.
Park, M. H., Y. B. Lee & Y. A. Joe (1997) Hypusine is essential for eukaryotic cell proliferation. *Biol Signals,* 6**,** 115-23.
Park, M. H., K. Nishimura, C. F. Zanelli & S. R. Valentini (2010) Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids,* 38**,** 491-500.
Passos, D. O., M. K. Doma, C. J. Shoemaker, D. Muhlrad, R. Green, J. Weissman, J. Hollien & R. Parker (2009) Analysis of Dom34 and its function in no-go decay. *Mol Biol Cell,* 20**,** 3025-32.
Paushkin, S. V., V. V. Kushnirov, V. N. Smirnov & M. D. Ter-Avanesyan (1996) Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *EMBO J,* 15**,** 3127-34.
Pisarev, A. V., C. U. Hellen & T. V. Pestova (2007) Recycling of eukaryotic posttermination ribosomal complexes. *Cell,* 131**,** 286-99.
Pisarev, A. V., M. A. Skabkin, V. P. Pisareva, O. V. Skabkina, A. M. Rakotondrafara, M. W. Hentze, C. U. Hellen & T. V. Pestova (2010) The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell,* 37**,** 196-210.
Pisareva, V. P., A. V. Pisarev, C. U. Hellen, M. V. Rodnina & T. V. Pestova (2006) Kinetic analysis of interaction of eukaryotic release factor 3 with guanine nucleotides. *J Biol Chem,* 281**,** 40224-35.
Pisareva, V. P., M. A. Skabkin, C. U. Hellen, T. V. Pestova & A. V. Pisarev (2011) Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J,* 30**,** 1804-17.
Plant, E. P., K. L. Jacobs, J. W. Harger, A. Meskauskas, J. L. Jacobs, J. L. Baxter, A. N. Petrov & J. D. Dinman (2003) The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA,* 9**,** 168-74.
Proudfoot, N. & J. O'Sullivan (2002) Polyadenylation: a tail of two complexes. *Curr Biol,* 12**,** R855-7.
Prusiner, S. B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science,* 216**,** 136-44.
Pure, G. A., G. W. Robinson, L. Naumovski & E. C. Friedberg (1985) Partial suppression of an ochre mutation in Saccharomyces cerevisiae by multicopy plasmids containing a normal yeast tRNAGln gene. *J Mol Biol,* 183**,** 31-42.
Qin, S. L., K. Moldave & C. S. McLaughlin (1987) Isolation of the yeast gene encoding elongation factor 3 for protein synthesis. *J Biol Chem,* 262**,** 7802-7.
Qu, X., J. D. Wen, L. Lancaster, H. F. Noller, C. Bustamante & I. Tinoco (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature,* 475**,** 118-21.
Rasmussen, L. C., B. S. Laursen, K. K. Mortensen & H. U. Sperling-Petersen. 2009. Initiator tRNAs in Bacteria and Eukaryotes.
Ratje, A. H., J. Loerke, A. Mikolajka, M. Brünner, P. W. Hildebrand, A. L. Starosta, A. Dönhöfer, S. R. Connell, P. Fucini, T. Mielke, P. C. Whitford, J. N. Onuchic, Y. Yu, K. Y. Sanbonmatsu, R. K. Hartmann, P. A. Penczek, D. N. Wilson & C. M. Spahn (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature,* 468**,** 713-6.
Rich, B. E. & J. A. Steitz (1987) Human acidic ribosomal phosphoproteins P0, P1, and P2: analysis of cDNA clones, in vitro synthesis, and assembly. *Mol Cell Biol,* 7**,** 4065-74.
Rodnina, M. V., K. B. Gromadski, U. Kothe & H. J. Wieden (2005) Recognition and selection of tRNA in translation. *FEBS Lett,* 579**,** 938-42.
Rodnina, M. V. & W. Wintermeyer (2011) The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation. *Biochem Soc Trans,* 39**,** 658-62.
Saini, P., D. E. Eyler, R. Green & T. E. Dever (2009) Hypusine-containing protein eIF5A promotes translation elongation. *Nature,* 459**,** 118-21.
Salas-Marco, J. & D. M. Bedwell (2004) GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol Cell Biol,* 24**,** 7769-78.
Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi & A. Yonath (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell,* 102**,** 615-23.
Schmeing, T. M. & V. Ramakrishnan (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature,* 461**,** 1234-42.

Schmeing, T. M., R. M. Voorhees, A. C. Kelley & V. Ramakrishnan (2011) How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol,* 18**,** 432-6.

Schuette, J. C., F. V. Murphy, A. C. Kelley, J. R. Weir, J. Giesebrecht, S. R. Connell, J. Loerke, T. Mielke, W. Zhang, P. A. Penczek, V. Ramakrishnan & C. M. Spahn (2009) GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J,* 28**,** 755-65.

Schuwirth, B. S., M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton & J. H. Cate (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science,* 310**,** 827-34.

Seit-Nebi, A., L. Frolova & L. Kisselev (2002) Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep,* 3**,** 881-6.

Selmer, M., C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir & V. Ramakrishnan (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science,* 313**,** 1935-42.

Shiba, T., H. Mizote, T. Kaneko, T. Nakajima & Y. Kakimoto (1971) Hypusine, a new amino acid occurring in bovine brain. Isolation and structural determination. *Biochim Biophys Acta,* 244**,** 523-31.

Shoemaker, C. J., D. E. Eyler & R. Green (2010) Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science,* 330**,** 369-72.

Shoemaker, C. J. & R. Green (2011) Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A,* 108**,** E1392-8.

Shorter, J. & S. Lindquist (2005) Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet,* 6**,** 435-50.

Simonović, M. & T. A. Steitz (2009) A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim Biophys Acta,* 1789**,** 612-23.

Skuzeski, J. M., L. M. Nichols, R. F. Gesteland & J. F. Atkins (1991) The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *J Mol Biol,* 218**,** 365-73.

Somogyi, P., A. J. Jenner, I. Brierley & S. C. Inglis (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol,* 13**,** 6931-40.

Sonenberg, N. & A. G. Hinnebusch (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell,* 136**,** 731-45.

Song, H., P. Mugnier, A. K. Das, H. M. Webb, D. R. Evans, M. F. Tuite, B. A. Hemmings & D. Barford (2000) The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell,* 100**,** 311-21.

Spahn, C. M., R. Beckmann, N. Eswar, P. A. Penczek, A. Sali, G. Blobel & J. Frank (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. *Cell,* 107**,** 373-86.

Spahn, C. M., M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jørgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta & J. Frank (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J,* 23**,** 1008-19.

Spiegel, P. C., D. N. Ermolenko & H. F. Noller (2007) Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. *RNA,* 13**,** 1473-82.

Spirin, A. S. (1969) A model of the functioning ribosome: locking and unlocking of the ribosome subparticles. *Cold Spring Harb Symp Quant Biol,* 34**,** 197-207.

--- (2009) The ribosome as a conveying thermal ratchet machine. *J Biol Chem,* 284**,** 21103-19.

Squires, J. E. & M. J. Berry (2008) Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors. *IUBMB Life,* 60**,** 232-5.

Stahl, G., L. Bidou, J. P. Rousset & M. Cassan (1995) Versatile vectors to study recoding: conservation of rules between yeast and mammalian cells. *Nucleic Acids Res,* 23**,** 1557-60.

Stahl, G., G. P. McCarty & P. J. Farabaugh (2002) Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci,* 27**,** 178-83.

Stansfield, I., K. M. Jones, V. V. Kushnirov, A. R. Dagkesamanskaya, A. I. Poznyakovski, S. V. Paushkin, C. R. Nierras, B. S. Cox, M. D. Ter-Avanesyan & M. F. Tuite (1995) The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in Saccharomyces cerevisiae. *EMBO J,* 14**,** 4365-73.

Staple, D. W. & S. E. Butcher (2005) Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J Mol Biol,* 349**,** 1011-23.

Sundararajan, A., W. A. Michaud, Q. Qian, G. Stahl & P. J. Farabaugh (1999) Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshifting in yeast. *Mol Cell,* 4**,** 1005-15.

Söll, D., D. S. Jones, E. Ohtsuka, R. D. Faulkner, R. Lohrmann, H. Hayatsu & H. G. Khorana (1966) Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique. *J Mol Biol,* 19**,** 556-73.

Takyar, S., R. P. Hickerson & H. F. Noller (2005) mRNA helicase activity of the ribosome. *Cell,* 120**,** 49-58.

Taliaferro, D. & P. J. Farabaugh (2007) An mRNA sequence derived from the yeast EST3 gene stimulates programmed +1 translational frameshifting. *RNA,* 13**,** 606-13.

Tanaka, M., P. Chien, N. Naber, R. Cooke & J. S. Weissman (2004) Conformational variations in an infectious protein determine prion strain differences. *Nature,* 428**,** 323-8.

Taylor, D. J., B. Devkota, A. D. Huang, M. Topf, E. Narayanan, A. Sali, S. C. Harvey & J. Frank (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure,* 17**,** 1591-604.

Taylor, D. J., J. Nilsson, A. R. Merrill, G. R. Andersen, P. Nissen & J. Frank (2007) Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J,* 26**,** 2421-31.

Toh, Y., H. Hori, K. Tomita, T. Ueda & K. Watanabe. 2009. Transfer RNA Synthesis and Regulation. Encyclopedia of Life Sciences.

Trabuco, L. G., E. Schreiner, J. Eargle, P. Cornish, T. Ha, Z. Luthey-Schulten & K. Schulten (2010) The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. *J Mol Biol,* 402**,** 741-60.

Triana-Alonso, F. J., K. Chakraburtty & K. H. Nierhaus (1995) The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. *J Biol Chem,* 270**,** 20473-8.

True, H. L., I. Berlin & S. L. Lindquist (2004) Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature,* 431**,** 184-7.

True, H. L. & S. L. Lindquist (2000) A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature,* 407**,** 477-83.

Tsuboi, T., K. Kuroha, K. Kudo, S. Makino, E. Inoue, I. Kashima & T. Inada (2012) Dom34:Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3' End of Aberrant mRNA. *Mol Cell*.

Tsuchihashi, Z. & A. Kornberg (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A,* 87**,** 2516-20.

Tu, C., T. H. Tzeng & J. A. Bruenn (1992) Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci U S A,* 89**,** 8636-40.

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

Vallabhaneni, H., H. Fan-Minogue, D. M. Bedwell & P. J. Farabaugh (2009) Connection between stop codon reassignment and frequent use of shifty stop frameshifting. *RNA,* 15**,** 889-97.

Valle, M., A. Zavialov, W. Li, S. M. Stagg, J. Sengupta, R. C. Nielsen, P. Nissen, S. C. Harvey, M. Ehrenberg & J. Frank (2003a) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol,* 10**,** 899-906.

Valle, M., A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg & J. Frank (2003b) Locking and unlocking of ribosomal motions. *Cell,* 114**,** 123-34.

van den Elzen, A. M., J. Henri, N. Lazar, M. E. Gas, D. Durand, F. Lacroute, M. Nicaise, H. van Tilbeurgh, B. Séraphin & M. Graille (2010) Dissection of Dom34-Hbs1 reveals independent functions in two RNA quality control pathways. *Nat Struct Mol Biol,* 17**,** 1446-52.

van Hoof, A., P. A. Frischmeyer, H. C. Dietz & R. Parker (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science,* 295**,** 2262-4.

Villa, E., J. Sengupta, L. G. Trabuco, J. LeBarron, W. T. Baxter, T. R. Shaikh, R. A. Grassucci, P. Nissen, M. Ehrenberg, K. Schulten & J. Frank (2009) Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A,* 106**,** 1063-8.

Vimaladithan, A. & P. J. Farabaugh (1994) Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol Cell Biol,* 14**,** 8107-16.

Viphakone, N., F. Voisinet-Hakil & L. Minvielle-Sebastia (2008) Molecular dissection of mRNA poly(A) tail length control in yeast. *Nucleic Acids Res,* 36**,** 2418-33.

Warner, J. R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci,* 24**,** 437-40.

Weiss, R. B., D. M. Dunn, M. Shuh, J. F. Atkins & R. F. Gesteland (1989) E. coli ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol,* 1**,** 159-69.

Weiss, R. B., W. M. Huang & D. M. Dunn (1990) A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell,* 62**,** 117-26.

Weixel, K. & N. A. Bradbury (2002) Analysis of CFTR endocytosis by cell surface biotinylation. *Methods Mol Med,* 70**,** 323-40.

Weixlbaumer, A., H. Jin, C. Neubauer, R. M. Voorhees, S. Petry, A. C. Kelley & V. Ramakrishnan (2008) Insights into translational termination from the structure of RF2 bound to the ribosome. *Science,* 322**,** 953-6.

Wickner, R. B., H. K. Edskes, F. Shewmaker & T. Nakayashiki (2007) Prions of fungi: inherited structures and biological roles. *Nat Rev Microbiol,* 5**,** 611-8.

Wieden, H. J., W. Wintermeyer & M. V. Rodnina (2001) A common structural motif in elongation factor Ts and ribosomal protein L7/12 may be involved in the interaction with elongation factor Tu. *J Mol Evol,* 52**,** 129-36.

Williams, I., J. Richardson, A. Starkey & I. Stansfield (2004) Genome-wide prediction of stop codon readthrough during translation in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res,* 32**,** 6605-16.

Wills, N. M., B. Moore, A. Hammer, R. F. Gesteland & J. F. Atkins (2006) A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *J Biol Chem,* 281**,** 7082-8.

Wills, N. M., M. O'Connor, C. C. Nelson, C. C. Rettberg, W. M. Huang, R. F. Gesteland & J. F. Atkins (2008) Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *EMBO J,* 27**,** 2533-44.

Wilson, M. A., S. Meaux & A. van Hoof (2007) A genomic screen in yeast reveals novel aspects of nonstop mRNA metabolism. *Genetics,* 177**,** 773-84.

Wilusz, C. J., M. Wormington & S. W. Peltz (2001) The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol,* 2**,** 237-46.

Wolff, E. C., K. R. Kang, Y. S. Kim & M. H. Park (2007) Posttranslational synthesis of hypusine: evolutionary progression and specificity of the hypusine modification. *Amino Acids,* 33**,** 341-50.

Yonath, A., B. Tesche, S. Lorenz, J. Müssig, V. A. Erdmann & H. G. Wittmann (1983) Several crystal forms of the Bacillus stearothermophilus 50 S ribosomal particles. *FEBS Lett,* 154**,** 15-20.

Yoshizawa, S., D. Fourmy & J. D. Puglisi (1999) Recognition of the codon-anticodon helix by ribosomal RNA. *Science,* 285**,** 1722-5.

Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate & H. F. Noller (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science,* 292**,** 883-96.

Yusupova, G. Z., M. M. Yusupov, J. H. Cate & H. F. Noller (2001) The path of messenger RNA through the ribosome. *Cell,* 106**,** 233-41.

Zaher, H. S. & R. Green (2009) Fidelity at the molecular level: lessons from protein synthesis. *Cell,* 136**,** 746-62.

Zanelli, C. F., A. L. Maragno, A. P. Gregio, S. Komili, J. R. Pandolfi, C. A. Mestriner, W. R. Lustri & S. R. Valentini (2006) eIF5A binds to translational machinery components and affects translation in yeast. *Biochem Biophys Res Commun,* 348**,** 1358-66.

Zhang, F. & A. G. Hinnebusch (2011) An upstream ORF with non-AUG start codon is translated in vivo but dispensable for translational control of GCN4 mRNA. *Nucleic Acids Res,* 39**,** 3128-40.

Références bibliographiques des figures

Agirrezabala, X., J. Lei, J. L. Brunelle, R. F. Ortiz-Meoz, R. Green & J. Frank (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol Cell,* 32**,** 190-7.

Agris, P. F. (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep,* 9**,** 629-35.

Andersen, C. B., T. Becker, M. Blau, M. Anand, M. Halic, B. Balar, T. Mielke, T. Boesen, J. S. Pedersen, C. M. Spahn, T. G. Kinzy, G. R. Andersen & R. Beckmann (2006) Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature,* 443**,** 663-8.

Becker, T., J. P. Armache, A. Jarasch, A. M. Anger, E. Villa, H. Sieber, B. A. Motaal, T. Mielke, O. Berninghausen & R. Beckmann (2011) Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol,* 18**,** 715-20.

Becker, T., S. Franckenberg, S. Wickles, C. J. Shoemaker, A. M. Anger, J. P. Armache, H. Sieber, C. Ungewickell, O. Berninghausen, I. Daberkow, A. Karcher, M. Thomm, K. P. Hopfner, R. Green & R. Beckmann (2012) Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature,* 482**,** 501-6.

Ben-Shem, A., N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova & M. Yusupov (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science,* 334**,** 1524-9.

Cheng, Z., K. Saito, A. V. Pisarev, M. Wada, V. P. Pisareva, T. V. Pestova, M. Gajda, A. Round, C. Kong, M. Lim, Y. Nakamura, D. I. Svergun, K. Ito & H. Song (2009) Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev,* 23**,** 1106-18.

Cornish, P. V., D. N. Ermolenko, D. W. Staple, L. Hoang, R. P. Hickerson, H. F. Noller & T. Ha (2009) Following movement of the L1 stalk between three functional states in single ribosomes. *Proc Natl Acad Sci U S A,* 106**,** 2571-6.

Cruz-Vera, L. R., M. S. Sachs, C. L. Squires & C. Yanofsky (2011) Nascent polypeptide sequences that influence ribosome function. *Curr Opin Microbiol,* 14**,** 160-6.

Demeshkina, N., L. Jenner, E. Westhof, M. Yusupov & G. Yusupova (2012) A new understanding of the decoding principle on the ribosome. *Nature*.

Dinman, J. D., M. J. Ruiz-Echevarria & S. W. Peltz (1998) Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol,* 16**,** 190-6.

Flanagan, J. F., O. Namy, I. Brierley & R. J. Gilbert (2010) Direct observation of distinct A/P hybrid-state tRNAs in translocating ribosomes. *Structure,* 18**,** 257-64.

Gao, Y. G., M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley & V. Ramakrishnan (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science,* 326**,** 694-9.

Giedroc, D. P. & P. V. Cornish (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res,* 139**,** 193-208.

Goldman, E. 2008. Transfer RNA. Encyclopedia of Life Sciences.

Graille, M., M. Chaillet & H. van Tilbeurgh (2008) Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. *J Biol Chem,* 283**,** 7145-54.

Jackson, R. J., C. U. Hellen & T. V. Pestova (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol,* 11**,** 113-27.

Knight, R. D., S. J. Freeland & L. F. Landweber (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet,* 2**,** 49-58.

Namy, O., A. Galopier, C. Martini, S. Matsufuji, C. Fabret & J. P. Rousset (2008) Epigenetic control of polyamines by the prion [PSI+]. *Nat Cell Biol,* 10**,** 1069-75.

Namy, O., J. P. Rousset, S. Napthine & I. Brierley (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell,* 13**,** 157-68.

Pisarev, A. V., C. U. Hellen & T. V. Pestova (2007) Recycling of eukaryotic posttermination ribosomal complexes. *Cell,* 131**,** 286-99.

Pisarev, A. V., M. A. Skabkin, V. P. Pisareva, O. V. Skabkina, A. M. Rakotondrafara, M. W. Hentze, C. U. Hellen & T. V. Pestova (2010) The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell,* 37**,** 196-210.

Qu, X., J. D. Wen, L. Lancaster, H. F. Noller, C. Bustamante & I. Tinoco (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature,* 475**,** 118-21.

Ratje, A. H., J. Loerke, A. Mikolajka, M. Brünner, P. W. Hildebrand, A. L. Starosta, A. Dönhöfer, S. R. Connell, P. Fucini, T. Mielke, P. C. Whitford, J. N. Onuchic, Y. Yu, K. Y. Sanbonmatsu, R. K. Hartmann, P. A. Penczek, D. N. Wilson & C. M. Spahn (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature,* 468**,** 713-6.

Schmeing, T. M. & V. Ramakrishnan (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature,* 461**,** 1234-42.

Shorter, J. & S. Lindquist (2005) Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet,* 6**,** 435-50.

Song, H., P. Mugnier, A. K. Das, H. M. Webb, D. R. Evans, M. F. Tuite, B. A. Hemmings & D. Barford (2000) The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell,* 100**,** 311-21.

Spahn, C. M., R. Beckmann, N. Eswar, P. A. Penczek, A. Sali, G. Blobel & J. Frank (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. *Cell,* 107**,** 373-86.

Taylor, D. J., J. Nilsson, A. R. Merrill, G. R. Andersen, P. Nissen & J. Frank (2007) Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J,* 26**,** 2421-31.

Valle, M., A. Zavialov, W. Li, S. M. Stagg, J. Sengupta, R. C. Nielsen, P. Nissen, S. C. Harvey, M. Ehrenberg & J. Frank (2003) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol,* 10**,** 899-906.

Wills, N. M., M. O'Connor, C. C. Nelson, C. C. Rettberg, W. M. Huang, R. F. Gesteland & J. F. Atkins (2008) Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *EMBO J,* 27**,** 2533-44.

Zaher, H. S. & R. Green (2009) Fidelity at the molecular level: lessons from protein synthesis. *Cell,* 136**,** 746-62.

Chapitre 1

Etude du rôle d'eEF2 dans le maintien du cadre de lecture.

Sommaire

I. Introduction 55
1. Le facteur d'élongation eEF2 55
A. Les changements structuraux au cœur de la fonction d'eEF2 55
B. Les modifications post-traductionnelles présentes dans eEF2. 56
2. Le contexte du projet 57
II. Résultats 59
1. Le projet de mutagénèse aléatoire d'eEF2 59
2. Construction de la souche eft1Δ eft2Δ et des plasmides pour la mutagénèse aléatoire. 60
A. La souche eft1Δ eft2Δ 60
B. Les plasmides pour la mutagénèse aléatoire d'EFT2 63
3. Mise au point des conditions de mutagénèse aléatoire d'*EFT2* 63
4. Mise au point du crible génétique 64
5. Test du mutant eft2 (H699N) 65
III. Discussion 67
1. Construction de la souche FY1619 ; *eft2Δ::KanR* ; *eft1Δ::TRP1* ; pFL38-EFT2 67
2. Le rôle d'eEF2 dans le décalage du cadre de lecture en -1 68
3. Le mutant eEF2 (H699N) : diphthamide et recodage 68
Références bibliographiques 70

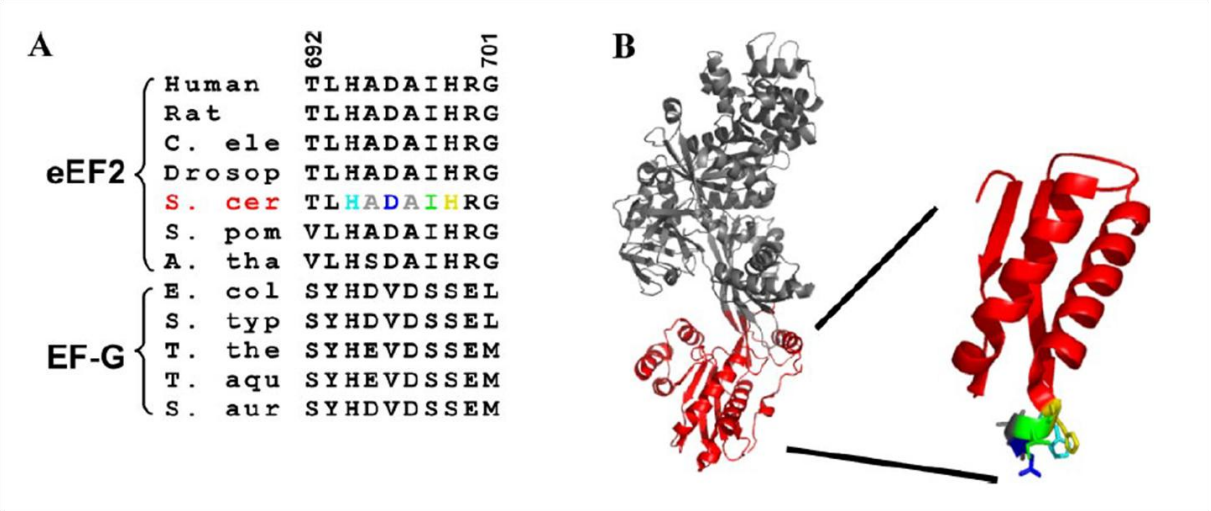


Figure 1 : Conservation du domaine IV d'eEF2

A) Alignement de séquence des acides aminés 692 à 701 d'eEF2 chez les eucaryotes et des acides aminés d'EF-G correspondants chez les bactéries. **B)** Structure d'eEF2 avec une vue détaillée du domaine IV (rouge). Les acides aminés 692 à 701 sont situés à la pointe d'eEF2. Figure issue de (Ortiz et al. 2006).

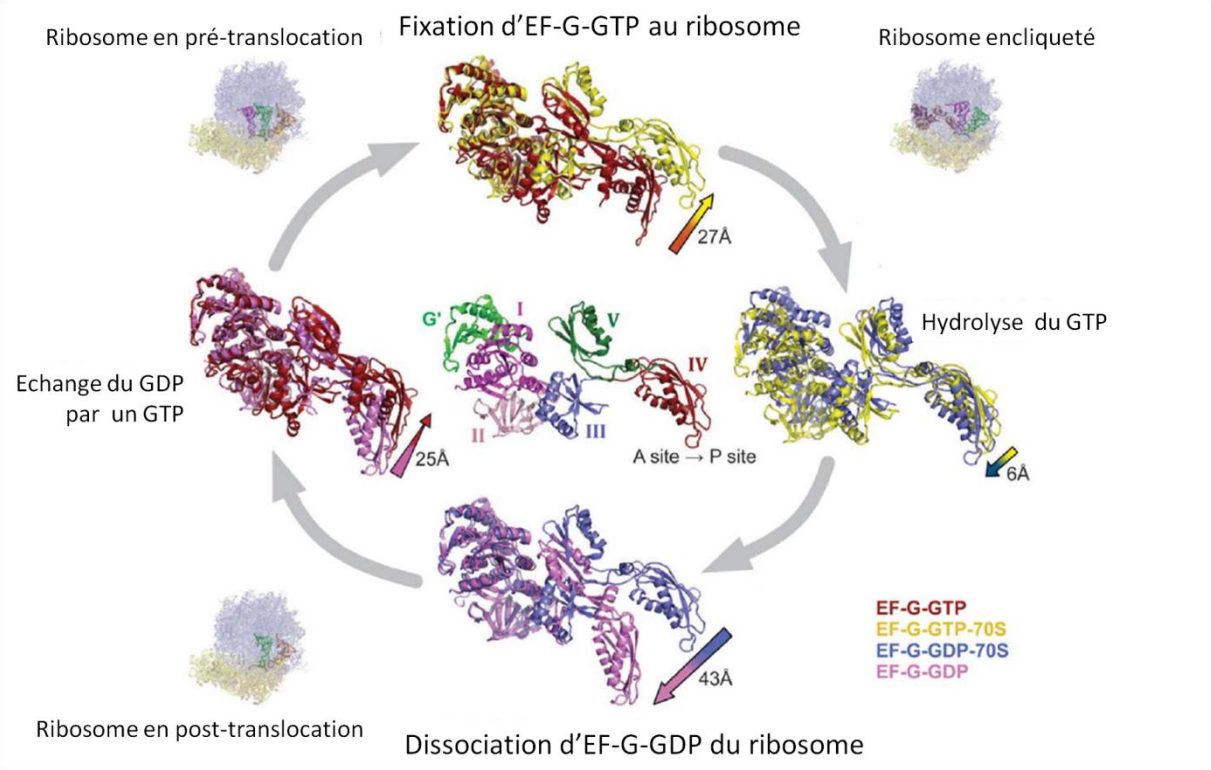


Figure 2 : Différentes conformations d'EF-G au cours du cycle de translocation

La flexibilité d'eEF2 est essentielle à son rôle dans l'élongation. Au cours d'un cycle de translocation, les domaines IV et V se réorientent par rapport aux autres domaines. Le domaine IV est celui qui est le plus mobile. Figure adaptée de (Li et al. 2011)

I. Introduction

Ce 1er chapitre porte sur une étude du facteur d'élongation eEF2 dans le maintien du cadre de lecture. Afin d'introduire ce projet que j'ai développé au cours de ma thèse, je vais présenter plus en détails la protéine eEF2 et resituer le contexte scientifique.

1. Le facteur d'élongation eEF2

Le facteur d'élongation eucaryote eEF2 et son homologue bactérien EF-G catalysent la translocation du ribosome lors d'un cycle d'élongation.

Chez *S. cerevisiae*, eEF2 est codé par 2 copies du même gène : *EFT1* (YOR133W) et *EFT2* (YDR385W) (Perentesis et al. 1992). La présence d'au moins une de ces 2 copies est essentielle pour la survie de la cellule. La transcription d'*EFT1* et d'*EFT2* est régulée par des facteurs de transcription différents. En phase exponentielle de croissance dans des conditions de culture standard, la transcription issue d'*EFT2* représente 70% des ARNm codant eEF2 (Veldman, Rao and Bodley 1994).

A. Les changements structuraux au cœur de la fonction d'eEF2

Le facteur eEF2 est une protéine de 842 acides aminés. Il comprend 6 domaines structuraux : le domaine I ou domaine G, le domaine G' et les domaines II à IV (Jørgensen et al. 2003). Le domaine G (GTPase) forme une poche de fixation pour la molécule de GTP ou de GDP (figure 1).

Dans le ribosome 80S, les domaines I et V d'eEF2 interagissent avec la sous-unité 60S. En particulier, la poche de fixation du GTP contacte la boucle Sarcine-Ricine (SRL) de l'ARNr 28S (Spahn et al. 2004). Cette boucle SRL servirait seulement de site d'ancrage pour les GTPase et n'aurait donc pas de rôle direct dans l'hydrolyse du GTP (Shi et al. 2012). Le domaine II d'eEF2 interagit avec la sous-unité 40S alors que les domaines III et IV rentrent dans le ribosome et interagissent avec les 2 sous-unités ribosomiques (Spahn et al. 2004).

Très peu d'études structurales sont disponibles pour eEF2. L'essentiel des informations sur les changements structuraux associés à ce facteur lors de la translocation a été apporté par l'étude du facteur bactérien EF-G.

Lors de la découverte de l'encliquetage des sous-unités ribosomiques après la transpeptidation, il avait été proposé qu'EF-G induisait cette conformation (Frank and Agrawal 2000, Valle et al. 2003). Cette vision a été abandonnée après l'observation de rotations spontanées des sous-unités et des états hybrides des ARNt en absence d'EF-G (Cornish et al. 2008, Agirrezabala et al. 2008, Julián et al. 2008). L'hypothèse actuellement favorisée est qu'il permettrait de stabiliser la conformation « encliqueté » pour favoriser la translocation.

D'après les différentes structures obtenues à différents stades de la translocation, il apparaît clairement que la fonction d'EF-G repose sur sa flexibilité (figure 2). Le mouvement le plus prononcé est la rotation du domaine IV par rapport au reste de la molécule (Li et al. 2011). Ce changement se produit avant la translocation des ARNt et de l'ARNm (Savelsbergh et al. 2003). L'importance de la

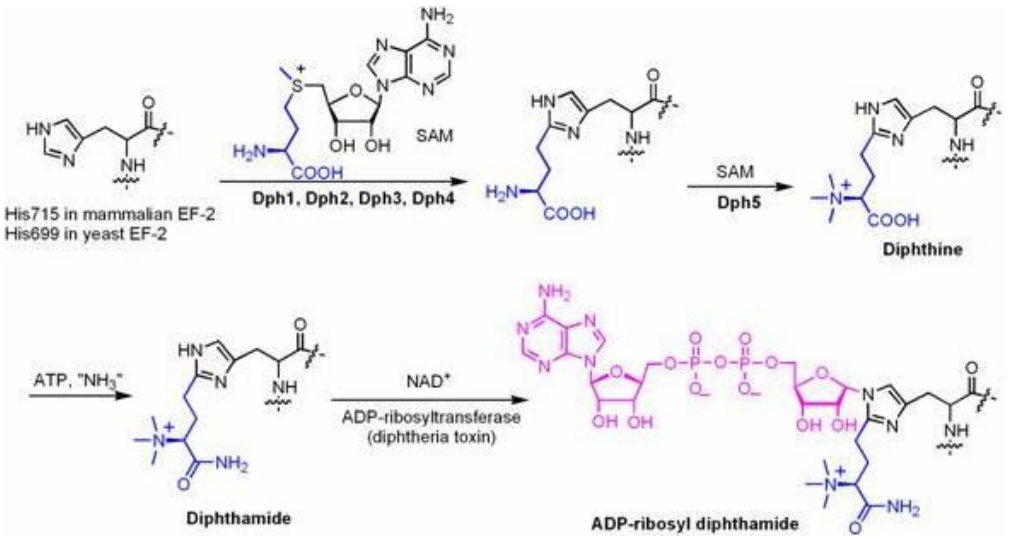


Figure 3 : Diphthamide et ADP ribosylation

L'ADP-ribosylation d'eEF2, par exemple par la toxine diphtérique, entraîne l'ajout d'une grande molécule (en rose) sur le diphthamide. Figure issue de http://www.chem.cornell.edu/hl379/Research.html

flexibilité a été démontrée en accrochant chimiquement les domaines I et V d'EF-G. L'hydrolyse du GTP était alors conservée mais l'échange du GDP ainsi que la translocation étaient abolis (Peske et al. 2000). Ce résultat a été confirmé chez les eucaryotes où une étude structurale du ribosome a montré que le changement de conformation d'eEF2 était dû à l'hydrolyse du GTP (Taylor et al. 2007).

La présence du domaine IV est un des éléments nécessaires à la translocation. Une délétion du domaine IV n'affecte pas l'activité GTPase d'EF-G. Par contre, la translocation se produit 1000 fois plus lentement et à la fin EF-G ne se détache pas du ribosome (Rodnina et al. 1997). De plus, une mutation de l'histidine conservée H583 à la pointe du domaine IV diminue d'un facteur 100 l'efficacité de la translocation (Savelsbergh et al. 2000). Dans le ribosome, cette pointe du domaine IV est située dans le centre de décodage. L'interaction entre la pointe du domaine IV et l'hélice codon:anticodon n'a jamais été encore clairement établie (Gao et al. 2009). La proximité du domaine IV avec le centre de décodage et l'hélice codon anticodon (Spahn et al. 2004, Taylor et al. 2007) a conduit à l'hypothèse que la pointe pouvait prévenir un décalage du cadre de lecture, en particulier en -1, en maintenant l'appariement codon:anticodon de l'ARNt peptidyl (A/P) lors de la translocation (Ortiz et al. 2006).

Bien que les mécanismes de traduction soient conservés entre les eucaryotes et les bactéries, les protéines EF-G et eEF2 ne sont pas totalement équivalentes. Le facteur eEF2 possède 2 modifications post-traductionnelles qui vont moduler son activité. Le travail sur EF-G est donc un modèle limité pour la compréhension des modifications d'eEF2 dans la traduction chez les eucaryotes. Je vais à présent décrire ces 2 modifications et leur rôle dans la régulation d'eEF2.

B. Les modifications post-traductionnelles présentes dans eEF2.

Le facteur eEF2 est modifié post-traductionnellement par l'addition d'un diphthamide et parfois par l'ajout d'un groupement phosphate.

❖ *Le diphthamide*

Le facteur eEF2 possède un résidu diphthamide sur une histidine à la pointe du domaine IV. Chez *S. cerevisiae*, il s'agit de l'histidine en position 699 (H699), et chez les mammifères de l'histidine à la position 715 (Jørgensen et al. 2003, Ortiz et al. 2006) (figure 1). La présence du diphthamide est conservée parmi les eucaryotes et est aussi retrouvée chez les archées (Pappenheimer et al. 1983). L'ajout du diphthamide sur eEF2 nécessite l'activité concertée de 5 enzymes codées par les gènes *DPH* (DiPHthamide biosynthesis) 1 à 5 (Liu et al. 2004). Une 6ème enzyme codée par *DPH6* semblerait aussi impliquée (communication de Uthman S, et al. (2012) au site SGD).

Le diphthamide est la cible de certaines toxines, comme la toxine diphtérique et l'endotoxine A, qui vont l'ADP-ribosyler. L'ADP-ribosylation d'eEF2 est létale chez les eucaryotes, son mécanisme d'action est inconnu. La taille du motif qui est ajouté pourrait créer un encombrement stérique à la pointe du domaine IV, ce qui serait incompatible avec la fonction d'eEF2 (figure 3).

Le diphthamide est impliqué dans diverses fonctions biologiques. Une étude l'impliquerait dans une résistance indirecte à la toxine ricine qui inhibe la traduction en dépurinant une base de la boucle SRL (Gupta et al. 2008). Une autre étude suggère que sa présence confère une affinité plus grande de la

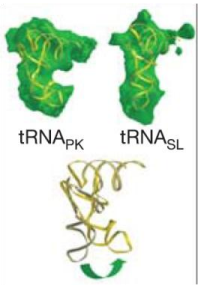


Figure 4 : Conformation de l'ARNt dans un ribosome pausant au pseudonoeud du site de recodage du virus IBV

Dans un ribosome pausant au pseudonoeud (PK) du site de recodage du virus IBV, l'ARNt au site P est déformé (en haut à gauche). Comparé à un ARNt control (SL, à droite), la tige anticodon de l'ARNt est pliée en direction du site A. Figure issue de (Namy et al. 2006)

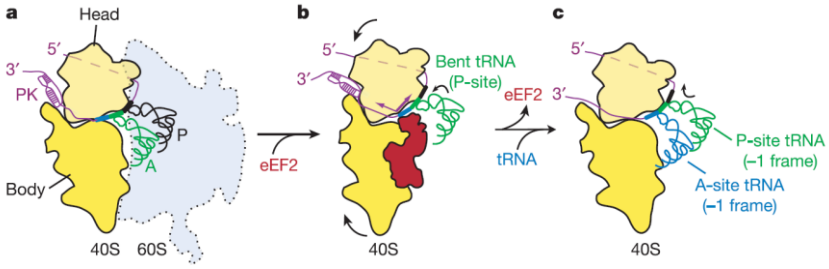


Figure 5 : Modèle mécanique du rôle du pseudonoeud et d'eEF2 dans le décalage du cadre de lecture en -1

Le modèle est détaillé dans le paragraphe I.2 . Figure issue de (Namy et al. 2006)

sordarine, une molécule antifongique, pour sa cible eEF2 (Botet et al. 2008). Enfin, une dernière étude corrèle la présence du diphthamide avec l'expression de gènes exprimés précocement lors du développement embryonnaire de la souris (Webb et al. 2008). Ces études ne permettent cependant pas d'établir un rôle précis pour le diphthamide.

Une substitution de l'histidine à la position 699 (H699) par un des 19 autres acides aminés conduit le plus souvent à une protéine fonctionnelle compatible avec une viabilité cellulaire (Phan, Perentesis and Bodley 1993, Kimata and Kohno 1994). Cependant, les mutations H699G, H699P, H699S et H699D donnent des protéines non fonctionnelles conduisant à une létalité des cellules. Dans le cas de ces mutants, le phénotype est récessif. L'inactivation ne résulte pas d'une baisse de fixation au ribosome ni d'une baisse de l'activité GTPase. D'autre part, les mutations H699R et H699K sont létales *in vivo* et ce phénotype est dominant négatif comme dans le cas où eEF2 est ADP-ribosylé. (Kimata and Kohno 1994, Ortiz et al. 2006). D'autres mutations dans la boucle II à la pointe du domaine IV peuvent abolir la biosynthèse du diphthamide et la fonction catalytique d'eEF2 (Foley, Moehring and Moehring 1995, Ortiz et al. 2006). Ces observations indiquent que le diphthamide n'est pas essentiel pour la fonction d'eEF2 (Phan et al. 1993, Ortiz and Kinzy 2005) et que la nature du résidu à cette position peut changer totalement l'activité d'eEF2.
Les souches avec les mutants eEF2 I698A et H699N, ainsi que les souches Δ*dph2* et Δ*dph5,* ont une efficacité de décalage en -1 plus grande sur la séquence glissante **G GGT TTA** du virus L-A de *S. cerevisiae*. Par contre, l'efficacité de décalage en +1 ne change pas sur la séquence **CTT AGG C** du rétrotransposon Ty1. L'efficacité de translecture sur les 3 stops n'est pas non plus affectée (Ortiz et al. 2006). La pointe du domaine IV, en particulier le diphthamide, semble donc avoir un rôle dans le maintien du cadre de lecture lors de la translocation.

- ***Phosphorylation de la Thr 57.***

Chez *S. cerevisiae*, la thréonine 57 (domaine I) est phosphorylée par la Ser/Thr kinase Rck2p (Teige et al. 2001). Chez les mammifères, cette modification est réalisée par la kinase eEF2k dépendante du Ca^{2+} et de la calmoduline (Nairn and Palfrey 1987). La phosphorylation réduit l'affinité d'eEF2 pour le GTP mais pas pour le GDP. Elle décroît aussi la fixation d'eEF2 sur le ribosome (Dumont-Miscopein et al. 1994). La phosphorylation ne semble pas impliquée dans la fidélité de la traduction.

2. Le contexte du projet

Une cartographie par cryo-microscopie électronique (cryoEM) de ribosomes pausant sur une structure secondaire du virus IBV (Infectious Bronchitis Virus) a été publiée en 2006. La structure secondaire employée est impliquée dans la stimulation du décalage en -1 sur la séquence glissante **U UUA AAC**. Dans l'étude, cette séquence a été remplacée par la séquence non glissante **C GAG GCA**. Les auteurs ont montré qu'eEF2 était présent dans les ribosomes, ces derniers étant dans un état « encliqueté ». Outre la présence d'eEF2, les autres éléments importants étaient la position et la conformation de l'ARNt au site P. L'ARNt est haut dans le site P et sa tige anticodon est courbée en direction du site A (Namy et al. 2006) (figure 4).

Le mécanisme proposé alors pour expliquer le décalage en -1 était qu'il existerait 2 forces antagonistes : la résistance du pseudonoeud à l'ouverture et la présence d'eEF2 qui pousserait les ARNt lors de la translocation. Ces 2 tensions induiraient une torsion dans l'ARNt, ce qui figerait le

système dans un état de pause. La solution pour résoudre cette situation serait de casser la liaison codon:anticodon. L'ARNt reprendrait alors une conformation normale et se rapparierait dans la phase -1 (Namy et al. 2006) (figure 5). En 2010, les cartes ont été affinées. Les résultats sont décrits dans le chapitre I au point V.2.b.

Les données de cryoEM, ainsi que celles sur l'augmentation du décalage en -1 observée pour des mutants du domaine IV d'eEF2, nous ont conduits à nous poser la question suivante : **quel est le rôle du facteur d'élongation eEF2 dans le maintien du cadre de lecture ?**

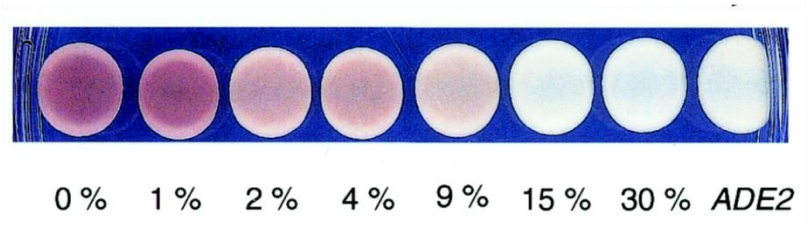


Figure 6 : Corrélation entre l'efficacité de recodage du gène rapporteur ADE2 et la couleur des levures.

Lorsqu'un site de recodage est inséré dans le gène rapporteur ADE2, il existe une corrélation entre l'efficacité de recodage de la séquence et la couleur de la souche. Figure issue de (Namy, Hatin and Rousset 2001)

II. Résultats

1. Le projet de mutagénèse aléatoire d'eEF2

Afin d'identifier des régions d'eEF2 impliquées dans le maintien du cadre de lecture, nous avons décidé de réaliser une mutagénèse aléatoire d'*EFT2* chez *Saccharomyces cerevisiae* car de nombreux outils génétiques et moléculaires sont disponibles pour cette espèce. La première étape a été de construire une souche délétée pour les copies endogènes *EFT1* et *EFT2*. Ces gènes étant essentiels, une copie d'*EFT2* a été apportée sur un vecteur pour permettre la viabilité des cellules.

La phase codante d'*EFT2* fait 2529nt. Une mutagénèse complète du gène où chaque position serait substituée par une des 3 autres bases conduirait idéalement à créer (2529 x 3) combinaisons soit 7587 mutants. Dans les conditions de l'expérience, nous avons décidé d'analyser 100 fois plus de candidats soit environ 700 000 candidats, pour s'assurer que la mutagénèse soit bien saturante. Face à un nombre aussi élevé, il est important d'isoler efficacement, grâce à un crible, les mutants diminuant ou augmentant l'efficacité de décalage du cadre de lecture.

L'utilisation des signaux d'ARNm induisant un décalage en -1 est le moyen le plus approprié pour isoler des mutants d'eEF2 affectant le maintien du cadre de lecture. Contrairement au décalage du cadre de lecture spontané, la fréquence de l'événement de recodage est bien plus élevée. Ce point est important car on ne s'attend pas à trouver un nombre élevé de mutants. Donc si la limite de détection est trop basse, ceux-ci seront difficiles à détecter. De plus, le site de recodage peut être inséré dans un gène rapporteur et son efficacité de recodage peut être modulée comme nous le verrons par la suite. Enfin, il a été montré que le facteur eEF2 était directement impliqué dans le décalage en -1. Nous avons donc décidé de travailler avec le site de décalage en -1 du virus IBV. La séquence glissante et le pseudonoeud ont été introduits dans le gène rapporteur *ADE2*. Les mutants d'eEF2 diminuant ou augmentant l'efficacité de décalage sont sélectionnés dans 2 cribles différents basés sur le même gène rapporteur.

Le gène *ADE2* code une P-ribosyl-amino-imidazole-carboxylase. Son inactivation conduit à une accumulation d'un pigment rouge dans la vacuole des cellules. Lors d'une étude sur la translecture, un codon stop a été introduit dans le 1er tiers de la phase codant d'*ADE2* au site de restriction HpaI. Il a été montré que la couleur des colonies de levures dépendait alors de l'efficacité de translecture (figure 6). En dessous de 2% de translecture, il n'y a pas assez de protéines synthétisées et les colonies sont [rouges]. Entre 2 et 10%, les colonies présentent différents degrés de [rose]. Au dessus de 15% de translecture, les colonies sont [blanches] (Namy, Hatin and Rousset 2001). De la même manière, des sites de glissement avec des efficacités de décalage en -1 différentes ont été clonés dans *ADE2* afin de réaliser un crible blanc/rouge que je vais à présent détailler.

Suite à une mutagénèse aléatoire du gène *EFT2*, les copies mutantes *eft2** portées par un vecteur d'expression sont utilisées pour transformer des cellules initialement [blanches] ou [rouges]. Le principe est qu'à partir de colonies initialement [blanches] (efficacité de décalage élevée), il est possible de sélectionner des colonies [roses] ou [rouges] après transformation avec *eft2**. Si les phénotypes observés sont bien liés à l'introduction d'*eft2**, on obtiendra des allèles diminuant l'efficacité de décalage. Les allèles eft2* isolés en présence de l'allèle sauvage *EFT2* conféreront un

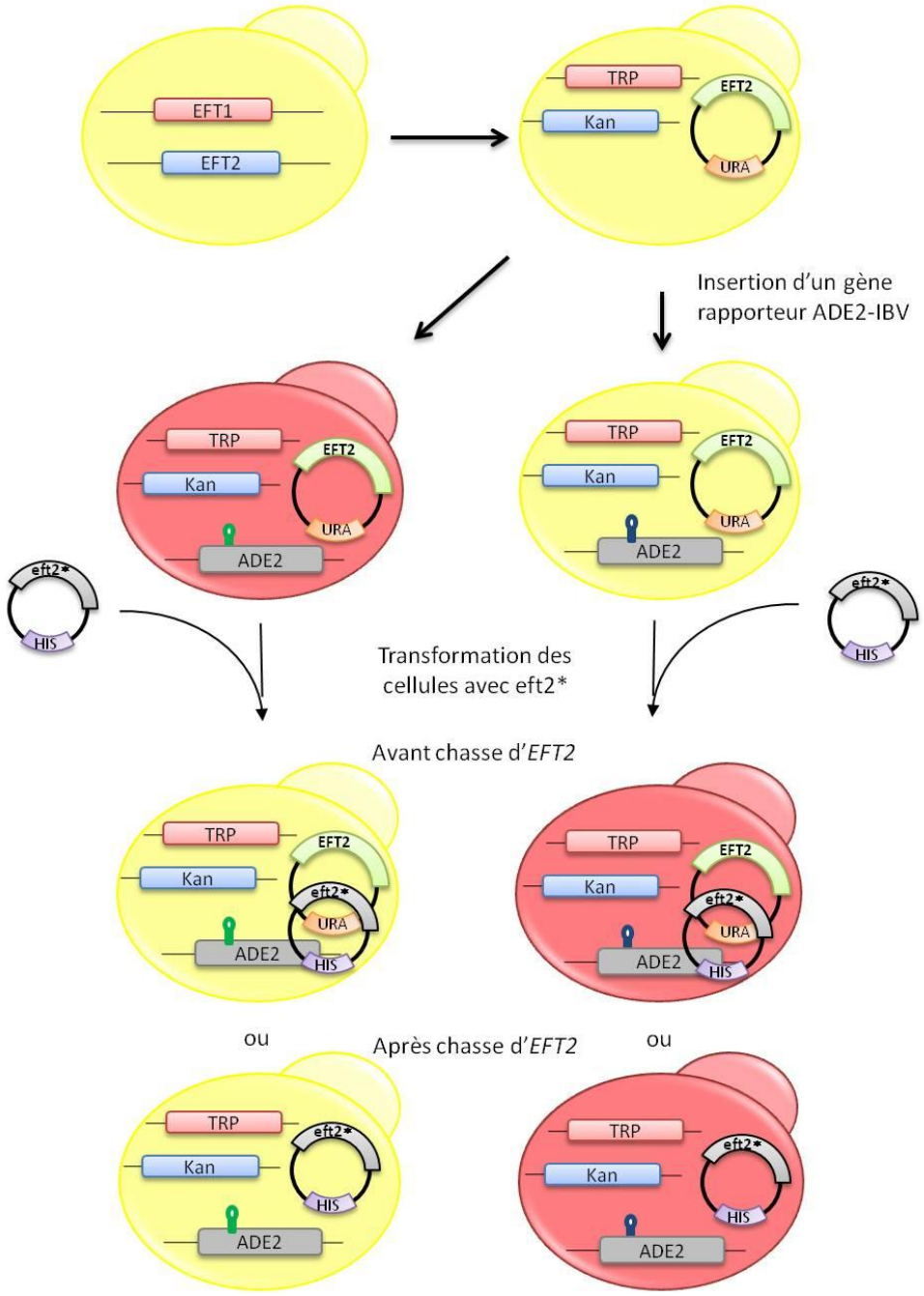


Figure 7 : Schéma présentant le crible génétique pour l'isolation de mutants d'eEF2 affectant le maintien du cadre de lecture.

phénotype dominant alors que ceux isolés après la chasse de l'allèle sauvage conféreront un phénotype récessif par rapport à l'allèle sauvage (figure 7).

A l'inverse, à partir de colonies [rouges], il sera possible d'isoler des allèles d'*eft2** qui vont augmenter l'efficacité de décalage. En fonction de la présence ou de l'absence de l'allèle sauvage *EFT2*, des allèles conférant des phénotypes dominants ou récessifs par rapport à l'allèle sauvage pourront être isolés.

La position des différentes mutations dans la protéine eEF2 devrait permettre d'identifier des régions impliquées dans le maintien du cadre de lecture. Elles seront alors étudiées plus en détails par des analyses génétiques et biochimiques appropriées.

2. Construction de la souche eft1Δ eft2Δ et des plasmides pour la mutagénèse aléatoire.

A. La souche eft1Δ eft2Δ

Afin de réaliser le crible, la stratégie initiale était de construire une souche de levure dans laquelle les gènes *EFT1* et *EFT2* sont délétés. Le fond génétique utilisé est celui de la levure haploïde *S. cerevisiae* FY1679. Comme le facteur d'élongation eEF2 est essentiel pour la cellule, la souche doit être complémentée par un plasmide portant une copie sauvage du gène codant eEF2 (figure 7). Pour cela, j'ai choisi d'exprimer le gène *EFT2* sous le contrôle de son propre promoteur car il est le plus exprimé *in vivo*. La copie est portée sur un vecteur pRS316 possédant une origine de réplication centromérique, ce qui permet d'avoir un faible nombre de copies du plasmide dans la cellule. D'autre part, le pRS316 a une cassette de sélection *URA3*. Cette cassette de sélection permet une contre-sélection du plasmide sur un milieu contenant du 5-fluoroorotic acid (5-FOA). Le 5-FOA est converti par le produit du gène *URA3* en 5-fluorouracil, un produit toxique pour la cellule.

Les étapes de la construction de la souche sont :

1. Délétion du gène *EFT1* par la cassette de sélection *TRP1* et sélection des transformants sur milieux YNB[1] CSM[2]-Trp.
2. Transformation des cellules *eft1Δ::TRP1* avec le pRS316-EFT2 et sélection des transformants sur milieu YNB CSM-Ura.
3. Délétion du gène *EFT2* par la cassette de résistance Kan^R et sélection des transformants sur milieux YPG[3] + G418. Afin d'éviter l'insertion du plasmide pRS316-EFT2 et une recombinaison homologue de la séquence *eft2** au locus EFT2, j'ai pris soin de déléter une région chromosomique plus grande que celle présente dans le plasmide ou la séquence mutante.
4. Test de l'essentialité du plasmide pRS316-EFT2 sur milieux YNB CSM-Ura +5-FOA pour vérifier que la copie du gène *EFT2* apportée par le vecteur est la seule présente dans la cellule.

[1] YNB (Yeast Nitrogen Base) est un milieu minimum pour la croissance de *S. cerevisiae*. Il doit être complété en acides aminés en fonction des auxotrophies des cellules.
[2] CSM ensemble des acides aminés ajoutés au milieu YNB. Le ou les acides aminés indiqués après sont ceux absent du milieu.
[3] YPG (Yeast , Peptone , Glucose) est un milieu riche.

La construction de la souche a été plus compliquée que prévu. J'ai dû mettre en œuvre différentes stratégies pour réussir cette construction. Je ne donnerai que les grandes lignes de chacune d'elles en expliquant le lien avec l'étape suivante. Seule la stratégie finale sera détaillée dans la partie « Matériels et Méthodes ».

Dans le fond génétique FY1619 :

- *Délétion d'EFT1 avec la cassette de sélection TRP1, complémentation avec le plasmide pRS316-EFT2 et délétion d'EFT2 avec la cassette de résistance Kan^R.*

  La région délétée par la cassette de sélection *TRP1* inclut *EFT1* ainsi que 40pb en amont et 18pb en aval. Les zones d'homologie de séquence pour la recombinaison homologue sont de 60pb de part et d'autre de la cassette.

  Le plasmide pRS316 porte le gène *EFT2* sous le contrôle d'une région promotrice de 400pb.

  La région délétée par la cassette de résistance *Kan^R* comprend *EFT2* ainsi que 400pb de région promotrice et 52pb de région terminatrice. Les zones d'homologie de séquence pour la recombinaison homologue sont de 526pb en 5' de la cassette et 423pb en 3'. De larges zones d'homologie de séquence permettent une recombinaison homologue plus efficace au locus d'un gène fortement exprimé.

  Tous les clones obtenus après sélection sur milieu YPG+G418 poussaient aussi sur milieu contenant du 5-FOA. Cette croissance indique que le pRS316-EFT2 peut être perdu et donc qu'une copie *d'EFT2* est encore présente dans les cellules. Ces clones étaient donc des faux positifs.

  Comme l'utilisation de la cassette *Kan^R* conduit à de nombreux faux positifs et que la délétion *d'EFT2* paraît plus difficile que celle d'*EFT1*, j'ai changé la stratégie de construction en délétant *EFT2* en premier. Là encore, tous les clones finalement obtenus étaient des faux positifs. L'hypothèse était alors que la complémentation du pRS316-EFT2 était insuffisante, j'ai donc construit un nouveau plasmide de complémentation.

- *Délétion d'EFT2 avec la cassette de résistance Kan^R, complémentation avec le plasmide pYX212-EFT2 et délétion d'EFT1 avec la cassette de sélection TRP1.*

  L'objectif est ici de construire un plasmide permettant une expression maximale d'EFT2 afin d'être certain que ce plasmide pourra complémenter la souche délétée. Le plasmide pYX212-EFT2 est un plasmide d'expression multicopie avec une origine de réplication 2µm. *EFT2* est exprimé sous le contrôle du promoteur et terminateur *TPI1*. Le promoteur *TPI1* est un promoteur fort, le gène *TPI1* (Triose-Phosphate Isomerase 1) est l'un des plus exprimé dans la cellule.

  Suite aux délétions d'*EFT2* et d'*EFT1*, tous les clones obtenus après sélection sur milieux YNB CSM-Trp étaient des faux positifs.

  Les régions d'homologie de séquence avec EFT1 pouvant être trop courtes, j'ai réessayé avec des régions d'homologie plus grandes. Malheureusement, même en ayant allongé les régions d'homologie pour améliorer l'efficacité de recombinaison, je n'ai pas obtenu de clone positif.

Puisque ces stratégies de délétions séquentielles ont échoué, j'ai opté pour une stratégie dans laquelle la souche double délétée serait obtenue par croisement de souches délétées individuellement pour chacun des gènes. Comme il n'est pas possible d'obtenir des spores à partir de diploïdes FY1619 car cette souche sporule très difficilement, j'ai utilisé les souches de la banque de délétion systématique Euroscarf dans le fond génétique BY4742.

Dans le fond génétique BY4742 (souches Euroscarf):

- *Croisement des souches Y12389 (eft1Δ::Kan^R) et Y04221 (eft2Δ::Kan^R).*
  Les souches Y12389 et Y04221 ont été croisées afin d'obtenir des diploïdes. Ces diploïdes ont été transformés avec le pYX212-EFT2 et mis à sporuler. Des tétrades de spores ont été disséquées de manière ordonnée et mises à pousser sur milieu YPG+G418. Les gènes *EFT1* et *EFT2* sont indépendants, donc d'après les règles de ségrégation mendélienne, les spores ayant les 2 délétions pousseront dans les tétrades ditypes recombinés et tétratypes. En présence de G418, 2 spores sur 4 (2:2) pousseront pour les ditypes recombinés et auront le génotype attendu. Pour les tétratypes, 3 spores pousseront (3:1) mais seulement une aura le bon génotype.
  Finalement, toutes les colonies provenant de tétrades (2:2) sont viables sur 5-FOA. Il est probable qu'il s'agisse en fait de tétrades (3:1) dans lesquelles la souche double délétée ne serait pas viable à cause d'un problème lié au pYX212-EFT2. Aucun clone positif n'a donc pu être isolé.

  Puisque cette stratégie s'est révélée infructueuse, j'en ai conclu que le problème venait de la complémentation avec le plasmide. La quantité d'eEF2 est très régulée in vivo, une surexpression peut être tout aussi délétère qu'une sous-expression. J'ai donc reconstruit 2 nouveaux plasmides portant *EFT2* sous le contrôle de 990bp de son propre promoteur.

Dans le fond génétique FY1619 :

- *Délétion d'EFT2 avec la cassette de résistance Kan^R, complémentation avec le plasmide pFL38-EFT2 ou pFL44L-EFT2 et délétion d'EFT1 avec la cassette de sélection TRP1.*
  Afin d'inclure tous les signaux de transcription *d'EFT2*, le gène avec 990pb de région promotrice et 572pb de région terminatrice a été inséré dans les plasmides pFL38 et pFL44L. Ces 2 plasmides ont une cassette de sélection *URA3* mais diffèrent par leur origine de réplication. Le pFL38 est un plasmide centromérique et le pFL44L est un plasmide 2µm (multicopie).
  Après sélection sur milieu YNB CSM-Trp, seules les colonies avec le plasmide pL38-EFT2 ont poussé. Comme attendu dans le cas de clones positifs, ces colonies ne sont pas viables sur milieu YNB CSM-Ura+5-FOA. Les délétions ont aussi été vérifiées par PCR sur colonies.

Finalement, j'ai réussi à construire la souche HC07 (FY1619 ; *eft2Δ::KanR* ; *eft1Δ::TRP1* ; pFL38-EFT2) par une délétion séquentielle des gènes EFT2 et EFT1. Le plasmide pFL38-EFT2 incluant de grandes régions régulatrices d'EFT2 a permis de complémenter la souche en exprimant une quantité d'eEF2 compatible avec la viabilité cellulaire. Avec ce plasmide, il existe néanmoins un risque de recombinaison homologue entre les régions régulatrices d'*EFT2* et le locus chromosomique *d'EFT2*.

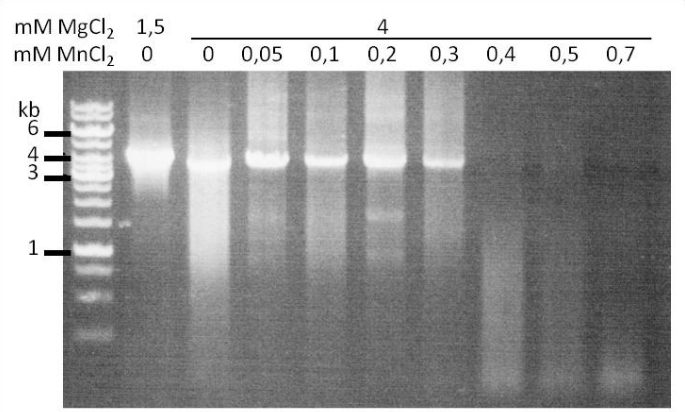


Figure 8 : Mise au point des conditions de mutagénèse aléatoire d'*EFT2*

Produits d'amplification par PCR de $Prom_{PGK1}$-*EFT2*-Ter_{PGK1} obtenus avec différentes concentrations en $MgCl_2$ et $MnCl_2$. Le premier puits correspond à une amplification dans des conditions standard (1,5mM $MgCl_2$ et 0mM $MnCl_2$).

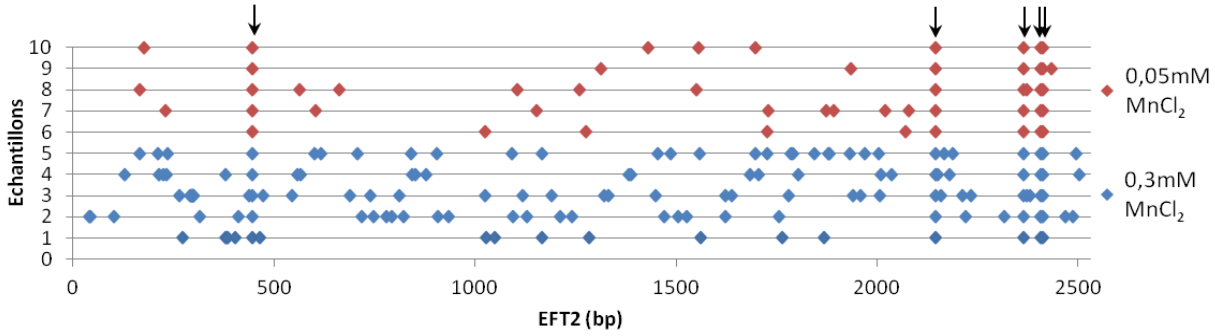


Figure 9: Détermination du taux de mutagénèse d'EFT2

Représentation de la position des mutations indentifiées pour chacun des gènes *eft2**. Les échantillons 1 à 5 (bleu) ont été obtenus suite à une mutagénèse aléatoire par PCR avec 4mM $MgCl_2$ et 0,3mM $MnCl_2$; et les échantillons 6 à 10 (rouge) avec 4mM $MgCl_2$ et 0,05mM $MnCl_2$.

B. Les plasmides pour la mutagénèse aléatoire d'EFT2

Afin de réaliser une mutagénèse aléatoire *d'EFT2* par PCR, j'ai construit 2 plasmides.

Le 1er est le pUC19-PGK-EFT2 dans lequel *EFT2* est sous le contrôle du promoteur et du terminateur *PGK1*. Le promoteur *PGK1* est un promoteur fort. Ce plasmide sert de matrice ADN pour amplifier le fragment $Prom_{PGK1}$-*EFT2*-Ter_{PGK1} lors de la PCR.

Le 2nd est le pFL44H-PGK. Il s'agit d'un plasmide multicopie (2µm) qui porte une cassette de sélection *HIS3*. Il sert de plasmide receveur pour exprimer *eft2* in vivo*. L'insertion *d'eft2** dans le pFL44H-PGK se fera *in vivo* par recombinaison homologue entre les régions promotrice et terminatrice PGK1. Afin que la recombinaison homologue se produise spécifiquement avec le plasmide receveur, il était important que les régions promotrices et terminatrices soient différentes de celles permettant l'expression d'*EFT2* dans le plasmide sauveur pFL38-EFT2.
Afin de s'assurer que le plasmide pFL44H-PGK sera capable d'exprimer eEF2 *in vivo* et d'assurer la viabilité cellulaire, j'ai créé le pFL44H-PGK-EFT2. Les cellules HC07 ont été transformées avec ce plasmide et cultivées sur milieu YNB CSM-His. Les transformants ont ensuite été mis en croissance sur milieu YNB CSM+5-FOA afin de contre sélectionner le plasmide pFL38-EFT2. Toutes les colonies étaient viables, ce qui indiquait que le plasmide pFL38-EFT2 pouvait être perdu et que le plasmide pFL44H-PGK était suffisant pour assurer la viabilité cellulaire. Le plasmide pFL44H-PGK peut donc servir de base pour l'expression *d'eft2**.

3. Mise au point des conditions de mutagénèse aléatoire d'*EFT2*

Pour induire aléatoirement des mutations dans *EFT2*, j'ai mis au point des conditions de PCR comprenant dans la réaction du chlorure de manganèse ($MnCl_2$). Le manganèse Mn^{2+} permet d'induire des erreurs ponctuelles d'incorporation lors de la PCR.

Tout d'abord, j'ai déterminé une gamme de concentrations compatibles avec l'amplification d'un fragment de 3789pb à partir du pUC19-PGK-EFT2. J'ai réalisé une série de PCR avec les concentrations finales en $MnCl_2$ suivantes : 0 ; 0,05 ; 0,1 ; 0,2 ; 0,3 ; 0,4 ; 0,5 et 0,7mM. Afin de stabiliser les mésappariements, j'ai augmenté la concentration en $MgCl_2$ dans la réaction en passant de 1,5mM à 4mM. Les concentrations comprises entre 0,05 et 0,3mM permettent d'obtenir des amplicons de la taille attendue alors que les concentrations supérieures à 0,4mM inhibent complètement la réaction d'amplification (figure 8).

J'ai ensuite déterminé le taux de mutagénèse de la PCR pour les conditions avec 0,05 et 0,3mM de $MnCl_2$. Pour cela, j'ai co-transformé les cellules HC07 avec le produit de PCR et le plasmide pFL44H-PGK linéarisé par digestion avec l'enzyme de restriction NotI. Les transformants ont été sélectionnés sur milieu YNB CSM-His. Afin d'isoler et analyser le plasmide pFL44H-PGK-eft2*, l'ADN de certaines colonies a été extrait et utilisé pour transformer *Escherichia coli*. Pour chacune des conditions, 10 colonies de levures ont été testées. Un profil de digestion avec l'enzyme EcoRI a ensuite été réalisé sur les plasmides extraits d'*E. coli*. Les 10 plasmides analysés étaient des pFL44H-PGK ayant recombiné avec *eft2**. Pour chaque condition de mutagénèse, 5 plasmides ont été séquencés et les mutations analysées (figure 9).

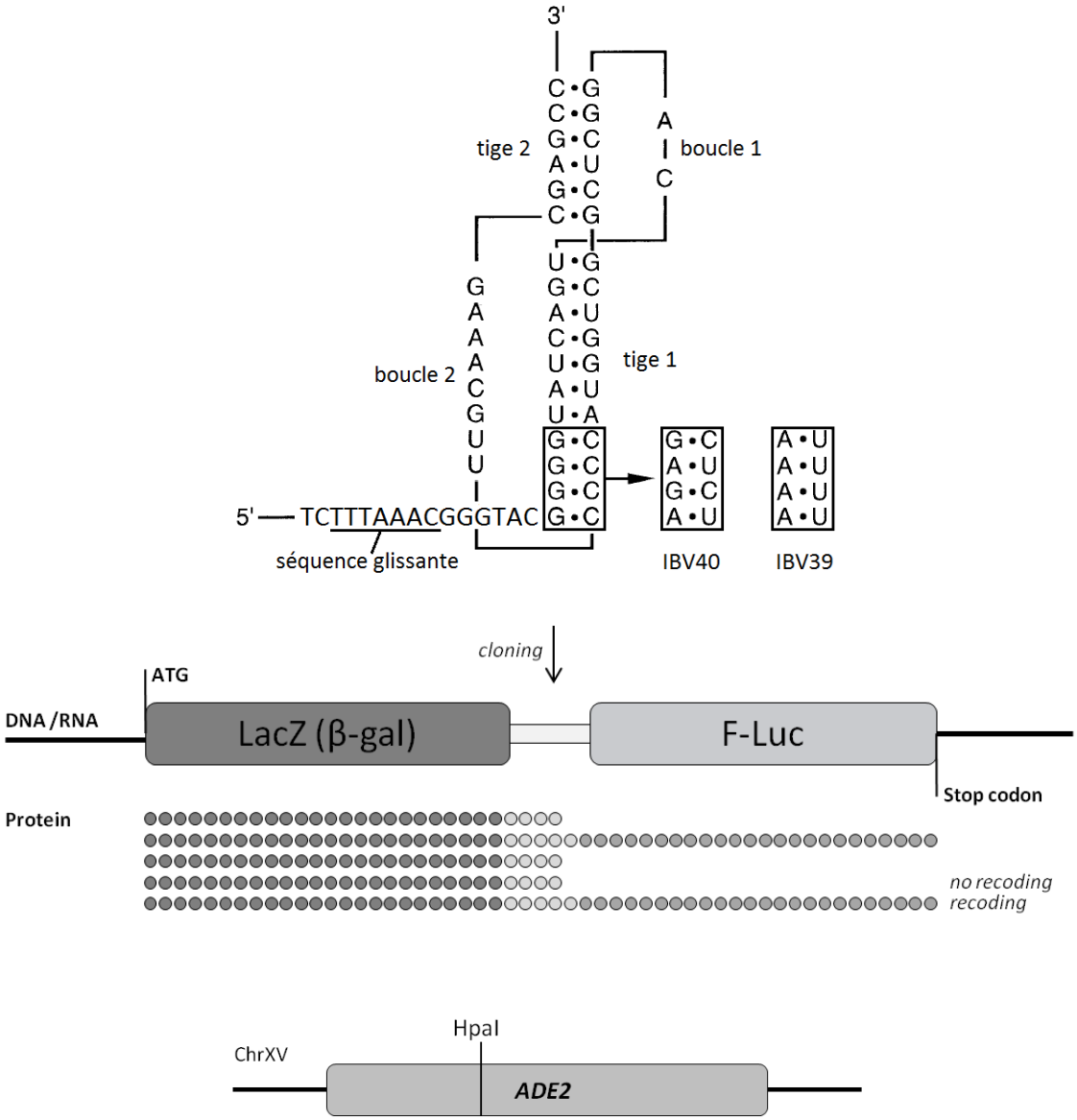


Figure 10 : Séquences de recodage IBV39 et IBV 40

Schéma de la séquence de recodage IBV avec sa séquence glissante (soulignée) et son pseudonoeud. Les mutations réalisées pour obtenir les séquences IBV39 et IBV40 sont indiquées dans les encarts. Ces séquences ont été clonées dans le vecteur double rapporteur pAC-PGK007 entre les phases codantes de lacZ et F-luc afin de mesurer leurs efficacités de recodage respectives. Ces séquences ont aussi été clonées au site de restriction HpaI présent dans le gène *ADE2*.

Le graphique de la figure 9 représente, pour chaque séquence, la position des mutations dans le gène *eft2**. Cette représentation permet d'isoler rapidement des groupes de mutations redondantes. Par exemple, 5 mutations (A447G, C2145T, C2364T, T2406C, et G2412A) sont retrouvées dans toutes les séquences. Ces mutations sont en fait des mutations silencieuses traduisant l'existence d'un polymorphisme entre les 2 souches FY et 74D-694 utilisées lors des constructions. En effet, le pUC-PGK-EFT2 a été construit à partir d'ADN génomique (ADNg) de la souche HFY et les séquences des mutants *eft2** ont été comparées avec celle du pFL44H-PGK-EFT2 dont *EFT2* provient de la souche 74-D694.

L'analyse des mutants montre qu'avec 0,05mM $MnCl_2$, il y a en moyenne 5 mutations (variation comprise entre 4 et 8 mutations) sur l'ensemble du gène avec une fréquence de 2 mutations/kb (variation comprise entre 1 et 3 mutations/kb). Avec 0,3mM $MnCl_2$, on constate en moyenne 24 mutations (variation comprise entre 13 et 33 mutations) sur l'ensemble du gène avec une fréquence de 9 mutations/kb (variation comprise entre 5 et 13 mutations/kb). En plus de mutations ponctuelles, cette condition induit des délétions de quelques nucléotides dans la séquence.

Comme EFT2 est un gène très long, il ne serait pas efficace en terme de temps et de coût d'analyser chaque position individuellement. Il faut donc obtenir assez de mutations pour pouvoir isoler des régions susceptibles d'êtres analysées plus en détails par la suite. A l'inverse, trop de mutations rendraient l'analyse difficile et augmenteraient surtout le nombre de protéines non fonctionnelles au détriment de mutants intéressants. La condition de mutagénèse avec 0,05mM $MnCl_2$ apporte un bon compromis sur le nombre de mutations car elle en induit un nombre limité réparti de manière homogène le long du gène. Cette condition de PCR est donc celle retenue pour la suite.

4. Mise au point du crible génétique

Pour sélectionner les mutants influençant l'efficacité de décalage, j'ai utilisé le gène *ADE2* qui permet de réaliser un crible génétique basé sur la couleur des colonies de levures. Le site de recodage utilisé est celui du virus IBV. Lorsque ce site est inséré au site HpaI du gène *ADE2* (*ADE2-IBV*) les colonies sont blanches. Cette construction a été réalisée précédemment au laboratoire par Agnès Baudin-Baillieu, et utilisée comme contrôle pour un crible similaire sur une protéine ribosomique. Cette construction permet donc la sélection de mutants conduisant à des colonies roses ou rouges. Le taux de recodage associé à l'allèle *ADE2-IBV* est de 14%, ce qui est une valeur élevée comparée aux 10% qui permettent la transition entre les phénotypes [blanc] et [rose]. Les colonies [rose, rouge] observées devraient être, dans la plupart des cas, liées à de rares mutants d'eEF2* qui diminuent fortement l'efficacité de décalage en -1. Dans le but d'isoler un plus grand nombre de mutants avec des effets plus subtils sur le maintien du cadre de lecture, différents mutants du motif de recodage de l'IBV ont été créés par Agnès. Afin de ne pas modifier la nature moléculaire de l'événement, à savoir l'identité des ARNt impliqués dans le glissement, ce n'est pas la séquence glissante qui a été modifiée mais la structure du pseudonoeud (Brierley et al. 1991).

Les mutants IBV39 et IBV40 possèdent des mutations diminuant la stabilité de la 1ère tige à la base du pseudonoeud (figure 10). Ces séquences ont été clonées dans le vecteur double rapporteur pAC-PGK007 (*CEN*, *LEU2*) entre les phases codantes de *lacZ* et *F-luc* (figure 10). La mesure des activités enzymatiques de la β-galactosidase (βgal) et de la luciférase (F-Luc) permet d'estimer l'efficacité de

recodage (Luc/βgal). Un plasmide contrôle pAC-PGK-TQ exprime la fusion βgal-Luc, ce qui permet de normaliser les valeurs obtenues pour les différentes cibles. Le taux de recodage de la séquence IBV39 est de 1,6% et celui d'IBV40 est de 9%. Les séquences ont aussi été insérées au site HpaI du gène *ADE2* dans le pFL38-ADE2 afin de vérifier le phénotype associé à chacune des constructions (figure 10). Ces plasmides ont été utilisés pour transformer une souche (*ade2*-) [rouge]. Les colonies restent rouges avec le pFL38-ade2-IBV39 alors qu'avec le pFL38-ade2-IBV40 les colonies sont blanches. Ces 2 sites de décalage semblent donc parfaitement adaptés pour être utilisés comme systèmes rapporteurs pour le crible génétique.

A partir de la souche HC07, j'ai construit des souches avec les allèles *ade2-ibv39* et *ADE2-IBV40* intégrés au locus *ADE2* par recombinaison homologue. De manière inattendue, le phénotype associé à chacun des allèles était instable. Par exemple, dans le cas de la souche HC07 *ade2-ibv39*, un nombre non négligeable de colonies blanches apparaissaient après repiquage d'une colonie rouge. Certaines colonies rouges présentaient des secteurs blancs indiquant que la transmission du phénotype [blanc] était clonale. Ces colonies blanches ont été repiquées et de nombreuses colonies rouges étaient présentes parmi les blanches indiquant qu'une réversion du phénotype [blanc] était possible. De la même manière, des colonies rouges instables apparaissaient pour la souche HC07 (*ADE2-IBV40*). Afin de vérifier si l'instabilité du phénotype était due au fond génétique, l'allèle ade2-ibv39 a été intégré dans la souche Y349 au locus ADE2 par recombinaison homologue. Le même problème d'instabilité est apparu. La proportion et l'origine inexpliquée de cette instabilité ont rendu ces souches inutilisables pour le crible.

Le site d'insertion HpaI dans *ADE2* a été précédemment utilisé pour un crible sur la translecture (Hatin et al. 2009). Les séquences insérées alors au site HpaI étaient plus courtes que celle du site de recodage de l'IBV. L'ajout de trop d'acides aminés dans Ade2p pourrait entraver sa fonctionnalité. Nous avons donc décidé d'insérer les séquences de recodage IBV au début du gène *ADE2*. Ces différentes constructions ont été réalisées par Isabelle Hatin. Les 4 premiers codons du gène *TPI1* et le site de restriction Eco47III ont été insérés juste en aval du codon d'initiation d'*ADE2*. Les séquences IBV39 et IBV40 ont ensuite été clonées au site Eco47III. Les nouvelles constructions ont été intégrées par recombinaison homologue dans la souche FS1. Les phénotypes associés à ces nouveaux allèles (*ibv39-ade2* et *IBV40-ADE2*) sont stables. La mise au point de ce crible a été bien plus compliquée que prévu mais maintenant tous les éléments sont prêts. A présent, il reste à intégrer ces constructions dans la souche HC07 afin de réaliser le projet sur la mutagénèse aléatoire d'eEF2.

5. Test du mutant eft2 (H699N)

En attendant d'obtenir des constructions stables dans le gène *ADE2*, j'ai utilisé la souche HC07 pour tester l'influence de l'absence du diphthamide sur l'efficacité de recodage de cibles disponibles dans le laboratoire. Auparavant, il a été montré que le mutant eEF2(H699N) augmentait l'efficacité du décalage en -1 sur la séquence de décalage du virus L-A de *S. cerevisiae,* mais ne changeait pas l'efficacité de décalage en +1 du rétrotransposon Ty1 (Ortiz and Kinzy 2005). Les séquences de recodage que j'ai testées sont : TMV (translecture), *OAZ1* (décalage en +1 stimulé par un codon stop), *EST3* (décalage en +1 stimulé par un codon rare), et IBV (décalage en -1).

	Médianes des rapports Luc/B-gal		% de recodage		Mann-Whitney
	WT	H699N	WT	H699N	p value
TQ	1,0E+09	1,1E+09	100,0%	100,0%	0,251
TMV	6,0E+07	5,8E+07	18,4%	13,5%	0,175
EST3	1,9E+08	1,5E+08	5,8%	5,2%	0,465
OAZ1	7,0E+07	3,2E+07	6,9%	2,8%	**0,008**
IBV	1,2E+08	8,7E+07	11,7%	7,7%	0,251

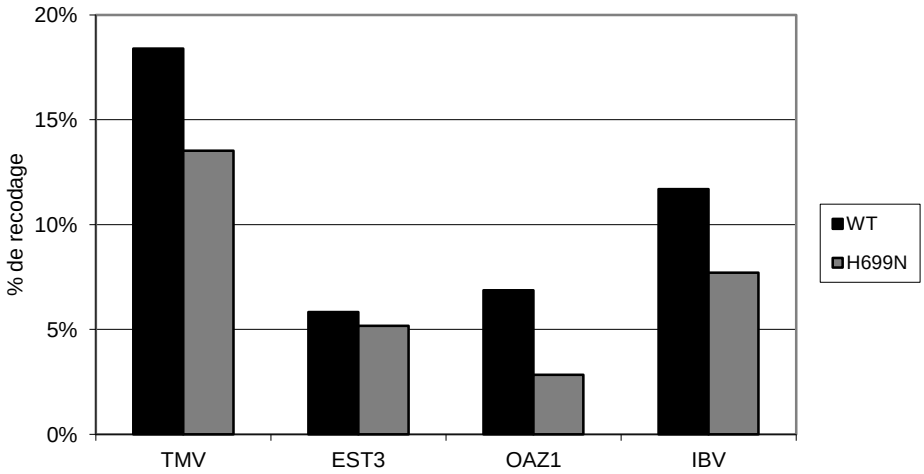


Figure 11 : La mutation H699N dans eEF2 diminue l'efficacité de recodage de la séquence OAZ1.

Le tableau regroupe les résultats de dosages des pAC-PGK-TQ ; TMV ; EST3 ; OAZ1 et IBV en présence du pFL44H-PGK-EFT2 ou du pFL44H-PGK-eft2(H699N) dans la souche HC07.

Afin de comparer les allèles *EFT2* et *eft2(H699N)*, j'ai construit le pFL44H-PGK-eft2(H699N) en mutant le codon CAC n°699 en codon AAC. Chaque plasmide a été utilisé pour transformer la souche HC07, les transformants ont été sélectionnés sur milieu YNB CSM-His. Ensuite chaque souche a été transformée avec le pAC-PGK007 contenant une des cibles de recodage. Pour chaque construction, 6 mesures indépendantes ont été réalisées pour quantifier le % de recodage. Les résultats sont regroupés dans le graphique de la figure 11. On constate que la mutation H699N n'a pas d'effet sur l'efficacité de traduction du témoin TQ. Comme il y a 6 mesures indépendantes, le test non paramétrique Mann-Whitney a été utilisé pour estimer si les valeurs de recodage étaient statistiquement différentes pour chaque cible entre les 2 souches. On constate qu'il n'y a pas d'effet de la mutation H699N sur le recodage du TMV, d'*EST3* et d'IBV. Par contre pour *OAZ1*, l'efficacité de recodage est diminuée d'un facteur 2,5 dans une souche possédant *eft2(H699N)* (p value = 0,008).

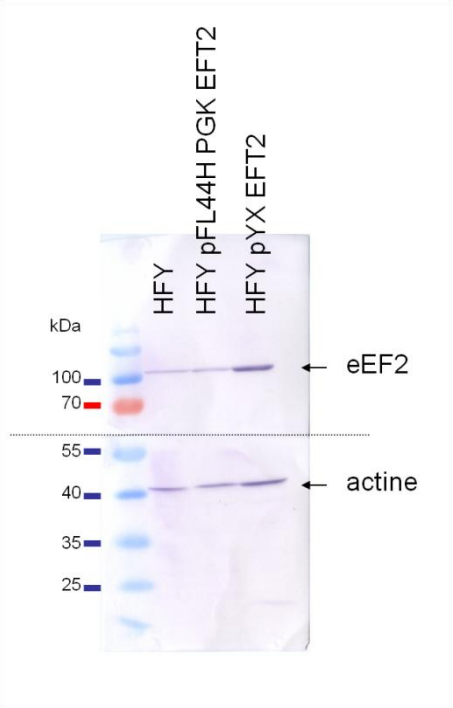


Figure 12 : Surexpression du facteur eEF2

Western blot contre les protéines eEF2 et actines. Les extraits cellulaires proviennent de souches HFY120, HFY1200 pFL44H-PGK-EFT2 et HFY1200 pYX212EFT2.

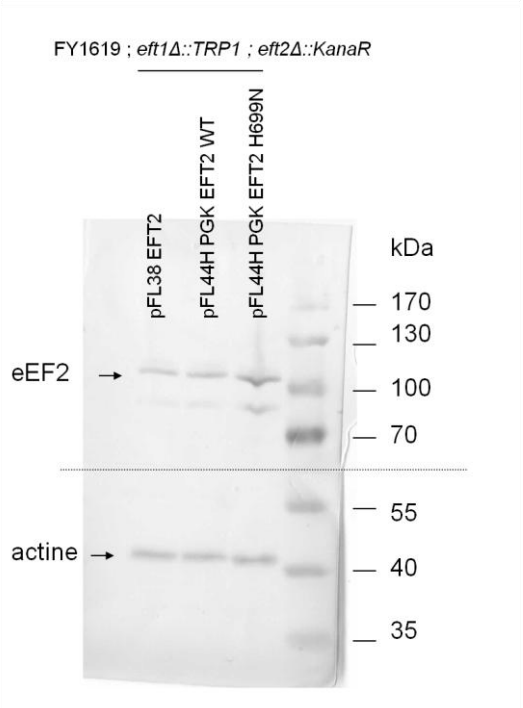


Figure 13 : Expression d'eEF2 à partir de différents plasmides

Western blot contre les protéines eEF2 et actines. Les extraits cellulaires proviennent de souches FY1619 *eft1Δ ::TRP1 eft2Δ ::KanR* transformées avec le plasmide pFL38-EFT2, pFL44H-PGK-EFT2 ou pFL44H-PGK-eft2(H699N)

III. Discussion

1. Construction de la souche FY1619 ; *eft2Δ::KanR* ; *eft1Δ::TRP1* ; pFL38-EFT2

La construction de la souche a été bien plus compliquée que prévu initialement. Comme eEF2 est une protéine essentielle, il est important qu'elle soit exprimée en quantité suffisante pour assurer une viabilité cellulaire normale. Contrairement à d'autres protéines, elle est difficilement sur-exprimable (communication personnelle Barry Cooperman) car son abondance est finement régulée *in vivo* (Ortiz and Kinzy 2005). Pour illustrer ce point, j'ai transformé des levures du fond génétique HFY1200 avec le plasmide pYX212-EFT2 ou le plasmide pFL44H-PGK EFT2 et j'ai réalisé un western blot contre les protéines eEF2. Le niveau d'expression d'eEF2 dans la souche avec le plasmide pFL44H-PGK EFT2 (figure 12 puits 2) est le même que dans souche sauvage (puits 1). Ce résultat n'est pas dû à un défaut d'expression du plasmide car ce dernier est capable de complémenter une souche délétée pour les gènes *EFT1* et *EFT2*. Pour la souche avec le plasmide pYX212-EFT2, il y a une surexpression du facteur eEF2 mais elle n'est que d'un facteur 2. L'expression d'eEF2 est donc une question de « ni trop, ni trop peu ».

Dans les différentes constructions de plasmides, il a fallu trouver la bonne combinaison entre la force du promoteur et la quantité de plasmides par cellule (origine de réplication centromérique ou 2µm). La principale surprise a été de constater que 500pb de la région promotrice n'était pas suffisant pour exprimer eEF2 et qu'il fallait utiliser une région de 1000pb. Une étude du promoteur d'*EFT2* utilisant le gène LacZ comme rapporteur avait pourtant montré que les 414bp juste en amont du gène étaient suffisants pour exprimer autant de β-galactosidase qu'avec un promoteur de 1300bp (Veldman et al. 1994). Comme on ne peut pas exclure que la délétion de la 2nde copie codant eEF2 était le point limitant de la stratégie, il serait intéressant de transformer les cellules FY1619 ; *eft2Δ::KanR* ; *eft1Δ::TRP1* avec le plasmide pRS316-EFT2. Si effectivement ce plasmide comprenant 500pb de promoteur ne suffit pas à assurer la viabilité cellulaire, on pourra supposer qu'il existe une ou des régions régulatrices assez éloignées qui contrôlent la transcription d'*EFT2* en fonction de la quantité d'eEF2 intracellulaire. Le promoteur de 500pb serait un élément minimum nécessaire à l'expression d'eEF2 mais ne serait pas systématiquement suffisant pour assurer la viabilité cellulaire.

Les promoteurs des gènes *TPI1* et *PGK1* sont connus comme appartenant aux promoteurs les plus forts de *Saccharomyces cerevisiae*. Tous 2 ont été utilisés pour sur-exprimer *EFT2* dans des plasmides multicopies (2µm) mais seul le promoteur *PGK1* a permis d'assurer une viabilité cellulaire. Il semble donc que ce promoteur soit un peu moins fort que celui de *TPI1*. J'ai réalisé des western blots contre eEF2 avec la souche FY1619 ; *eft2Δ::KanR* ; *eft1Δ::TRP1* transformée avec le plasmide pFL38-EFT2 ou le plasmide pFL44H-PGK-EFT2 (figure 13). La quantification des signaux correspondant à eEF2 par rapport à ceux correspondant à l'actine indique que la quantité d'eEF2 dans ces souches est comparable à celle obtenue pour la souche HFY1200 sauvage (figure 12 puits 1). Ce qui montre que, soit le promoteur PGK1 ne permet pas une forte sur-expression d'EFT2, soit que le promoteur endogène d'EFT2 est lui aussi un promoteur très fort. Nous favorisons cette 2ème hypothèse étant donnée l'abondance de la protéine eEF2 dans la cellule (78100 molécules/cellule) (Ghaemmaghami et al. 2003).

L'autre point délicat dans la construction de cette souche a été la délétion des gènes endogènes *EFT1* et *EFT2*. D'après l'expérience acquise lors de cette construction, il est vraiment important d'inclure de larges zones d'homologie de séquence de part et d'autre de la cassette de sélection pour que la recombinaison homologue soit efficace à un locus fortement exprimé. Le choix de la cassette de résistance est aussi important. Pour des délétions délicates, il est préférable d'utiliser des cassettes de sélection qui confèrent une autotrophie plutôt qu'une cassette comme *KanR* qui confère une résistance à un antibiotique. Les cassettes d'autotrophie génèrent moins de colonies « faux positifs » après sélection sur le milieu approprié. Dans notre cas, les marqueurs disponibles dans la souche et les différents plasmides nécessaires pour l'expérience ne permettaient pas de choisir cette stratégie.

2. Le rôle d'eEF2 dans le décalage du cadre de lecture en -1

A présent que les gènes rapporteurs *ibv39-ade2* et *IBV40-ADE2* induisent des phénotypes stables, ils pourront être insérés dans la souche HC07 et le projet de mutagénèse va pouvoir être réalisé. Il est probable que la mutagénèse aléatoire fasse ressortir des régions dans le domaine IV d'eEF2 ainsi que d'autres régions impliquées dans la flexibilité d'eEF2. Ces mutants nous permettraient de confirmer ou d'infirmer l'hypothèse d'un rôle direct d'eEF2 dans le maintient de l'appariement codon:anticodon lors de la translocation.

D'autre part, les découvertes de ces dernières années permettent d'affiner le modèle proposé en 2006 en ce qui concerne le rôle d'eEF2 dans le décalage en -1. De manière générale, il apparait qu'eEF2 ne pousserait pas directement l'ARNm et les ARNt lors de la translocation. Il faciliterait les changements de conformation des sous-unités ribosomiques pour diminuer la distance que l'ARNm et les ARNt ont à parcourir, et imposerait la direction du mouvement.
Dans le décalage en -1, les 2 forces antagonistes seraient alors la résistance du pseudonoeud d'une part, et l'affinité des ARNt pour les sites P et E et les changements de conformation du ribosome d'autre part. Bien qu'on sache que les sous-unités ribosomiques sont dans un état « encliqueté », il serait intéressant de vérifier l'état de rotation de la tête de la sous-unité 40S pour déterminer plus précisément à quelle étape de la translocation les ribosomes sont bloqués. Dans la carte, il apparait que l'ARNt est positionné assez haut dans le site P. Hélas, l'ARNm n'est pas visible donc il n'est pas possible de savoir si l'appariement codon:anticodon est encore présent. On peut envisager que la présence d'eEF2 contraigne l'ARNm à rester dans son canal facilitant la rupture de l'hélice codon:anticodon. Sa présence empêcherait aussi un retour à une conformation antérieure et figerait ainsi le système assez longtemps pour laisser le temps à l'ARNt de se rapparier en phase -1.

3. Le mutant eEF2 (H699N) : diphthamide et recodage

En accord avec les résultats publiés par Ortiz et al. (Ortiz and Kinzy 2005), le mutant eEF2(H699N) n'a pas d'influence sur la translecture. Par contre pour le décalage en -1, l'efficacité de recodage pour la cible IBV ne varie pas en présence du mutant eEF2(H699N). La nature de la séquence de recodage (virus L-A ou IBV) pourrait-être la cause cette différence. De plus, le système rapporteur utilisé est différent. Dans cette étude, le recodage est mesuré grâce à un vecteur double rapporteur pAC-PGK007 alors que dans l'étude précédente il s'agit d'un système simple rapporteur lacZ. En comparaison avec un

système double rapporteur, un simple rapporteur peut être plus sensible à de nombreux artéfacts. Il serait donc intéressant de confirmer nos observations en construisant le pAC-PGK-LA.

D'après Ortiz et al., le mutant eEF2(H699N) n'aurait pas d'effet sur le décalage en +1 pour la séquence du Ty1 (**CUU AGG C**). Nous obtenons le même résultat pour la séquence EST3 (**CUU AGU U**) qui utilise un mécanisme de décalage en +1 similaire à celui du Ty1. Il s'agit d'un glissement de l'ARNt au site P de CUU à UUA, le glissement est stimulé par un codon rare AGG ou AGU. De manière étonnante, eEF2(H699N) diminue l'efficacité de décalage en +1 sur la séquence d'OAZ1. Ce résultat est intriguant car le décalage en +1 (**GCG TGA C)** est stimulé par un codon stop. L'influence de la terminaison avait déjà été étudiée (Namy et al. 2008, Karamysheva et al. 2003) mais jusqu'à présent le facteur eEF2 n'avait jamais été impliqué dans un décalage du cadre de lecture en +1. Afin de confirmer ou infirmer ce résultat, plusieurs expériences pourraient être réalisées. Concernant eEF2, il serait intéressant de tester d'autres mutants de la position H699 pour vérifier si l'observation sur le décalage est due à la présence spécifique d'un résidu histidine ou à la présence de mutations qui désorganiseraient localement la protéine eEF2. D'autre part, le décalage en +1 sur la séquence OAZ1 est stimulé par un codon stop. Afin de tester si l'effet observé est dû à une compétition entre la translocation et la terminaison, nous pourrions refaire la quantification dans une souche [*PSI+*] dans laquelle l'efficacité de terminaison est diminuée (cf « Introduction »). En abaissant l'efficacité de terminaison, on pourrait s'attendre à ce que l'efficacité de décalage en +1 soit plus importante dans une souche [*PSI+*] que dans une souche [*psi-*] malgré la présence d'eEF2 H699N ce qui confirmerait que l'efficacité de décalage est non seulement dépendante de l'étape d'élongation mais aussi en partie de l'efficacité de translocation.

Dans le mutant eEF2(H699N), le diphthamide est absent. Les effets observés de ce mutant pourrait être dus soit à l'absence du diphthamide, soit à la présence de l'acide aminé asparagine (N) qui déstructurerait localement la protéine. Toutes ces données doivent être confirmées par d'autres mutations affectant soit la synthèse du diphthamide soit d'autres acides aminés de la région. Ce dernier point est délicat car il faut cibler des acides aminés qui ne sont pas impliqués dans l'accrochage du diphthamide sur son résidu.

Pour le moment, le rôle du diphthamide dans le maintien du cadre de lecture est encore loin d'être éclairci.

Références bibliographiques

Agirrezabala, X., J. Lei, J. L. Brunelle, R. F. Ortiz-Meoz, R. Green & J. Frank (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol Cell,* 32**,** 190-7.

Botet, J., M. Rodríguez-Mateos, J. P. Ballesta, J. L. Revuelta & M. Remacha (2008) A chemical genomic screen in Saccharomyces cerevisiae reveals a role for diphthamidation of translation elongation factor 2 in inhibition of protein synthesis by sordarin. *Antimicrob Agents Chemother,* 52**,** 1623-9.

Brachmann, C. B., A. Davies, G. J. Cost, E. Caputo, J. Li, P. Hieter & J. D. Boeke (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast,* 14**,** 115-32.

Brierley, I., N. J. Rolley, A. J. Jenner & S. C. Inglis (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol,* 220**,** 889-902.

Chernoff, Y. O., S. L. Lindquist, B. Ono, S. G. Inge-Vechtomov & S. W. Liebman (1995) Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science,* 268**,** 880-4.

Cornish, P. V., D. N. Ermolenko, H. F. Noller & T. Ha (2008) Spontaneous intersubunit rotation in single ribosomes. *Mol Cell,* 30**,** 578-88.

Dumont-Miscopein, A., J. P. Lavergne, D. Guillot, B. Sontag & J. P. Reboud (1994) Interaction of phosphorylated elongation factor EF-2 with nucleotides and ribosomes. *FEBS Lett,* 356**,** 283-6.

Foley, B. T., J. M. Moehring & T. J. Moehring (1995) Mutations in the elongation factor 2 gene which confer resistance to diphtheria toxin and Pseudomonas exotoxin A. Genetic and biochemical analyses. *J Biol Chem,* 270**,** 23218-25.

Frank, J. & R. K. Agrawal (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature,* 406**,** 318-22.

Gao, Y. G., M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley & V. Ramakrishnan (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science,* 326**,** 694-9.

Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea & J. S. Weissman (2003) Global analysis of protein expression in yeast. *Nature,* 425**,** 737-41.

Gupta, P. K., S. Liu, M. P. Batavia & S. H. Leppla (2008) The diphthamide modification on elongation factor-2 renders mammalian cells resistant to ricin. *Cell Microbiol,* 10**,** 1687-94.

Hatin, I., C. Fabret, J. P. Rousset & O. Namy (2009) Molecular dissection of translation termination mechanism identifies two new critical regions in eRF1. *Nucleic Acids Res,* 37**,** 1789-98.

He, F. & A. Jacobson (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev,* 9**,** 437-54.

Julián, P., A. L. Konevega, S. H. Scheres, M. Lázaro, D. Gil, W. Wintermeyer, M. V. Rodnina & M. Valle (2008) Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc Natl Acad Sci U S A,* 105**,** 16924-7.

Jørgensen, R., P. A. Ortiz, A. Carr-Schmid, P. Nissen, T. G. Kinzy & G. R. Andersen (2003) Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. *Nat Struct Biol,* 10**,** 379-85.

Karamysheva, Z. N., A. L. Karamyshev, K. Ito, T. Yokogawa, K. Nishikawa, Y. Nakamura & S. Matsufuji (2003) Antizyme frameshifting as a functional probe of eukaryotic translational termination. *Nucleic Acids Res,* 31**,** 5949-56.

Kimata, Y. & K. Kohno (1994) Elongation factor 2 mutants deficient in diphthamide formation show temperature-sensitive cell growth. *J Biol Chem,* 269**,** 13497-501.

Li, W., L. G. Trabuco, K. Schulten & J. Frank (2011) Molecular dynamics of EF-G during translocation. *Proteins,* 79**,** 1478-86.

Liu, S., G. T. Milne, J. G. Kuremsky, G. R. Fink & S. H. Leppla (2004) Identification of the proteins required for biosynthesis of diphthamide, the target of bacterial ADP-ribosylating toxins on translation elongation factor 2. *Mol Cell Biol,* 24**,** 9487-97.

Nairn, A. C. & H. C. Palfrey (1987) Identification of the major Mr 100,000 substrate for calmodulin-dependent protein kinase III in mammalian cells as elongation factor-2. *J Biol Chem,* 262**,** 17299-303.

Namy, O., A. Galopier, C. Martini, S. Matsufuji, C. Fabret & J. P. Rousset (2008) Epigenetic control of polyamines by the prion [PSI+]. *Nat Cell Biol,* 10**,** 1069-75.

Namy, O., I. Hatin & J. P. Rousset (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep,* 2**,** 787-93.

Namy, O., S. J. Moran, D. I. Stuart, R. J. Gilbert & I. Brierley (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature,* 441**,** 244-7.

Ortiz, P. A. & T. G. Kinzy (2005) Dominant-negative mutant phenotypes and the regulation of translation elongation factor 2 levels in yeast. *Nucleic Acids Res,* 33**,** 5740-8.

Ortiz, P. A., R. Ulloque, G. K. Kihara, H. Zheng & T. G. Kinzy (2006) Translation elongation factor 2 anticodon mimicry domain mutants affect fidelity and diphtheria toxin resistance. *J Biol Chem,* 281**,** 32639-48.

Pappenheimer, A. M., P. C. Dunlop, K. W. Adolph & J. W. Bodley (1983) Occurrence of diphthamide in archaebacteria. *J Bacteriol,* 153**,** 1342-7.

Perentesis, J. P., L. D. Phan, W. B. Gleason, D. C. LaPorte, D. M. Livingston & J. W. Bodley (1992) Saccharomyces cerevisiae elongation factor 2. Genetic cloning, characterization of expression, and G-domain modeling. *J Biol Chem,* 267**,** 1190-7.

Peske, F., N. B. Matassova, A. Savelsbergh, M. V. Rodnina & W. Wintermeyer (2000) Conformationally restricted elongation factor G retains GTPase activity but is inactive in translocation on the ribosome. *Mol Cell,* 6**,** 501-5.

Phan, L. D., J. P. Perentesis & J. W. Bodley (1993) Saccharomyces cerevisiae elongation factor 2. Mutagenesis of the histidine precursor of diphthamide yields a functional protein that is resistant to diphtheria toxin. *J Biol Chem,* 268**,** 8665-8.

Rodnina, M. V., A. Savelsbergh, V. I. Katunin & W. Wintermeyer (1997) Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature,* 385**,** 37-41.

Savelsbergh, A., V. I. Katunin, D. Mohr, F. Peske, M. V. Rodnina & W. Wintermeyer (2003) An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation. *Mol Cell,* 11**,** 1517-23.

Savelsbergh, A., N. B. Matassova, M. V. Rodnina & W. Wintermeyer (2000) Role of domains 4 and 5 in elongation factor G functions on the ribosome. *J Mol Biol,* 300**,** 951-61.

Shi, X., P. K. Khade, K. Y. Sanbonmatsu & S. Joseph (2012) Functional Role of the Sarcin-Ricin Loop of the 23S rRNA in the Elongation Cycle of Protein Synthesis. *J Mol Biol*.

Spahn, C. M., M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jørgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta & J. Frank (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J,* 23**,** 1008-19.

Taylor, D. J., J. Nilsson, A. R. Merrill, G. R. Andersen, P. Nissen & J. Frank (2007) Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J,* 26**,** 2421-31.

Teige, M., E. Scheikl, V. Reiser, H. Ruis & G. Ammerer (2001) Rck2, a member of the calmodulin-protein kinase family, links protein synthesis to high osmolarity MAP kinase signaling in budding yeast. *Proc Natl Acad Sci U S A,* 98**,** 5625-30.

Valle, M., A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg & J. Frank (2003) Locking and unlocking of ribosomal motions. *Cell,* 114**,** 123-34.

Veldman, S., S. Rao & J. W. Bodley (1994) Differential transcription of the two Saccharomyces cerevisiae genes encoding elongation factor 2. *Gene,* 148**,** 143-7.

Webb, T. R., S. H. Cross, L. McKie, R. Edgar, L. Vizor, J. Harrison, J. Peters & I. J. Jackson (2008) Diphthamide modification of eEF2 requires a J-domain protein and is essential for normal development. *J Cell Sci,* 121**,** 3140-5.

Winston, F., C. Dollard & S. L. Ricupero-Hovasse (1995) Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast,* 11**,** 53-5.

Chapitre 2

Etude de l'élongation en molécule unique.

Sommaire

I. Introduction 72
1. Dynamique du ribosome : apports des études en molécule unique. 72
2. Le projet 73
II. Méthodologie 74
1. Le système optique 74
A. Microscopie à onde évanescente ou TIRFM 74
B. Chimie de surface 75
C. La cellule microfluidique et le dispositif chauffant 75
2. Le système biologique 76
A. L'ARNm 76
B. Le ribosome, les ARNt et les facteurs d'élongation 77
3. L'analyse d'images et traitement des données 77
III. Résultats et discussion 79
1. L'ARNm opt ATTO 79
2. Fixation des ribosomes sur l'ARNm opt ATTO 79
3. Traduction sur une surface en mesure d'ensemble 80
4. Traduction en molécule unique 81
IV. Perspectives 88
1. L'utilisation l'ARNm opt ATTO 88
2. Marquage de facteurs de la traduction : Système de traduction purifié et Zero Mode Wave Guide 88
A. Système de traduction purifié 88
B. Zero Mode Waveguide 89
Références bibliographiques 90
Références bibliographiques des figures 91

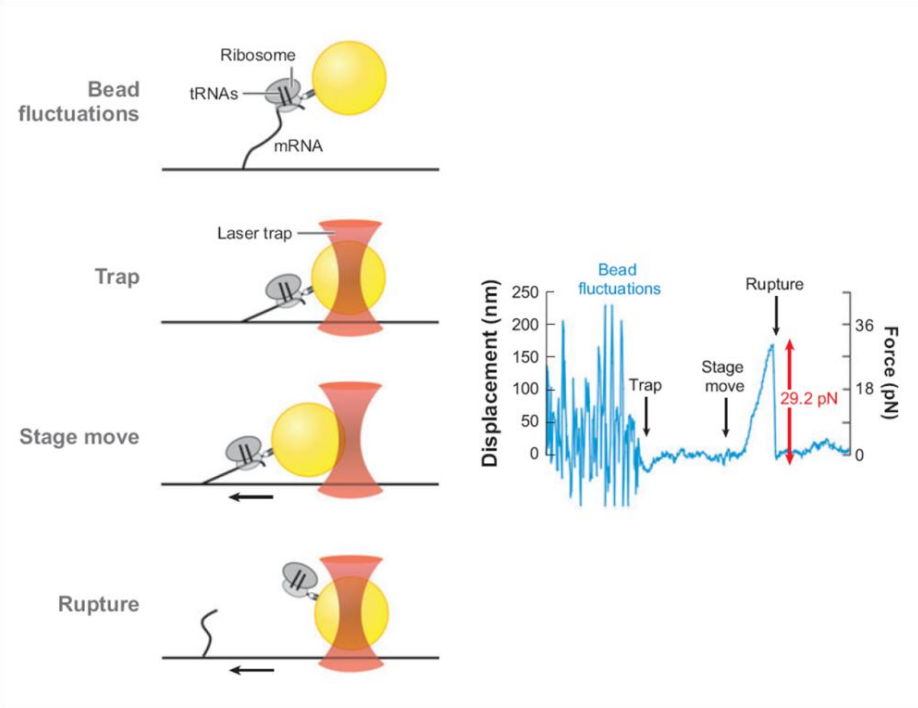


Figure 1 : Exemple d'une expérience de piégeage optique

Schéma de l'expérience. Un ARNm est en complexe avec un ribosome et des ARNt. L'ARNm est accroché à la surface de la lamelle et la petite sous-unité du ribosome à la bille (bead fluctuation). Une fois le laser allumé, la bille est piégée par le faisceau ce qui met l'ARNm en tension (Trap). Lorsque la platine s'écarte (stage move), la bille s'écarte de sa position d'équilibre à l'intérieur du piège. Lorsque l'ARNm subit une tension trop grande l'ARNm et le ribosome se dissocient. C'est le déplacement de la bille par rapport à la position d'équilibre qui est mesuré (graphique à droite). La mesure permet d'évaluer la force en jeu entre l'ARNm et les ARNt. Figure issue de (Marshall et al. 2008)

I. Introduction

Mon 2ème projet a consisté en l'étude de l'élongation chez les eucaryotes en molécule unique. Afin d'introduire ce projet je vais présenter les intérêts des approches en molécule unique et décrire leur apport dans la compréhension de la traduction.

1. Dynamique du ribosome : apports des études en molécule unique.

Comme nous avons pu le voir dans la partie « Introduction », la traduction est un processus complexe multi-étapes. Ses différents composants assurent à la synthèse des protéines les caractères de vitesse, de fidélité et de régulation. La compréhension de ces mécanismes nécessite l'utilisation combinée de différentes techniques.
Les études de génétique et de biochimie ont permis d'identifier les composants de la traduction, de décrire les différentes étapes du processus et d'en estimer les paramètres cinétiques. Les études de biochimie classiques analysent la transition d'un système chimique entre 2 états d'équilibre. Il s'agit toujours de mesures d'ensemble dans lesquelles le comportement individuel de chaque réaction est moyenné. Les étapes transitoires et/ou rares ne peuvent donc pas être détectées.
Les études structurales (diffraction de cristaux aux rayons X et cryo électro-microscopie) apportent une représentation physique du ribosome à une échelle moléculaire voire atomique. Ces études décrivent de plus les interactions entre le ribosome et ses partenaires. Elles mettent aussi en évidence des changements de conformations de chacun des acteurs. L'ensemble de ces images statiques supporte l'idée que la traduction est un processus dynamique qui permet de coordonner l'activité de sites actifs éloignés. Comme pour les études précédentes, les études structurales sont aussi des mesures d'ensemble. Elles ne saisissent par ailleurs que des instantanés d'un processus dont la nature, l'ampleur, la vitesse et la fonction des mouvements restent largement inconnues.

En suivant le comportement individuel de chaque réaction dans le temps, les études en molécule unique apportent le lien manquant entre les données apportées par les structures moléculaires statiques et celles sur la dynamique de l'enzyme. Comme les résultats associés à chaque réaction peuvent être triés et synchronisés *a posteriori*, ces études évitent le problème de moyenne dans une population. L'intérêt est de pouvoir observer des évènements rares, asynchrones et stochastiques, ainsi que des chemins de réaction alternatifs.

La sensibilité des techniques utilisées dans les études en molécule unique en fait des outils adaptés pour étudier la traduction. Par exemple, la micromanipulation de molécules par piégeage optique (appelé aussi pince optique) peut mesurer des forces comprises entre 1 et 100 pico Newton ($1pN=10^{-12}N$) (figure 1). Cette gamme de forces est appropriée pour l'étude des forces mises en jeux lors de la traduction d'un ARNm structuré. Quant aux techniques de microscopie de fluorescence, elles permettent de localiser des molécules fluorescentes avec une résolution pouvant aller jusqu'à de l'ordre du nanomètre. Elles peuvent mesurer aussi des variations de distance de l'ordre de l'angström ($1Å = 10^{-10}m$). Je vais me concentrer plus en détail sur ce dernier aspect.

Le smFRET (single molecule Forster Resonance Energy Transfer) est une technique de biophysique qui permet de mesurer des variations de distance entre 2 molécules fluorescentes compatibles. Les

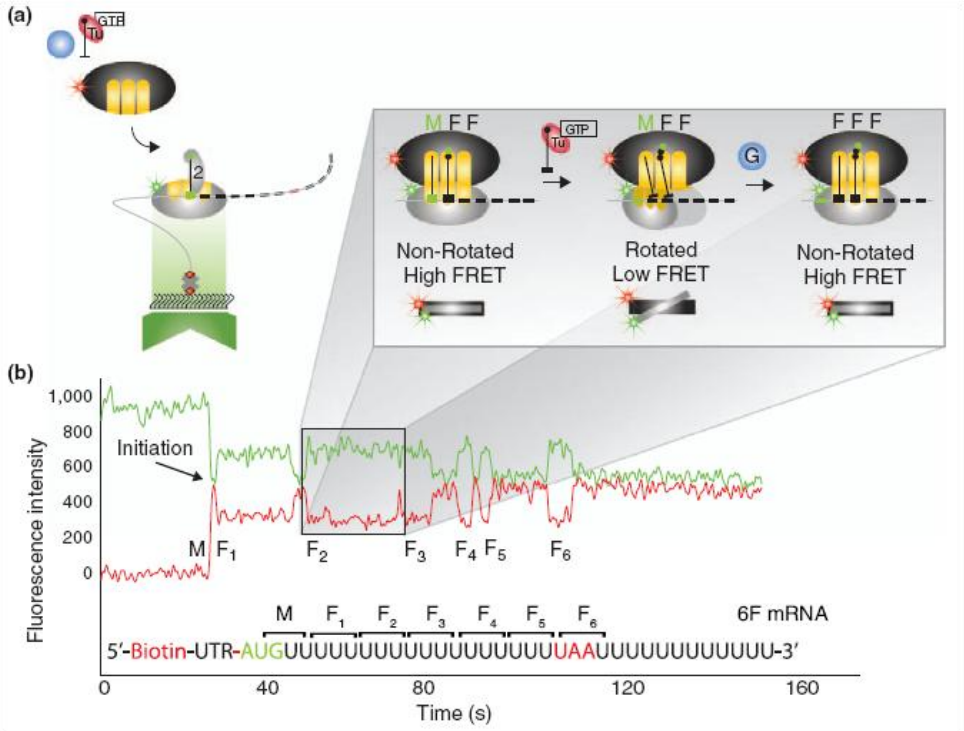


Figure 2 : Exemple d'expérience en smFRET

A) Schéma de l'expérience. L'ARNm est accroché à la surface et chacune des sous-unités ribosomiques est marquée spécifiquement avec un fluorophore. La sous-unité 60S est marquée avec un Cy5 et la sous-unité 40S avec un Cy3. Le signal de FRET entre les 2 fluorophores renseigne sur l'état d'encliquetage des sous-unités. B) suivi du signal de FRET dans le temps. Il permet de suivre la conformation du ribosome sur plusieurs cycles de translocation. Figure issue de (Petrov et al. 2011)

fluorophores sont compatibles si le spectre d'émission du fluorophore dit « donneur » recouvre le spectre d'absorption de « l'accepteur ». Lorsque le donneur est excité, il peut transmettre son énergie à l'accepteur si ce dernier est assez proche spatialement. Ce transfert d'énergie non radiatif, appelé FRET (Forster Resonance Energy Transfer), se traduit par une émission de fluorescence de la part de la molécule acceptrice. Comme l'intensité du signal de FRET est directement dépendante de la distance entre les 2 molécules, et ce suivant une loi en $1/R^6$, où R est la distance entre les 2 fluorophores, elle permet de quantifier très précisément des variations de distance de l'ordre de 1 à 10nm (figure 2).

Grâce à la disponibilité des structures atomiques du ribosome et de ses partenaires à une échelle atomique, il est possible de marquer spécifiquement et de façon rationnelle des molécules avec des fluorophores inorganiques et de concevoir des expériences de smFRET. Le choix de la localisation des marquages est un critère important pour observer un changement de conformation précis et pertinent entre 2 éléments et d'associer les variations de signal de smFRET à des étapes du cycle de traduction. En couplant les expériences de smFRET avec l'utilisation de protéines ou d'ARN mutés, et des molécules inhibitrices (comme des antibiotiques), il devient possible d'avoir accès aux différents états transitoires du ribosome.

Durant la dernière décennie, le smFRET a été appliqué à l'étude de la traduction. D'un point de vue purement technique, 3 études ont marqué des avancés majeures dans l'aptitude à faire fonctionner des ribosomes dans des études en molécule unique. Tout d'abord, en 2004 il y a la 1ère expérience de smFRET appliquée à la traduction. Le marquage des ARNt aux sites A et P a permis d'observer leur dynamique dans le ribosome avant translocation (Blanchard et al. 2004b). Ensuite en 2008, une protéine fluorescente GFP entière a pu être synthétisée et sa fluorescence détectée (Uemura et al. 2008). Enfin en 2010, l'utilisation d'une technique d'imagerie décrite au point IV B a permis de suivre l'incorporation d'ARNt fluorescents en temps réel sur plusieurs cycles d'élongation (Uemura et al. 2010).

D'un point de vue biologique, diverses études de smFRET ont permis de caractériser entre autres la sélection des ARNt (Blanchard et al. 2004a) et les états classique et hybrides des ARNt après la transpeptidation (Blanchard et al. 2004b, Munro et al. 2007). La rotation spontanée des sous-unités ribosomiques a pu être observée en marquant chacune des sous-unités ribosomiques (Cornish et al. 2008). Enfin le mouvement de la protubérance L1 a pu être corrélé avec l'état de l'ARNt aux sites P et E (Cornish et al. 2009, Fei et al. 2009).

2. Le projet

Jusqu'à présent, les études en molécule unique se sont majoritairement focalisées sur l'étude du ribosome bactérien. Seules deux études rapportent l'utilisation de ribosomes eucaryotes. La première décrit le marquage spécifique du ribosome de *Saccharomyces cerevisiae* en vue de futures études en molécule unique (Petrov and Puglisi 2010). La seconde observe la dynamique des ARNt aux sites A et P dans un ribosome avant translocation (Budkevich et al. 2011).

Dans le but d'étudier l'élongation chez les eucaryotes, nous avons décidé de mettre au point le premier système de traduction en molécule unique pour les eucaryotes.

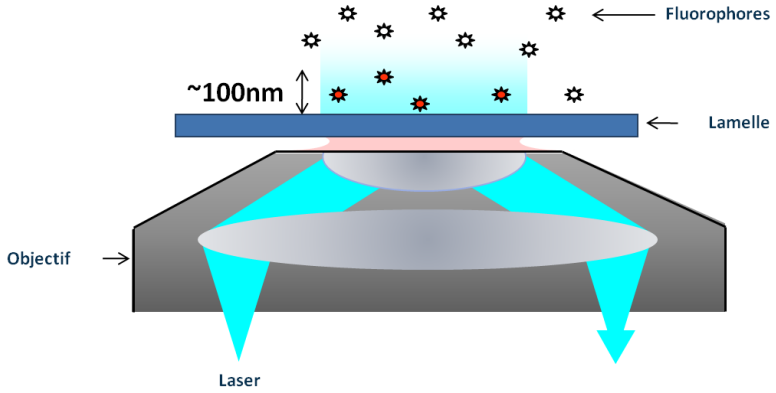


Figure 3 : Microscopie à onde évanescente

Le faisceau incident du laser est envoyé sur la lamelle de manière à être totalement réfléchi. Seuls les fluorophores présents dans le champ évanescent (en bleu) sont excités (étoiles rouges), leur fluorescence est collectée par l'objectif. Figure issue de (Le Gall 2011)

II. Méthodologie

Ce projet a été réalisé en étroite collaboration avec l'équipe de biophysique en molécule unique dirigée par Nathalie Westbrook au Laboratoire Charles Fabry (Institut d'Optique). Les études en molécule unique ne sont pas des techniques courantes de biochimie, je m'attacherai donc à décrire dans cette partie « Méthodologie » les différents aspects qui sont à la base de notre projet.

Nicolas Fiszman, Antoine Le Gall et Karen Perronet ont réalisé le montage optique, la chimie de surface, les cellules microfluidiques et les analyses d'images. Pour ma part, j'ai réalisé toute la partie biologie. Quant aux expériences en molécule unique, elles ont été réalisées à « 4 mains ».

1. Le système optique

A. Microscopie à onde évanescente ou TIRFM

Le TIRFM (Total Internal Reflexion Fluorescence Microscopy) est une technique de microscopie de fluorescence qui consiste à exciter seulement les fluorophores présents à proximité d'une surface solide comme une lamelle de microscope (figure 3).

Un laser est envoyé sur lamelle-échantillon de manière à être totalement réfléchi. Cette orientation du faisceau conduit à la formation d'une onde électromagnétique dite « évanescente » si le milieu en contact avec la surface a un indice de réfraction inférieur a celui de la lamelle. L'intensité de cette onde non propagative décroît exponentiellement lorsqu'on pénètre dans l'échantillon ce qui permet d'exciter des fluorophores sur une profondeur de ~ 100-200 nm environ. Un élargisseur de faisceau est utilisé pour augmenter la surface d'illumination et l'adapter à la taille du champ d'observation.

Cette technique de microscopie permet d'obtenir des images contrastées avec un bon rapport signal sur bruit. Elle permet d'observer simultanément plusieurs fluorophores dans un même champ d'observation sans exciter les zones voisines. De plus, le TIRFM minimise la part de contamination due à la fluorescence de molécules présentes hors du plan focal et préserve ces dernières du photoblanchiment (Axelrod 2001).

Dans nos conditions d'observation, 1 pixel correspond à 260nm dans le plan de l'échantillon et le champ d'observation fait une taille de 133 µm x 133 µm. Pour comparaison, le ribosome eucaryote à un diamètre de 25-30nm et un codon mesure 1nm. Quant à la tâche de diffraction d'un fluorophore ayant une longueur d'onde d'émission de 669nm, elle a dans notre cas un diamètre d'environ 563nm[1]. En fonction de la manière dont la tâche est centrée sur les pixels, cela correspond à un carré de 4 ou 9 pixels. Dans des conditions favorables, on peut observer jusqu'à 2000 molécules uniques en parallèle. L'utilisation d'une caméra emCCD (electron multiplying Charge-Coupled Device) rapide permet d'obtenir une résolution temporelle de 10ms et surtout de détecter un très faible flux de photons.

[1] Sachant que l'objectif de microscope a une ouverture numérique de 1,45.

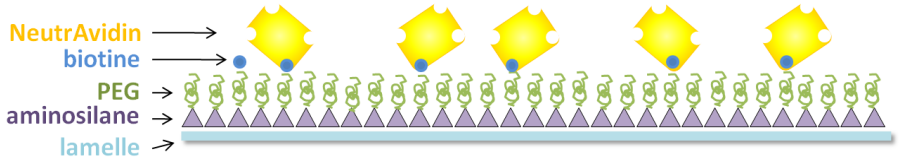


Figure 4 : Chimie de surface

Les lamelles sont traitées pour fixer spécifiquement les molécules à étudier. Elles sont recouvertes d'une couche d'aminosilane qui permet l'ancrage d'une couche de PEG. Certaines molécules de PEG sont couplées à une biotine qui interagit fortement avec des protéines NeutrAvidin. Comme ces protéines possèdent 4 sites de fixation pour des biotines, elles pourront aussi accrocher la molécule d'intérêt par l'intermédiaire d'une biotine.

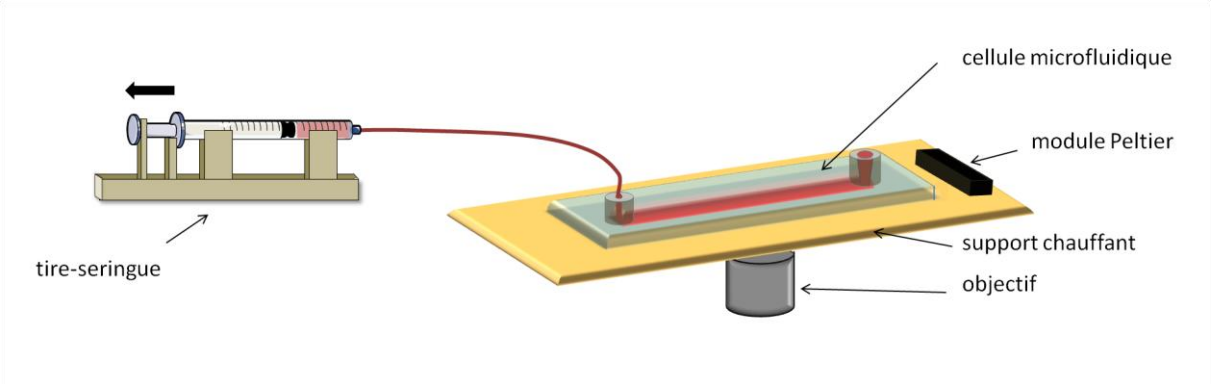


Figure 5 : Cellule microfluidique, dispositif chauffant et système d'injection

Les échantillons sont déposés dans le réservoir de la cellule microfluidique. Ils sont aspirés dans la chambre par aspiration grâce au tire seringue. La cellule microfluidique est thermostatée par le module Peltier qui est fixé à l'extrémité du support en cuivre. L'objectif est aussi chauffé pour éviter les pertes thermiques de l'échantillon à son contact.

B. Chimie de surface

La microscopie à onde évanescente impose que le système biologique étudié soit fixé à la surface. Pour cela, il est important que la lamelle soit traitée de façon à fixer les molécules étudiées de manière spécifique.

Suite à un nettoyage avec un plasma d'oxygène, la lamelle est fonctionnalisée par l'ajout d'amino-silane. La silanisation permet la fixation covalente d'une couche de polyéthylène glycol (PEG). Les molécules de PEG sont électriquement neutres et permettent donc de passiver la surface. Elles possèdent une extrémité NHS qui réagit avec la fonction amine présente à la surface de la lamelle. Parmi les molécules de PEG, environ une sur 4 a son extrémité libre couplée à une biotine. Ces biotines permettent la fixation avec une très haute affinité de protéines NeutrAvidin (K_D=10^{-15}M). Cette protéine possède 4 sites de fixation pour des biotines ce qui permet d'y accrocher aussi les molécules étudiées à l'aide d'une biotine (figure 4).

C. La cellule microfluidique et le dispositif chauffant

Afin d'être analysé, l'échantillon est injecté dans une cellule microfluidique (figure 5). Cela permet de travailler avec de petits volumes d'échantillon (~ 20µL) et de contrôler leur injection dans la chambre de la cellule microfluidique.

Cette chambre est réalisée en assemblant une lamelle traitée à une lame de verre par un ruban adhésif double face. Le canal[2] servant de chambre pour l'échantillon est créé en découpant une fine lamelle dans le ruban adhésif. Aux 2 extrémités du canal, la lame de verre est percée[3] pour permettre l'injection et l'évacuation des solutions. A une des ouvertures, un réservoir moulé en polymère PDMS (Poly Di Methyl Siloxane) permet de déposer l'échantillon. A l'autre, un plot en PDMS ancre un capillaire[4] relié à une seringue. La seringue permet d'aspirer l'échantillon présent dans le réservoir et de l'injecter dans la cellule. Dans le cas de notre étude de nombreuses molécules d'ARN sont présentes, il est donc crucial de travailler dans un environnement sans RNAse. Un soin important est donc consacré pour le montage de la cellule microfluidique. Ces cellules à usage unique sont longues et délicates à préparer, 5 sont fabriquées pour chaque journée d'expérience.

Le débit d'injection est constant, il est appliqué grâce un dispositif pousse-seringue convertit en tire-seringue. Le débit est réglé à 100µL/min. L'injection par « pression négative » évite la présence de volume mort à l'entrée de la cellule, diminue le risque d'injection de bulles et le risque de fuites.

La cellule est thermostatée pour réaliser des traductions à une température stable. Pour cela, la cellule est disposée sur un support en cuivre chauffé par un module Peltier placé à une des extrémités. Le contrôle de la température est précis au degré près.

La cellule peut subir de légères déformations à cause des variations thermiques et des pressions appliquées lors de l'injection. Du point de vue du montage optique, ces déformations peuvent entrainer des pertes de mises au point. Un système autofocus a été développé au laboratoire pour

[2] Dimensions du canal : (L) 35mm x (l) 3mm x (h) 0,2mm
[3] Orifices dans la lame de verre : Ø 0,8mm espacés de 33mm
[4] Capillaire : (L) 300mm, $Ø_{interne}$ 0,15mm

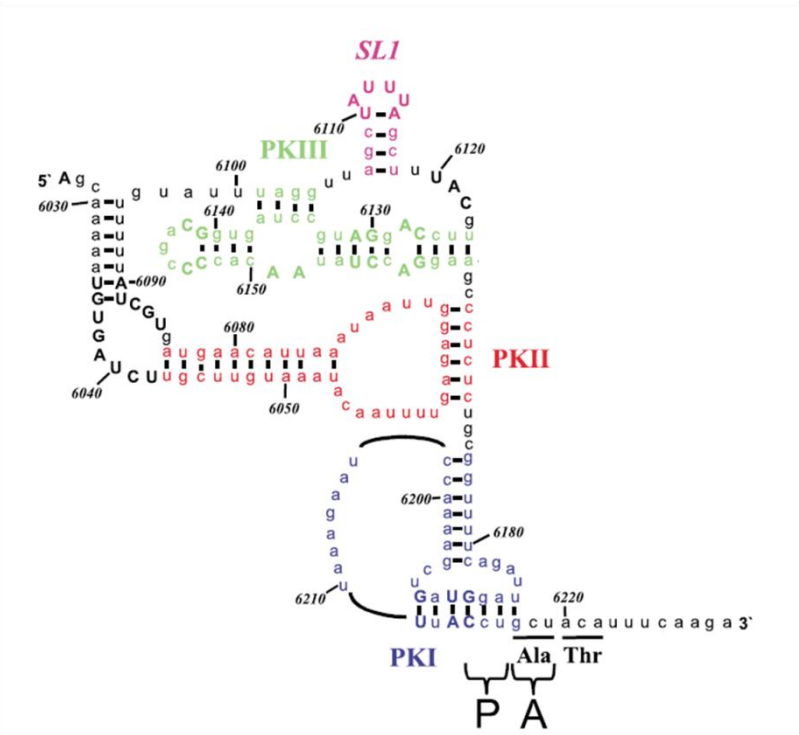


Figure 6 : Structure secondaire de l'IRES IGR CrPV

L'IRES est composé de 3 pseudonoeuds (PK 1 à 3) et d'une tige-boucle (SL). Le pseudonoeud n°1 mime l'ARNt initiateur en se logeant au site P. Figure issue de (Jan and Sarnow 2002).

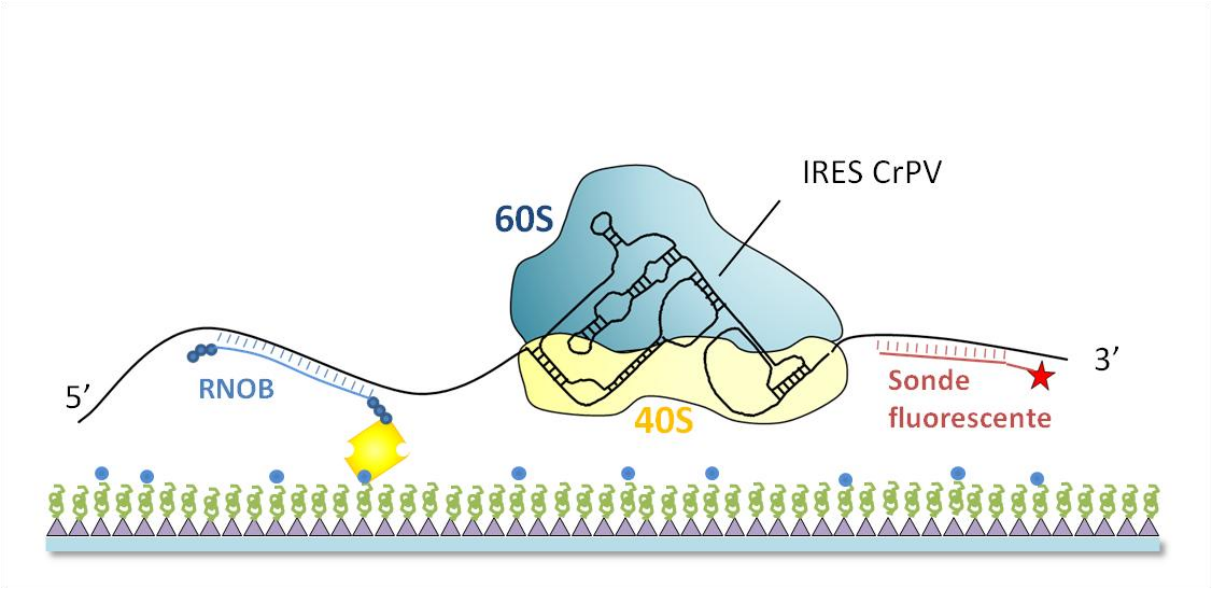


Figure 7 : Initiation du ribosome sur l'ARNm et accroche à la surface

Les ARNm utilisés pour les expériences en molécule unique possèdent une IRES CrPV pour fixer les sous-unités ribosomiques. Les ARNm sont accrochés par l'intermédiaire de l'oligonucléotide biotinylé RNOB. Les ARNm sont visualisés grâce à une sonde fluorescente hybridée en aval de l'IRES.

pallier ce problème afin de garder la mise au point à la surface de l'échantillon pendant toute la durée d'acquisition des images.

2. Le système biologique

Afin de suivre les échantillons biologiques dans le temps, la technique de microscopie utilisée implique qu'ils soient marqués par des fluorophores et soient fixés à la surface. Comme nous allons le voir cet impératif impose des contraintes dans l'utilisation et donc la conception du système biologique.

A. L'ARNm

Afin d'étudier l'étape d'élongation chez les eucaryotes, il convient de fixer le ribosome au préalable sur l'ARNm. Reconstituer le système d'initiation dépendante de la coiffe *in vitro* revient à purifier 13 complexes protéiques et à transcrire un ARNm avec une coiffe. Ces étapes de purification et de synthèse sont longues, compliquées et sans garantie de résultat alors qu'il existe des alternatives astucieuses pour initier la traduction.

L'alternative à l'initiation canonique est d'utiliser un site interne d'entrée pour le ribosome (IRES : Internal Ribosome Entry Site). Notre choix s'est porté sur l'IRES IGR (Intergenic region) du virus CrPV (Cricket Paralysis Virus) car elle a pour particularité de recruter le ribosome 80S indépendamment de tous facteurs d'initiation et de l'ARNt initiateur (Pestova and Hellen 2003). D'une taille de 192nt, elle est constituée de 3 pseudonoeuds et d'une tige-boucle (Jan and Sarnow 2002) (figure 6). Ces éléments se structurent en présence d'ions Mg^{2+} pour recruter la sous-unité 40S (Costantino and Kieft 2005). Dans le ribosome 80S, l'IRES se loge aux sites P et E (Spahn et al. 2004). Le pseudonoeud n°1 mime l'$ARNt_i$ et initie la traduction à un codon de démarrage non AUG. Le premier acide aminé incorporé sera celui du premier ARNt aminoacylé accommodé au site A.

L'utilisation de cette IRES confère plusieurs autres avantages techniques. L'ARNm n'a pas besoin d'être coiffé ce qui facilite la transcription *in vitro*. Il peut aussi être accroché par son extrémité 5' à la lamelle pour permettre au ribosome de s'éloigner de la surface en traduisant. De plus, l'initiation se fait dans un cadre de lecture défini ce qui permet de concevoir des expériences basées sur le maintient du cadre de lecture. Enfin, cette IRES permet de synchroniser le démarrage de la traduction de plusieurs ARNm. Pour cela, les sous-unités ribosomiques sont dans un premier temps fixées sur l'IRES avant d'apporter dans un second temps les facteurs d'élongation et les ARNt aminoacylés.

Dans les différentes expériences que je vais présenter où l'ARNm est accroché à une surface, il est fixé par son extrémité 5' par l'intermédiaire d'un oligonucléotide appelé RNOB (figure 7). Cet oligonucléotide ARN est composé des bases modifiées 2'-O-allyl. Ces bases modifiées le rendent résistant à d'éventuelles nucléases et surtout, elles stabilisent fortement son appariement avec l'ARNm. Chaque extrémité du RNOB est couplée à une chaîne de 3 biotines. Cet oligonucléotide avait été précédemment utilisé pour purifier des ARNm en cours de traduction avec des billes magnétiques recouvertes de protéines streptavidine (Namy et al. 2006). Son efficacité d'appariement a donc déjà été démontrée.

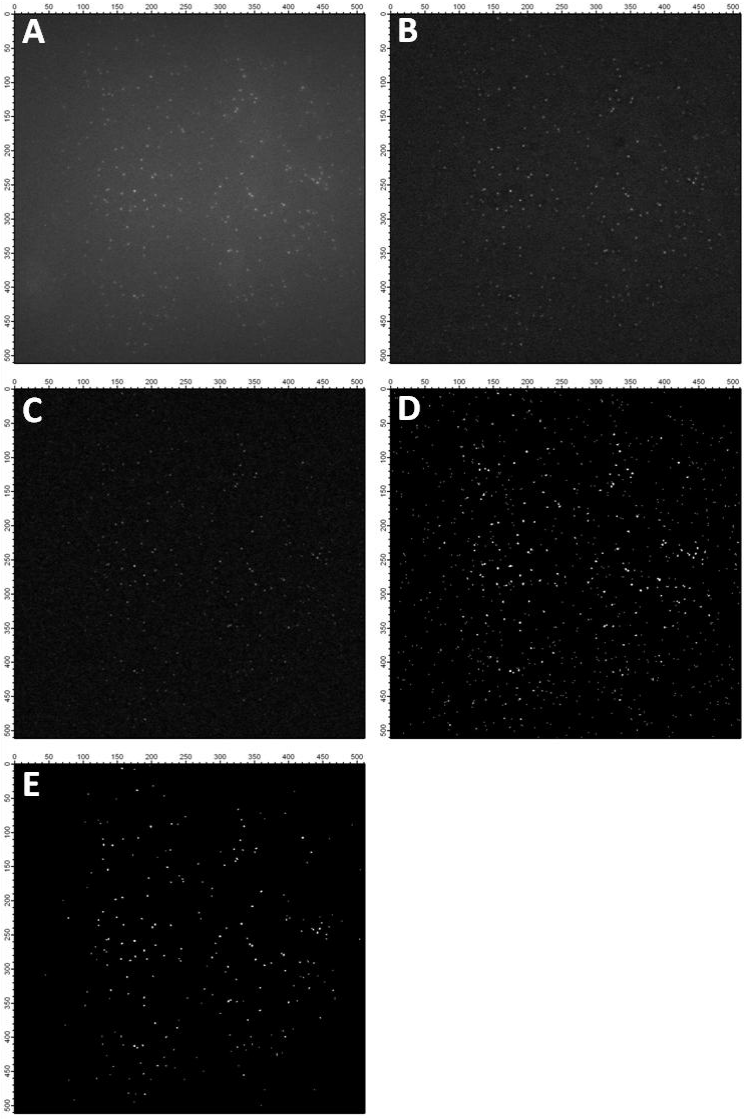


Figure 8 : Détection des particules dans une image

A) Image originale issue du film. **B)** Image après soustraction des oscillations spatiales rapides du bruit de fond **C)** Image après retranchement des oscillations spatiales lentes du bruit de fond. **D)** Image contenant toutes les particules détectées. **E)** Image contenant toutes les particules retenues pour l'analyse.

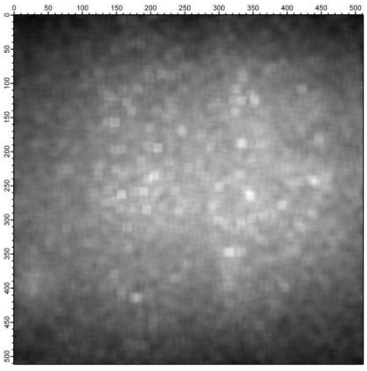


Figure 9 : Image "lissée"

Image « type » où chaque valeur de pixel est moyennée par les valeurs d'intensité présentes localement dans un carré de 7x7 pixels

Lors des expériences en molécule unique, l'ARNm est détecté grâce à un ou des oligonucléotides fluorescents hybridés en aval de l'IRES. Les fluorophores utilisés sont des ATTO. Ils ont été choisis pour leurs excellentes propriétés de brillance et de stabilité.

B. Le ribosome, les ARNt et les facteurs d'élongation

Les ribosomes utilisés dans cette étude sont purifiés à partir d'une préparation de lysat de réticulocytes[5] de lapin (RRL : Rabbit Reticulocyte Lysat). Les sous-unités ribosomiques sont dissociées dans le RRL par un traitement avec l'antibiotique puromycine avant d'être purifiées sur un coussin de sucrose. Le culot est repris dans un tampon qui contient au final les sous-unités 40S et 60S dissociées.

Les ARNt et les facteurs d'élongation ainsi que l'énergie nécessaire à la traduction sont apportés par injection de RRL dans la cellule microfluidique. Le RRL a été prétraité et enrichi par le fabricant pour améliorer les rendements de traduction. Les extraits ont été traités à la nucléase micrococcale pour éliminer les ARNm endogènes et l'enzyme a ensuite été inactivée. La population d'ARNt a été enrichie avec des ARNt purifiés provenant de foie de veau afin d'optimiser l'usage de tous les codons. Enfin, de la créatine kinase et de la créatine phosphate ont été ajoutées pour régénérer l'énergie liée à l'ATP dans le RRL.

Le RRL est un extrait rouge sang à cause de la présence d'hémoglobines. Nous nous sommes donc assurés au préalable qu'il était possible d'utiliser le RRL dans les différentes longueurs d'ondes que nous comptions utiliser. Etonnement, son utilisation est compatible avec nos expériences de microscopie. Dans le rouge (λex 640nm, λem > 665nm), il y a tout de même un léger fond de fluorescence qui donne l'avantage de détecter l'arrivée du RRL lors de l'injection dans la cellule.

3. L'analyse d'images et traitement des données

L'analyse des films comprend 2 étapes :

- La détection des particules (figure 8)
- L'analyse temporelle de l'intensité des particules

La 1ère étape pour détecter les particules présentes dans l'image consiste à filtrer le bruit de fond. Ce dernier n'est pas homogène dans le champ d'observation, il est composé d'oscillation spatiales rapides (variations locales) et lentes (tendance générale) (figure 10A).
Pour retrancher les oscillations rapides, une image «type » est créée. Il s'agit d'une cartographie de l'amplitude du bruit au cours du temps. Les valeurs d'intensité de chaque pixel de cette image correspondent aux variances d'intensité de chaque pixel dans le temps. A cette étape, des pics locaux d'intensité sont toujours présents dans l'image à cause des particules fluorescentes. Pour lisser l'image « type », chaque valeur de pixel est moyennée par les valeurs d'intensité présentes localement dans un carré de 7x7 pixels (figure 9). Les valeurs de l'image « lissée » servent à uniformiser le bruit de fond des 2 premières images du film. Pour chaque image, l'intensité de chaque pixel est divisée par la valeur qui lui est associée dans l'image « lissée » (figure 8B, figure 10B).

[5] Les réticulocytes sont les cellules précurseurs des globules rouges.

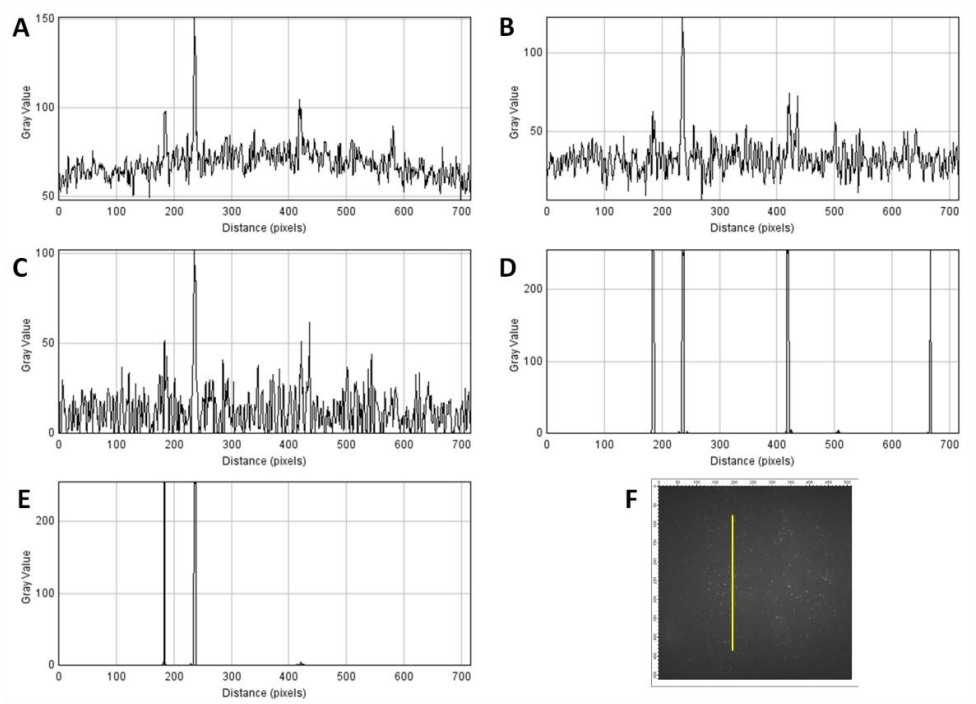


Figure 10 : Retrait du bruit de fond

Graphiques représentant l'intensité de chaque pixel présent dans la section analysée (ligne jaune image F). Les graphiques **A** à **E** correspondent respectivement aux analyses des images **A** à **E** de la figure 8.

Au fur et à mesure des traitements d'image, les oscillations rapides du bruit de fond sont retranchées entre les images **A** et **B** et les oscillations lentes sont retranchées entre les images **B** et **C**. Toutes les particules détectées ressortent dans le graphiques **D** alors que seules les particules retenues pour l'analyse apparaissent dans le graphique **E**

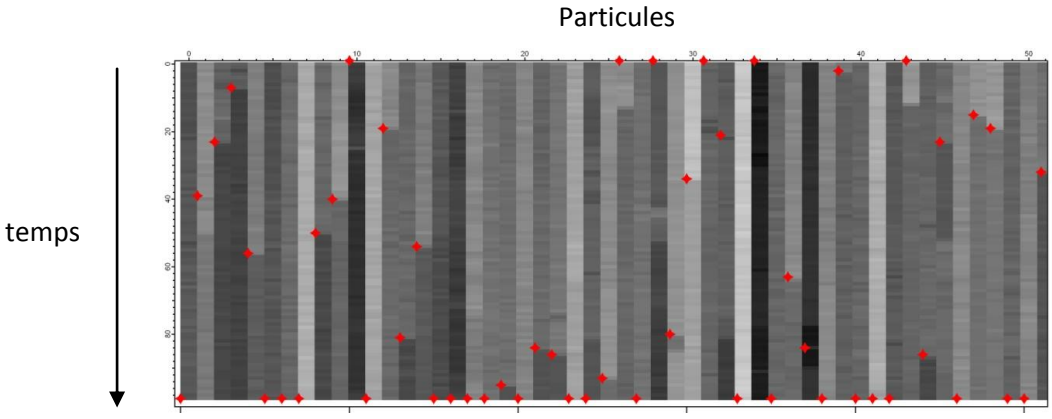


Figure 11 : Vérification de la détection de disparition des particules

Représentation de l'intensité de chaque particule en fonction du temps. Une colonne correspond à l'intensité d'une particule au cours du temps, la marque rouge indique le moment où le programme détecte que la particule s'éteint.

Les oscillations spatiales lentes sont retranchées pour chacune des 2 images grâce à un procédé basé sur celui du « rolling ball ». A cette étape, les minimums locaux d'intensité sont ramenés à 0 (figure 8C, figure 10C).
La 2nde étape est de détecter les particules de l'image qui correspondent aux fluorophores. Pour être retenues, les particules doivent avoir une taille maximale de 3*3 pixels et avoir une intensité supérieure à la valeur seuil choisie. La valeur de seuil est déterminée expérimentalement à partir de différents films (figure 8D, figure 10D). D'autre part, les particules doivent être présentes sur les 2 premières images du film (figure 8E, figure 10E). Cette étape élimine les particules en suspension qui passent transitoirement dans le champ d'observation. La position et l'aire de chaque particule sont alors répertoriées pour pouvoir les retrouver dans le film original.

Le film original est utilisé pour déterminer l'intensité dans le temps de chaque particule retenue. L'intensité associée à chaque particule correspond à la moyenne des valeurs d'intensités de l'aire de la particule. Un programme permet ensuite de détecter pour chaque particule le moment auquel elle s'éteint. Dans le cas d'une molécule unique, l'extinction doit se faire en une seule étape. Afin de vérifier la pertinence de l'analyse, il est possible de visualiser *a posteriori* pour chaque particule si la détection de l'extinction correspond bien à une extinction de la particule dans le film (figure 11).

Les instants de disparition de chaque particule sont utilisés pour tracer un graphique représentant le pourcentage de disparition des fluorophores en fonction du temps.

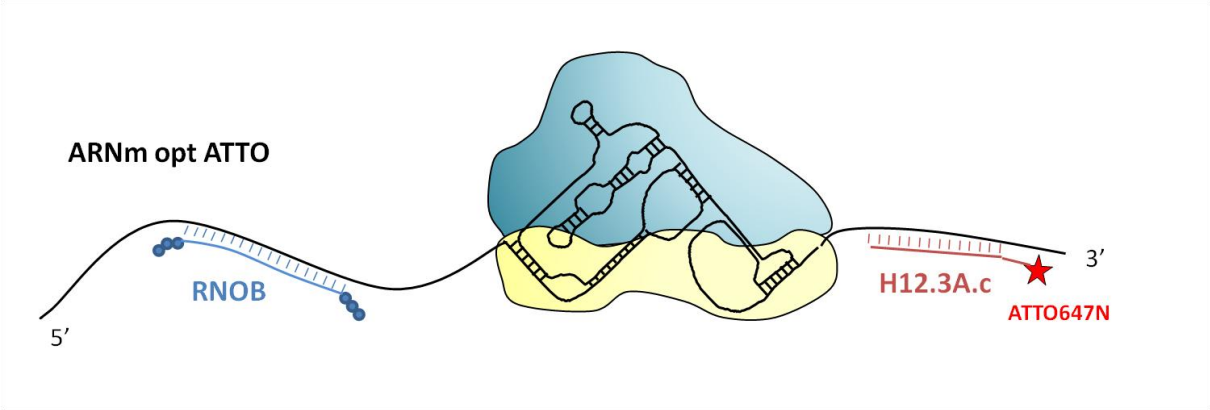


Figure 12 : ARNm opt ATTO

L'ARNm opt ATTO possède un site d'hybridation pour l'oligonucléotide RNOB, une IRES CrPV et un site d'hybridation pour l'oligonucléotide H12.3A.c. Cet oligonucléotide possède un fluorophore ATTO647N en 5'. L'ARNm opt ATTO est dépourvu de codon stop.

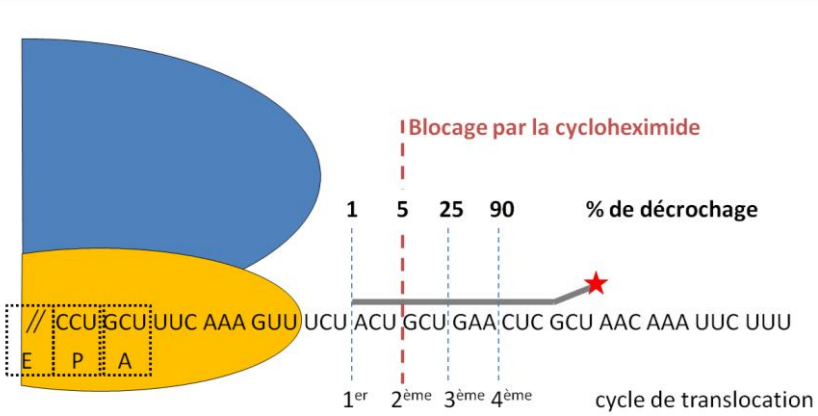


Figure 13 : Principe de l'expérience de traduction en molécule unique

Lors de l'élongation, l'activité hélicase du ribosome va décrocher l'oligonucléotide H12.3A.c. Grâce aux expériences de (Takyar, Hickerson and Noller 2005), il est possible de prédire le pourcentage de départ l'oligonucléotide à chaque cycle de translocation. Les expériences de (Pestova and Hellen 2003) permettent de prédire à quelle étape la cycloheximide bloquera le ribosome.

III. Résultats et discussion

1. L'ARNm opt ATTO

Le 1er ARNm conçu pour les expériences de traduction en molécule unique est l'ARNm opt ATTO. Cet ARNm n'est pas coiffé. Il possède, de l'extrémité 5' vers l'extrémité 3', une séquence d'appariement pour l'oligonucléotide d'accroche RNOB, une IRES CrPV et un site d'accroche pour l'oligonucléotide fluorescent H12.3A.c. De plus, il est dépourvu de codon stop dans la phase de lecture définie par le codon d'initiation de l'IRES (figure 12).

La séquence de l'oligonucléotide H12.3A.c est basée sur la séquence de l'oligonucléotide H12 utilisé dans l'étude sur l'activité hélicase du ribosome bactérien (Takyar, Hickerson and Noller 2005). Cette étude a montré que 4 cycles de translocation sont nécessaires pour décrocher l'oligonucléotide H12 lorsqu'il est hybridé à partir du 6ème codon en aval du site P du ribosome (figure 13). L'oligonucléotide H12.3A.c possède 3 adénines non appariées à l'extrémité 5'. Leur rôle est d'éloigner le fluorophore ATTO647N pour qu'il ne perturbe pas l'appariement entre l'oligonucléotide et l'ARNm. L'ATTO647N possède un maximum d'absorption à 644nm et un maximum d'émission à 669nm.

2. Fixation des ribosomes sur l'ARNm opt ATTO

Afin de vérifier que des ribosomes 80S peuvent se former spécifiquement sur les IRES de ces ARNm, j'ai réalisé des expériences de retard sur gel. Dans ces expériences, la position des ARNm dans le gel d'agarose est détectée grâce à la fluorescence des oligonucléotides H12.3A.c.

- ***L'oligonucléotide H12.3A.c versus le complexe ARNm opt ATTO /oligonucléotide H12.3A.c***

Comme on peut le constater dans les puits 1 et 2 de la figure 14, les oligonucléotides H12.3A.c migrent à la même hauteur que les complexes ARNm/oligonucléotide alors que leurs tailles sont bien différentes. Une explication est que leur migration lors d'une électrophorèse n'est pas comparable car elle ne dépend pas seulement de leur taille. Pour une longueur donnée, un ARN simple brin et un ARN double brin ne migrent pas de la même façon car la charge électrique de l'ARN double brin est plus grande que celle de l'ARN simple brin. Malgré le fait que les oligonucléotides seuls et les complexes ARNm/oligonucléotide migrent à la même hauteur, ils se distinguent par la forme de leur signal. L'oligonucléotide produit un signal diffus alors que le complexe produit une bande nette. Dans ce cas, la petite taille de l'oligonucléotide H12.3A.c explique la diffusion locale du signal.

- ***Fixation des ribosomes sur les ARNm opt ATTO***

Afin de vérifier que les sous-unités ribosomiques étaient capables de se fixer sur les ARNm opt ATTO, j'ai mis des complexes ARNm/oligonucléotide en présence d'une concentration croissante de sous-unités ribosomiques. On constate sur la figure 14 que l'ajout de sous unités ribosomiques (puits 3 à 7) entraine dès 5nM l'apparition d'un signal plus haut que dans le puits contrôle (puits 2). L'intensité du signal augmente avec la concentration en sous-unités ribosomiques. Cette bande, dite retardée, correspond au complexe ARNm/oligonucléotide/ribosome. Sa présence indique que les ribosomes seraient capables de fixer l'ARNm opt ATTO en présence de l'oligonucléotide H12.3A.c. L'autre hypothèse pour expliquer ce résultat, serait que les oligonucléotides H12.3A.c marqueraient non-spécifiquement les sous-unités ribosomiques.

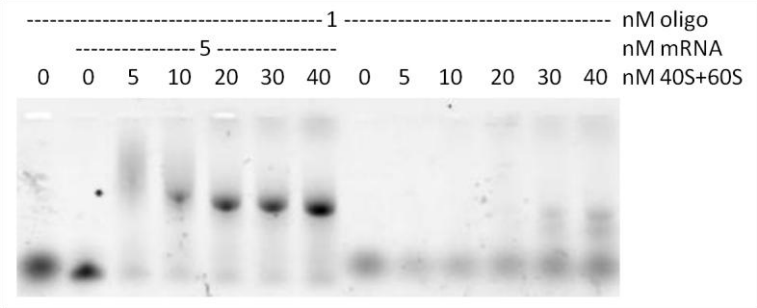


Figure 14 : Fixation des sous-unités ribosomiques sur les ARNm opt ATTO

Expérience de gel retard, les signaux correspondent à la fluorescence de l'oligonucléotide H12.3A.c. Les puits 3 à 7 testent la fixation des sous-unités ribosomiques sur l'ARNm opt ATTO. Les puits 8 à 13 testent le marquage spécifique de l'oligonucléotide H12.3A.c.

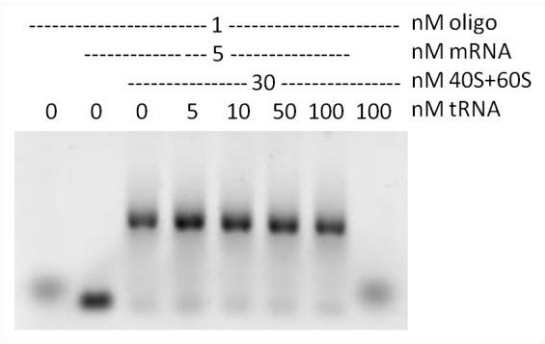


Figure 15 : Spécificité de la fixation des sous-unités ribosomiques sur l'ARNm opt ATTO

Expérience de gel retard, les signaux correspondent à la fluorescence de l'oligonucléotide H12.3A.c. Les puits 3 à 7 testent la fixation des sous-unités ribosomiques sur l'ARNm opt ATTO en présence d'ARN compétiteurs non spécifiques (ARNt).

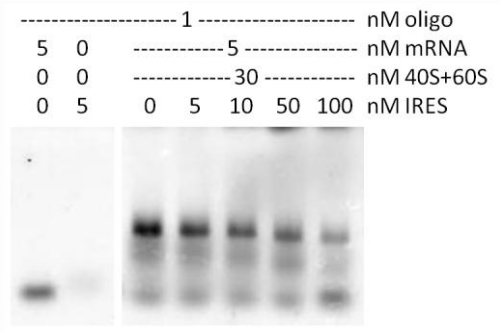


Figure 16 : Spécificité de la fixation des sous-unités ribosomiques sur l'IRES CrPV de l'ARNm opt ATTO

Expérience de gel retard, les signaux correspondent à la fluorescence de l'oligonucléotide H12.3A.c. Les puits 3 à 7 testent la fixation des sous-unités ribosomiques sur l'ARNm opt ATTO en présence d'IRES CrPV compétitrices non marquées (ARNm opt QF).

Afin de tester cette dernière hypothèse, une concentration croissante de sous-unités ribosomiques a été mise en présence d'oligonucléotides H12.3A.c (figure 14, puits 8 à 13). Des signaux diffus et retardés apparaissent lorsque les sous-unités ribosomiques sont en large excès (20-40nM) par rapport à la sonde (1nM). L'apparition et la forme du signal étant différentes de celles constatées en présence d'ARNm, on en conclut que ce ne sont pas les sous-unités ribosomiques qui sont marquées dans les puits 3 à 7 mais bien les ARNm opt ATTO dans les complexes ARNm/oligonucléotide/ribosome.

❖ ***Spécificité de la fixation des ribosomes sur les ARNm opt ATTO***

J'ai ensuite vérifié que la fixation des ribosomes sur les ARNm opt ATTO était spécifique. Une concentration croissante d'ARN compétiteurs non spécifiques, en l'occurrence des ARNt, a été ajoutée dans le mélange contenant les ARNm, les oligonucléotides et les sous-unités ribosomiques (figure 15 puits 4 à 7). L'ajout des ARNt compétiteurs ne change pas l'intensité du signal retardé et n'induit pas d'apparition d'un nouveau signal. Comme l'équilibre n'est pas déplacé, on en conclut que la fixation des ribosomes sur l'ARNm opt ATTO est bien spécifique.

❖ ***Spécificité de la fixation des ribosomes sur les IRES CrPV***

Pour finir, j'ai vérifié que la fixation des ribosomes sur les ARNm opt ATTO se faisait grâce aux IRES CrPV. A cet effet, une concentration croissante d'ARNm compétiteur spécifique a été ajoutée au mélange contenant les ARNm, les oligonucléotides et les sous-unités ribosomiques. Cet ARNm (ARNm opt QF) est quasiment identique à l'ARNm opt ATTO sauf qu'il ne contient pas de site d'appariement avec l'oligonucléotide H12.3A.c. Sur la figure 16 (puits 3 à 7), on constate que l'augmentation de concentration en ARNm compétiteurs est corrélée avec une diminution d'intensité du signal retardé. Puisque l'équilibre du signal est déplacé en présence d'ARNm compétiteur spécifique non marqué, on en conclut que les ribosomes se fixent spécifiquement à l'ARNm sur l'IRES CrPV.

L'oligonucléotide H12.3A.c est un marquage indirect, son utilisation de ne permet pas de calculer une constante de dissociation K_d pour les complexes IRES CrPV-80S. Cependant, comme les concentrations utilisées lors des expériences de gels retard sont de l'ordre de grandeur de celles utilisées lors des études en molécule unique (nM), on peut en conclure qu'il sera possible de former des complexes IRES CrPV-80S avec des concentrations aussi faibles.

3. Traduction sur une surface en mesure d'ensemble

Après avoir vérifié que les ribosomes purifiés pouvaient se fixer sur l'IRES CrPV, j'ai testé si ces ribosomes étaient capables de traduire les ARNm lorsque ces derniers sont ancrés à une surface. La phase codante de l'ARNm opt ATTO est trop petite (16 acides aminés) pour produire une protéine détectable par SDS-PAGE. Matthieu Saguy a construit un ARNm plus long en insérant la phase codante PPS0 sans codon initiateur ATG en aval de l'IRES CrPV. L'ARNm opt PPS0 ne contient pas de codon stop, il code une protéine de 261 acides aminés (~ 30 kDa) qui comprend 22 acides aminés lysines.

Lors de l'expérience, les ribosomes sont préfixés sur les appariements ARNm opt ATTO/RNOB avant de démarrer la traduction. J'ai testé les quantités en sous-unités ribosomiques suivantes : 0, 10, 50, 250 10^{-15}mol. Les complexes sont ensuite accrochés à des billes magnétiques recouvertes de

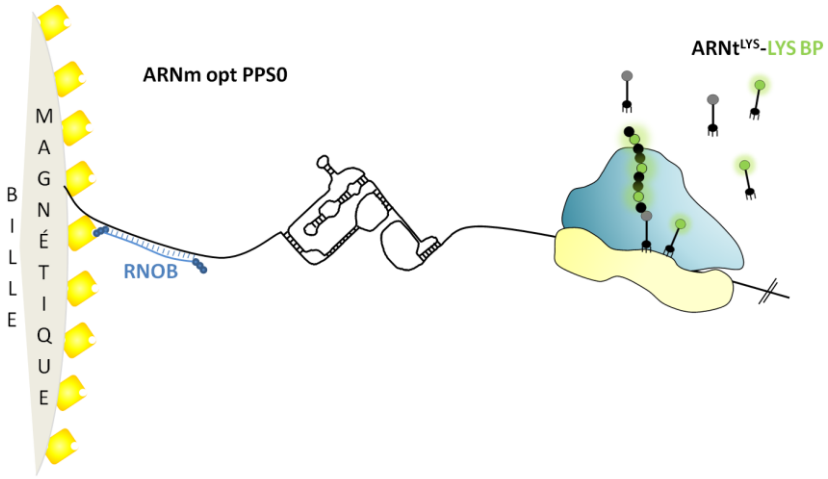


Figure 17 : Principe de l'expérience de traduction sur une surface en mesure d'ensemble

Les ARNm opt PPS0 sont fixés grâce à l'oligonucléotide RNOB sur des billes magnétiques couvertes de streptavidines. Les protéines PPS0 synthétisées sont marquées par l'incorporation d'acides aminés Lys couplés à des fluorophores BODIPY-FL (BP). Comme l'ARNm opt PPS0 est dépourvu de codon stop, les ribosomes atteignant la fin de l'ARNm ne sont pas dissociés et les peptides ne sont pas relargués.

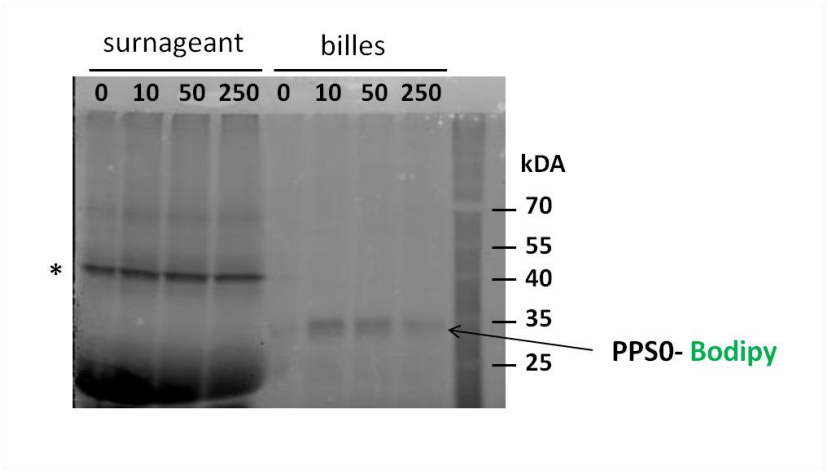


Figure 18 : Traduction de l'ARNm opt ATTO sur une surface en mesure d'ensemble

Electrophorèse dans un gel SDS-PAGE. La fluorescence du BODIPY-FL est détectée en scannant le gel. Une bande contaminante est présente dans le surnageant, il s'agit d'un contaminant caractérisé dans le RRL. Les protéines PPS0 fluorescentes sont détectées dans les puits correspondant aux billes.

streptavidines. Tous les facteurs nécessaires à l'élongation sont apportés par l'ajout de RRL. Afin de détecter les protéines PPS0 synthétisées, des $ARNt^{LYS}$ aminoacylés avec des lysines marquées avec le fluorophore BODIPY®-FL (LYS-BP) sont ajoutés au mélange réactionnel (figure 17).

L'ARNm opt PPS0 ne contient pas de codon stop. Donc quand un ribosome atteint la fin de l'ARNm, il n'est pas dissocié par les facteurs de terminaison et il reste sur l'ARNm. En fin de traduction, les billes magnétiques permettent d'isoler des complexes comprenant des ribosomes avec les peptides naissants marqués avec des acides aminés LYS-BP. Les protéines du surnageant (RRL) et celles associées aux billes sont analysées par SDS-PAGE. Le gel est scanné avec un laser (λ 488nm) pour détecter la fluorescence associée au fluorophore BODIPY-FL (λ510nm) entre 500 et 540nm.

Dans les puits correspondant aux billes (figure 18, puits 5 à 8), une protéine fluorescente autour de 35kDa est détectée. La protéine PPS0 a une taille attendue de 30kDa, mais avec les fluorophores BODIPY-FL la taille apparente peut augmenter. Le fluorophore BODIPY-FL a une masse molaire de 250 g/mol. Si chacune des 22 lysines est marquées, la taille de la protéine augmente de 5,5kDa ce qui est compatible avec la taille de la protéine observée[6]. Dans le premier puits, il n'y a pas de bande détectable. Cette observation montre que malgré le fait que des ribosomes soient présents dans le RRL, il n'y a pas de traduction détectable lorsque les ribosomes ne sont pas préfixés sur l'ARNm. Cela est probablement dû à la présence des facteurs d'initiation eIF1, eIF1A et eIF3 dans le RRL. Ces facteurs réduisent de 95% l'efficacité de fixation des sous-unités 60S sur les complexes d'initiation 48S comprenant une sous-unité 40S sur une IRES CrPV (Pestova, Lomakin and Hellen 2004).

En conclusion, les ribosomes préfixés sur des IRES CrPV sont capables de traduire des ARNm accrochés sur une surface.

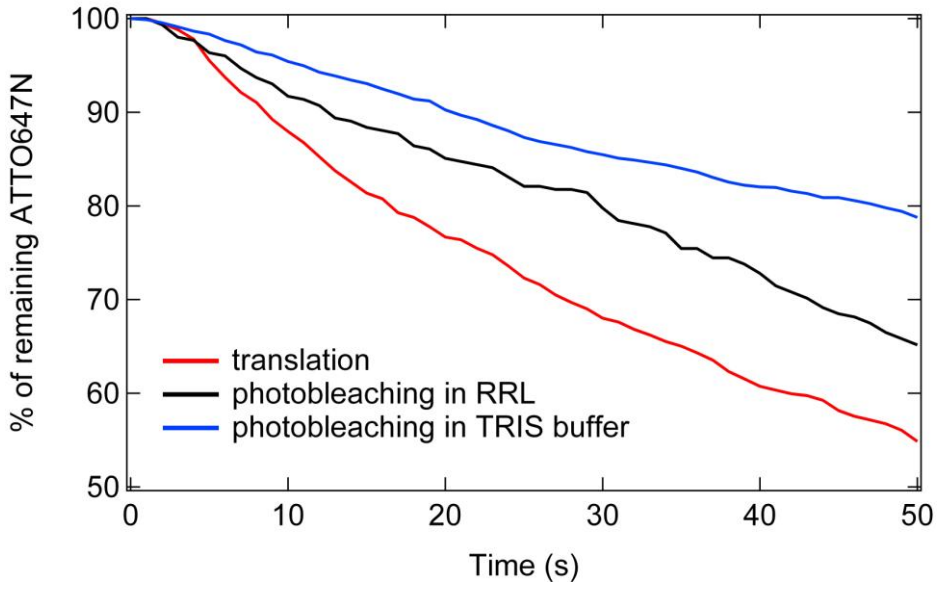
4. Traduction en molécule unique

❖ *Expérience de traduction*

Les ARNm opt ATTO ont été utilisés pour réaliser les expériences de traduction en molécule unique. Les ARNm sont hybridés avec les oligonucléotides RNOB et H12.3A.c et incubés en présence de sous-unités ribosomiques. Les complexes sont injectés dans la cellule microfluidique pour être fixés à la lamelle. La chambre est ensuite lavée pour enlever les molécules non accrochées à la surface et les fluorophores en excès. La traduction est déclenchée par l'injection de RRL juste après avoir lancé l'acquisition des images. La réaction est réalisée à 31°C.

Les résultats de l'expérience sont présentés sous forme d'un graphique représentant la cinétique de disparition de la fluorescence des ATT0647N (graphique 1). Dans cette expérience de traduction, la disparition des signaux de fluorescence peut s'expliquer par 2 composantes principales. La première composante est le décrochement de l'oligonucléotide H12.3A.c lors de la traduction, la seconde composante est le photoblanchiment de l'ATTO647N. Le photoblanchiment correspond à une perte irréversible des propriétés de fluorescence, tous les fluorophores excités y sont soumis.

[6] En biochimie, les unités de masse molaire (g/mol) et de masse en dalton (Da) sont équivalentes. 1Da = 1g/mol



Graphique 1 : Traduction et photoblanchiment

Cinétiques de disparition des ATTO647N lors d'une expérience de traduction en molécule unique (courbe en rouge), et lors d'expériences de photoblanchiment dans le RRL (courbe en noir). Les expériences sont réalisées à 31°C. La courbe bleue correspond à la cinétique de disparition des ATTO647N dans un tampon Tris.

❖ *Photoblanchiment de l'ATTO647N dans le RRL*

Pour évaluer la part de photoblanchiment dans l'expérience, nous avons réalisé des expériences dans lesquelles seul le photoblanchiment des ATTO647N est observé. Le photoblanchiment a été mesuré dans le RRL ainsi que dans un tampon Tris[7] qui nous sert de référence. Les mesures dans le RRL sont effectuées systématiquement après chaque expérience de traduction sur une région de la lamelle qui n'a pas été éclairée auparavant.

D'après le graphique n°1, on constate que le photoblanchiment des ATTO647N est environ 2 fois plus rapide dans le RRL que dans le tampon de référence. Cette différence peut s'expliquer par l'abondance d'hémoglobines dans le RRL. Ces protéines ont une forte affinité pour le dioxygène (O_2) et augmentent donc la quantité d'O_2 dissoute dans la solution. Or, l'oxygène accélère le photoblanchiment en oxydant les fluorophores qui sont dans un état excité. Cette réaction est photoinduite.

Malgré un photoblanchiment des ATTO647N accéléré dans le RRL, la disparition des fluorophores est encore plus rapide dans l'expérience de traduction. Cette disparition plus rapide révèle l'existence d'un autre processus.

La disparition de la fluorescence observée au cours du temps lors de la traduction semble suivre une loi de décroissance exponentielle. Elle peut être modélisée par la formule suivante :

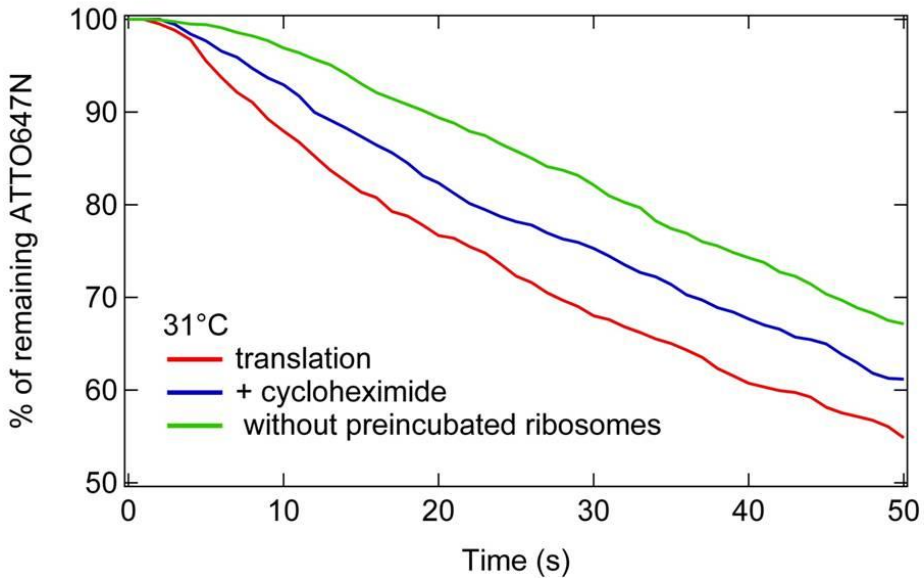
$$\boldsymbol{n(t) = e^{-t\,kobs} = Na\, e^{-t\ ktrad}\, e^{-t\ kb} + Ni\ e^{-t\,kb}}$$

Le terme « Na » correspond à la fraction de population d'ATTO647N associée avec des ribosomes actifs. Le terme « Ni » correspond au reste de la population d'ATTO647N qui est considéré comme inactive. Les taux de disparition k_{obs}, k_{trad} et k_b correspondent respectivement à la décroissance de fluorescence observée lors de l'expérience, à la décroissance associée à la traduction et à la décroissance associée au photoblanchiment. Le membre de l'équation en bleu représente la disparition de fluorescence des ATT0647N associés à des ribosomes actifs. Ces particules sont soumises à la fois à la traduction et au photoblanchiment. Le membre en rouge représente la disparition de fluorescence du reste des ATTO647N qui sont soumis seulement au photoblanchiment.

En calculant la pente de la tangente à l'origine, on obtient une valeur $k_{obs} = 1{,}30.10^{-2} \pm 0{,}03\ s^{-1}$. Cela signifie que 1,3% de la population restante d'ATTO647N disparait à chaque seconde.

Pour le moment, la cinétique de disparition de la fluorescence due à la traduction est modélisée comme suivant une loi de décroissance exponentielle. Ce parti pris est dû au photoblanchiment qui compte pour une part importante de nos mesures et dont la cinétique suit une loi de décroissance exponentielle. Le modèle a donc été établi pour des raisons mathématiques afin d'estimer au mieux la part de la traduction dans nos mesures. En réalité nous pensons que la cinétique de disparition due à la traduction doit suivre une loi de décroissance sigmoïde. Suite à l'injection du RRL, il y doit y avoir une courte phase de latence avant que les ribosomes commencent à décrocher les premiers oligonucléotides. Suit alors une phase de décroissance due au départ des oligonucléotides dont la pente dépend de la vitesse de décrochement. Cette décroissance doit finir par atteindre un plateau

[7] Tampon Tris standard : 100mM Tris, 250mM NaCl, pH=7,5.



Graphique 2 : Cinétiques de disparitions des ATTO647N dans les expériences de traduction

Cinétiques de disparitions des ATTO647N lors d'une expérience de traduction en molécule unique (courbe en rouge). Des expériences contrôles ont été réalisées en présence de cycloheximide (courbe en bleu) et sans pré-incubation préalable avec des sous-unités ribosomiques (courbe en vert). Les expériences sont réalisées à 31°C.

lorsque tous les ribosomes actifs ont décroché leurs oligonucléotides. Les phases d'accélération au début et de ralentissement à la fin dépendent de la synchronisation de l'activité des ribosomes. Dans ce modèle, la représentation du nombre de disparitions d'ATTO647N dues à la traduction en fonction du temps aurait l'aspect d'une fonction gaussienne ce qui est modélisable mathématiquement. Cette analyse sera accessible si nous parvenons à diminuer fortement la part du photoblanchiment dans nos mesures. Ce point sera abordé dans la partie « Améliorations à court terme ».

❖ *Expériences contrôles et évaluation de la vitesse de traduction*

Dans le but de vérifier que l'accélération de la disparition des ATTO647N est bien due à la traduction, nous avons réalisé 2 expériences contrôles.

Dans le 1er témoin, nous avons tiré parti du fait que dans le RRL il y a des facteurs d'initiation qui inhibent la fixation des sous-unités ribosomiques sur l'IRES CrPV. Les ARNm opt ATTO hybridés avec les oligonucléotides RNOB et H12.3A.c ont été injectés dans la cellule microfluidique sans incubation préalable avec des sous-unités ribosomiques. Après rinçage, le RRL est injecté dans la chambre. Les résultats sont présentés dans le graphique n°2. On constate que comme pour le photoblanchiment, la disparition des ATTO647N est plus lente que lors de la traduction. Le calcul de la pente de la tangente à l'origine donne une valeur k_{obs} de $0{,}73.10^{-2} \pm 0.01\ s^{-1}$. Cette valeur est 2 fois plus faible que celle observée lors de la traduction.

Dans le 2nd témoin, la traduction a été réalisée en présence de cycloheximide. Cet antibiotique bloque la traduction chez les eucaryotes en inhibant la translocation par le facteur eEF2. La molécule se fixerait au site E de la sous-unité 60S seulement en présence d'un ARNt dans ce site. Lors d'une initiation canonique, le ribosome est donc bloqué après le premier cycle de translocation lorsque l'ARNt initiateur se déplace du site P au site E (Schneider-Poetsch et al. 2010). Dans le cas de l'IRES CrPV, il a été montré que la cycloheximide bloque le ribosome après le 2ème cycle de translocation lorsque le 1er ARNt entre au site E (figure 13) (Pestova and Hellen 2003). Les résultats sont présentés dans le graphique n°2. On constate que là encore, la disparition des ATTO est plus lente que lors de la traduction. Néanmoins, ces départs sont plus rapides que pour le témoin précédent. Le calcul de la pente de la tangente à l'origine donne une valeur k_{obs} de $0{,}99.10^{-2} \pm 0{,}02\ s^{-1}$. Cette valeur est comprise entre les 2 valeurs mesurées auparavant. D'après les prédictions, la cycloheximide bloque le ribosome après 2 cycles de translocation alors que 5% des oligonucléotides H12.33A.c devraient être décrochés. La différence entre les 2 constantes des expériences contrôles peut donc s'expliquer par la proportion d'oligonucléotides H12.3A.c s'étant décrochée avant l'inhibition par la cycloheximide.

Afin d'estimer une vitesse de traduction moyenne, la valeur de la constante k_{trad} a été extrapolée d'après la formule : $n(t) = e^{-t\,k\mathrm{obs}} = Na\, e^{-t\ ktrad}\, e^{-t\ kb} + Ni\ e^{-t\,kb}$

Lorsque $(t\ k_{obs})$, $(t\ k_{trad})$, $(t\ k_b) << 1$, on obtient le développement mathématique suivant :

$$1 - k_{obs} = Na\,(1 - K_{trad}\,t)(1 - k_b\,t) + Ni\,(1 - k_b\,t)$$

$$1 - k_{obs} = Na\,(1 - K_{trad}\,t - k_b\,t) + Ni\,(1 - k_b\,t)$$

$$k_{obs} = Na\,k_{trad}\,t + (Na + Ni)\,k_b\,t$$

Comme Na +Ni =1, on obtient $k_{trad} = (k_{obs}-k_b)/ Na$

La valeur k_{obs} correspond à la valeur obtenue pour l'expérience de traduction. Pour la valeur k_b, c'est la valeur (k_{obs} = $0{,}73.10^{-2}$ ±0,01 s^{-1}) obtenue pour le témoin sans pré-incubation de sous-unités ribosomiques qui a été utilisée. Cette valeur a été choisie car contrairement aux expériences contrôles de blanchiment, elle tient aussi compte des quelques disparitions qui peuvent être dues à des décrochements lors de l'injection du RRL. Le tampon présent dans la chambre et le RRL sont deux solutions ayant des concentrations salines différentes. L'échange de ces solutions lors de l'injection peut donc déplacer légèrement l'équilibre d'appariement entre les oligonucléotides et l'ARNm.

Comme nous ne pouvons pas détecter directement les ribosomes actifs à la surface, il est impossible d'estimer précisément la valeur Na mais il est cependant possible de l'approximer. D'après l'étude sur l'affinité de la sous-unité 40S pour l'IRES CrPV (Jan and Sarnow 2002), la constante de dissociation (Kd) du complexe IRES CrPV / 40S est de 24 ±6 nM. Dans nos conditions, la concentration des sous-unités 40S est en large excès par rapport à la concentration en ARNm. On peut donc estimer que :

[IRES/40S] = [ARNm total]*[40S] / (Kd + [40S])

Soit **[IRES/40S]** = 1nM * 28nM/ (24nM + 28nM) **= 0.53nM**

En faisant l'hypothèse que les complexes IRES CrPV/80S se comportent de la même manière que les complexes IRES CrPV/ 40S, on peut estimer que 0,53nM des ARNm opt ATTO à la surface sont liés à un ribosome. Cela correspond donc à un taux d'occupation d'environ 50%.

D'après l'étude sur les premiers cycles d'élongation à partir de l'IRES CrPV dans du RRL (Pestova and Hellen 2005), la présence des facteurs d'initiation eIF1, eIF1A et eIF3 diminuerait d'au moins 20% l'efficacité des premiers cycles d'élongation donc seulement 40% des ARNm à la surface seraient traduits.

Nous avons donc fixé la valeur de Na à 0,4.

On obtient donc : $\mathbf{k_{trad}}$ = $(1{,}30.10^{-2} - 0{,}73.10^{-2})/0{,}4$ **= $1{,}42.10^{-2}$ s^{-1}** et un temps de disparition caractéristique de $\mathbf{T_{50\%\ de\ disparitions}}$ = ln (0,5) / $(-\ 1{,}5.10^{-2})$ **= 46s**.

Comme 4 cycles de translocation sont nécessaires pour détacher efficacement l'oligonucléotide H12.3A.c, la vitesse de translocation est d'environ un 1 cycle toutes les ~ 11 secondes. Comparé à une vitesse de traduction *in vitro* dans du RRL de l'ordre de 1 à 2 cycles/seconde (Scott et al. 1979), cette estimation donne une vitesse de translocation lente. Cette observation peut s'expliquer par 3 effets. Tout d'abord, il s'agit d'une mesure in vitro dans laquelle l'accroche à la surface peut inhiber la traduction en ralentissant le recrutement des facteurs d'élongation. D'autre part, le ribosome doit quitter l'IRES CrPV alors qu'il n'y a pas d'ARNt aminoacylé au site P. Il n'est pas exclus que cette étape se fasse plus lentement que lors d'une initiation canonique. A ce jour, aucune information n'est disponible sur la cinétique de cette étape. Enfin, le départ du ribosome de l'IRES se fait de manière concomitante avec la « butée » sur l'oligonucléotide. Le décrochement de l'oligonucléotide par l'activité hélicase du ribosome est une étape qui ralentit la traduction. Ces 3 effets cumulés font que les ribosomes ne sont pas dans une configuration favorable pour traduire rapidement les premiers codons.

Dans l'étude de TIRFM où des protéines GFP ont été traduites à partir de ribosomes fixés spécifiquement à la surface, 12% des ribosomes étaient actifs (Uemura et al. 2008). Dans le cas où seulement 10% des ribosomes seraient actifs (Na = 0,1), la constante k_{trad} vaut $6{,}0.10^{-2}$ s^{-1} et le temps de disparition caractéristique est de 11s. La vitesse de translocation est alors d'un cycle toutes les 3 secondes.

En l'état actuel, les ordres de grandeur des vitesses estimées ne pas aberrants mais il reste difficile d'établir un scénario sur la vitesse de traduction. Nous comptons donc améliorer certains paramètres pour renforcer l'analyse des résultats.

❖ *Améliorations à court terme*

➢ Le photoblanchiment

Dans nos expériences, le photoblanchiment est un paramètre non négligeable qui complique la détection des événements de traduction ainsi que l'analyse des données. Il s'agit donc du paramètre à améliorer en premier lieu.

Il existe des réactions enzymatiques qui consomment l'O_2 dissout dans une solution. Par exemple, le système PCA/PCD est basé sur l'activité de l'enzyme PCD (protocatechuate-3,4-dioxygenase) et de son substrat le PCA (acide protocatechuique) (Aitken, Marshall and Puglisi 2008). L'utilisation du système PCA/PCD permet d'augmenter d'un facteur 30 le temps de demi-vie des ATTO647N dans un tampon standard. Malheureusement, l'utilisation du PCA/PCD dans le RRL est moins évidente car le photoblanchiment n'est plus constant au cours du temps. Il est nécessaire d'attendre 30 min pour que le temps de demi-vie des ATTO647N soit augmenté seulement d'un facteur 4 (Le Gall 2011).

Pour l'expérience que nous réalisons, il est important d'obtenir un facteur d'augmentation supérieur à 4. L'idéal serait d'obtenir au moins un facteur 15 afin que les ATTO647N deviennent suffisamment stables pour que le photoblanchiment devienne négligeable.

Afin de contourner le problème du photoblanchiment, nous comptons tester 2 nouvelles stratégies. Dans la première, nous allons utiliser des extraits de cellules humaines HeLa (Pierce) au lieu du RRL. Le système PCA/PCD sera ajouté aussi aux extraits.
Si cette stratégie échoue, nous prévoyons dans la seconde stratégie d'ôter l'hémoglobine des extraits de RRL en passant les extraits sur une colonne contenant une résine Ni-NTA (protocole communiqué par Ramanujan Hedge). Les extraits ainsi purifiés seront aussi complémentés avec le système PCA/PCD. Nous espérons qu'une de ces stratégies permettra de réduire le photoblanchiment sans affecter l'efficacité de traduction.

➢ Le marquage des ribosomes

L'idéal pour l'expérience de traduction que nous avons développée serait de pouvoir aussi détecter les ribosomes actifs. Afin de s'assurer de la présence d'un ribosome entier, il faudrait un marquage de la sous-unité 60S qui soit spécifique et sans effet sur l'activité du ribosome. Ce point n'est pas réalisable avec les sous-unités ribosomiques purifiées à partir de RRL car elles ne peuvent être modifiées pour créer un site d'accroche spécifique.

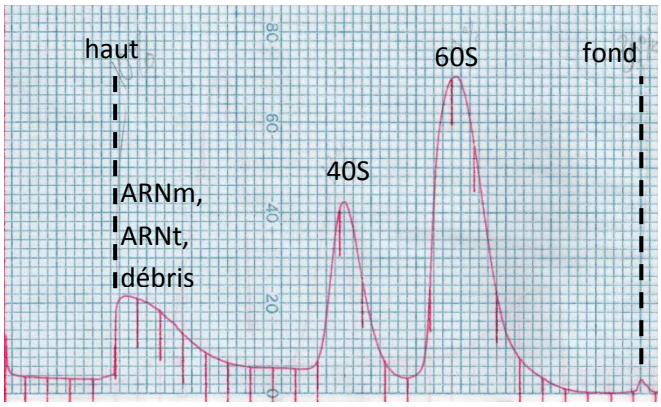


Figure 19 : Purification de sous-unités ribosomiques à partir d'extrait cellulaire de *S. cerevisiae*

Isolation et collecte des sous-unités ribosomiques 40S et 60S dans un gradient de sucrose 10-30%.

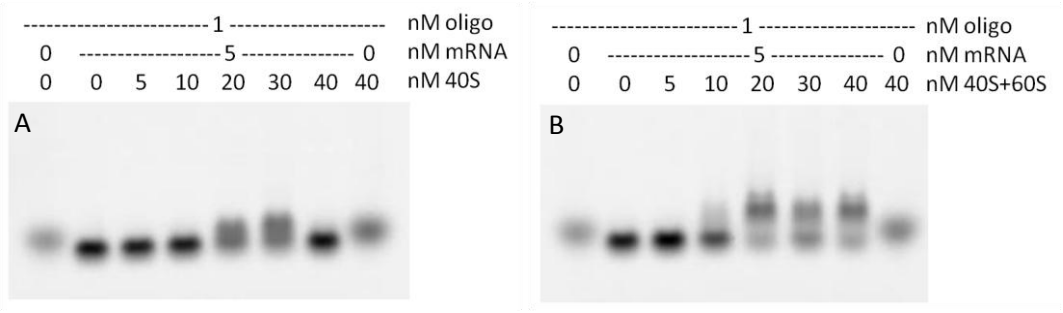


Figure 20 : Fixation des sous-unités ribosomiques de levure sur l'IRES CrPV de l'ARNm opt ATTO

Expérience de gel retard, les signaux correspondent à la fluorescence de l'oligonucléotide H12.3A.c. **A)** Les puits 3 à 7 testent la fixation des sous-unités 40S sur l'ARNm opt ATTO. **B)** Les puits 3 à 7 testent la formation de ribosomes 80S sur l'ARNm opt ATTO.

Dans un premier temps, un marquage aspécifique des sous-unités ribosomiques avec des ATTO550 serait suffisant pour les co-localiser avec l'oligonucléotide H12.3A.c. Cette expérience permettra de déterminer la proportion exacte d'ARNm opt ATTO associée à un ribosome 80S sur la surface. La valeur obtenue ne donnera pas la proportion de ribosomes actifs mais permettra de déterminer une valeur maximale pour la constante Na dans nos équations.

Afin de marquer indépendamment les sous-unités ribosomiques avec des fluorophores, j'ai mis au point au laboratoire la technique de purification des sous-unités ribosomiques à partir de levures *Saccharomyces cerevisiae* (Acker et al. 2007). A ce jour, j'ai obtenu des fractions de sous-unités 40S et 60S (figure 19). Des expériences de retard sur gel avec l'ARNm opt ATTO montrent que les sous-unités 40S se fixent sur l'ARNm et que l'ajout de sous-unités 60S permet de former des ribosomes 80S (figure 20). A présent, il reste à tester si ces ribosomes sont fonctionnels dans du RRL. Ces expériences préliminaires ouvrent la porte au marquage spécifique du ribosome au laboratoire. A présent que la structure atomique du ribosome de la levure *S. cerevisiae* est disponible (Ben-Shem et al. 2011), il est possible de concevoir des marquages de protéines ribosomiques ou de modifier localement un ARNr pour y hybrider un oligonucléotide fluorescent (Petrov and Puglisi 2010).

- Autres expériences contrôles

Dans l'expérience que nous avons conçue, l'élongation de la traduction est détectée par la perte d'un signal de fluorescence ce qui est toujours plus délicat à analyser que l'apparition d'un signal.

Nous avons déjà essayé de traduire les ARNm opt ATTO en présence d'ARNt^{LYS} couplés à des LYS-BODIPY-FL. Comme dans l'ARNm opt PPS0 utilisé pour les mesures d'ensemble, l'ARNm opt ATTO contient de nombreux codons Lys et il est dépourvu de codon stop. Le principe de l'expérience était que lors de la traduction, des acides aminés LYS-BP sont incorporés ce qui rend le peptide naissant détectable à la surface de la lamelle. Ces expériences n'ont pas fonctionné car la concentration maximale utilisable d'ARNt^{LYS} couplés à des LYS-BODIPY-FL en suspension est de 5-10nM. A cette concentration dans le RRL, les ARNt aminoacylés fluorescents sont en compétition avec les ARNt déjà présents. De plus, leur accommodation est plus lente que des ARNt normaux (communication personnelle de Barry Cooperman) donc leur incorporation est fortement défavorisée.

Cette expérience illustre l'intérêt de notre système initial avec l'oligonucléotide H12.3A.c pour la microscopie à onde évanescente. Bien qu'il s'agisse de disparition de signaux de fluorescence, il n'y a pas de problème de concentration limitante pour le marquage des ARNm et le marquage ne rentre pas en compétition avec un des facteurs de la traduction. De plus, comme ces facteurs ne sont pas marqués il est possible d'utiliser un extrait cellulaire comme le RRL dans lequel les concentrations des facteurs ne sont pas limitantes pour une étude cinétique de l'élongation.

Afin de montrer qu'une partie des disparitions d'ATTO647N est due à de la traduction de l'ARNm opt ATTO, j'ai construit un autre ARNm dans lequel 3 codons ont été insérés juste en amont du site d'hybridation de l'oligonucléotide H12.3A.c. Il s'agit de l'ARNm opt ATTO bis (figure 21). S'il y a effectivement de la traduction, dans cette configuration la disparition des oligonucléotides sera retardée par rapport à la construction initiale. Ces expériences sont actuellement en cours. Les résultats préliminaires indiquent que l'importance du photoblanchiment empêche de tirer des

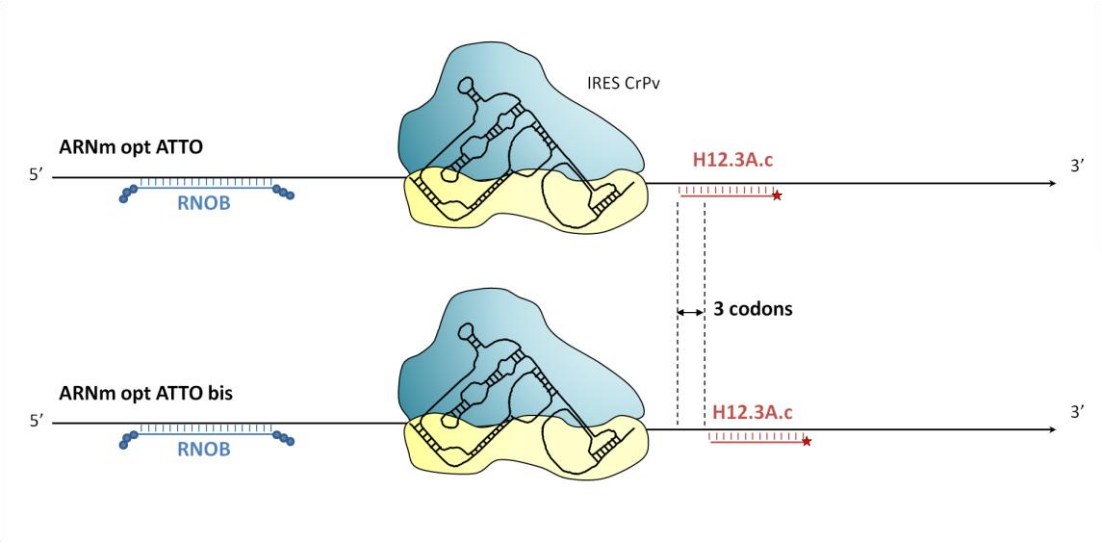


Figure 21 : ARNm opt ATTO bis

Par rapport à l'ARNm opt ATTO, l'oligonucléotide H12.3A.c est décalé de 3 codons en aval dans l'ARNm opt ATTO bis. Le but est de détecter un délai dans la disparition de la fluorescence de l'ATTO647N entre les 2 constructions. Ce délai prouvera que la disparition de fluorescence est bien liée à l'élongation de la traduction.

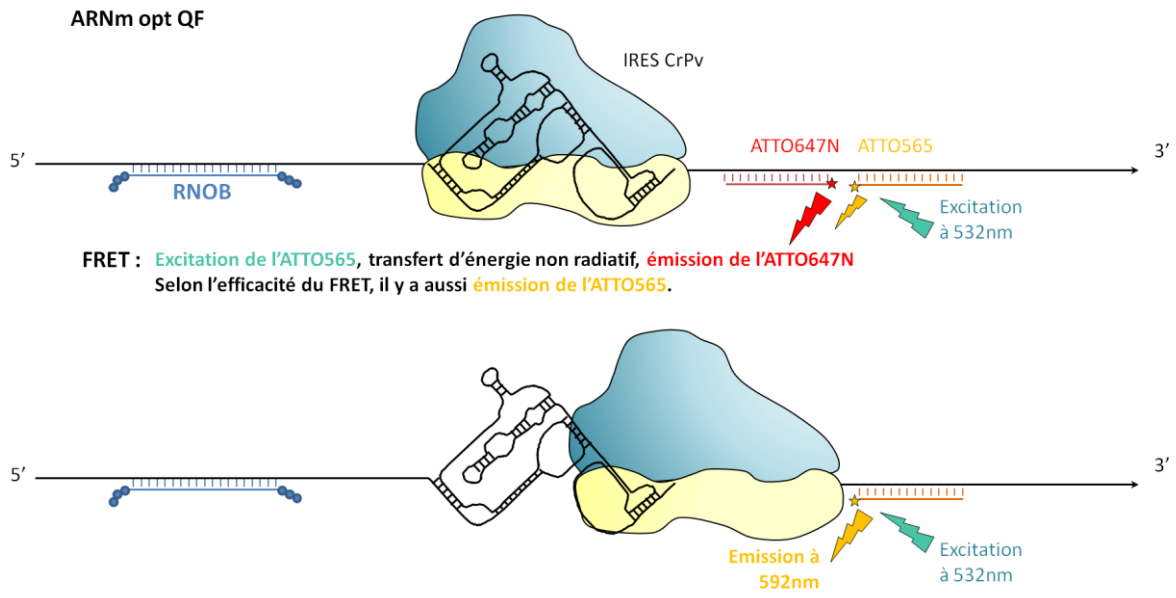


Figure 22 : ARNm opt QF

Dans l'ARNm opt QF, 2 oligonucléotides sont hybridés en aval de l'IRES CrPV. L'oligonucléotide amont est marqué en 5' avec un ATTO647N et celui en aval marqué en 3' avec un ATTO565. Lorsque les 2 oligonucléotides sont hybridés sur l'ARNm, il y a un signal de FRET entre les 2 fluorophores. Si le FRET n'est pas efficace à 100%, les 2 fluorophores peuvent être détectés lorsque l'ATTO565 est excité.

Lors de l'élongation de la traduction, le premier oligonucléotide est décroché ce qui fait disparaitre le signal de FRET. Seule la fluorescence de l'ATTO565 est détectée lorsqu'il est excité.

conclusions claires quant à la cinétique de traduction. Il est donc crucial de diminuer en premier lieu le photoblanchiment avant de répéter ces expériences.

J'ai aussi construit un autre ARNm dans lequel l'élongation est détectée par la variation d'un signal de smFRET. Dans cet ARNm opt QF (figure 22), 2 oligonucléotides sont hybridés en aval de l'IRES CrPV. Les séquences d'appariement pour ces 2 oligonucléotides ont été conçues de manière à ce qu'ils aient la même stabilité d'appariement. Les oligonucléotides sont marqués avec des fluorophores compatibles pour faire du FRET. Dans cette configuration, lorsque l'ATTO565 est excité il transmet une partie de son énergie par FRET à l'ATTO647N qui émet alors de la fluorescence. Ce signal de smFRET indique que les 2 oligonucléotides sont présents sur l'ARNm. Au cours de la traduction, le ribosome décroche le 1er oligonucléotide portant l'ATTO647N. Seule la fluorescence du 2nd oligonucléotide portant l'ATTO550 est alors directement détecté. La perte du signal de smFRET est alors le témoin de la traduction.
Bien qu'élégante sur le papier, cette expérience se heurte à différents écueils techniques. Il est notamment crucial d'apparier les oligonucléotides de telle manière que l'écart entre les 2 fluorophores soit optimal pour détecter un smFRET. Les expériences préliminaires réalisées avec cet ARNm ont montré qu'hybrider 2 sondes fluorescentes sur un même ARNm n'était pas aisé. Cette expérience nécessitera d'optimiser les conditions d'hybridation pour obtenir un maximum d'ARNm portant les deux sondes simultanément.

IV. Perspectives

Les expériences que nous avons réalisées constituent une preuve de principe démontrant qu'il est possible de réaliser des expériences en molécule unique pour étudier le ribosome eucaryote. Il s'agit de techniques lourdes à mettre en place qu'il ne faut utiliser que lorsqu'elles sont les seules à être adaptées pour répondre à la question biologique posée.

1. L'utilisation l'ARNm opt ATTO

Une fois que les expériences seront optimisées, il sera possible de concevoir des expériences pour poser des questions biologiques précises. La force des études en molécule unique est de pouvoir post synchroniser les événements lors de l'analyse des données. Pour cela, il est nécessaire d'avoir un événement fixe qui soit présent dans chaque réaction et qui servira de repère temporel.

Il est difficile de poser des questions biologiques directement à partir de l'ARNm opt ATTO et de la traduction dans un extrait cellulaire. Cependant, il est possible d'adapter la séquence de l'ARNm pour faire des études cinétiques. Par exemple, il serait possible d'étudier la cinétique de translecture d'un codon stop dans différents contextes nucléotidiques. Pour cela, l'ARNm est marqué avec 2 sondes portant des fluorophores différents. Ces sondes sont séparées par une courte séquence comprenant la séquence à analyser. La disparition du signal de fluorescence associée au 1er oligonucléotide sert de témoin pour sélectionner les particules dont l'ARNm est traduit et postsynchroniser les événements. Quant à la disparition du signal de fluorescence associée au 2nd oligonucléotide, elle sert de témoin pour identifier la fraction de particules dont l'ARNm est translu. Le temps entre les disparitions des 2 signaux de fluorescence permet ensuite de calculer précisément des cinétiques de translecture propres à chaque séquence. La comparaison des cinétiques de translecture dans différents contextes nucléotidiques en présence ou absence de molécules inhibitrices permettraient de mieux comprendre les dynamiques associées à l'étape de terminaison.

Dans l'ARNm, l'IRES CrPV est un outil efficace pour initier l'élongation de la traduction sans avoir à passer par une initiation canonique. Cette IRES pourrait aussi être utilisée comme sujet d'étude. Peu d'informations sont disponibles pour les 1ers cycles de translocation lorsque l'IRES traverse le ribosome. Cette initiation est vraiment particulière car il n'y a pas d'ARNt initiateur donc le premier acide aminé incorporé est celui apporté par le 1er ARNt accommodé au site A. Des expériences de smFRET entre le pseudonoeud n°1 de l'IRES et le premier ARNt accommodé permettraient d'obtenir des informations cinétiques précises sur cette étape.

2. Marquage de facteurs de la traduction : Système de traduction purifié et Zero Mode Wave Guide

A. Système de traduction purifié

Pour le moment, les extraits cellulaires comme le RRL apportent l'énergie et les molécules nécessaires à l'élongation et la terminaison de la traduction. Cependant, il est difficile d'en modifier les paramètres pour agir sur la traduction autrement qu'en y ajoutant des molécules inhibitrices. Comme nous l'avons vu avec les ARNt-LYS BODIPY FL, le problème de compétition entre molécules

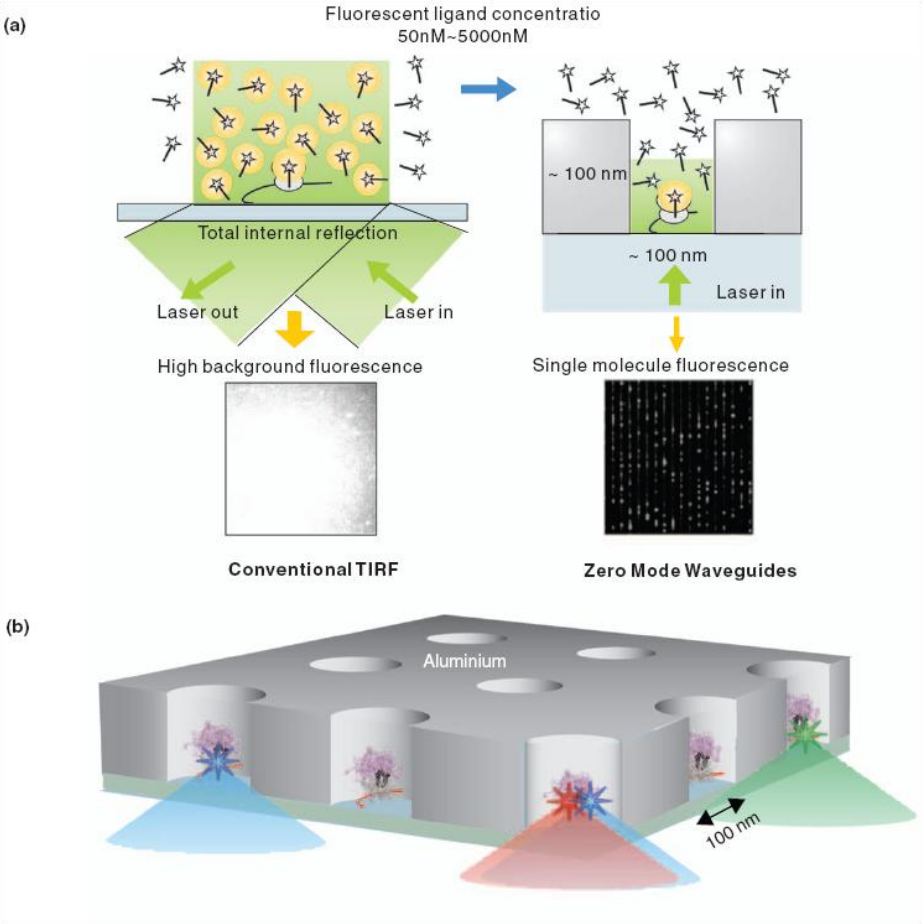


Figure 23 : Intérêt du Zero Mode Waveguide

La microscopie à onde évanescente (TIRF) n'est pas compatible avec l'utilisation de concentrations en fluorophores supérieures à 50nM. Les cavités Zero Mode Waveguide à la surface de la lamelle permettent de travailler avec de telles concentrations. Chaque cavité a un volume permettant de n'accepter qu'un seul complexe à l'intérieur. Le traitement du fond du puits permet d'accrocher le complexe spécifiquement. Figure issue de (Petrov et al. 2011)

marquées et non-marquées dans le RRL exclut les expériences qui nécessitent l'utilisation de molécules mutées.

A terme, il sera donc nécessaire de mettre au point au laboratoire un système de traduction reconstitué à partir des facteurs purifiés. Cela offrira l'avantage de pouvoir inclure dans le mélange des molécules marquées avec des fluorophores. Les questions biologiques sur des interactions entre molécules et les changements de conformation seront alors envisageables.

B. Zero Mode Waveguide

La microscopie en TIRF impose une concentration maximale des espèces fluorescentes en suspension au-dessus de la surface. Le volume minimum d'observation pour un fluorophore est typiquement de $0,1 .10^{-15}$ L. Si chaque volume minimal contient un fluorophore, cela correspond à une solution avec une concentration de 10^{16} molécules/L soit 16nM [8]. Pour être en mesure de distinguer 2 fluorophores facilement il ne faut donc pas excéder 5nM de fluorophore en suspension.

Dans le cas où l'étude nécessite des concentrations supérieures en molécules marquées, une autre technique de microscopie est disponible. Le « Zero Mode Waveguide » (ZMW) utilise une lamelle recouverte d'une pellicule en aluminium perforée de 100nm d'épaisseur (figure 23). Les puits ont un diamètre variable compris entre 50 et 200nm. Dans notre cas, cela permet de déposer et d'ancrer un seul ribosome par puits si on travaille dans ceux ayant un diamètre compris entre 50 et 100nm. L'intérêt majeur du ZMW est que le diamètre des puits est inférieur à la longueur d'onde d'émission des fluorophores utilisés. Cette configuration permet de diminuer le volume minimal d'observation qui est alors de 10^{-20} L. Avec cette condition, si chaque volume minimal contient un fluorophore, cela correspond à une solution avec une concentration de 10^{20} molécules/L soit 160µM. Il est alors possible de travailler avec des concentrations en fluorophores de l'ordre du µM ce qui est l'ordre de grandeur des concentrations en protéines utilisés lors des expériences de biochimie classique.

Le ZMW permet donc d'étudier des réactions biologiques comme l'incorporation d'ARNt à chaque cycle de translocation (Uemura et al. 2010). La limite des expériences n'est plus la concentration en molécules mais la résolution temporelle. Si l'événement observé est trop rapide par rapport aux limites de résolution temporelle, il faudra tout de même adapter les concentrations pour ralentir la réaction. Par exemple, il faudra adapter la concentration en ARNt pour être capable de détecter chaque cycle de translocation.

Malgré l'intérêt de cette technologie, elle n'est pas encore couramment utilisée. Bien que les ZMW soient commercialisés, ces lamelles à usage unique sont encore très chères et leur utilisation demande un appareillage adapté. L'étape limitante n'est pas tant la partie optique mais la partie microfluidique. Les lamelles vendues actuellement ne sont pas associées à une cellule microfluidique. Comme il s'agit d'une nanotechnologie, l'usinage et la chimie de surface des ZMW n'est pas à la portée de n'importe quel laboratoire. L'utilisation courante de cette technique nécessite encore de la mise au point avant d'être utilisable au laboratoire.

[8] 1mole contient $6,02.10^{23}$ molécules.

Références bibliographiques

Acker, M. G., S. E. Kolitz, S. F. Mitchell, J. S. Nanda & J. R. Lorsch (2007) Reconstitution of yeast translation initiation. *Methods Enzymol,* 430**,** 111-45.

Aitken, C. E., R. A. Marshall & J. D. Puglisi (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J,* 94**,** 1826-35.

Axelrod, D. (2001) Total internal reflection fluorescence microscopy in cell biology. *Traffic,* 2**,** 764-74.

Ben-Shem, A., N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova & M. Yusupov (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science,* 334**,** 1524-9.

Blanchard, S. C., R. L. Gonzalez, H. D. Kim, S. Chu & J. D. Puglisi (2004a) tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol,* 11**,** 1008-14.

Blanchard, S. C., H. D. Kim, R. L. Gonzalez, J. D. Puglisi & S. Chu (2004b) tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A,* 101**,** 12893-8.

Budkevich, T., J. Giesebrecht, R. B. Altman, J. B. Munro, T. Mielke, K. H. Nierhaus, S. C. Blanchard & C. M. Spahn (2011) Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol Cell,* 44**,** 214-24.

Cornish, P. V., D. N. Ermolenko, H. F. Noller & T. Ha (2008) Spontaneous intersubunit rotation in single ribosomes. *Mol Cell,* 30**,** 578-88.

Cornish, P. V., D. N. Ermolenko, D. W. Staple, L. Hoang, R. P. Hickerson, H. F. Noller & T. Ha (2009) Following movement of the L1 stalk between three functional states in single ribosomes. *Proc Natl Acad Sci U S A,* 106**,** 2571-6.

Costantino, D. & J. S. Kieft (2005) A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. *RNA,* 11**,** 332-43.

Fei, J., J. E. Bronson, J. M. Hofman, R. L. Srinivas, C. H. Wiggins & R. L. Gonzalez (2009) Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc Natl Acad Sci U S A,* 106**,** 15702-7.

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

Le Gall, A. 2011. Pince optique et microscopie de fluorescence pour l'étude de la synthèse des protéines en molécule unique.

Munro, J. B., R. B. Altman, N. O'Connor & S. C. Blanchard (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell,* 25**,** 505-17.

Namy, O., S. J. Moran, D. I. Stuart, R. J. Gilbert & I. Brierley (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature,* 441**,** 244-7.

Pestova, T. V. & C. U. Hellen (2003) Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev,* 17**,** 181-6.

--- (2005) Reconstitution of eukaryotic translation elongation in vitro following initiation by internal ribosomal entry. *Methods,* 36**,** 261-9.

Pestova, T. V., I. B. Lomakin & C. U. Hellen (2004) Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep,* 5**,** 906-13.

Petrov, A. & J. D. Puglisi (2010) Site-specific labeling of Saccharomyces cerevisiae ribosomes for single-molecule manipulations. *Nucleic Acids Res,* 38**,** e143.

Schneider-Poetsch, T., J. Ju, D. E. Eyler, Y. Dang, S. Bhat, W. C. Merrick, R. Green, B. Shen & J. O. Liu (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol,* 6**,** 209-217.

Scott, M. P., R. V. Storti, M. L. Pardue & A. Rich (1979) Cell-free protein synthesis in lysates of Drosophila melanogaster cells. *Biochemistry,* 18**,** 1588-94.

Spahn, C. M., E. Jan, A. Mulder, R. A. Grassucci, P. Sarnow & J. Frank (2004) Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell,* 118**,** 465-75.

Takyar, S., R. P. Hickerson & H. F. Noller (2005) mRNA helicase activity of the ribosome. *Cell,* 120**,** 49-58.

Uemura, S., C. E. Aitken, J. Korlach, B. A. Flusberg, S. W. Turner & J. D. Puglisi (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature,* 464**,** 1012-7.

Uemura, S., R. Iizuka, T. Ueno, Y. Shimizu, H. Taguchi, T. Ueda, J. D. Puglisi & T. Funatsu (2008) Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic Acids Res,* 36**,** e70.

Références bibliographiques des figures

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

Le Gall, A. 2011. Pince optique et microscopie de fluorescence pour l'étude de la synthèse des protéines en molécule unique.

Marshall, R. A., C. E. Aitken, M. Dorywalska & J. D. Puglisi (2008) Translation at the single-molecule level. *Annu Rev Biochem,* 77**,** 177-203.

Pestova, T. V. & C. U. Hellen (2003) Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev,* 17**,** 181-6.

Petrov, A., G. Kornberg, S. O'Leary, A. Tsai, S. Uemura & J. D. Puglisi (2011) Dynamics of the translational machinery. *Curr Opin Struct Biol,* 21**,** 137-45.

Takyar, S., R. P. Hickerson & H. F. Noller (2005) mRNA helicase activity of the ribosome. *Cell,* 120**,** 49-58.

Chapitre 3

Recherche d'événements de recodage dans le génome de *S. cerevisiae* et étude de *VOA1*.

Sommaire

I. Recherche d'événements non-conventionnels de traduction conservés chez les génomes de levures. 92

II. Détermination de l'état traductionnel des candidats grâce aux données de traductomique (« ribosome profiling »). 125

1. Méthodologie 125

2. Profils traductionnels des gènes candidats 126

A. Les jeux de données analysés 126

B. Les critères d'analyse 126

C. Résultats et discussion 127

III. Analyse moléculaire de *VOA1/ORF2* : résultats et discussion 130

1. Expression de Voa1-orf2p 130

A. Marquage de l'extrémité C-terminale de Orf2p avec l'étiquette 6HIS 130

B. Double marquage des extrémités N-terminale de Voa1p et C-terminale de Orf2p 130

2. Analyse de la région de chevauchement 132

C. Les constructions 132

D. Efficacité de recodage 132

E. Expression de la β-galactosidase et de la luciférase 133

3. Discussion 137

IV. Références bibliographiques 139

I. Recherche d'événements non-conventionnels de traduction conservés chez les génomes de levures.

Ce dernier chapitre porte sur la recherche d'événements non-conventionnels de la traduction chez la levure *Saccharomyces cerevisiae*.

Ce projet, initié avant mon arrivé dans l'équipe, a été réalisé en collaboration avec l'équipe de bioinformatique dirigée par Alain Denise (Institut de Génétique et Microbiologie, Laboratoire de Recherche en Informatique, Institut National de Recherche en Informatique et en Automatique). La recherche bioinformatique a été développée par Christine Drevet et Florent Bourassé.

Cette première partie présente une version préliminaire du papier que nous comptons soumettre à la revue "*Genome Research*".

J'y présente une méthode de génomique comparative, que nous avons développée, pour identifier des événements de recodage et potentiellement d'autres événements non conventionnels de la traduction dans le génome de *S. cerevisiae*. Cette approche est basée sur une recherche d'organisations génomiques dans lesquelles 2 ORFs se suivent ou se chevauchent. Seuls les événements conservés au sein de différents génomes de levures sont conservés dans l'analyse. L'efficacité de recodage des gènes candidats identifiés par cette analyse a été quantifiée *in vivo* grâce à un vecteur double rapporteur. Cela nous a permis d'identifier le gène *VOA1* qui est impliqué dans l'assemblage des complexes V-ATPase. Pour finir, j'ai démontré que l'extension en aval du gène *VOA1* est exprimée chez *S. cerevisiae* et qu'elle assure une fonction biologique.

Les 2 parties suivantes présentent les expériences additionnelles qui ont été réalisées pour ce projet et qui feront l'objet d'une publication ultérieure.

Article

Unconventional translation events conserved among yeast genomes

En vue d'une soumission à la revue "*Genome Research*"

Unconventional translation events conserved among yeast genomes

Hélène Chommy[1], Christine Drevet[1], Isabelle Hatin[1], Florent Bourassé[2], Agnès Baudin-Baillieu[1], Jean-Pierre Rousset[1], Alain Denise[1], Olivier Namy[1,3,]*

[1] Institut de Génétique et Microbiologie, Université Paris-Sud, 91400 Orsay

[2] Consultant at Nell'Armonia

[3] CNRS UMR8621, 91400 Orsay

* To whom correspondence should be addressed:

Dr Olivier Namy

IGM

Université Paris-SUD, CNRS UMR8621

Batiment 400.

91400 Orsay.

France

Phone: +33 1 69 15 46 34

Email: olivier.namy@igmors.u-psud.fr

Running title: Recoding in yeasts

Key words: Translation fidelity, yeast genomes, recoding, programmed frameshifting, ribosome profiling, genome comparison.

Abstract (<250 words)

Proteomic diversity is increased by numbers of dual coding events. Among them, recoding allow translation of two different proteins from the same mRNA. Here, we present a genomic comparative approach to search for unconventional translational events in the genome of *Saccharomyces cerevisiae*. The approach relies on a genomic organization analysis. Events containing 2 consecutive ORFs sharing a stop codon or 2 overlapping ORFs were collected and retained if they are conserved among yeast species. Translational bypass of candidate genes was quantified *in vivo* using a dual reporter system. This allowed us to identify *VOA1* gene which is involved in V-ATPase assembly. We further demonstrated that the extension of *VOA1* is expressed and plays a biological role in yeast.

INTRODUCTION

Proteomic diversity of a cell is generated by multiple mechanisms altering either the mRNA or the protein, such as RNA editing (Wulff et al. 2011), alternative splicing (Matlin et al. 2005), post-translational modifications (Schwartz et al. 2009). Another possibility to increase proteomic diversity is to express different proteins from the same genomic region. These events, known as dual coding, occur when a genomic sequence contains overlapping coding sequences (Liang 2010). Dual-coding events also include recoding events such as stop codon readthrough, +1 and -1 ribosomal frameshifting (Namy et al. 2004).

Recoding events can be divided into two major classes: i) stop codon readthrough which occurs when the ribosome incorporates a near cognat tRNA at the stop codon; ii) -1 and +1 programmed frameshifting which occur on a slippery sequence and are stimulated by slowly decoded regions. The latter can be a poorly recognized codon (hungry codon), a stop codon or a secondary structure like a pseudoknot.

Initially found in viral genomes and transposable elements (Goelet et al. 1982; Jacks and Varmus 1985; Wilson et al. 1986), recoding events are also known to regulate the expression of cellular genes. Known unconventional programmed translational events are described in Recode2 database (http://recode.ucc.ie/) (Bekaert et al. 2010). These recoding can be involved in key regulations of cell physiology. For instance, +1 frameshifting of *OAZ1,* which encodes the ornithine decarboxylase antizyme Oaz1p, controls cellular concentration of polyamines in eukaryotes (Matsufuji et al. 1995; Ivanov et al. 2000), and stop codon readthrough of *PDE2,* which encodes the cyclic AMP (cAMP) phosphodiesterase, controls cellular concentration of cAMP in *S. cerevisiae* (Namy et al. 2002).

During the last decade many bioinformatics approaches have been deployed to search for unconventional translational events that regulate gene expression. Most approaches are specialized of one type of recoding because predictions are based on consensus sequences and detection of stimulatory elements. The KnotInFrame program (Theis et al. 2008) and the PRFdb project (Belew et al. 2008) are dedicated to -1 frameshifting prediction. They survey nucleotide sequences for a slippery site XXXYYYZ, a spacer and a pseudoknot (Jacks et al. 1988; Brierley et al. 1989). With FSFINDER (Moon et al. 2004; Byun et al. 2007), user chooses the type of events (+1 or -1 frameshifting) and a related species to apply specific model to the search. The FSscan program looks for +1 frameshifting hotspot in the genome of *Escherichia coli* and confirm some candidates by crossing information with mass

spectrometry data (Liao et al. 2009). The main limit of these approaches is that they can only detect events fitting with specifically built models.

The large amount of sequenced genomes, as well of transcriptomic, proteomic and translatomic studies now allow identification of putative recoding sites. In Drosophila, genome sequencing and further analysis predict a high readthrough usage in cellular genes (Jungreis et al. 2011). Nonetheless, one important drawback in these two interesting studies is that candidate genes are not validated *in vivo* (Ribrioux et al. 2008). In *Euplotes ssp.*, comparison of genomic and proteomic data reveals that 10% of genes could use +1 frameshifting events (Klobutcher 2005). Recently, search for such unconventional translational events have been done using previously published ribosome profiling data of the human genome. This approach is based on genomic organizations and most probable reading frame detection (Michel et al. 2012). Unfortunately, this exciting approach is currently limited both by the depth of sequencing and by the level of gene expression, which makes difficult the identification of translational events occurring at low level.

Here, we propose a general method to identify recoding events, and other unconventional translational events, in genomes with minimal *a priori* onto the molecular mechanism. The approach is based on the phylogenetic conservation of genomic organizations. We took advantage of the availability of a large number of yeast genomes to apply this method to *S. cerevisiae.* Events containing 2 consecutive ORFs sharing a stop codon or 2 overlapping ORFs are collected and retained if they are conserved among several yeast species. A subset of 19 candidate genes has been selected and tested for translational recoding. Seven regions displayed significant bypass efficiency either in +1 or -1 reading frame. Among them we found all genes already known to use programmed frameshifting in *S. cerevisiae*. Furthermore we demonstrated by phenotypic profiling that the extension of *VOA1* gene plays a physiological role leading to sensitivity to the antibiotic blasticidine S, known to inhibit peptide bond formation by the ribosome. Overall, these results demonstrate that adding phylogenic criteria is relevant to enhance the detection of recoding sites by bioinformatics approaches, and that alternative reading of the genetic code is still underestimated in genomes.

RESULTS

CEV pipeline

To identify genomic regions compatible with unconventional decoding events in the genome of *S. cerevisiae*, we developed regular expressions to search particular triplet arrangement in sequences. We chose the following parameters to identify genomic organization: The region starts with i) a coding DNA sequence (named CDS-A) beginning with a start codon ATG and ending with one of the 3 stop codons (TAA, TGA or TAG); and ends by ii) an open reading frame (called ORF-B) marked off with stop codons. In the case of readthrough compatible events (R-CEV), CDS-A and ORF-B are contiguous in the same reading frame. CDS-A is at least 100bp long and ORF-B is at least 50bp long (figure 1). As small overlapping ORFs are very frequent in *S. cerevisiae* genome we increased the size limit for ORF-B to decrease the number of candidate sequences. In the case of frameshifting compatible events, ORF-B is overlapping CDS-A either in the +1 or the -1 reading frame (respectively +1CEV and -1CEV). CDS-A is at least 100bp long and ORF B is at least 100bp long. Minimal length of the event (CDS-A + ORF-B) is 150 bp long and the size of the overlap region is between 3 and 1000bp (figure 1).

Step 1: Regular expressions

Analysis of each type of organization events was performed independently. The pipeline used to build the database is described in figure 2. Nineteen genomes of Fungi detailed in (table 1) were submitted to the pipeline (figure 2). Among them, 15 belong to Hemiascomycetes including *S. cerevisiae*. At this stage, 15972 +1CEV events, 14753 -1CEV events and 17840 R-CEV events were identified in *S. cerevisiae* genome. For other genomes, data are presented in table 1.

Step 2: Genomic comparison of compatible events.

To study conservation of events among analyzed species, CDS-A and ORF-B were translated in putative protein sequence according the standard genetic code. Each event composed of CDSA-ORFB pairs and identified in *S. cerevisiae* was used as query to run 2 BLASTP. When both CDS-A and ORF-B from the query event have a hit with the 2 proteins of another event from the databases, BLASTP results were stored in CEV database for further analysis.

Step 3: Building groups containing *S. cerevisiae* compatible events.

Queries in PostgreSQL database allowed visualization of the conservation of each *S. cerevisiae* event with other events localized elsewhere in the *S. cerevisiae* genome or in the sequences of other species. Finally, 1053 +1CEV, 885 -1CEV and 955 R-CEV compatible events, identified in *S. cerevisiae* genome, show similarities with at least one event registered in the database. The hit repartition among species is shown in table 1. Each query and its hits have been grouped.

Filtering groups.

To restrict study to groups of events having the most biological relevance, we applied several filters based on genomic characteristics. Groups of events were retained for further analysis when: (i) CDS-A was annotated in the genome of *S. cerevisiae;* (ii) CDS-A and ORF-B had no other hit in the genome of *S. cerevisiae*. The filter (i) was made to restrain analysis to documented genes in which potential recoding events extend a characterized protein. The second filter (ii) was applied to discard multigenic family or transposable elements. Both filters reduced the number of conserved groups to 891 +1CEV, 447 -1CEV and 54 R-CEV.

Survey of *S. cerevisiae* characterized recoding events in CEV database.

To verify that this approach was able to detect already known recoding events in *S. cerevisiae*, we decided to search for *BSC1*, *BSC2, BSC4*, *BSC5*, *BSC6*, *IMP3*, *PDE2*, *ABP140*, *EST3* and *OAZ1* in CEV database using their genomic coordinates.

As expected, *BSC1*, *BSC2, BSC4*, *BSC5*, *BSC6*, *IMP3* and *PDE2* are found in the database as R-CEV events. Only *BSC1* is retained in the final list after applying filters with conservation in 4 species: *K. lactis*, *S. castellii*, *S. cerevisiae, Y lipolytica*. The 3 characterized +1 frameshifting genes *ABP140*, *EST3* and *OAZ1* are present in the database as +1CEV and show a strong conservation among genomes as previously described (Farabaugh et al. 2006). *ABP140* is conserved in *A. gossypii*, *C. glabrata*, *K. lactis*, *K. waltii*, *S. bayanus*, *S. cerevisiae*, *S. kluyveri*, *S. mikatae*, *S. paradoxus*. *EST3* is conserved in *C. glabrata*, *K. waltii*, *S. bayanus*, *S. castellii*, *S. cerevisiae*, *S. kluyveri*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus*. *OAZ1* is conserved in *C. glabrata*, *S. bayanus*, *S. castellii*, *S. cerevisiae*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus*. Nonetheless, they were not found in the final selection because both CDS-A and ORF-B are officially annotated as a single CDS in GenBank file.

Selection of candidate genes

To further analyze some candidates *in vivo*, we manually analyzed candidates and selected some on biological expertise.

Finally, 19 candidates were retained: 10 for -1CEV (*AAD6*, *ATF2*, *CRC1*, *GAT3*, *ROT1*, *STE18*, *VOA1*, *YML003w*, *YJL107c* and *YLR042c*) and 9 for +1CEV (*ABP140*, *ATO2*, *ATP10*, *BUD31*, *CCC1*, *EST3, OAZ1, UGP1* and YOL159c). No R-CEV candidate has been retained for R-CEV. Information about these candidates are presented in table 2. Five of 19 candidate genes included an annotated gene in ORF B, pairs were *AAD6-AAD16*, *STE18-YJR087w*, *YML003w-YML002w*, *YJL107c-PMR10* and *ATP10-CST9*. Both genes actually overlapped in two cases. In *STE18-YJR087w*, the last base of *STE18* was the first base of *YJR087w*. In *YJL107c-PMR10*, genes overlap was 3bp long. For following experiments and discussion, we will refer to the events according to the name of CDS-A.

Ribosomes can bypass CDS-A stop codon

To translate the ORF-B, ribosomes must evade the stop codon of the CDS-A. This can occur through different mechanisms (recoding, alternative splicing, editing) grouped under the term "bypassing" which is neutral regarding the molecular mechanism involved. To determine whether ribosomes translating CDS-A can also translate the ORF-B, we quantified the bypass efficiency of each candidate using a dual reporter system based on a translational fusion between LacZ and F-luc genes (pACPGK007). Ribosome must shift to the new reading frame in the overlapping region between CDS-A and ORF-B. Recoding often requires the presence of a stimulatory element like a secondary structure. When overlapping region was shorter than 100bp, we included further downstream DNA sequence in the region to analyze to include potential secondary structure. To avoid measurement of putative translational re-initiation, if ATG codons are present in this overlapping region they are mutated into sense codon (supplementary table 2).

Among the 19 candidates tested, 12 displayed a background level around 0.2% and 7 displayed a bypass efficiency > 5 fold of the background (≥ 1%) (figure 3). *VOA1* displayed the highest bypassing efficiency (16%) of newly identified candidates, it was retained for further analysis.

Genomic organization of *VOA1* and its extension in yeast chromosome

VOA1 belongs to the -1CEV group and is conserved in *Saccharomyces cerevisiae*, *Saccharomyces bayanus*, *Saccharomyces mikatae* and *Saccharomyces kudriavzevii*. This

genomic region displays a very peculiar organization: *VOA1* (CDS-A) and the extension named *ORF2* (ORF-B) are respectively 798bp and 336bp long and overlap on 64bp. No ATG codon is present in *ORF2* explaining why this *ORF2* is unannotated despite its size. Moreover the *ORF2* also overlaps 113bp of *VMA21* located on the complementary strand.

VOA1 displays canonical -1 frameshifting signals

To gain insights into the molecular mechanism allowing the expression of the *ORF2* we searched for motifs classically found in -1 programmed frameshifting. We first submitted *VOA1/ORF2* sequence to KnotInFrame web server (Theis et al. 2008). It that two overlapping sequences near the end of VOA1, **AAAAAAA** and **AAAAAAC,** could be -1 frameshifting slippery sites according to the consensus **X XXY YYZ** (Jacks et al. 1988). In addition, a downstream pseudoknot is predicted both by KnotInFrame and Kinefold (Xayaphoummine et al. 2005). Thus, *VOA1* possesses all documented features of -1 frameshifting event (figure 4).

VOA1* extension is expressed *in vivo* in *S. cerevisiae

Next, we tested if *ORF2* is actually expressed *in vivo*. For this purpose, translational fusion between *VOA1-ORF2* and F-luc ORF was realized. This fusion, under the control of 641bp of *VOA1* endogenous promoter, is inserted in the plasmid pFL44L. We detected a median luciferase activity of 61 RLU/µg of proteins.

***VOA1* extension plays a physiological role in yeast**

One of the major drawbacks classically associated with genome wide bioinformatics analysis to search for unconventional expression mechanisms is the absence of physiological role associated with the events identified.

To address this question and to demonstrate that *ORF2* plays a functional role in yeast we performed a high-throughput phenotypic screening using two strains, one carrying the wild-type organization *VOA1/ORF2*, ant the other carrying a deletion of *ORF2 (VOA1ΔORF2)*. Strains were built in FL100 background to avoid any auxotrophy complementation. The high-throughput phenotypic screening was performed by Biolog Inc. (Bochner et al. 2001; Homann et al. 2005). Strains were assayed for growth in 1440 different conditions relative to carbon, nitrogen, phosphorus and sulfur sources, osmolarity, pH and chemical sensitivity to various inhibitors. Among 1440 conditions tested, blasticidin S strongly inhibits the growth of FL100 *VOA1/ORF2* strain compared to FL100 *VOA1ΔORF2* strain (supplementary figure 3). We

confirmed this data by reproducing growth experiments with 26µg/ml blasticin S in YPD (figure 5). The FL100 *VOA1/ORF2* strain does not grow in presence of blasticidine S whereas the strain deleted for *ORF2* shows only a reduction of the growth rate (7h50 vs 2h34 in absence of drug). We also tested 16 growth conditions producing ambiguous results during the high-throughput phenotypic screening: pH4, pH5, pH6, apramycin, amphotericin B, kanamycin, myltefosine, neomycin, nystatin, pentamidin isethionate, polymixin, thioridazine, tobramicin, citric acid, chromium III chloride and sodium fluoride. However, no detectable or reproducible growth difference has been observed in these conditions.

DISCUSSION

This works describes a comparative genomic approach of 19 Fungi genomes to identify genomic organizations compatible with an unusual decoding event. The candidate genes are either composed by 2 ORFs sharing a stop codon or by 2 overlapping ORFs. In first instance we searched for programmed unconventional translational events (recoding) by looking for annotated genes with an extended coding potential without hypothesis about the mechanism. Readthrough compatible organizations are named R-CEV and genomic organizations compatible with +1 or -1 frameshifting are respectively named +1CEV and -1CEV.

In the first step of the pipeline, collection of compatible events allowed us to retrieve all genes indexed in Recode2 database characterized for recoding in *S. cerevisiae* (Bekaert et al. 2010). *BSC1-6*, *IMP3* and *PDE2* are characterized for readthrough, and *ABP140*, *EST3* and *OAZ1* are submitted to +1 frameshifting events (Morris and Lundblad 1997; Asakura et al. 1998; Ivanov et al. 2000; Namy et al. 2002; Namy et al. 2003).

Building group of conserved events, at the protein level, retained only 4 of the above control genes (*BSC1*, *ABP140*, *EST3* and *OAZ1)*. The non-conservation of known readthrough events might indicate that there is no selection pressure on ORFs B and that these regions quickly evolve. This would support the hypothesis that readthrough in yeast might be a temporary adaptive mechanism as suggested by studies on [*PSI*+] prion (Shorter and Lindquist 2005). The conservation criterion appears to be a good filter to detect +1 frameshifing events. Although conservation of proteins encoded by 2 consecutive ORFs may correspond to actual protein conservation like readthrough, frameshifting, hopping, translational re-initiation, it may also correspond only to DNA conservation like highly conserved DNA motif. Selection based on systematic annotation of CDS-A and exclusion of multigenic and transposon

families reduced the number of conserved groups to 891 +1CEV, 447 -1CEV and 54 R-CEV. Considering 5879 referenced CDS in *S. cerevisiae* genome, 15%, 7% and 0.09% of annotated CDS fitted +1CEV, -1CEV and R-CEV criteria respectively. The low percentage of putative conserved readthrough is clearly a consequence of the accumulation of stop codons within the 50bp downstream the end of the CDS-A, which is fully consistent with previously published data on stop codon accumulation in 3'of CDSs (Williams et al. 2004).

We analyzed 19 candidate genes to determine whether the overlapping region can support bypassing of the CDS-A stop codon *in vivo* (for -1CEV: *AAD6*, *ATF2*, *CRC1*, *GAT3*, *ROT1*, *STE18*, *VOA1*, *YML003c*, *YJL107c* and *YLR042c*; for +1CEV: *ABP140*, *ATO2*, *ATP10*, *BUD31*, *CCC1*, *EST3, OAZ1, UGP1* and *YOL156c-A*). We quantified bypass efficiency for each event using a dual reporter system. We observed that 7 regions (*CCC1*, *YML003W*, *ATF2*, *OAZ1*, *EST3*, *VOA1* and *ABP140*) display bypass efficiency from 1% to 22%. It is important to highlight that we performed these assays in minimal media during exponential growth. One could propose that it would be worth to quantify bypass efficiency in various growth conditions or under stress conditions. However without preliminary indication about which stress or which growth condition to test it will be especially difficult to find the relevant condition to observe a change in bypass efficiency. Despite this limitation, we interestingly identified all the previously frameshifted genes described in yeast (EST3, *OAZ1*, ABP140). As there is no constraint on the molecular mechanisms allowing the expression of the ORF-B during our search, it is highly significant to be able to identify all the previously known frameshifted genes. These 3 genes are among the 4 more efficient sequences to promote CDS-A stop codon bypass. The fourth sequence corresponds to *VOA1* gene, which displays the second highest percentage of bypassing after *ABP140*. We focused our analysis to this candidate.

The *CDS-A* corresponding to *VOA1* encodes a membrane protein involved in V-ATPase assembly at the endoplasmic reticulum. Voa1p harbors 2 transmembrane domains and a C-terminal consensus motif involved in endoplasmic reticulum (ER) retention (Ryan et al. 2008). *VOA1* is 800pb long and the *ORF2* extends this coding sequence by 333pb, this represents a significant increase in length. The *ORF2* located in the -1 reading frame regarding *VOA1* is unannotated because of the absence of any ATG codon. To gain insights about the biological function of the protein encoded by *ORF2* we searched for protein motif but no specific consensus motif was found using either ProDom (Servant et al. 2002) or Pfam (Punta et al. 2012). However, 2 transmembrane domains were predicted with a strong

probability (> 0.6) using TMHMM v2.0 server (supplementary figure 2) (http://www.cbs.dtu.dk/services/TMHMM/). In this situation, the expression of a fusion protein Voa1-orf2p would have a strong impact on the topology of Voa1p by adding new transmembrane domains, and also by masking the ER retention signal which needs to be at the very end of the C-terminal part of the protein to be functional. One can propose that this new protein will have different function than the shorter Voa1p form. We verified that *ORF2* is indeed expressed in *S. cerevisiae* by constructing a translational fusion between *ORF2* and the F-luc reporter ORF. Measurement of luciferase activity indicated a low but significant level *ORF2* expression when expressed from a plasmid.

How the *ORF2* is translated remains an interesting mechanistic question. Several mechanisms could explain how ribosomes reach the *ORF2*. Among them programmed -1 frameshifting is one of the most evident. However one can also proposes alternative possibilities like, internal non-ATG initiation (IRES), stop and start mechanism or ribosomal hopping. Most of these events rely on specific motifs present on the mRNA that interacts with the ribosome. We searched for such motifs and identified two putative slippery sequences followed by a downstream pseudoknot. We did not put any constraint onto the molecular mechanism to express the *ORF2*, so the identification by chance of all the classical elements needed to induce translational frameshifting in the overlapping region is very improbable. All these elements argue for a -1 frameshifting event to express *ORF2*. However in absence of direct evidence for programmed -1 frameshifting, we cannot eliminate alternative explanations for *ORF2* translation. More work is needed to determine the translational mechanism used for *ORF2* expression.

The most important part was to demonstrate that whatever the mechanism is, ORF2 plays a physiological function in yeast. This was not an easy task as the protein does not display any motif which could help us to figure its biological function. We decided to perform a high-through-put phenotypic screening comparing strains expressing either *VOA1/ORF2* or only *VOA1* (PM service by Biolog Inc.). Results of this large screening clearly revealed that, among 1440 conditions tested in duplicate, deletion of the ORF2 confers a strong resistance to blasticidin S. Moreover results were ambiguous for 16 other conditions. These results prompted us to reproduce them individually for each ambiguous condition. We confirmed that deletion of *ORF2* improves resistance of yeast cells to the antibiotic blasticin S (figure 5), however none of the other phenotypes has been confirmed. It is currently difficult to link this phenotype to a specific biological function.

Overall, the comparative genomic approach allowed us to identify a cellular gene whose ORF-B is expressed. It would be also interesting to further analyze the other positive candidate from the dual reporter assay. For example, in tilling array study of the *S. cerevisiae* transcriptome the intensities of the signals are very similar for CDS-A and ORF-B for YML003w and ATF2 (Xu et al. 2009) which could suggest that CDS-A and ORF-B are transcribed as a single mRNA. In addition, we found the same transcription profile for candidates YJL107c-PMR10, and ATP10-CST9 whereas no bypassing was quantified using the dual reporter system (figure 3). Thus, it would be interesting, in these cases, to determine if both CDS are expressed together or separately.

This comparative approach has the potentiality to discover a large panel of un-annotated coding sequences and unknown unconventional events. Although the pipeline was designed to identify unconventional translational events, we cannot exclude that the conservation of close CDS-A and ORF-B sequences correspond to independent transcripts. In this work, we applied filters to focus on annotated genes in which translation of ORF-B would lead to a longer protein. Doing this we obviously missed all events that lead to the synthesis of a shorter protein by premature termination (Namy et al. 2002). This possibility will need to be investigated to obtain a complete overview of alternating expression mechanisms in yeast, and to determine the participation of these unconventional translation events to the diversity of the proteome.

MATERIALS AND METHODS

Plasmids

The pACPGK007 plasmid derives from pAC99 reporter plasmid previously described (Namy et al. 2002), it was obtained by inserting PGK promoter before LacZ gene and PGK terminator after F-luc gene. Constructs were obtained by inserting a PCR fragment containing the overlapping region into the cloning site, between the LacZ and F-luc genes in the plasmid pACPGK007. For recoding measurements, an in-frame control was used (pAC-TQ) which allowed the production of 100% fusion protein (β-galactosidase-luciferase) (Namy et al. 2003). The region including the inserted fragment was sequenced in the newly constructed plasmids.

The pFL44L-VOA1 was created by inserting VOA1/ORF2-Luc-URA3KL fragment from HC01 strain at EcoIcRI restriction site of pFL44L vector. VOA1/ORF2-Luc-URA3KL fragment contains VOA1/ORF2 under the control 641bp of the endogenous promoter. ORF2 is in translational fusion with F-luc. The URA3-KL cassette amplified from plasmids with oligonucleotides URA3KL.w and URA3KL.c is inserted just downstream F-Luc ORF.

Yeast strains

The strains used in this study are details in the following table.

Strains	Genotype and phenotype	Reference
74-D694	MATa; ade1-1CEV4; trp1-289; his3D200; leu2-3,112; ura3-52; [psi-]	
FL100	MATa ; ura3-2	(Lacroute 1968)
MRY14	SF838-1CEVDα; voa1::Hyg^R	(Ryan et al. 2008)
HC04	FL100; ura3Δ::VOA1-URA3; voa1::Hyg^R	This study
HC05	FL100; ura3Δ::VOA1/ORF2-URA3; voa1::Hyg^R	This study

To make HC4-5 strains, hphMW4 cassette and flanking *VOA1* region were amplified from MRY14 gDNA by PCR using oligos MRY14.w and MRY14.c. FL100 cells were transformed with the PCR product and selected on medium containing 100µg/ml HygromycinB. Correct DNA recombination was assayed by colony PRC with oligos in hphMX4 cassette (hphMX4.c) and flanking the recombination site (VOAprom.w). *VOA1* and *VOA1/ORF2* were amplified from gDNA by PCR by including 821bp of the promoter and adding PGK terminator in 3'of the stop codon. Each PCR product was cloned in pRS316 (Sikorski and Hieter 1989) at the Ecl136II restriction site. Plasmids were sequence in order to verify the construct. *VOA1-URA3KL* and *VOA1/ORF2-URA3KL* constructs were amplified by PCR from previous pRS316 plasmids including URA3 flanking regions using pRSVOA.w and

pRSVOA.c oligonucleotides. FL100; voa1::HygR cells were transformed with PCR products and selected on medium lacking uracil. Strains were verified by sequencing.

β-galactosidase and luciferase to establish recoding efficiency.

The overlapping region of each candidate were amplified by PCR from genomic DNA and cloned in pACPGK007 dual-reporter at MscI restriction site. In the final construction, the LacZ reporter was in frame with CDS A and the reporter F-luc was in frame with ORF B. To avoid measurement of putative translational re-initiation in ORF-B at an internal ATG codon, all ATG in the insert were mutated into other sens codon. Oligonucleotides used to amplify overlapping regions and ATG utations are described in supplementary table 2.

The 74D694 cells were transformed with the reporter plasmids using the lithium acetate method according to (Ito et al. 1983). In each case, at least five transformants, cultured in the same conditions, were assayed. Cells were broken using acid-washed glass beads; luciferase and β-galactosidase activities were assayed in the same crude extract, as previously described (Stahl et al. 1995). Bypass efficiency is defined as the ratio of luciferase activity to β-galactosidase activity. To establish the relative activities of β-galactosidase and luciferase when expressed in equimolar amounts, the ratio of luciferase activity to β -galactosidase from an in-frame control plasmid was taken as a reference. Readthrough frequency, expressed as percentage, was calculated by dividing the luciferase/ β-galactosidase ratio obtained from each test construct by the same ratio obtained with the in-frame control construct (Bidou et al. 2000).

Luciferase and Bradford assay

Luciferase assay were performed as described previously. No BSA was added in this case in buffer to allow quantification of protein concentration by the Bradford method (BioRad). Independent measurements from 5 independent colonies were carried out.

Growth analysis

Strains were cultured in YPD and inoculated at an optical density at 600 nm (OD_{600}) of 0.2 into YPD supplemented with the appropriate compound. Nystatin, neomycin miltefosine, polymyxin B, chromiuim (III) chloride, apramycin sulfate, blasticidin S, thioridazine hydrochloride, pentamidine isethionate, amphotericin B, sodium fluoride, tobramycin, kanamycin sulfate and citric acid trisodium salt were all bought from Sigma-Aldrich. The culture was inoculated into a 96-well plate and incubated in the TECCAN at 30 °C.

Measurements of OD_{600} were made at 60 min intervals. Doubling time (DT) was determined during exponential growth phase using the formula: $DT = (t_B-t_A)*\log(2) / (\log (C_B)-\log(C_A))$. The "t" correspond to 2 time points A and B with $t_B>t_A$, and "C" correspond to the OD_{600}.

CEV pipeline

Regular expressions

The regular expressions scan DNA sequences for the first ATG codon following a stop codon. This ATG codon will be the initiation codon of the CDS A and define phase 0 of the event. According to the considered genomic organization, the ATG codon will be used to detect the upstream stop codon of ORF B in the appropriate reading frame. This stop codon will anchor the search of the CDS A stop codon in 0 frame. Finally, this latter codon is used to define the ending stop codon of ORF B. Size criteria previously described were applied to collect events of interest for each type of genomic organization. Details of regular expressions integrated in Perl script are provided in figure 3. Selected events were stored in a PostgreSQL database called "CEV database". In addition of tables storing species, chromosomes and sequences, a table stores each event collected with its identification number, the kind of genomic organization and the DNA coordinates of its CDS A and ORF B within the reference sequence.

Genomic comparison of compatible events.

Putative protein sequences from CDS-A and ORF-B of all species were stored in 2 fasta formatted files. These files were used to build 2 databases named cdsA and orfB for BLASTP program search (Altschul et al. 1997). Each event composed of CDSA-ORFB pairs and identified in *S. cerevisiae* was used as query to run 2 BLASTP searches against the cdsa and the orfb databases respectively. The cutoff e-value was 0.001. When both CDS-A and ORF-B from the query event had a hit with the 2 proteins of another event from the databases, BLASTP results were stored in CEV database for further analysis.

Filtering groups

The CDS positions from the features of GenBank formatted files of the *S. cerevisiae* genome were crossed with the positions of the conserved events identified in the *S. cerevisiae* genome. Groups of events were retained for further analysis when: (i) ORF A was annotated in the genome of *S. cerevisiae;* (ii) ORFA had no other hit in the genome of *S. cerevisiae*.

ACKNOWLEDGMENTS

We are very grateful to Tom H. Stevens for sending the voa1::Hyg strain and for sharing unpublished results with us. Authors would like to thank all members of the GST lab for fruitful discussions. HC is supported by a FRM grant. This work was supported by a grant from ANR (ANR 2011 BSV6 011 01) to ON.

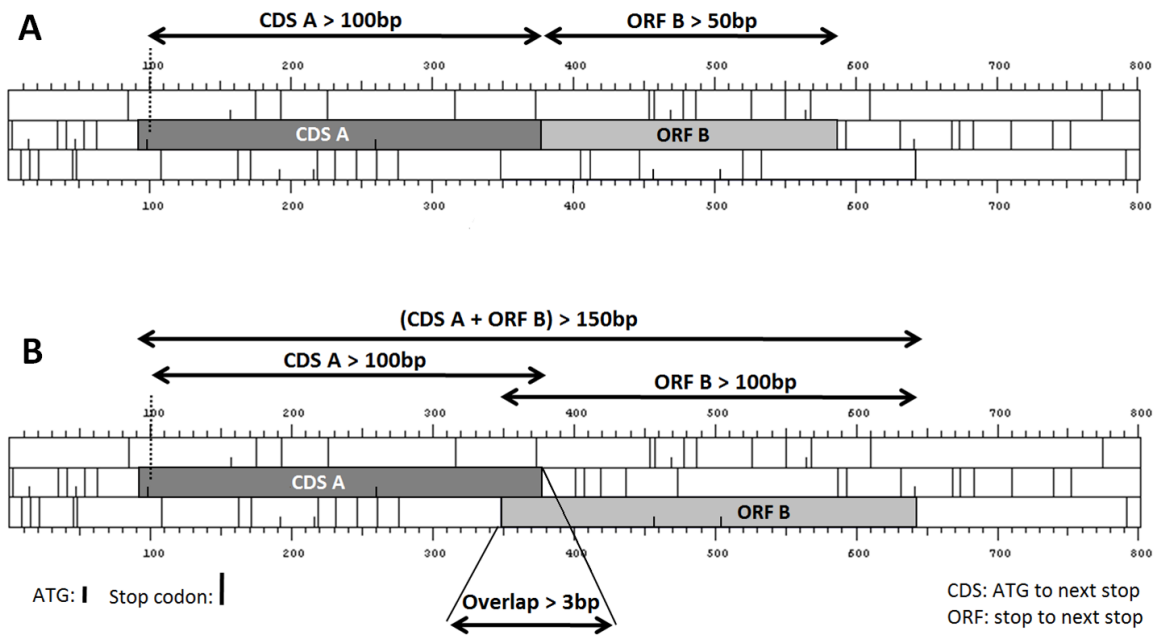


Figure 1. Schematic representation of genomic compatible events.

The three reading frame of a sequence are represented as horizontal lines. Vertical bars correspond to stop codons and semi bars correspond to ATG codons. Values above the arrows correspond to the size criteria applied in the pipeline. A. Genomic organization of R-CEV with 2 consecutive ORFs sharing a stop codon. B. Genomic organization of +1CEV with 2 overlapping ORFs.

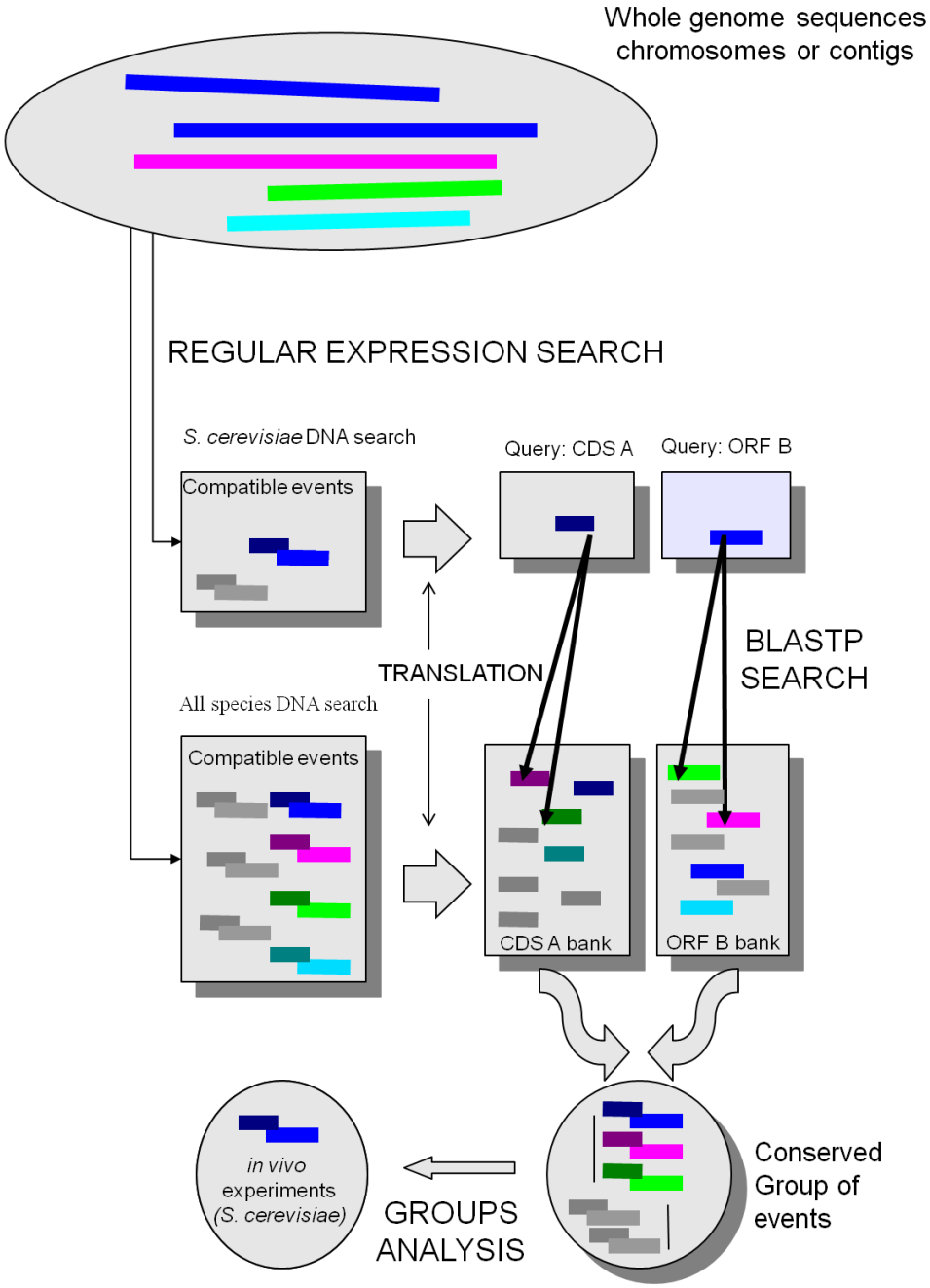


Figure 2. Pipeline for the database construction

Pipeline is composed of 3 steps. First, regular expressions search for compatible events in genomic sequences. Secondly, BLASTP searches for similarity between *S. cerevisiae* CDS-A and ORF-B proteins and their counterparts in all other species.

Organisms	-1CEV	conserved -1CEV	+1CEV	conserved +1CEV	R-CEV	conserved R-CEV
Ashbya gossypii	18186	122	19885	103	22245	163
Aspergillus fumigatus	77067	27	84198	19	86523	37
Candida albicans	12785	0	14183	7	16553	15
Candida glabrata	14785	226	15304	203	17621	217
Candida tropicalis	14600	70	16868	102	18143	61
Cryptococcus neoformans	53948	8	57182	8	57404	10
Debaryomyces hansenii	11441	51	12818	62	14813	61
Encephalitozoon cuniculi	6263	2	6660	4	7592	10
Kluyveromyces lactis	13108	149	14203	181	16299	182
Kluyveromyces waltii	16536	256	18162	315	19832	237
Saccharomyces bayanus	16956	1466	18105	1793	20118	1290
Saccharomyces castellii	12277	205	13553	280	15367	222
Saccharomyces cerevisiae	14753	885	15972	1053	17840	955
Saccharomyces kluyveri	18942	391	20382	517	22191	336
Saccharomyces kudriavzevii	16983	1650	18274	2191	19103	1395
Saccharomyces mikatae	14179	1994	15313	2502	17011	1598
Saccharomyces paradoxus	15241	2993	16404	3655	18317	2513
Schizosaccharomyces pombe	12507	19	14411	26	16336	19
Yarrowia lipolytica	48088	46	54711	73	58568	65

Table 1. Hits repartition of conserved events in CEV database

Genomic organization	CDS-A	Size CDS-A (bp)	ORF-B (Nbr of ATG)	Size ORF-B (bp)	Overlap (bp)	Conservation among genomes [#]
-1CEV	AAD6	639	AAD16 (2)	636	148	5
-1CEV	ATF2	1608	(1)	378	235	4
-1CEV	CRC1	984	(2)	480	283	4
-1CEV	GAT3	426	(1)	327	73	3
-1CEV	ROT1	771	(0)	276	97	5
-1CEV	STE18	333	YJR087w (1)	375	25 (1*)	4
-1CEV	VOA1	798	(0)	336	64	4
-1CEV	YML003w	873	YML002w (6)	2424	25	2
-1CEV	YJL107c	1164	PRM10 (4)	1188	40 (3*)	2
-1CEV	YLR042c	486	(1)	321	103	5
+1CEV	ABP140	840	ABP140 (5)	1092	44	9
+1CEV	ATO2	849	(1)	315	110	4
+1CEV	ATP10	840	CST9 (3)	1560	80	4
+1CEV	BUD31	474	(1)	324	89	5
+1CEV	CCC1	969	(0)	276	43	4
+1CEV	EST3	282	EST3 (2)	297	32	9
+1CEV	OAZ1	210	OAZ1 (3)	753	83	7
+1CEV	UGP1	1500	(1)	369	113	4
+1CEV	YOL159c-A	273	(1)	495	35	3

Table 2. The candidate genes

* Value corresponds to the size of the overlapping region between the 2 CDS.

Conservation among genomes includes *Saccharomyces cerevisiae* query.

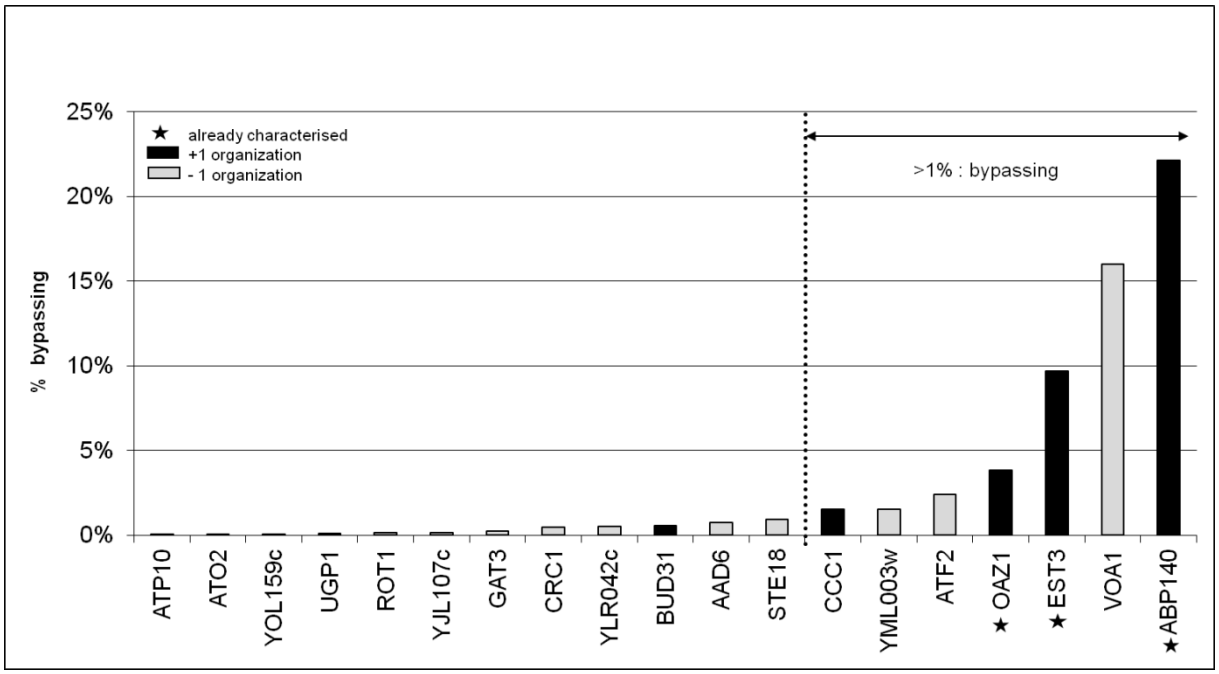


Figure 3. Bypassing efficiency in the dual reporter system

Graphic displays the bypassing efficiency of each candidate. The percentage corresponds to the ratio of luciferase and β-galactosidase activities normalized by an in-frame control. Bars in black correspond to +1CEV and bars in grey correspond to -1CEV. Black stars indicate genes already characterized for +1 frameshifting.

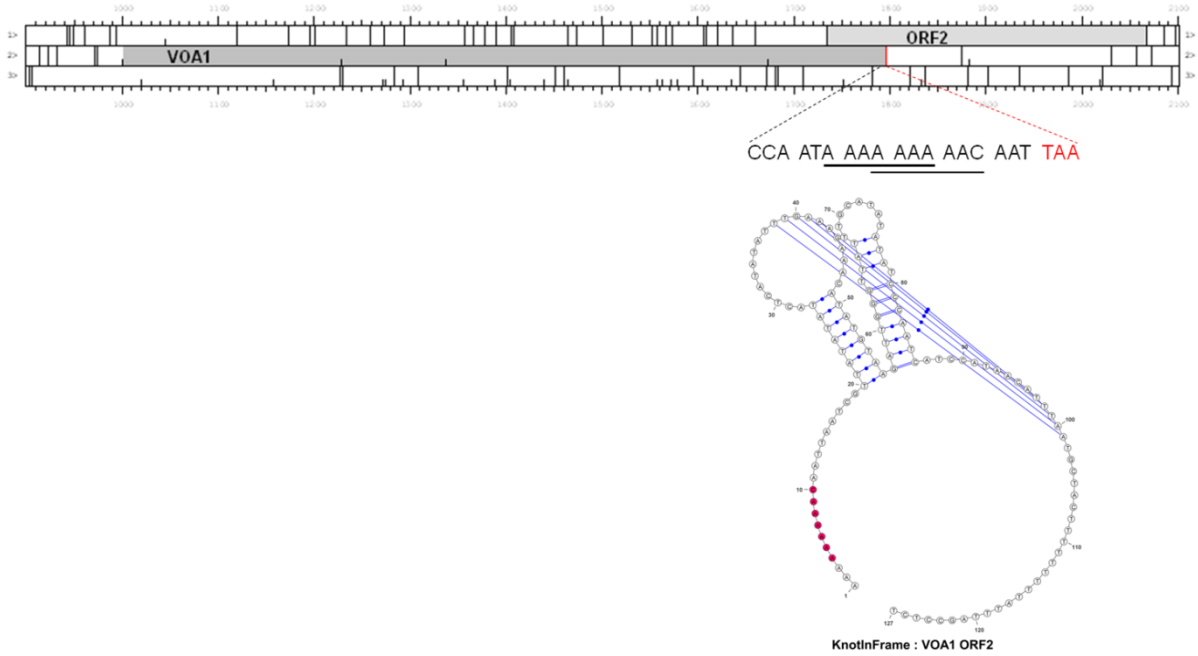
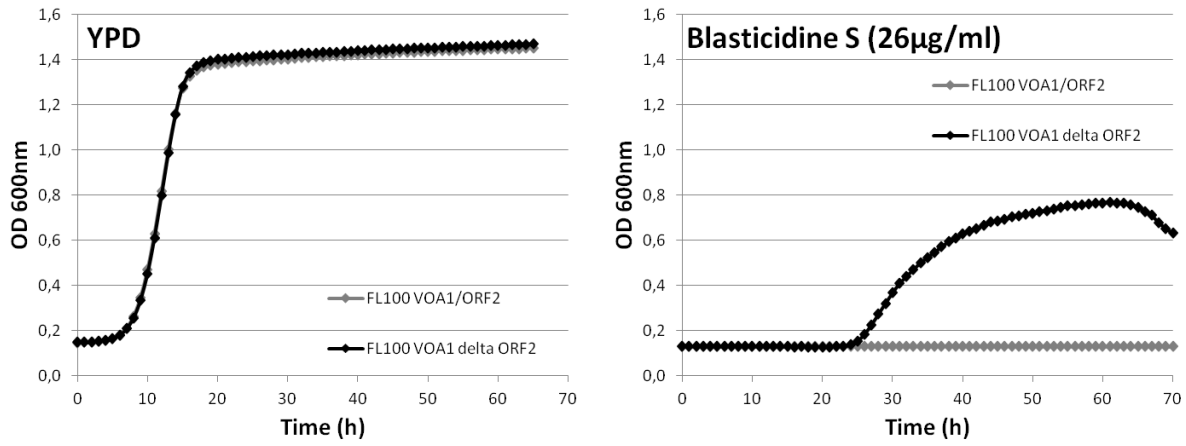


Figure 4. -1 frameshifting features in VOA1/ORF2

Schematic representation of *VOA1* and *ORF2* genomic organization (see Figure 1 for details about representation). Location and sequence of the 2 putative slippery sequences are indicated under the genomic organization. The pseudoknot structure predicted by the KnotInFrame program (Theis et al. 2008) was drawn with VARNA program (Darty et al. 2009). Putative slippery sequence is highlighted in red.



Strain	FL100 *VOA1/ORF2*	FL100 *VOA1ΔORF2*
YPD	3h 10min	2h 52min
Blasticidin S (26µg/ml)	-	7h 50min

Figure 5. Influence of *ORF 2* integrity on growth in different culture conditions.

Graphics display the growth curves of FL100 *VOA1/ORF2* strain (grey line) and FL100 *VOA1ΔORF2* strain (black line) in presence or absence of blasticidin S. Doubling times during exponential growth phase are indicated in the table.

Genome	Source	Sequences
Ashbya gossypii ATCC 10895	RefSeq accession	NC_005782, NC_005783, NC_005784, NC_005785, NC_005786, NC_005787, NC_005788
Aspergillus fumigatus Af293	RefSeq accession	NC_007194, NC_007195, NC_007196, NC_007197, NC_007198, NC_007199, NC_007200, NC_007201
Candida albicans SC5314	EMBL drafts Assembly 21 (2008-04-1CEV5)	Ca21chr1, Ca21chr2, Ca21chr3, Ca21chr4, Ca21chr5, Ca21chr6, Ca21chr7, Ca21chrR
Candida glabrata CBS 138	RefSeq accession	NC_005967, NC_005968, NC_006026, NC_006027, NC_006028, NC_006029, NC_006030, NC_006031, NC_006032, NC_006033, NC_006034, NC_006035, NC_006036
Candida tropicalis MYA-3404	Broad Institute draft	pool of contigs 01-JAN-2008
Cryptococcus neoformans var. neoformans JEC21	RefSeq accession	NC_006670, NC_006679, NC_006680, NC_006681, NC_006682, NC_006683, NC_006684, NC_006685, NC_006686, NC_006687, NC_006691, NC_006692, NC_006693, NC_006694
Debaryomyces hansenii CBS767	RefSeq accession	NC_006043, NC_006044, NC_006045, NC_006046, NC_006047, NC_006048, NC_006049
Encephalitozoon cuniculi GB-M1	RefSeq accession	NC_003229, NC_003230, NC_003231, NC_003232, NC_003233, NC_003234, NC_003235, NC_003236, NC_003237, NC_003238 NC_003242
Kluyveromyces lactis NRRL Y-1CEV140	RefSeq accession	NC_006037, NC_006038, NC_006039, NC_006040, NC_006041, NC_006042
Kluyveromyces waltii NCYC_2640	Broad Institute draft	pool des contigs 01-JAN-2008
Saccharomyces bayanus 623-6C	Broad Institute draft	pool des contigs 01-JAN-2008
Saccharomyces castellii NRRL_Y-1CEV2630_YM476	Whashington U draft	pool des contigs 01-JAN-2008
Saccharomyces cerevisiae	RefSeq accession	NC_001133, NC_001134, NC_001135, NC_001136, NC_001137, NC_001138, NC_001139, NC_001140, NC_001141, NC_001142, NC_001143, NC_001144, NC_001145, NC_001146, NC_001147, NC_001148, NC_001224
Saccharomyces kluyveri NRRL Y-1CEV2651	Whashington U draft	pool of contigs 01-JAN-2008
Saccharomyces kudriavzevii IFO_1802_YM6553	Whashington U draft	pool of contigs 01-JAN-2008
Saccharomyces mikatae IFO1815	Broad institute draft	pool of contigs 01-JAN-2008
Saccharomyces paradoxus NRRLY-1CEV721	Broad Institute draft	pool of contigs (01-JAN-2008)
Schizosaccharomyces pombe	RefSeq accession	NC_006067,NC_006068, NC_006069, NC_006070, NC_006071, NC_006072
Yarrowia lipolytica CLIB122	RefSeq accession	NC_006067,NC_006068, NC_006069, NC_006070, NC_006071, NC_006072

Supplementary table 1. Genomes and sequences

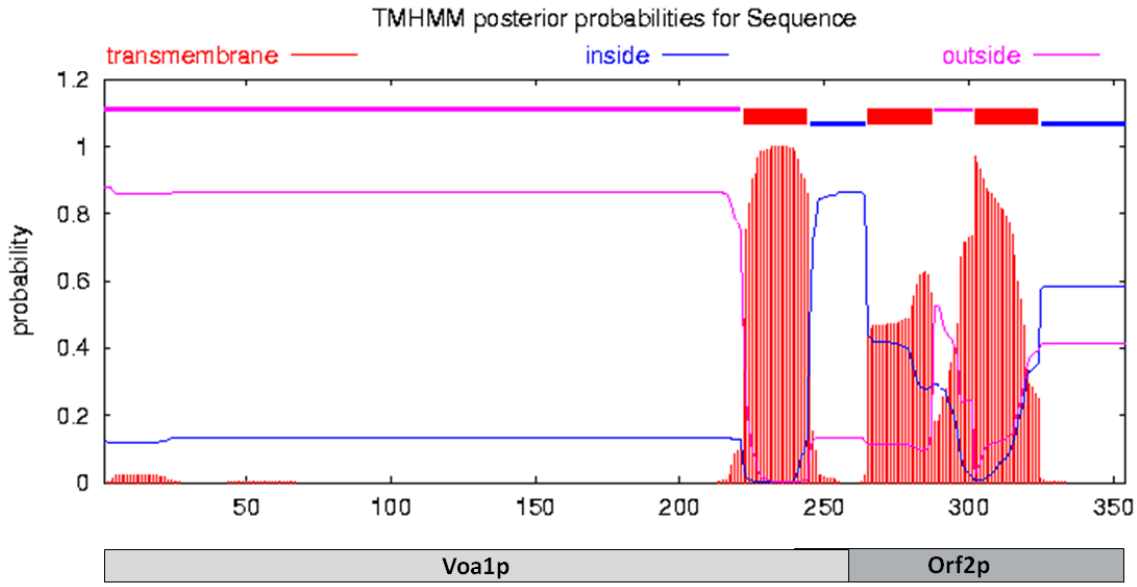
```
$codon_start="ATG";
$rev_start = "CAT";
$codon_stop="TAG|TGA|TAA";
$rev_stop= "CTA|TCA|TTA";

                %regexpr=(
                                "frameshift_minus" =>
"($codon_stop)((?(?!$codon_start)(?!$codon_stop)[ATCG]{3})*)($codon_start)((?(?!$codon_stop)[ATCG]{3})*[
ATCG]{2})($codon_stop)([ATCG](?(?!$codon_stop)[ATCG]{3})*)($codon_stop)([ATCG]{2}(?(?!$codon_stop)[A
TCG]{3})*)($codon_stop)",
                                "frameshift_minus_rev" =>
"($rev_stop)((?(?!$rev_stop)[ATCG]{3})*[ATCG]{2})($rev_stop)((?(?!$rev_stop)[ATCG]{3})*?[ATCG])($rev_sto
p)([ATCG]{2}(?(?!$rev_stop)[ATCG]{3})*)($rev_start)((?(?!$rev_start)(?!$rev_stop)[ATCG]{3})*)($rev_stop)",
                                "frameshift_plus" =>
"($codon_stop)((?(?!$codon_start)(?!$codon_stop)[ATCG]{3})*)($codon_start)((?(?!$codon_stop)[ATCG]{3})*[
ATCG])($codon_stop)([ATCG]{2}(?(?!$codon_stop)[ATCG]{3})*)($codon_stop)([ATCG](?(?!$codon_stop)[ATC
G]{3})*)($codon_stop)",
                                "frameshift_plus_rev" =>
"($rev_stop)((?(?!$rev_stop)[ATCG]{3})*[ATCG])($rev_stop)((?(?!$rev_stop)[ATCG]{3})*?[ATCG]{2})($rev_sto
p)([ATCG](?(?!$rev_stop)[ATCG]{3})*)($rev_start)((?(?!$rev_start)(?!$rev_stop)[ATCG]{3})*)($rev_stop)",
                                "readthrough" =>
"($codon_stop)((?(?!$codon_start)(?!$codon_stop)[ATCG]{3})*)($codon_start)(((?!$codon_stop)[ATCG]{3}){$
a,})($codon_stop)(((?!$codon_stop)[ATCG]{3}){$b,})($codon_stop)",

                                "readthrough_rev" =>
"($rev_stop)(((?!$rev_stop)[ATCG]{3}){$b,})($rev_stop)(((?!$rev_stop)[ATCG]{3}){$a,})($rev_start)((?(?!$rev_
start)(?!$rev_stop)[ATCG]{3})*)($rev_stop)"
                );
```

Supplementary figure 1. Implementation of regular expression used to search the compatible events in genomic sequences as Perl code.

Genomic organization	Gene	ATG mutagenesis in ORF B	Oligo watson	Oligo crick
-1CEV	AAD6	no	GATAGTTTACACATTCTAGTGC	GTTCAACAGGTTCCATTTACCTTG
-1CEV	ATF2	1 (ATG to ATT)	AGTAATACAGGACTTGTCTCGC	CATTAAAGCGACGCAAATTCGC
-1CEV	CRC1	no	GATGAACCGGTCAATATCTTG	AATCAAATGCCATACTTCTTG
-1CEV	GAT3	no	TTTAACGAAATAAAAGGTGTG	CTGTTTAAAGTCGATAAATG
-1CEV	ROT1	no	ATTAAAAGACAAAATACCTCAT	ACATGGGGCCTCTTTCCC
-1CEV	STE18	1 (ATG to TAC)	AATAGTGTTTGTTGTACGCTTATGTAATGATAGTAATAGAATCC	GGATTCTATTACTATCATTACATAAGCGTACAACAAACACTATT
-1CEV	VOA1	no	TCTAATTTGGATATCACATATGG	AAATGTTATGGATGATTGGGAT
-1CEV	YJL107c	1 (ATG to CTG)	CTTGAAATAGATGGTCAG	CCTAACCAATTGAACTTCG
-1CEV	YLR042c	1 (ATG to ATT)	ACTAATACTATAAGTTCAAG	GTTTAAGAAAACAAATAAGCG
-1CEV	YML003w	no	CACGAAGTAGACACTGAAAA	TTGAAGATGTTTGGACGCTAG
+1CEV	ABP140	2(ATG n°1 to TTG and ATG n°2 to ATT)	GTGATTGATTTGGAAGTTGTTGATTACTCTTGTCTTAGGCATTGAT	ATCAATGCCTAAGACAAGAGTAATCAACAACTTCCAAATCAATCAC
+1CEV	ATO2	1 (ATG to ATT)	GTAGCCTTCATTGCCTGGTACAACG	TATAATAGGGGTAAAAGTTATTTTAGCC
+1CEV	ATP10	no	TTAAAAATTAGGTGGGCAGC	TTGGGGAATGAACACGATCGCC
+1CEV	BUD31	no	TTAGAGGAAGAAGCACGC	TTTAGTCTGTGCTTGCAC
+1CEV	CCC1	no	GTAGCAGCAGGTGCGGCTTGG	GAAATGTATGTCGAGAAAATC
+1CEV	EST3	no	CAAATACTTAGTTGAGTTTTCC	GGAAAACTCAACTAAGTATTTG
+1CEV	OAZ1	no	TGGTGCGCGTGACATCCCTCTA	TAGAGGGATGTCACGCGCACCA
+1CEV	UGP1	1 (ATG to CTG)	TTGAGGGGTACTGTCATCATC	TTCAATGTTCCAAGATTTGC
+1CEV	YOL159c-A	no	ATAATCTTGCGTTCATGCTGG	ATCAAAGCCAATGGTAGTACC

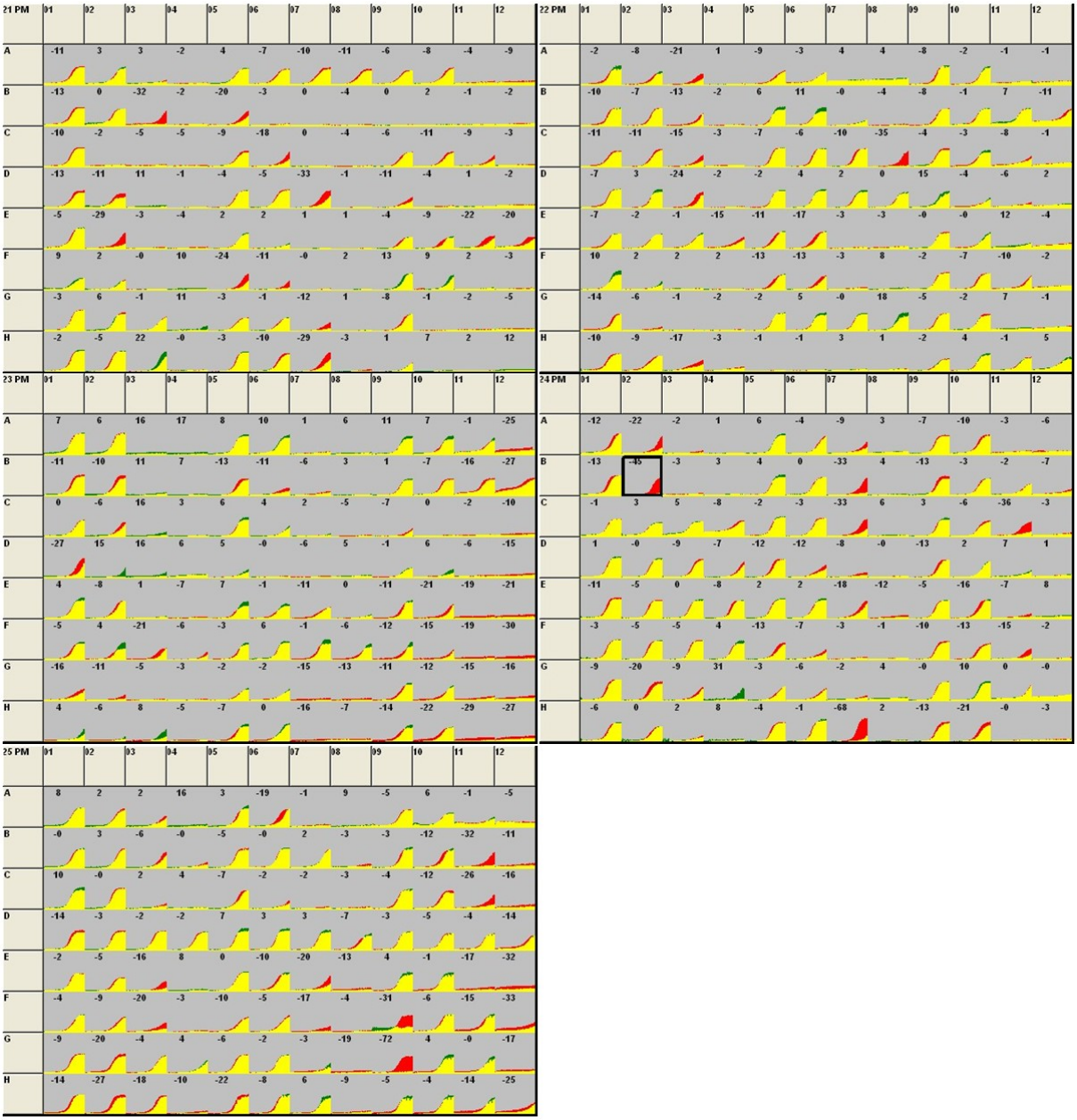
Supplementary table 2. The candidate regions analyzed the pAC PGK007.

For each candidate, the table summarized the type of compatible events, the name of the candidate, the number of ATG codons in ORF B reading frame of the analyzed region, and the oligonucleotides used to amplify the region of interest. If any, ATG were mutated into other sens codon. Mutations are described between brackets.



Supplementary figure 2. Prediction of transmembrane domains in the putative Voa1-orf2p protein.

Transmembrane domain prediction using TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/) indicates with a strong probability (> 0.6) that *ORF2* translation would add 2 transmembrane domains to Voa1p. Relative positions of transmembrane domains in Voa1-orf2p are indicated by the bottom scheme.



Supplementary figure 3. High throughput phenotypic screening results

Comparative growth curve analysis from PM plates 21 to 25 (Homann et al. 2005) between FL100 *VOA1/ORF2* (green) strain and FL100 *VOA1ΔORF2* strain (red). Overlays are displayed in yellow. The number in each well characterizes the growth difference between both strains. Negative number indicates … The black square high

Oligonucléotides	Sequences
URA3KL.w	ATACGACTCACTATAGGG
URA3KL.c	GGTGATGCTTCTCTTTTCCATGGAT
MRY14.w	CCTCTCATAGTTGCTGTGAACCC
MRY14.c	GATTAATTGTTTTTTTTTATTGGGTTTGTTGA
hphMX4.c	ATCTTAGCCAGACGAGCGGGTTCGG
VOAprom.w	ATCTGCTTGCACAGTTTTGG
pRSVOA.w	GGTTAATGTGGCTGTGGTTTCAGGGTCCATAAAGCTTTTCAATTCATCTTAAGC TTTGCAAAATCAAGTC
pRSVOA.c	GCTCTAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTGTAC TGAGAGTGCACCACGC

Supplementary table 3. Oligonucleotides

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402.

Asakura T, Sasaki T, Nagano F, Satoh A, Obaishi H, Nishioka H, Imamura H, Hotta K, Tanaka K, Nakanishi H et al. 1998. Isolation and characterization of a novel actin filament-binding protein from Saccharomyces cerevisiae. *Oncogene* **16**(1): 121-130.

Bekaert M, Firth AE, Zhang Y, Gladyshev VN, Atkins JF, Baranov PV. 2010. Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res* **38**(Database issue): D69-74.

Belew AT, Hepler NL, Jacobs JL, Dinman JD. 2008. PRFdb: a database of computationally predicted eukaryotic programmed -1 ribosomal frameshift signals. *BMC Genomics* **9**: 339.

Bidou L, Stahl G, Hatin I, Namy O, Rousset JP, Farabaugh PJ. 2000. Nonsense-mediated decay mutants do not affect programmed -1 frameshifting. *RNA* **6**(7): 952-961.

Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* **11**(7): 1246-1255.

Brierley I, Digard P, Inglis SC. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**(4): 537-547.

Byun Y, Moon S, Han K. 2007. A general computational model for predicting ribosomal frameshifts in genome sequences. *Comput Biol Med* **37**(12): 1796-1801.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**(15): 1974-1975.

Farabaugh PJ, Kramer E, Vallabhaneni H, Raman A. 2006. Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. *J Mol Evol* **63**(4): 545-561.

Goelet P, Lomonossoff GP, Butler PJ, Akam ME, Gait MJ, Karn J. 1982. Nucleotide sequence of tobacco mosaic virus RNA. *Proc Natl Acad Sci U S A* **79**(19): 5818-5822.

Homann OR, Cai H, Becker JM, Lindquist SL. 2005. Harnessing natural diversity to probe metabolic pathways. *PLoS Genet* **1**(6): e80.

Ito H, Fukuda Y, Murata K, Kimura A. 1983. Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* **153**(1): 163-168.

Ivanov IP, Matsufuji S, Murakami Y, Gesteland RF, Atkins JF. 2000. Conservation of polyamine regulation by translational frameshifting from yeast to mammals. *EMBO J* **19**(8): 1907-1917.

Jacks T, Madhani HD, Masiarz FR, Varmus HE. 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**(3): 447-458.

Jacks T, Varmus HE. 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**(4731): 1237-1242.

Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M. 2011. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res* **21**(12): 2096-2113.

Kane PM. 2006. The where, when, and how of organelle acidification by the yeast vacuolar H+-ATPase. *Microbiol Mol Biol Rev* **70**(1): 177-191.

Klobutcher LA. 2005. Sequencing of random Euplotes crassus macronuclear genes supports a high frequency of +1 translational frameshifting. *Eukaryot Cell* **4**(12): 2098-2105.

Lacroute F. 1968. Regulation of pyrimidine biosynthesis in Saccharomyces cerevisiae. *J Bacteriol* **95**(3): 824-832.

Lawrence CL, Botting CH, Antrobus R, Coote PJ. 2004. Evidence of a new role for the high-osmolarity glycerol mitogen-activated protein kinase pathway in yeast: regulating adaptation to citric acid stress. *Mol Cell Biol* **24**(8): 3307-3323.

Liang H. 2010. Decoding the dual-coding region: key factors influencing the translational potential of a two-ORF-containing transcript. *Cell Res* **20**(5): 508-509.

Liao PY, Choi YS, Lee KH. 2009. FSscan: a mechanism-based program to identify +1 ribosomal frameshift hotspots. *Nucleic Acids Res* **37**(21): 7302-7311.

Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**(5): 386-398.

Matsufuji S, Matsufuji T, Miyazaki Y, Murakami Y, Atkins JF, Gesteland RF, Hayashi S. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**(1): 51-60.

Michel AM, Roy Choudhury K, Firth AE, Ingolia NT, Atkins JF, Baranov PV. 2012. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res*.

Moon S, Byun Y, Kim HJ, Jeong S, Han K. 2004. Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res* **32**(16): 4884-4892.

Morris DK, Lundblad V. 1997. Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol* **7**(12): 969-976.

Namy O, Duchateau-Nguyen G, Hatin I, Hermann-Le Denmat S, Termier M, Rousset JP. 2003. Identification of stop codon readthrough genes in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**(9): 2289-2296.

Namy O, Duchateau-Nguyen G, Rousset JP. 2002. Translational readthrough of the PDE2 stop codon modulates cAMP levels in Saccharomyces cerevisiae. *Mol Microbiol* **43**(3): 641-652.

Namy O, Rousset JP, Napthine S, Brierley I. 2004. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* **13**(2): 157-168.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J et al. 2012. The Pfam protein families database. *Nucleic Acids Res* **40**(Database issue): D290-301.

Ribrioux S, Brüngger A, Baumgarten B, Seuwen K, John MR. 2008. Bioinformatics prediction of overlapping frameshifted translation products in mammalian transcripts. *BMC Genomics* **9**: 122.

Ryan M, Graham LA, Stevens TH. 2008. Voa1p functions in V-ATPase assembly in the yeast endoplasmic reticulum. *Mol Biol Cell* **19**(12): 5131-5142.

Schwartz D, Chou MF, Church GM. 2009. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol Cell Proteomics* **8**(2): 365-379.

Servant F, Bru C, Carrère S, Courcelle E, Gouzy J, Peyruc D, Kahn D. 2002. ProDom: automated clustering of homologous domains. *Brief Bioinform* **3**(3): 246-251.

Shorter J, Lindquist S. 2005. Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet* **6**(6): 435-450.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**(1): 19-27.

Stahl G, Bidou L, Rousset JP, Cassan M. 1995. Versatile vectors to study recoding: conservation of rules between yeast and mammalian cells. *Nucleic Acids Res* **23**(9): 1557-1560.

Theis C, Reeder J, Giegerich R. 2008. KnotInFrame: prediction of -1 ribosomal frameshift events. *Nucleic Acids Res* **36**(18): 6013-6020.

Williams I, Richardson J, Starkey A, Stansfield I. 2004. Genome-wide prediction of stop codon readthrough during translation in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* **32**(22): 6605-6616.

Wilson W, Malim MH, Mellor J, Kingsman AJ, Kingsman SM. 1986. Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. *Nucleic Acids Res* **14**(17): 7001-7016.

Wulff BE, Sakurai M, Nishikura K. 2011. Elucidating the inosinome: global approaches to adenosine-to-inosine RNA editing. *Nat Rev Genet* **12**(2): 81-85.

Xayaphoummine A, Bucher T, Isambert H. 2005. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res* **33**(Web Server issue): W605-610.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**(7232): 1033-1037.

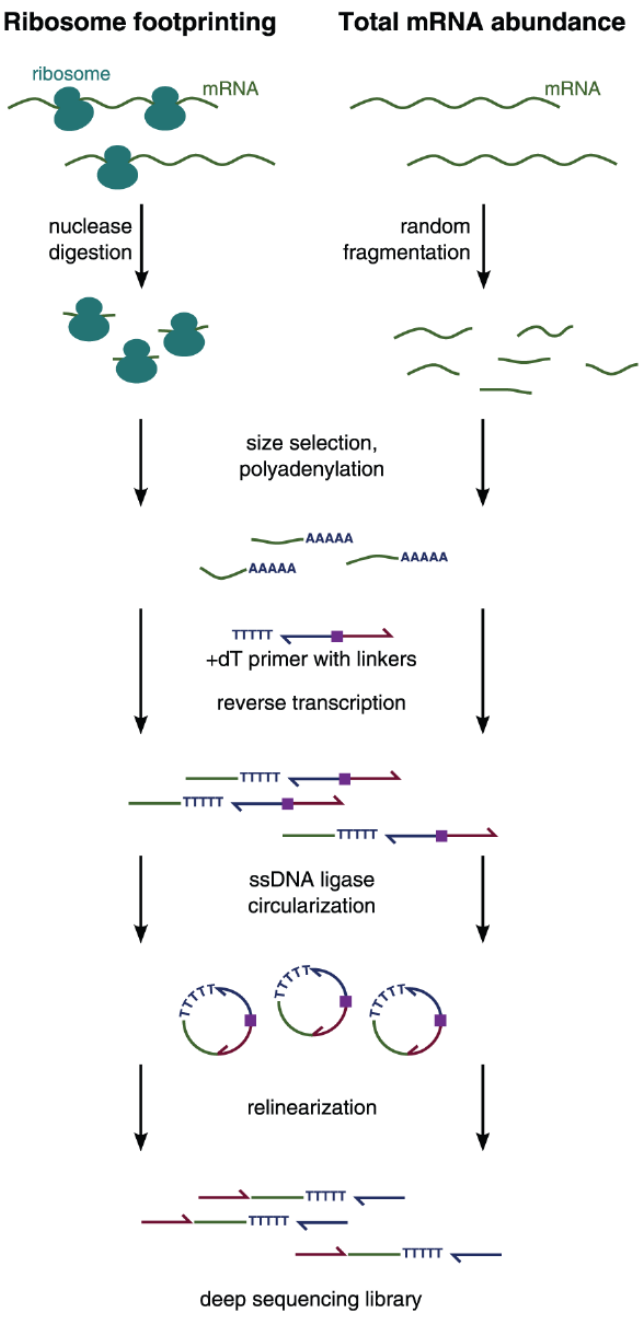


Figure 1 : Principe du profil traductionnel

Le profil traductionnel ou « ribosome profilling » est une étude à grande échelle de la synthèse des protéines d'une population cellulaire dans un état donné (ribosome footprinting). Elle est associée à une approche transcriptomique (total mRNA abundance). Après traitement avec une nucléase, les séquences d'ARNm protégées par des ribosomes sont purifiées et séquencées massivement. Les reads obtenus sont alignés sur le génome. Figure issue de (Ingolia et al. 2009)

II. Détermination de l'état traductionnel des candidats grâce aux données de traductomique (« ribosome profiling »).

L'état d'expression d'un gène peut être déterminé à 2 niveaux : celui de l'ARNm et celui de la protéine. Les études de transcriptomique et de protéomique fournissent des données qualitatives et quantitatives sur une population cellulaire à un état donné. Ces études décrivent l'observation des états d'équilibre entre synthèse et dégradation de chaque ARNm ou protéine.

Depuis peu, ces observations peuvent être complétées grâce à une nouvelle technique dite de traductomique. Elle permet de caractériser l'état traductionnel d'une population cellulaire à un instant donné. Cette technique initialement publiée en 2009 est appelée en anglais « ribosome profiling » ce qui peut se traduire en français par profil traductionnel (Ingolia et al. 2009). L'intérêt de cette technique est de pouvoir observer directement quels sont les ARNm en cours de traduction, sans le biais des régulations post-traductionnelles et de la dégradation des protéines.

1. Méthodologie

Le principe expérimental de cette technique est d'isoler les fragments d'ARNm qui sont protégés par des ribosomes 80S en cours de traduction (figure 1). Pour cela, la traduction d'une population de cellules est figée à un instant donné, soit par une congélation rapide dans l'azote liquide, soit par l'ajout de cycloheximide dans le milieu de culture. La cycloheximide est un inhibiteur de la traduction qui permet de bloquer les ribosomes en cours d'élongation sur l'ARNm. Les extraits cellulaires contenant des ARNm polysomaux sont traités à la RNaseI afin de dégrader les portions d'ARNm non protégés par des ribosomes. Les ribosomes 80S sont ensuite isolés afin de pouvoir purifier les fragments d'ARNm qu'ils protègent. Typiquement, un ribosome recouvre et protège une séquence d'environ 30 nucléotides.

Ces courts fragments d'ARNm sont séquencés massivement après une étape de rétro-transcription pour produire un ADNc[1]. Le séquençage génère des fragments d'une quarantaine de nucléotides auxquels il faut retirer les amorces utilisées pour construire la banque. Chaque séquence obtenue, appelée "read" correspond à la position d'un ribosome. Tous les reads sont ensuite alignés sur la séquence d'un génome de référence afin de déterminer quelles régions sont traduites. A partir d'un read de 30 nucléotides, il est possible de prédire quels sont les codons présents au site A, P et E du ribosome.

L'étude traductomique est réalisée en parallèle d'une étude transcriptomique. Les mêmes extraits cellulaires sont utilisés pour réaliser un séquençage massif des ARNm qui ont été préalablement fragmentés. Les études transcriptomiques et traductomiques sont donc différentes : la 1ère correspond à une expérience de fractionnement, alors que la 2nde est une expérience de protection.
En transcriptomique, le nombre de reads à chaque position permet de calculer une densité représentative de l'abondance de chaque ARNm dans les cellules. Dans le cas du profil traductionnel, la position de chaque read révèle quelles sont les régions traduites comme les uORF, les exons ou les zones de recodage. Quant à la densité de reads à une position, elle reflète entre autre la dynamique

[1] ADNc : ADN complémentaire

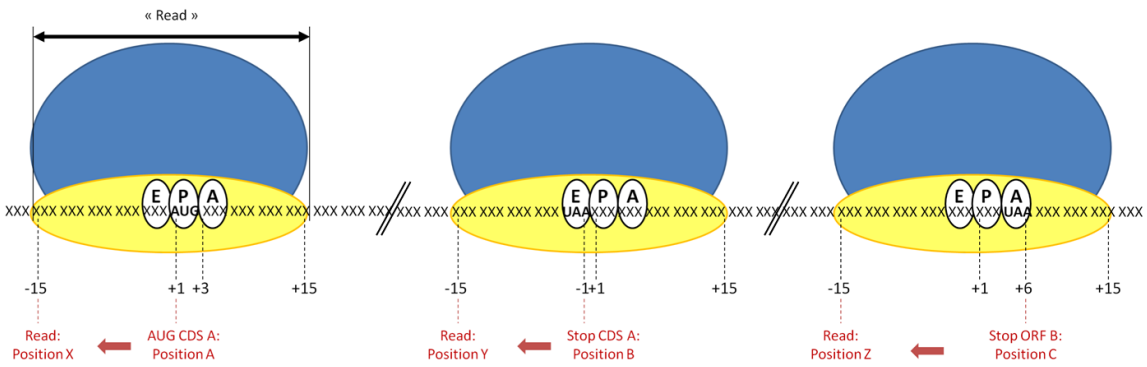


Figure 2 : Positions des reads par rapport aux coordonnées de l'événement

Chaque schéma représente la position du ribosome sur le codon initiation de la CDS-A (position A), 2 codons en aval du codon stop de la CDS-A (position B) et sur le 2ème codon stop de l'ORF-B (position C). La coordonnée 5' du read associé à chaque position est calculée en soustrayant le nombre de basse séparant la séparant de la position de référence.

de traduction en indiquant potentiellement des sites de pauses. Lorsque la profondeur de séquençage est suffisamment importante, il est possible de délimiter précisément les bornes des régions traduites et de prédire le cadre de lecture probable de la traduction.

L'analyse en parallèle des données de transcriptomique et de traductomique permet de faire un lien entre le niveau de transcription d'un ARNm et son niveau de traduction. Dans le cas de l'étude du recodage, l'intérêt du profil traductionnel est de pouvoir rechercher des régions traduites hors des régions annotées.

2. Profils traductionnels des gènes candidats

A. Les jeux de données analysés

Afin de caractériser les 19 candidats sélectionnés lors de l'approche de génomique comparative, j'ai décidé de regarder leur état traductionnel dans différentes conditions de culture. Pour cela, j'ai utilisé les données de profils traductionnels disponibles chez *S. cerevisiae*. Avec Christine Drevet, nous avons finalement étudié 4 jeux de données. Les 2 premiers sont issus d'une étude de Nicholas T. Ingolia à partir du fond génétique BY4741 (Ingolia et al. 2009). Cette étude compare l'état traductionnel d'une culture cellulaire dans 2 conditions de croissance différentes : l'une en milieu riche et l'autre après 20 min en milieu carencé en acides aminés. Les 2 autres jeux de données sont issus d'une étude de Gloria Brar sur la méiose dans le fond génétique SK1 (Brar et al. 2012). Les données proviennent de cellules synchronisées et isolées 4h après l'induction de la méiose.

B. Les critères d'analyse

Afin d'éviter les ambigüités liées au reads s'alignant de façon multiple sur le génome, nous avons décidé de ne travailler qu'avec les reads qui s'alignaient de façon unique. Pour chacun des 4 jeux de données, les reads ont été alignés sur le génome de référence de la souche S288C, car c'est celui qui a été analysé dans l'approche de génomique comparative et pour lequel nous avons les coordonnées génomiques des candidats.

Nous avons décidé de rechercher des reads dans la CDS-A et l'ORF-B de chaque candidat. Afin d'établir leur taux de traduction, tous les reads présents sur la totalité de la région (CDS-A + ORF-B) sont dénombrés. Pour détecter si l'ORF-B est traduite, nous nous intéressons en particulier aux reads qui correspondent à des ribosomes ayant franchi le stop de la CDS-A et qui traduisent l'ORF-B. Afin de ne pas inclure les reads qui correspondent à des ribosomes en cours de terminaison sur le codon stop de la CDS-A, nous avons décidé de compter ceux qui avaient poursuivi la traduction sur au moins 2 cycles de translocation. Afin de réaliser l'analyse, 3 coordonnées sont utilisées : celle de la 1ère base du codon d'initiation, celle de la 3ème base du second codon en aval du stop de la CDS-A, et celle de la 3ème base du dernier stop de l'ORF-B. Elles sont respectivement notées A, B et C.

L'alignement de chaque read sur le génome est référencé à partir du brin sur lequel il s'apparie et de la position de la 1ère base alignée sur la séquence chromosomique. Dans notre analyse, il n'est pas possible de rechercher directement des reads sur une région d'intérêt simplement en utilisant les coordonnées des positions A, B et C définies précédemment les codons qu'elles représentent sont à

	BY4741 milieu riche (2.096.758 total reads)							BY4741 milieu carencé (2.096.758 total reads)							
	CDS A et ORF B			ORF B			Génome*	CDS A et ORF B			ORF B			Génome*	
candidats	nb reads	length (bp)	nb reads/50bp	nb reads	length (bp)	nb reads/50bp	frequence	nb reads	length (bp)	nb reads/50bp	nb reads	length (bp)	nb reads/50bp	frequence	candidats
ugp1	1065	1756	30,33	1	254	0,20	1	3421	1756	97,41	5	254	0,99	1	ugp1
voa1	589	1070	27,53	5	270	0,93	1	495	1070	23,13	5	270	0,93	1	voa1
abp140	949	1888	25,13	356	1046	17,02	0,003	190	683	13,91	0	348	0,00	1	ste18
ste18	213	683	15,60	0	348	0,00	1	495	1888	13,11	186	1046	8,89	0,019	abp140
rot1	194	950	10,21	3	177	0,85	1	235	950	12,37	7	177	1,98	0,258	rot1
ccc1	187	1201	7,79	1	230	0,22	1	204	1201	8,50	0	230	0,00	1	ccc1
yrl042c	78	704	5,54	3	216	0,70	1	55	547	5,03	2	263	0,38	1	est3
atf2	137	1751	3,91	0	141	0,00	1	37	880	2,10	4	668	0,30	1	oaz1
oaz1	53	880	3,01	3	668	0,23	1	25	704	1,78	2	216	0,47	1	yrl042c
bud31	37	709	2,61	0	233	0,00	1	25	733	1,71	1	458	0,11	1	yol159c
est3	28	547	2,56	5	263	0,95	1	17	709	1,20	0	233	0,00	1	bud31
yol159c	34	733	2,32	2	458	0,22	1	45	2320	0,97	2	1478	0,07	1	atp10
atp10	54	2320	1,17	0	1478	0,00	1	24	1751	0,69	1	141	0,36	1	atf2
yjl107c	27	2312	0,59	0	1146	0,00	1	29	3272	0,45	8	2397	0,17	1	yml002w
yml002w	11	3272	0,17	3	2397	0,07	1	10	1181	0,43	0	195	0,00	1	crc1
aad6	0	1127	0,00	0	486	0,00	1	14	2312	0,31	0	1146	0,00	1	yjl107c
ato2	0	1054	0,00	0	203	0,00	1	6	1054	0,29	0	203	0,00	1	ato2
crc1	0	1181	0,00	0	195	0,00	1	2	1127	0,09	1	486	0,11	1	aad6
gat3	0	680	0,00	0	252	0,00	1	1	680	0,08	1	252	0,20	1	gat3

	SK1 A (14.304.058 total reads)							SK1 B (15.225.876 total reads)							
	CDS A et ORF B			ORF B			Génome*	CDS A et ORF B			ORF B			Génome*	
candidats	nb reads	length (bp)	nb reads/50bp	nb reads	length (bp)	nb reads/50bp	frequence	nb reads	length (bp)	nb reads/50bp	nb reads	length (bp)	nb reads/50bp	frequence	candidats
crc1	10783	1181	456,52	18	195	4,62	0,155	9366	1181	396,53	15	195	3,85	0,206	crc1
rot1	3220	950	169,48	10	177	2,83	0,282	2770	950	145,79	8	177	2,26	0,297	rot1
voa1	2072	1070	96,82	5	270	0,93	1	1723	1070	80,52	1	270	0,19	1	voa1
bud31	894	709	63,05	0	233	0,00	1	1140	1201	47,46	4	230	0,87	1	ccc1
ccc1	1299	1201	54,08	1	230	0,22	1	666	709	46,97	0	233	0,00	1	bud31
oaz1	412	880	23,41	91	668	6,81	0,101	368	880	20,91	80	668	5,99	0,129	oaz1
atp10	799	2320	17,22	148	1478	5,01	0,122	702	2320	15,13	130	1478	4,40	0,158	atp10
ato2	286	1054	13,57	3	203	0,74	1	256	1054	12,15	2	203	0,50	1	ato2
est3	117	547	10,70	9	263	1,71	0.432	105	547	9,60	5	263	0,95	1	est3

Figure 3 : Profil traductionnel des candidats

Le tableau synthétise les résultats obtenus pour chaque candidat pour chaque jeu de données. Les gènes caractérisés comme soumis au décalage en +1 sont surlignés en gris. Le nombre de reads est calculé soit pour l'ensemble de l'événement (CDS-A + ORF-B) soit pour l'extension ORF-B CDS-A exclue (ORFB). Le nombre de read est normalisé par la taille de la région analysée. Ce résultat est présenté en nombre de reads/ 50pb. La colonne génome précise pour chaque valeur « nombre de reads/50bp » de l'ORF B la fréquence à laquelle cette valeur est retrouvée sur les 50bp en aval de l'ensemble des CDS.

l'intérieur d'un read et non à son extrémité (figure 2). Dans notre cas, les positions correspondent à des codons qui ne sont pas dans les mêmes sites du ribosome (A, P et E). L'analyse nécessite donc de déduire les coordonnées des reads correspondant à des ribosomes situés aux bornes A, B et C. Ces positions sont respectivement annotées X, Y et Z.

Dans un read dont les positions sont numérotées de -15 à +15, les sites E, P et A correspondent en théorie aux positions -3 à -1, +1 à +3, et +4 à +6 (figure 2). Dans le cas d'un candidat présent sur le brin watson, lorsque le ribosome est sur un codon d'initiation, ce dernier est au site P, ce qui correspond aux positions +1 à +3 du read. Tous les reads d'intérêt correspondant à cette position du ribosome auront donc comme coordonnées 5' : X = A- 15 nucléotides. Dans le cas où le codon stop se retrouve au site E, la coordonnée 5' du read correspond à : Y = B - 14 nucléotides. Quant au cas où le ribosome est en cours de terminaison sur le second stop de l'ORF-B, la position 5' du read correspond à : Z = C - 20 nucléotides. Lorsque les événements sont situés sur le brin crick, le calcul est un peu différent. Au lieu de déterminer la position 5 ' des reads, c'est la coordonnée 3' qui est considérée. Les bornes des régions analysées sont alors : X = A-14, Y = B - 15 et Z= C – 18. Tous les reads compris sur la totalité de l'événement entre les positions (X ; Z) ou ceux situés seulement sur l'extension (Y ; Z) sont retenus pour l'analyse.

C. Résultats et discussion

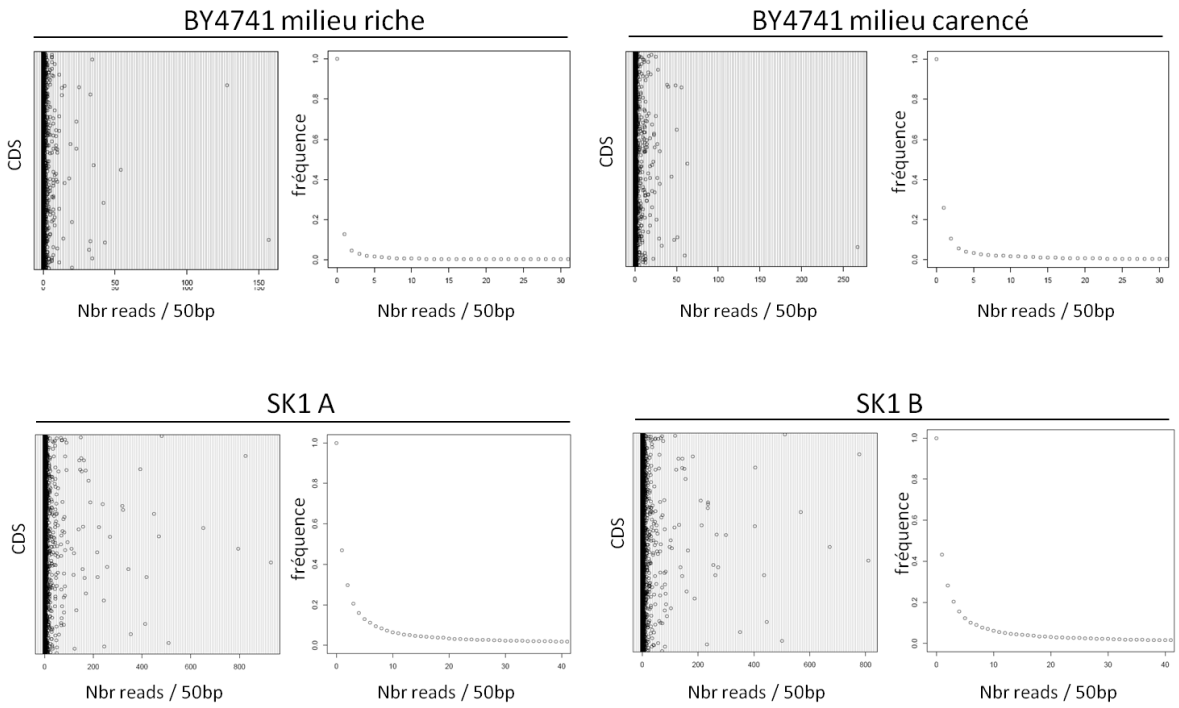
Il apparaît clairement que la profondeur de séquençage est différente entre les 2 études. Pour la souche BY4741, environ 2 millions de reads sont disponibles pour chacun des jeux de données alors que pour la souche SK1, entre 14 et 15 millions de reads le sont. Le résultat est présenté dans le tableau figure 3.

❖ *Les candidats*

Dans le cas où les reads proviennent du fond génétique BY4741, il n'y a pas de problème d'alignement, les structures génomiques de tous les candidats sont conservées. A l'inverse, des insertions ou des délétions ainsi que des mésappariements sont détectés pour certains candidats lorsque les reads proviennent du fond génétique SK1. Nous avons donc décidé de ne retenir pour l'analyse que les candidats ne possédant aucune délétion ou insertion, ainsi que ceux ayant moins de 6 mésappariements sur la totalité de l'événement. Les candidats retenus avec ces critères sont : *ATO2*, *ATP10*, *BUD31*, *CCC1*, *CRC1*, *EST3*, *OAZ1*, *ROT1* et *VOA1*.

Comme on pouvait s'y attendre, en fonction des différentes conditions de culture, certains candidats ne sont pas exprimés. A l'inverse, le gène qui est le mieux exprimé dans toutes les conditions est *ABP140*. De nombreux reads sont détectés sur la région (CDS-A + ORF-B), et une proportion importante est positionnée dans l'extension de l'ORF-B. Dans le cas d'un gène exprimé fortement et dont l'efficacité de décalage du cadre de lecture en +1 est efficace, il est donc possible de détecter du recodage grâce à un profil traductionnel.

En regardant *EST3* et *OAZ1*, on constate que le nombre de reads dans l'ORF-B compte pour une faible part des reads présents sur la région (CDS-A + ORF-B). Ces 2 gènes ont été caractérisés comme étant soumis à du décalage en +1 ; on peut donc estimer que les reads détectés dans l'ORF-B sont associés



Conditions	BY4741 milieu riche	BY4741 milieu carencé	SK1 A	SK1 B
Valeur min (reads/50pb)	0	0	0	0
Valeur max (reads/50pb)	157	267	929	810
Moyenne (reads/50pb)	0,38	0,78	4,66	4,25
Ecart-type (reads/50pb)	3,26	4,60	30,89	28,15
Médiane (reads/50pb)	0	0	0	0

Figure 4 : Détermination du bruit de fond

Le bruit de fond est déterminé est calculant le nombre de reads dans les 50pb en aval de chaque CDS du génome. Les graphiques représentent pour chaque jeu de données le nombre de gènes en fonction du nombre de reads/50pb ainsi que la fréquence de CDS associées à chaque valeur de « nombre de reads/50pb ».

Le tableau résume pour chaque jeu de données les valeurs caractéristiques de cette analyse.

à des ribosomes qui ont changés de cadre de lecture. Malheureusement, le faible nombre de reads ne permet pas de déterminer le cadre de lecture et de confirmer cette hypothèse.

D'autres candidats possèdent un faible nombre de reads dans l'ORF-B. La question est alors de savoir si la présence de ces reads correspond à de la traduction ou à du bruit de fond. Il en découle l'interrogation suivante : « quelle est la probabilité de trouver X read après une CDS quelconque dans le génome ? » Afin de normaliser les données, le nombre de reads est divisé par la longueur de la région analysée. Ce ratio est ensuite exprimé en nombre de reads/50pb pour être comparé avec une valeur de bruit de fond.

- ***Détermination du bruit de fond***

Dans notre cas, nous avons estimé que le bruit de fond correspondait à la probabilité de trouver des reads en aval des 5879 CDS déclarées. Le génome de *S. cerevisiae* est dense ; donc pour minimiser la probabilité de compter des reads appartenant à une CDS en aval, nous avons décidé de limiter l'analyse à une région de 50 nucléotides. Pour chaque jeu de données, les graphiques de la figure 4 présentent la distribution du nombre de CDS en fonction du nombre de reads/50pb en aval, ainsi que la fréquence des CDS en fonction du nombre de reads/50pb en aval.

Quels que soient les jeux de données considérés, on constate que la majorité des gènes ne possèdent pas de reads dans les 50 nucléotides en aval de leur codon stop. Seuls 12,6 % des gènes dans la condition « BY4741 milieu riche », 25,8 % dans la condition « BY4741 milieu carencé », 43,2 % dans la condition « SK1 A » et 46,9 % dans la condition « SK1 B » présentent des reads en 3' du codon stop. Les différences entre ces valeurs s'expliquent en partie par la profondeur de séquençage. Dans l'étude de la souche SK1, le nombre de reads est environ 7 fois plus important, il est donc normal que le bruit de fond soit plus important. En revanche les différences d'expression observées entre les conditions « BY4741 milieu riche » et « BY4741 milieu carencé », pourraient avoir une signification biologique.

- ***Les candidats versus le bruit de fond***

La fréquence associée à chaque nombre de reads/50pb permet de déterminer pour chaque candidat quelle est la probabilité d'observer un nombre de reads donné/50pb dans le génome (figure 3 et 4). Par exemple, il y a 17,3 reads/50pb dans l'ORF-B d'*ABP140* dans la condition « BY4741 milieu riche ». D'après le calcul sur les CDS, des gènes avec une valeur supérieure ou égale à 17 reads/50pb se rencontrent à une fréquence de 0,003 dans le génome. On peut donc estimer que dans le cas d'*ABP140* cette valeur reflète bien de la traduction et non du bruit de fond. A l'inverse, pour *OAZ1* et *EST3*, il n'est pas possible de conclure car la fréquence est égale à 1.

En fixant arbitrairement un seuil de fréquence à 0,2 et en regroupant toutes les données, on peut estimer que les reads présent dans l'ORF-B des gènes *ATP10, ABP140*, *CRC1* et *OAZ1* sont effectivement traduites. Dans le cas d'*ATP10*, cela pourrait être seulement le reflet de la traduction du gène *CST9* présent dans l'ORF-B. Un northern blot avec une sonde chevauchant les 2 CDS permettrait de déterminer si ces 2 CDS sont présentes sur le même ARNm comme le suggère une étude transcriptomique[2] (Xu et al. 2009). Si c'est effectivement le cas, cela soulève des interrogations sur les régulations traductionnelles de *CST9* : « est-ce que sa traduction est dépendante de celle

[2] http://steinmetzlab.embl.de/cgi-bin/viewNFRsharing.pl?showSamples=NFRsharing&gene=ATP10

d'ATP10 ? » ; « est-il traduit à partir d'une IRES ou par un mécanisme de réinitiation? » Pour *CRC1*, il serait intéressant d'étudier plus en détails ce candidat pour vérifier s'il y a effectivement, dans les conditions de l'expérience, traduction de l'ORF-B dans la souche SK1. Cette ORF contient un codon méthionine qui pourrait être le site d'initiation d'une protéine de 66 acides aminés.

- ***Limites et perspectives***

Cette analyse montre à quel point il est difficile de détecter de manière fiable des événements de recodage faiblement exprimés comme *EST3* et *OAZ1*. Nous avons utilisé les données de profils traductionnels pour étudier spécifiquement les gènes candidats. Finalement, aucun des candidats positifs nouvellement identifiés avec le vecteur double rapporteur ne ressort dans cette analyse.

Malgré tout, il serait intéressant de faire une analyse plus globale de ces données afin d'étudier les CDS qui possèdent de nombreux reads dans les 50pb en aval de leurs codons stop. Il faudrait déterminer si la traduction se poursuit au-delà des 50pb et trouver la cause de cette traduction. Une autre expérience à faire serait de regarder si ces CDS sont présentes dans la base de données CEV pour vérifier si notre approche aurait été capable de les identifier.

En conclusion, il serait vain de réaliser des profils traductionnels dans le seul but de détecter uniquement des événements de recodage. Pour que de telles études soient efficaces, il faudrait les réaliser sur des populations de cellules dans divers états de croissance afin que tous les gènes soient exprimés dans au moins une condition. Cela maximiserait la probabilité d'identifier de nouveaux événements. Cependant, cette nouvelle technique promet d'être largement utilisée dans les années à venir et donc de nouveaux profils caractérisant des états cellulaires différents seront alors disponibles. Une étude systématique et automatisée de ces nouvelles données pourrait aider à la détection d'événements de recodage mais aussi appréhender les régulations traductionnelles dans leur globalité.

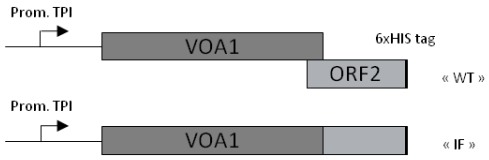


Figure 5 : Marquage de l'extrémité C-terminale de Voa1-orf2p

Schéma des constructions pYX-VOA1-ORF2(6HIS)-WT et pYX-VOA1-ORF2(6HIS)-IF

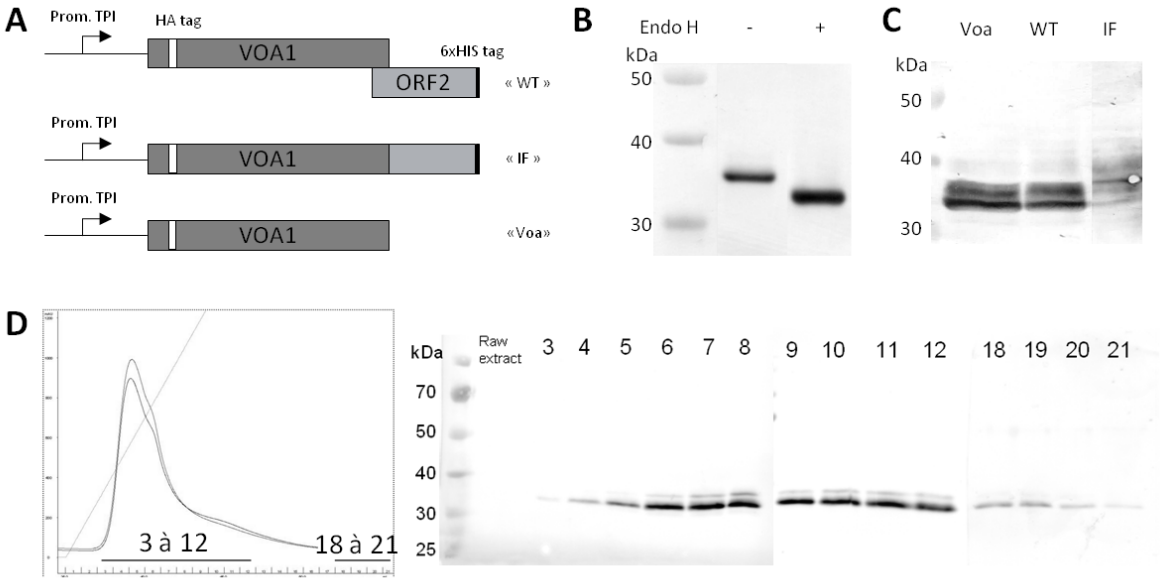


Figure 6 : Double marquage des extrémités N-terminale de Voa1p et C-terminale d'Orf2p : étiquettes HA et 6HIS

A) Schéma des constructions pYX-(HA)VOA1-ORF2(6HIS)-WT, pYX-(HA)VOA1-ORF2(6HIS)-IF et pYX-(HA)VOA1. **B)** Western blot avec des anticorps anti-HA. Les extraits exprimant la construction « WT » ont été traités ou non avec l'enzyme Endoglycosylase H (Endo H). **C)** Western blot avec des anticorps anti-HA à partir d'extraits cellulaires bruts exprimant les constructions Voa, WT ou IF. La piste IF a été révélée pendant un temps beaucoup plus long que pour les 2 autres pistes. **D)** Purification d'extrait cellulaire exprimant la construction WT sur une colonne Ni-NTA. Les fractions d'élution 3 à 12 et 18 à 21 sont analysées par western blot avec des anticorps anti-HA.

III. Analyse moléculaire de *VOA1/ORF2* : résultats et discussion

Au moment où j'ai repris ce projet, l'efficacité de recodage de *VOA1* avait déjà été quantifiée avec le vecteur double rapporteur et les motifs caractéristiques des événements de décalage en -1 avaient été identifiés. De plus, une expérience de RT-PCR avait permis de mettre en évidence la présence d'ARNm comprenant *VOA1* et l'*ORF2*. Ces arguments forts soutenaient l'hypothèse que VOA1 était soumis à un décalage du cadre de lecture en -1.

1. Expression de Voa1-orf2p

Dans le but de déterminer s'il existait une protéine recodée Voa1-orf2p, j'ai réalisé différentes constructions à partir de *VOA1/ORF2*. Elles sont exprimées à partir d'un promoteur fort et portées sur des plasmides multicopies possédant des origines de réplication 2µm.

A. Marquage de l'extrémité C-terminale de Orf2p avec l'étiquette 6HIS

Pour détecter la protéine Voa1-orf2p, le premier essai a consisté à marquer l'extrémité C-terminale d'Orf2p avec une étiquette composée de 6 histidines (6HIS) (figure 5). La construction est exprimée à partir du promoteur *TPI1* sur le plasmide pYX212. Cette construction se nomme pYX-VOA1-ORF2(6HIS)-WT. J'ai construit aussi un plasmide contrôle dans lequel *VOA1* et *ORF2* sont dans la même phase de lecture, il s'appelle pYX-VOA1-ORF2(6HIS)-IF. Pour cela, 2 adénines ont été délétées dans la région contenant les séquences potentielles de glissement (figure 8, voir Yif).

Des cellules HFY ont été transformées avec ces plasmides et les extraits cellulaires ont été utilisés pour réaliser des western blots avec des anticorps anti-6HIS. La protéine Voa1p est une protéine glycosylée qui a une taille apparente de 36 kDa (Ryan, Graham and Stevens 2008). La traduction de l'*ORF2* suite à un décalage en -1 conduirait à ajouter une extension de 12 kDa.

Finalement, aucun des western blots réalisés n'a permis de détecter spécifiquement, même pour la construction contrôle, de protéines correspondant à Voa1-orf2p. La détection était complexifiée par la présence de nombreuses bandes. Leur présence s'explique par le fait qu'il existe chez *S. cerevisiae* des protéines possédant 6 histidines d'affilée.

B. Double marquage des extrémités N-terminale de Voa1p et C-terminale de Orf2p

Afin d'avoir une étiquette plus spécifique pour la détection de Voa1-orf2p, j'ai utilisé l'étiquette HA provenant de la séquence de la protéine hémagglutinine du virus de l'influenza humaine. J'ai inséré cette étiquette dans les constructions précédentes au début de Voa1p (figure 6). Cette dernière possède un signal d'adressage au réticulum endoplasmique qui est clivé post-traductionnellement. J'ai intégré l'étiquette HA juste en aval du site de clivage comme décrit par M. Ryan en 2008 (Ryan et al. 2008) (figure 6). Les constructions sont alors nommées « WT » pour pYX-(HA)VOA1-ORF2(6HIS)-WT et « IF » pour pYX-(HA)VOA1-ORF2(6HIS)-IF. De plus, j'ai réalisé une construction supplémentaire dans laquelle l'extension de l'ORF2 est absente. Il s'agit du plasmide pYX-(HA)VOA1 qui sera nommé « VOA ».

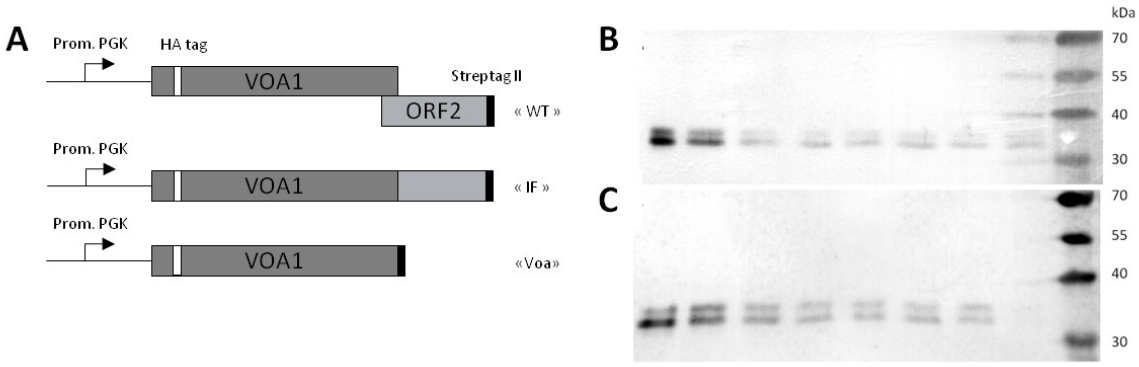


Figure 7 : Double marquage des extrémités N-terminale de Voa1p et C-terminale de Orf2p : étiquettes HA et Streptag II

A) Schéma des constructions pFL44L-(HA)VOA1-ORF2(streptag)-WT et pFL44L-(HA)VOA1-ORF2(streptag)-IF, et pFL44L-(HA)VOA1(streptag). **B) et C)** Purification d'extraits cellulaires exprimant la construction Voa (**B**) ou WT (**C**) sur une colonne StrepTactin. Les fractions d'élution sont analysées par western blot avec des anticorps anti-HA.

❖ *Extraits cellulaires bruts*

Le 1er western blot de la figure 6 a été réalisé avec la construction « WT ». L'anticorps anti-HA a permis de détecter une protéine aux environs de 35 kDa. Un traitement des extraits avec l'enzyme recombinante Endoglycosydase H (EndoH) permet de cliver certaines modifications post-traductionnelles liées à la glycosylation. Après traitement, la taille du signal observé diminue, indiquant que cette protéine était bien glycosylée. La spécificité de l'étiquette, la taille du signal et la présence de modifications post-traductionnelles de type glycosylation indiquent que la protéine observée correspond à Voa1p.

Le 2nd western blot a été réalisé avec les constructions « WT », « IF » et « VOA ». Les anticorps anti-HA ont permis de détecter 2 signaux compris entre 30 et 40 kDa pour les constructions « WT » et « VOA » (figure 6C). Ces signaux correspondent respectivement aux formes de Voa1p non-modifiée et modifiée post-traductionnellement. Pour la construction contrôle « IF », aucun signal reproductible n'a pu être détecté. Cela pourrait être dû au fait que la forme longue n'est pas fortement traduite ou qu'elle n'est pas stabilisée dans nos conditions de cultures.

❖ *Extraits cellulaires purifiés*

Afin d'enrichir en protéine d'intérêt les fractions analysées par western blot, j'ai utilisé les propriétés d'affinité de l'étiquette 6HIS pour l'ion nickel (Ni) pour réaliser des purifications d'extrait cellulaire sur des colonnes d'affinité Ni-NTA.

Les fractions d'élution obtenues à partir de cellules exprimant la construction « WT » ont ensuite été analysées par Western blot avec des anticorps anti-HA (figure 6D). Etonnamment, seules les formes non-glycosylées et glycosylées de Voa1p sont détectées alors qu'elles ne possèdent pas l'étiquette 6HIS. Pour s'assurer de leur identité, les bandes colorées au nitrate d'argent correspondant à ces poids moléculaires ont été analysées par spectrométrie de masse. Cette analyse a été réalisée par Manuela Argentini de la plateforme Imagif (Gif-sur-Yvette). Aucune des protéines identifiées ne correspondaient à Voa1p. Contrairement à l'impression donnée par la détection en western blot, les protéines Voa1p sont en très faible quantité dans le gel ce qui empêche de les détecter par spectrométrie de masse. Finalement, aucune des constructions « WT » et « IF » n'a permis de détecter de protéines correspondant à la fusion Voa1-orf2p.

Les expériences de purification ont été répétées en utilisant, à la place de l'étiquette 6HIS, une étiquette Strep-tag®II qui a une très grande affinité pour la Strep-Tactin®. Ces constructions ont été exprimées à partir d'un promoteur PGK sur un plasmide pFL44L. Là encore, aucune protéine autre que Voa1p n'a pu être détectée par western blot à partir d'extraits cellulaires bruts ou purifiés, en utilisant un anticorps anti HA ou anti Strep-tag®II (figure 7).

J'ai ensuite voulu déterminer avec précision où pouvait se produire l'événement de recodage.

Figure 8 : Séquences des constructions dérivées d'YGR106

La séquence de chaque construction est alignée par rapport à la séquence de référence YGR106. Les constructions sont réparties en 3 groupes : (1) constructions testant l'hypothèse du décalage en -1, (2) constructions délimitant la région minimale nécessaire à l'expression de l'*ORF2*, et (3) les constructions qui changent l'organisation génomique. Les mutations et délétions internes sont surlignées en rouge. Les codons stop de la phase 0 sont surlignés en vert.

2. Analyse de la région de chevauchement

Dans le but de caractériser le mécanisme à l'origine de la traduction de l'*ORF2* j'ai étudié plus en détails la région « YGR106 » analysée initialement dans le vecteur double rapporteur. Pour cela, j'ai réalisé des délétions et des mutations dans la séquence et j'ai mesuré l'efficacité de recodage avec le système double rapporteur.

C. Les constructions

Toutes les séquences des constructions sont indiquées dans la figure 8. L'ensemble de ces constructions peut se répartir dans les 3 groupes suivants.

Les séquences du 1er groupe ont pour but de tester l'hypothèse d'un décalage du cadre de lecture en -1. Ce groupe est constitué de séquences qui possèdent des mutations dans les séquences potentielles de glissement (Yss1 à 4, Yvoa). Ces mutations affectent la troisième base d'un codon situé dans la phase de lecture de *VOA1* (phase 0). Dans les constructions Yss1, Yss2 et Yss4, les mutations silencieuses ont été réalisées de manière à ce que le codon ne soit plus reconnu par le même ARNt, tout en conservant le même acide aminé. En revanche, pour la séquence Yss3, il ne s'agit pas d'une mutation silencieuse. En effet, seuls 2 codons, AAC et AAU, correspondent à l'acide aminé asparagine et tous les 2 sont reconnus par le même ARNt. C'est donc le codon AAG qui a été introduit afin de changer la nature de l'ARNt décodant cette position. Le groupe contient aussi 2 constructions, yibv1 et Yibv2, dans lesquelles chaque site potentiel de glissement a été remplacé par la séquence de décalage en -1 (**TTTAAAC**) du virus IBV (Inglis, Rolley and Brierley 1990). Pour la dernière construction (Ysibv), c'est la séquence en aval des 2 sites potentiels de glissement qui a été remplacée par celle du pseudonoeud du virus IBV qui stimule le décalage en -1.

Les séquences du 2nd groupe ont pour but de délimiter la région d'intérêt grâce à des délétions de la région 5' et/ou 3'. Il s'agit des constructions Ys, Y5, Y3, Y3-z, Y3-y, Y3-d et Y3-a. Ce groupe contient aussi la construction Y3del1 dans laquelle une partie de la région propre à l'ORFB a été délétée.

Enfin, le 3ème et dernier groupe comprend les constructions qui modifient l'organisation génomique de l'événement. Dans la construction Yif, 2 adénines ont été délétées dans la région contenant les 2 séquences potentielles de glissement. *VOA1* et l'*ORF2* sont alors dans la même phase de lecture. Dans la construction Yof, seule une adénine a été délétée. *VOA1* et l'*ORF2* sont dans une organisation génomique en +1 et le codon stop de VOA1 est décalé de 23 nucléotides en aval. Dans les constructions Ycaa et Ygaa, le codon stop de *VOA1* est muté en codon sens pour tester si la terminaison de la traduction était nécessaire à la traduction de l'*ORF2*. Dans le cas d'Ygaa, la mutation du codon TAA en GAA introduit la présence d'un codon stop TGA dans la phase de lecture -1. Dans ces 2 constructions, l'organisation génomique en -1 est conservée mais le codon stop de *VOA1* est décalé de 75 nucléotides en aval.

D. Efficacité de recodage

J'ai quantifié l'efficacité de recodage de ces séquences dans le système double rapporteur. Les résultats sont présentés dans le graphique figure 9.

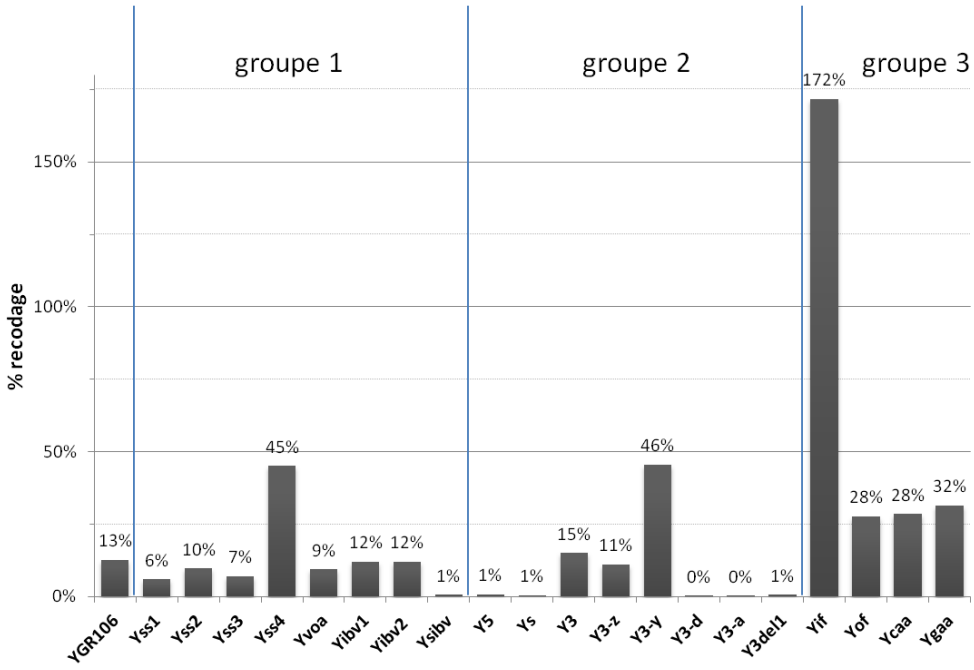


Figure 9 : Efficacité de recodage des constructions dérivées d'YGR106

L'efficacité de recodage de chaque construction est mesurée grâce au vecteur double rapporteur pACPGK007. Les séquences sont clonées entre les phases ouvertes de lecture de la β-galactosidase et de la luciférase.

Dans le groupe 1, on constate qu'il y a toujours de l'activité de recodage pour les constructions Yss1 à 4, Yvoa, Yibv1 et 2. Dans le cas d'Yss4, l'activité est ~3,5 fois supérieure à celle d'YGR106. Pour Ysibv, l'efficacité de recodage est drastiquement diminuée, elle tombe à 1%.
Dans le groupe 2, l'activité de recodage est nulle ou faible (1%) pour les constructions qui ont des délétions dans la partie 3' en aval du codon stop de *VOA1* (Y5, Ys, Y3-d, Y3-a et Y3del1). Pour les constructions tronquées en 5' du codon stop de *VOA1*, l'efficacité de recodage est conservée. Pour les constructions Y3 et Y3-z, l'efficacité est la même que pour YGR106 alors que l'efficacité de Y3-y est ~3,5 fois supérieure.
Dans le groupe 3, les activités de recodage de Yof, Ycaa et Ygaa sont de 2 à 2,5 fois supérieures à celles d'YGR106. Pour Yif, qui est une construction où la β-galactosidase et la luciférase sont en phase, l'efficacité de recodage est de 172%. Nous reviendrons sur cette valeur un peu plus tard.

L'ensemble des données indique que des mutations ou des délétions dans la région en 5' du stop de VOA1 n'abolissent pas l'activité luciférase. Sa traduction ne dépend donc pas de cette région, on peut donc exclure que sa traduction soit due à un décalage du cadre de lecture. De plus, le fait d'exprimer de la luciférase à partir de la construction Yof indique que l'expression de la luciférase n'est pas dépendante de la phase de lecture. Toutes les mutations ou délétions dans la région en 3' du codon stop de *VOA1* diminuent drastiquement l'activité de recodage. Cela indique donc que cette région est essentielle pour obtenir de l'expression de la luciférase.

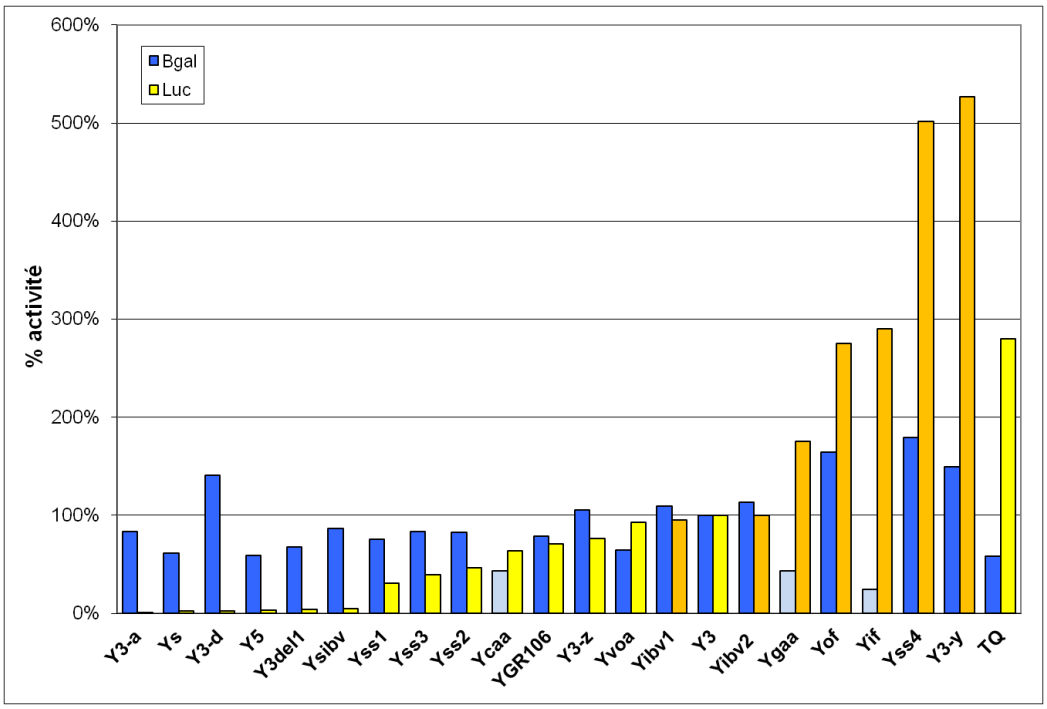
La valeur du rapport β-galactosidase/luciférase (β-gal/lucassociée à la construction Yif peut paraître surprenante car cette construction est équivalente au témoin en phase (TQ) utilisé pour normaliser les données. Ces constructions, Yif et TQ, ne varient que par la séquence qui est insérée entre la β-galactosidase et la luciférase, elles devraient donc avoir le même niveau d'expression pour les 2 enzymes. D'après cette hypothèse, la valeur du rapport β-gal/luc de Yif devrait être de 100% et non pas de 172%. La différence observée pourrait être due à 2 facteurs : soit l'activité β-galactosidase est moins importante dans le Yif que dans le TQ, ce qui conduirait à augmenter artificiellement la valeur du rapport β-gal/luc, soit l'activité luciférase est plus importante que celle du TQ. Dans les 2 hypothèses, la cause du changement peut être déterminée en regardant indépendamment les valeurs d'activité enzymatique de la β-galactosidase et de la luciférase.

E. Expression de la β-galactosidase et de la luciférase

❖ *Comparaison des activités enzymatiques*

Les activités enzymatiques de la β-galactosidase et de la luciférase pour chacun des candidats sont représentées dans le graphique figure 10A. Chaque barre correspond à la médiane des valeurs d'activité enzymatique qui a été normalisée par rapport à la valeur médiane de la construction Y3. Cette construction a été choisie comme référence car c'est la séquence minimale qui permet d'avoir un rapport β-gal/luc similaire à celui de la construction YGR106 d'origine. Les constructions ont été classées par ordre croissant de l'activité enzymatique de la luciférase.

Dans le cas des constructions TQ et Yif, on constate que les valeurs d'activité luciférase ne sont pas significativement différentes (p=0,5401) alors que pour la β-galactosidase l'activité est environ 2 fois moins importante pour Yif (p=0,0033). Cet écart d'activité montre que la différence d'efficacité de



test Mann-Whitney

	TQ vs Yif
β-galactosidase	0,0033
Luciférase	0,5401

test Kruskal -Wallis

	β-galactosidase		Luciférase
	vs TQ	vs Y3	vs Y3-y
YGR106	1,0000	1,0000	0,1407
Yss1	1,0000	1,0000	0,0114
Yss2	1,0000	1,0000	0,0186
Yss3	1,0000	1,0000	0,0298
Yss4	0,0012	1,0000	1,0000
Yvoa	1,0000	1,0000	0,2472
Yibv1	0,2369	1,0000	1,0000
Yibv2	0,3134	1,0000	1,0000
Ysibv	1,0000	1,0000	< 0,0001
Y5	1,0000	0,2521	< 0,0001
Ys	1,0000	0,1895	< 0,0001
Y3	< 0,0001	-	0,2844
Y3-z	1,0000	1,0000	1,0000
Y3-y	0.0569	1,0000	-
Y3-d	1,0000	1,0000	< 0,0001
Y3-a	1,0000	1,0000	< 0,0001
Y3del1	1,0000	1,0000	< 0,0001
Yif	1,0000	< 0,0001	1,0000
Yof	< 0,0001	1,0000	1,0000
Ycaa	1,0000	0,0028	0,2577
Ygaa	1,0000	0,0026	1,0000
TQ	-	< 0,0001	1,0000

Figure 10 : Activités β-galactosidase et luciférase des différentes constructions dérivées d'YGR106

Le graphique indique la valeur médiane des activités β-galactosidase et luciférase de chaque construction. Les données sont normalisées par rapport à celle de la construction Y3. Les valeurs d'activité des constructions TQ et Yif ont été comparé par le test non-paramétrique Mann-Whitney. Les valeurs d'activités de chaque candidat ont été comparées simultanément avec un test statistique non-paramétrique Kruskal-Wallis. Le tableau indique les valeurs de p value obtenues pour chaque construction lorsqu'elle est comparée à une construction de référence (TQ et Y3 pour la β-galactosidase, Y3-y pour la luciférase). Les valeurs de p value < 0,05 sont indiquées en rouge. Les valeurs en vert indiquent les constructions pour lesquelles les valeurs d'activité luciférase ne sont pas statistiquement différentes de celle d'Y3-y.

recodage est bien due à une baisse de l'activité β-galactosidase. La raison de cette baisse reste à élucidér.

En regardant l'ensemble des activités β-galactosidase, on s'aperçoit que l'activité n'est pas constante pour les différentes constructions. L'ensemble des valeurs d'activité β-galactosidase de chaque construction a été soumis au test statistique non-paramétrique Kruskal-Wallis (figure 10B). La comparaison simultanée de toutes les constructions fait apparaître que pour une p-value inférieure à 5 %. Les activités des constructions, Ycaa ($p = 0,0028$), Ygaa ($p = 0,0026$) et Yif ($p < 0,0001$) sont significativement inférieures à celles d'Y3.

En comparant les activités β-galactosidase et luciférase sur l'ensemble des candidats, on constate que ces valeurs ne semblent pas liées. Par exemple, dans les constructions Ygaa et Yif, l'activité β-galactosidase est faible alors que l'activité luciférase est importante. A l'inverse, dans le cas Ycaa, les 2 activités sont faibles.

- ***Détections des protéines β-galactosidase et luciférase***

Comme les activités ne semblent pas liées l'une à l'autre, j'ai décidé de réaliser un western blot contre les protéines β-galactosidase et luciférase afin de déterminer si les différences d'activité enzymatique sont liées à une différence dans la quantité d'enzymes ou à une différence d'activité spécifique. L'expérience a été effectuée sur les candidats pour lesquels il y a la plus grande différence entre les 2 activités enzymatiques, à savoir : Yif, Yof, Ycaa, Ygaa, Yss4 et Y3-y. Les résultats sont présentés dans la figure 11. Les tailles attendues des protéines sont de 180 kDa dans le cas d'une fusion traductionnelle, de 117 kDa pour la β-galactosidase et de 61 kDa pour la luciférase seule.

Pour la construction contrôle TQ, les anticorps anti-β-galactosidase détectent une bande faible comprise entre 110 et 130 kDa alors que les anticorps anti-luciférase permettent de détecter une bande au-delà de 170 kDa. Cette observation montre que les protéines se retrouvent sous 2 formes : soit une fusion traductionnelle, soit sous la forme de 2 protéines distinctes.

Pour les autres constructions, aucune protéine de fusion n'est détectée. Par contre, les protéines β-galactosidase et luciférase sont détectées séparément.
Le signal correspondant à la β-galactosidase est facilement détectable pour les constructions Y3, Yof, Yss4 et Y3-y. Pour les autres constructions, il est difficile de distinguer une bande. Pour chaque construction, l'abondance de protéines corrèle avec les valeurs d'activité enzymatique mesurées.
Dans le cas de la luciférase, le signal d'intérêt est situé juste au-dessus d'une bande contaminante qui est présente dans tous les échantillons. Le signal de la luciférase est surtout détectable pour les constructions Yss4 et Y3-y. Là encore, l'abondance de signal est en corrélation avec les valeurs d'activités luciférase mesurées.

Ces expériences indiquent que les écarts relevés entre les mesures d'activité enzymatique sont dus à des quantités d'enzymes différentes. D'autre part, aucune fusion traductionnelle n'a été détectée, pas même pour les constructions YSS4 et Y3-y. Cette observation suggère que les 2 enzymes sont soient séparées post-traductionnellement, soient traduites indépendamment. Dans ce dernier cas, leurs activités enzymatiques reflètent 2 choses différentes : la β-galactosidase caractériserait la traduction de *VOA1* alors que la luciférase caractériserait la traduction de l'*ORF2*.

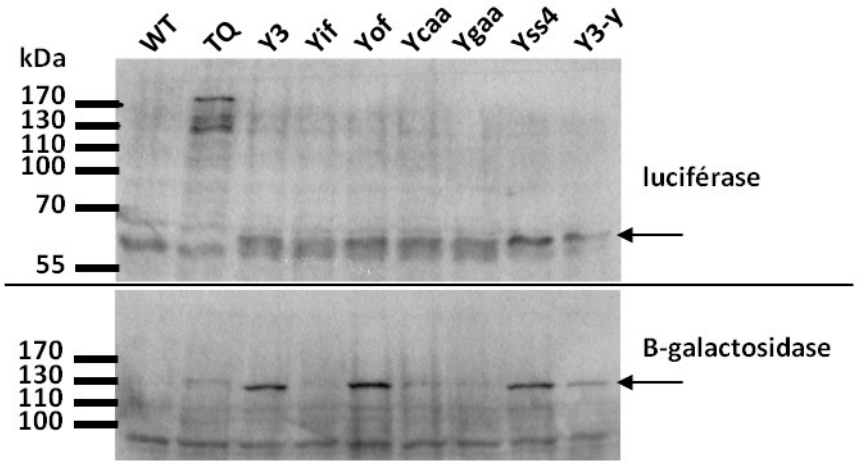


Figure 11 : Détection des protéines β-galactosidase et luciférase

Western blot à partir des constructions TQ, Y3, Yif, Yof, Ycaa, Ygaa, Yss4 et Y3-y insérées dans le vecteur double rapporteur.

Dans cette perspective où les 2 protéines sont indépendantes, les mesures des activités enzymatiques doivent être réinterprétées en répondant à 2 questions :

(i) Pourquoi la quantité de β-galactosidase n'est pas constante alors qu'elle est traduite de manière constitutive ?
(ii) Comment est traduite la luciférase puisqu'il n'y pas de codon AUG dans l'*ORF2* ?

❖ *Hypothèses sur l'expression de la β-galactosidase et de la luciférase*

β-galactosidase : site de polyadénylation et NSD

Pour commencer, je vais traiter la question de l'expression de la β-galactosidase.

La traduction de β-galactosidase est fortement diminuée pour 3 constructions : Yif, Ycaa, Ygaa. Ces 3 constructions ont pour point commun de ne pas avoir le codon stop de *VOA1* (figure 12). Dans le cas d'Yif, le codon stop est celui présent à la fin de la luciférase. Dans les cas d'Ycaa et d'Ygaa, le codon est décalé de 78 nucléotides en aval de sa position initiale. Une analyse de la séquence en aval du codon stop de *VOA1* révèle qu'il y aurait probablement un site de polyadénylation dans l'*ORF2*. Ce site serait composé d'une cassette « d'efficacité » UAUAUA située une dizaine de nucléotides en amont d'une cassette de « positionnement » (ou « reconnaissance ») **AAAGAAA** riche en adénine (Graber, McAllister and Smith 2002) (figure 12). Le site de clivage serait 20 à 30 nucléotides en aval de cette cassette.

Si cette hypothèse est avérée, l'ARNm de la β-galactosidase se terminerait dans l'*ORF2* en aval du codon stop de *VOA1*. Les ARNm issus des constructions Yif, Ycaa et Ygaa seraient alors, d'après la position du site de coupure, dépourvus de codon stop (figure 12). Lorsque de tels ARNm sont traduits, les ribosomes finissent leur traduction dans la queue poly A. Ils sont alors reconnus par la voie NSD (Non Stop Decay) qui va dégrader l'ARNm et les peptides synthétisés. Ce scénario expliquerait donc la faible quantité de β-galactosidase. Dans la mesure où un peu d'activité β-galactosidase est tout de même détectée, on peut penser que le site de polyadénylation n'est pas efficace à 100%. Les ARNm issus de cette « fuite » permettraient de traduire une faible quantité de β-galactosidases fonctionnelles et d'expliquer la faible activité détectée.

Une seule autre construction possède un stop différent de celui de *VOA1*. Il s'agit de la construction Yof. Dans ce cas, le codon stop est chevauchant avec la cassette potentielle de « positionnement » du site de polyadénylation. Les ARNm issus de cette construction possèderaient toujours un codon stop et ne seraient donc pas soumis à la voie NSD ce qui expliquerait l'importance de l'activité β-galactosidase mesurée à partir de ce plasmide.

Pour confirmer cette hypothèse, les mesures d'activité enzymatique ont été réalisées dans des souches délétées pour SKI7 ou SKI2. Les protéines Ski7p et Ski2p sont 2 facteurs qui sont nécessaires pour la dégradation des ARN par l'exosome, ils sont impliqués dans la voie NSD. Dans ces souches, l'activité β-galactosidase des constructions Yif, Ycaa et Ygaa est stabilisée. Ce résultat confirme que les ARNm β-galactosidase de ces constructions sont soumis à la voie NSD. Cependant, l'hypothèse du site de polyadénylation nécessite d'être confirmée par d'autres expériences. De plus, des expériences de northern blot renseigneraient sur la taille et l'abondance des ARNm. Pour cela, il faudrait comparer avec une construction témoin comme Y3, pour avoir la taille de l'ARNm β-

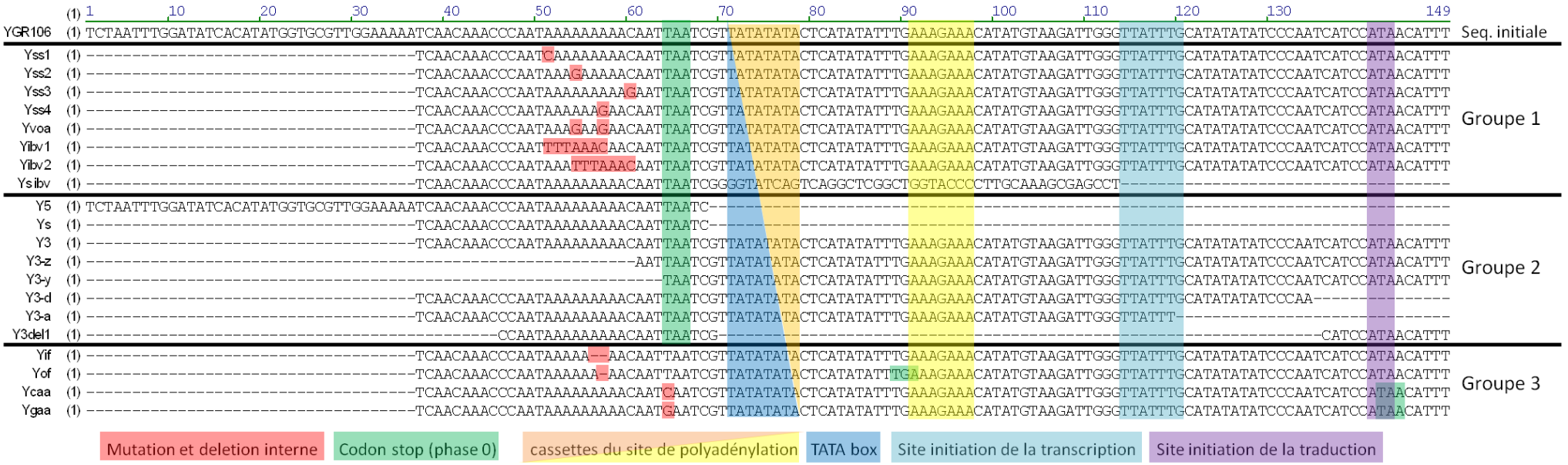


Figure 12 : Site de polyadénylation et promoteur dans l'*ORF2*

Comme pour la figure 8, la séquence de chaque construction est alignée sur la séquence YGR106 de référence.

Dans l'hypothèse d'un site de polyadénylation dans l'ORF2, les séquences des cassettes « d'efficacité » et de « positionnement » sont respectivement surlignées en orange et jaune.

Dans l'hypothèse d'un promoteur indépendant pour l'expression de l'ORF2, les séquences de la TATA box et du site d'initiation de la transcription sont surlignées respectivement en bleu foncé et bleu clair. Le potentiel codon d'initiation de la traduction est surligné en violet.

galactosidase, et les constructions Yif, Ycaa, Ygaa dans des cellules sauvages ou mutées pour la voie NSD. Si ces hypothèses sont avérées, alors les ARNm β-galactosidase issus de ces dernières constructions doivent être moins abondants que ceux d'Y3 dans des cellules sauvages. Leurs quantités devraient augmenter dans des mutants de la voie NSD.

Luciférase : fuite du site de polyadénylation ou promoteur indépendant ?

Si l'hypothèse du site de polyadénylation est vérifiée, alors il en découle directement une question sur l'origine transcriptionnelle de l'ARNm luciférase. Une première explication serait que le site de polyadénylation n'est pas très efficace. Il y aurait alors 2 ARNm, un court correspondant à la β-galactosidase et un long portant la β-galactosidase et la luciférase. Si ce n'est pas le cas, alors la luciférase doit être transcrite à partir d'une région promotrice différente de celle du promoteur *PGK1*. En parallèle des expériences précédentes, les expériences de Northern blot devront être aussi réalisées avec une sonde spécifique de la luciférase. La taille de l'ARNm renseignera sur la validité de l'une ou l'autre des hypothèses.

Dans le cas d'une traduction de la luciférase à partir d'un ARNm contenant la β-galactosidase et la luciférase, 2 mécanismes permettraient d'expliquer l'initiation de sa traduction : par une réinitiation de la traduction ou une initiation interne via une IRES (Internal ribosome Entry Site). Dans les 2 cas, l'initiation devra se produire sur un codon non-AUG, puisqu'il n'y a aucun codon AUG permettant la synthèse d'une luciférase active.
Dans le cas de la réinitiation traductionnelle, il est difficile de prédire un site d'initiation précis d'après les différentes constructions car différentes délétions abolissent l'activité luciférase. Il est peu probable que différents codons non-AUG servent de site d'initiation de la traduction.
Dans le cas d'une IRES, la région d'intérêt comprendrait toute la séquence à partir du codon stop de *VOA1*. Cette hypothèse implique que cette séquence assure une initiation de la traduction très efficace pour synthétiser autant de luciférase qu'une construction en phase comme le TQ.

Luciférase : promoteur indépendant

Suite à l'observation des activités luciférase des différentes constructions, il apparait clairement que la région en aval du codon stop de *VOA1* est essentielle pour sa traduction puisqu'il n'y a pas d'activité luciférase dans les mutants affectant cette région (Y3-a, Ys, Y3-d, Y5, Y3del1 et Ysibs).

D'après la forte activité luciférase observée avec la construction Y3-y, on peut déduire que toute la région en amont du codon stop de *VOA1* n'est pas nécessaire à l'expression de la luciférase. Cependant, si l'on compare l'activité luciférase associée à Y3-y à l'activité luciférase d'autres constructions, on s'aperçoit que la présence de cette région diminue significativement l'activité luciférase (figure 12). Malgré tout, la présence de délétions ou de mutations dans cette région permet aux constructions Yss4, Yof, Yif et Ygaa d'avoir une activité identique à celle d'Y3-y.

Dans l'hypothèse où la luciférase serait transcrite à partir d'un autre promoteur, il est plus que probable que celui-ci soit localisé dans le fragment qui est inséré dans le vecteur double rapporteur. D'après l'activité luciférase de la construction Y3-y, on en déduit que le site d'initiation de la transcription serait en aval du stop de *VOA1*. Pour que la luciférase soit traduite, il est nécessaire qu'il y ait un site d'initiation de la transcription et en aval un site d'initiation de la traduction. La comparaison des constructions Y3-d et Y3del1 pourrait dans ce cas apporter un élément de réponse.

Dans la construction Y3-d, les 15 derniers nucléotides de la construction ont été délétés. A l'inverse, dans la construction Y3del1, les 14 derniers nucléotides sont présents mais c'est la région comprise entre le codon en aval du stop de *VOA1* et les derniers 14 nucléotides qui a été délétée. D'après ces 2 constructions, on peut faire l'hypothèse que le site d'initiation de la transcription est compris dans la région délétée d'Y3del1 et que le site d'initiation de la traduction serait 1 des 5 codons manquant dans la construction Y3-d. La traduction de la luciférase s'initierait donc sur un codon non-AUG.

L'analyse de la région manquante dans Y3del1 révèle qu'il pourrait y avoir un promoteur composé d'une TATA box suivant le consensus TATAWAWR[3] et d'un site d'initiation INR proche du consensus YYANWYY (Dikstein 2011) (figure 12). Ce site potentiel d'initiation de la transcription serait compatible avec la présence d'un codon d'initiation de la traduction en aval dans les 14 derniers nucléotides. D'après la liste des codons d'initiation non-AUG permettant l'initiation de la traduction des uORF, le codon d'initiation AUA serait le plus probable car c'est le seul qui ne diffère que d'une base du codon AUG (Chang et al. 2010).

D'après les activités luciférase des constructions TQ et Y3-γ, on peut déduire que s'il y a un promoteur interne à l'*ORF2*, il doit s'agir d'un promoteur fort (autant que le promoteur *PGK1*), d'autant plus que la traduction commencerait sur un codon non-AUG, ce qui est généralement peu efficace. La présence de la région en amont du codon stop de *VOA1* semble diminuer l'activité luciférase. Ce pourrait être une région qui inhibe l'activité transcriptionnelle.

Ces hypothèses nécessitent d'être confirmées par d'autres expériences. Les expériences de northern blot permettraient de les confirmer en partie. Si la luciférase est transcrite à partir du promoteur potentiel, les ARNm devraient être absents dans le cas des constructions Y3del1, Ys, Y5, Ysibv et Y3-a, mais présents dans Y3-d. En changeant le promoteur *PGK1* du vecteur double rapporteur par un promoteur inductible, il serait possible d'éteindre la transcription de l'ARNm β-galactosidase. Si le promoteur de l'ORF2 est indépendant de la transcription en amont, il doit y avoir pour la construction Y3-γ de l'activité luciférase sans activité β-galactosidase. D'autre part, une mutagénèse dirigée des 5 derniers codons d'Y3-γ permettrait de déterminer, avec des dosages et des western blots, si l'un d'entre eux est utilisé comme codon d'initiation de la traduction.

3. Discussion

En conclusion, les résultats obtenus à partir des constructions dans le vecteur double rapporteur indiqueraient que les protéines gènes LacZ et F-luc sont transcrits et donc traduits de façon indépendante. Ces données nécessitent d'être confirmées mais elles n'excluent pas un mécanisme de régulation transcriptionnelle entre les 2 CDS.

Ce modèle d'expression conduit à s'interroger sur la traduction endogène de *VOA1* et de l'*ORF2*. Ces résultats expliqueraient le fait de ne jamais avoir détecté de protéine recodée Voa1-orf2p par western blot. Si la protéine Orf2p est traduite indépendamment, alors elle aurait une taille d'environ 10 kDa. Les gels SDS-PAGE utilisés pour réaliser les western blots n'étaient pas adaptés à la détection d'une petite protéine. Il serait donc intéressant de refaire les western blots de manière à ce qu'ils soient résolutifs pour cette gamme de poids moléculaires. L'autre alternative serait de réaliser un

[3] W = A ou T ; R = A ou G

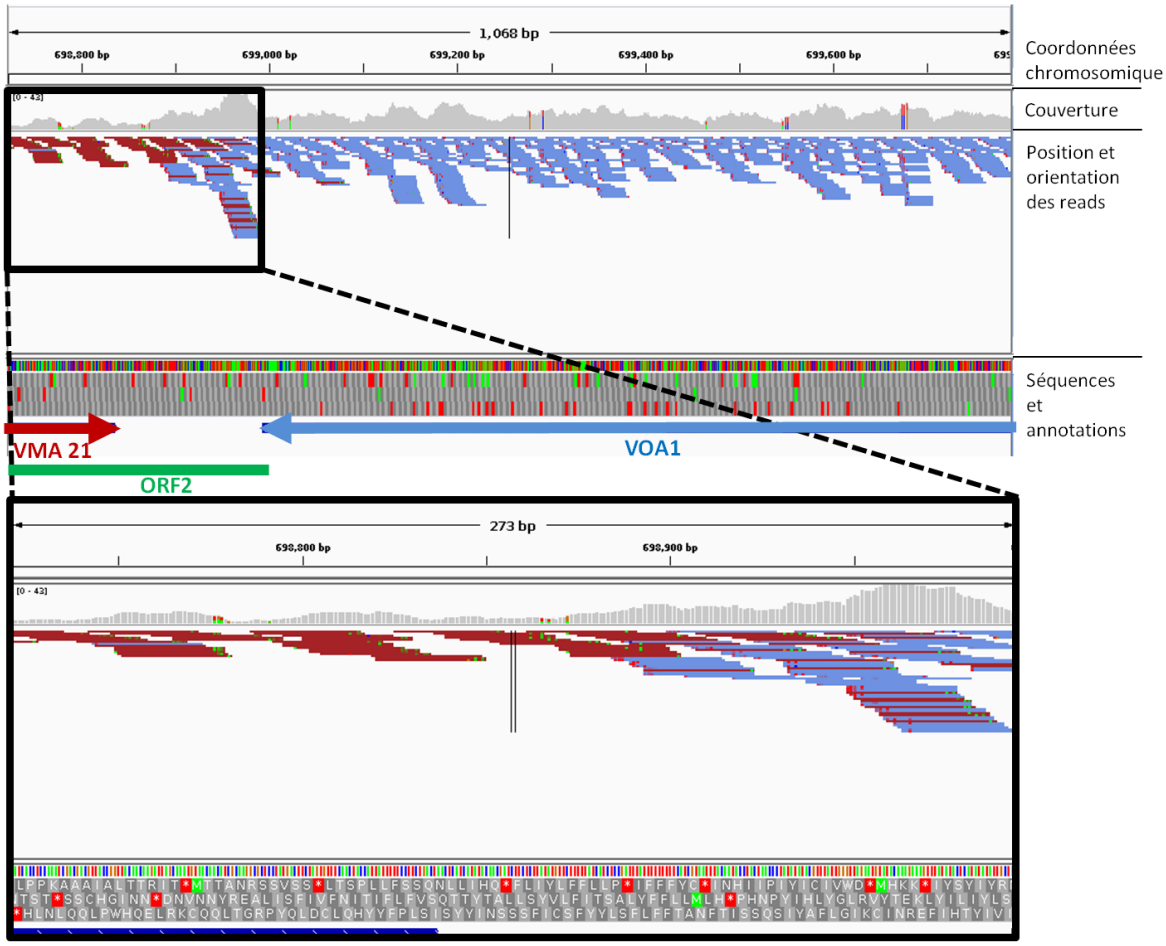


Figure 13 : Données transcriptomique sur VOA1

Représentation Genome viewer de la transcription sur la région *VOA1/ORF2* et sur l'extension *ORF2*. Chaque trait horizontal correspond à un read, les rouges s'alignent sur le brin de *VMA21* et les bleus sur s'alignent sur le brin de *VOA1*. Les données sont issues du jeu de données « BY4741 milieu riche »

western blot contre la luciférase à partir d'extraits de cellules exprimant la construction pFL44L-VOA1ORF2-Luc (cf papier).

Concernant l'origine de la transcription de l'*ORF2* au niveau endogène, la question nécessite d'être reposée. Avant que je commence ce projet, il avait été montré par RT-PCR qu'il existait des ARNm dans la cellule qui recouvraient à *VOA1* et l'*ORF2*. Cette technique est très sensible mais elle ne rend pas compte de l'abondance de ces ARNm dans la cellule. En regardant les données de transcriptome associées au profil traductionnel sur la BY4741 (cf partie II), on remarque que la couverture de reads dans la région de *VOA1* s'arrête majoritairement dans l'*ORF2* (figure 13). Le site de polyadénylation serait cependant un peu plus en aval que dans notre hypothèse. Cela est peut être dû à un contexte nucléotidique différent entre l'*ORF2* et *F-luc*. Ces données suggèrent qu'il y aurait un site de polyadénylation dans l'*ORF2*. Une faible fuite du site de polyadénylation pourrait être à l'origine d'une fraction minoritaire d'ARNm portant *VOA1* et l'*ORF2*.

Le fait de ne pas détecter de reads correspondant à des ARNm ou à de la traduction dans l'ORF2 tend à montrer que, dans les conditions de l'expérience, l'ORF2 n'était pas exprimée. Si l'hypothèse du promoteur indépendant est vérifiée, alors son activité pourrait être contrôlée par des séquences régulatrices présentes dans *VOA1* ou *VMA21*. Il est tout de même intriguant que cette région, qui semble impliquée dans la perméabilité cellulaire, soit incluse entre 2 CDS qui sont impliqués dans l'assemblage des complexes V-ATPase. Ce locus pourrait être un exemple de régulation originale de l'activité des pompes à protons. De nombreux travaux sont encore à réaliser pour répondre à l'ensemble des questions posées.

IV. Références bibliographiques

Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia & J. S. Weissman (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science,* 335**,** 552-7.

Chang, C. P., S. J. Chen, C. H. Lin, T. L. Wang & C. C. Wang (2010) A single sequence context cannot satisfy all non-AUG initiator codons in yeast. *BMC Microbiol,* 10**,** 188.

Dikstein, R. (2011) The unexpected traits associated with core promoter elements. *Transcription,* 2**,** 201-6.

Graber, J. H., G. D. McAllister & T. F. Smith (2002) Probabilistic prediction of Saccharomyces cerevisiae mRNA 3'-processing sites. *Nucleic Acids Res,* 30**,** 1851-8.

Inglis, S. C., N. Rolley & I. Brierley (1990) A ribosomal frameshift signal in the polymerase-encoding region of the IBV genome. *Adv Exp Med Biol,* 276**,** 269-73.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman & J. S. Weissman (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science,* 324**,** 218-23.

Ryan, M., L. A. Graham & T. H. Stevens (2008) Voa1p functions in V-ATPase assembly in the yeast endoplasmic reticulum. *Mol Biol Cell,* 19**,** 5131-42.

Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, E. Guffanti, F. Stutz, W. Huber & L. M. Steinmetz (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature,* 457**,** 1033-7.

Conclusions générales et perspectives.

Lors de mes 4 années de thèse, j'ai étudié les erreurs programmées de la traduction aussi appelées recodage. Ce terme regroupe différents mécanismes comme la translecture et le décalage du cadre de lecture en -1 ou en +1. Le recodage permet d'augmenter la diversité protéique en permettant la traduction de 2 protéines à partir du même ARNm et de réguler le ratio entre ces 2 protéines. L'étude de ces événements permet de comprendre les mécanismes moléculaires qui modifient le comportement du ribosome afin de lui faire faire une erreur. Par extension, ce sont donc des outils qui permettent d'appréhender la fidélité de la traduction de manière générale.

Au cours de ce travail, j'ai étudié la fidélité de la traduction chez les eucaryotes d'un point de vue mécanistique et fonctionnel.

Tout d'abord, j'ai tenté d'étudier le rôle d'eEF2 dans le maintien du cadre de lecture grâce à un crible génétique. Pour cela, j'ai développé une stratégie qui associe une mutagénèse aléatoire du gène *EFT2* à un criblage phénotypique permettant d'isoler des mutants capables d'augmenter ou de diminuer l'efficacité de recodage d'une séquence de décalage du cadre de lecture en -1. Des écueils méthodologiques ont compliqué la construction des outils génétiques permettant de répondre à cette question. Maintenant que ces outils sont construits, ils vont être utilisés pour réaliser ce projet et identifier des mutants qui permettront de mieux comprendre comment eEF2 contribue au mécanisme du décalage du cadre de lecture en -1.

Dans un second projet, j'ai mis en place, en collaboration avec l'équipe de biophysique de Nathalie Westbrook, des outils qui permettent l'étude de l'élongation de la traduction eucaryote en molécule unique. Afin de contourner les difficultés liées à l'initiation dépendante de la coiffe, j'ai développé un système de traduction basé sur un ARNm qui possède un site d'entrée interne pour le ribosome. Il s'agit de l'IRES du virus CrPV qui a comme caractéristique d'être totalement indépendante des nombreux facteurs d'initiation et de l'ARNt initiateur. Les facteurs de traduction sont apportés par un lysat cellulaire de réticulocyte de lapin. L'élongation de la traduction est détectée grâce au décrochement d'un oligonucléotide fluorescent par l'activité hélicase du ribosome. Nos expériences constituent une preuve de principe démontrant que l'étude de la traduction eucaryote en molécule unique est possible. L'optimisation du blanchiment permettra dans un premier temps de mesurer des cinétiques de traduction, par exemple le passage d'un codon stop lors de la translecture. A plus long terme, le marquage spécifique de différentes protéines permettra de concevoir des expériences étudiant l'interaction entre différents facteurs. Par exemple, il sera possible d'étudier le comportement de mutants d'eEF2 et des ARNt dans un ribosome lors d'un événement de décalage du cadre de lecture en -1.

D'autre part, de nombreuses études mettent en évidence le fait que les ribosomes ont une composition hétérogène en modifications post-transcriptionnelles des ARNr, en protéines et en modifications post-traductionnelles de ces protéines. Ces observations ont conduit à l'émergence du concept de « ribosomes spécialisés ». Ces variations ajouteraient un niveau de régulation supplémentaire à la traduction, en favorisant la traduction d'un sous-groupe d'ARNm en fonction d'un type cellulaire ou d'une étape d'un cycle de développement (Xue and Barna 2012). L'une des questions sous-jacentes est de déterminer s'il existe des ribosomes spécialisés dans le recodage. Dans la perspective où ces sous-groupes de ribosomes pourraient être marqués spécifiquement, les

approches en molécules uniques permettraient de déterminer si certaines de ces variations modifient la fidélité et la vitesse de traduction.

Le deuxième axe majeur de ma thèse a été d'étudier le rôle physiologique du recodage, par une approche génomique. En collaboration avec l'équipe de bioinformatique d'Alain Denise, nous avons développé une approche de génomique comparative dans le but d'identifier des événements de recodage, ainsi que d'autres événements non-conventionnels de la traduction, dans des génomes de levures. Pour cela, nous avons recherché des organisations génomiques compatibles avec des événements de recodage dans 19 génomes de levures et les avons inclus dans une base de données appelée CEV. Nous nous sommes intéressés plus particulièrement au génome de *Saccharomyces cerevisiae.* Seuls les événements conservés au niveau protéique dans au moins un autre génome de levure ont été conservés. Des filtres ont été appliqués sur la liste des candidats pour réduire leur nombre et analyser les plus pertinents. Ces derniers ont été analysés in vivo grâce à un vecteur double rapporteur, ce qui a permis de mettre en évidence le gène *VOA1*. Un criblage phénotypique haut-débit a permis de montrer que la CDS chevauchant *VOA1* est impliquée dans la sensibilité à la blasticidine S, qui est un antibiotique bloquant le fonctionnement du ribosome.

Les perspectives à court terme sont de continuer l'exploitation de la base de données CEV. L'analyse de candidats des autres génomes aiderait à l'annotation de ces derniers. De plus, ces études permettraient de mieux évaluer la prévalence du recodage dans les génomes et son impact sur la régulation de la physiologie cellulaire. L'étude de ces événements devrait aussi permettre d'améliorer leur annotation dans les bases de données qui reste encore actuellement très problématique.

Malgré l'amélioration constante de la sensibilité des analyses de transcriptomique, de traductomique et de protéomique, les approches bioinformatiques de ce type trouveront toujours leur place pour l'annotation de génomes peu étudiés et qui ne profiteront pas de ces analyses d'expression. Sur le long terme, il serait utile de développer des programmes permettant d'intégrer et d'analyser à la fois toutes les données d'expressions et de génomiques, ainsi que les différentes prédictions bioinformatiques. L'intérêt d'un tel logiciel serait de rendre l'analyse des études à grande échelle accessible à tous les laboratoires qui n'aurait besoin que d'un accès ponctuel à ces données et qui n'aurait pas de bioinformaticien pour les analyser.

L'étude du gène candidat *VOA1* et de son *ORF2* aura été un défi. Toutes les informations sur son apparence (l'organisation génomique, l'absence de codon ATG dans l'*ORF2* et les motifs consensus) plaidaient en faveur d'une régulation par un mécanisme du décalage du cadre de lecture en -1. Les différentes constructions réalisées dans le vecteur double rapporteur ont mis en évidence le fait que ce n'était pas le cas. Nous proposons maintenant que *VOA1* et l'*ORF2* soient transcrits et traduits de manière indépendante. Des expériences sont actuellement en cours dans l'équipe afin de vérifier cette hypothèse. L'organisation de cette région du chromosome est originale car l'ORF2 est chevauchante avec 2 gènes, *VOA1* et *VMA21*, impliqués dans l'assemblage des complexes V-ATPase. Il serait donc intéressant de déterminer le rôle biologique associé à l'expression de l'*ORF2*. Si son expression est aussi impliquée dans l'assemblage des complexes V-ATPase, alors nous serions en présence d'une organisation génomique originale permettant une régulation transcriptionnelle de ces gènes.

Cette recherche de gènes régulés par des événements non conventionnels pose la question générale de la taille minimale nécessaire pour modifier la fonction, ou la localisation d'une protéine. L'étude sur les protéines GAPDH et Pgk1p dans plusieurs génomes de levures montre que non. La localisation différentielle de ces 2 protéines dans le cytoplasme ou dans les peroxysomes est régulée soit par épissage alternatif soir par translecture conduisant à l'ajout d'une courte séquence peptidique (Freitag et al. 2012). Le masquage ou l'ajout d'une courte séquence d'adressage peut donc changer la stabilité, la localisation d'une protéine. Il serait donc intéressant de poursuivre le travail d'analyse des génomes en recherchant spécifiquement des organisations génomiques compatibles avec ce type d'événement. L'analyse sera à coup sûr délicate car ces types d'organisation sont nombreux dans un génome. De plus la petite taille des motifs impliqués rends difficile leur identification par une approche bioinformatique. Cependant en recherchant spécifiquement des motifs protéiques et en croisant la liste des candidats avec des données de profils traductionnels, il devrait être possible d'identifier ce genre d'événements.

Freitag J, Ast J, Bölker M. 2012. Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. *Nature* **485**(7399): 522-525.

Xue S, Barna M. 2012. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol* **13**(6): 355-369.

Annexe

Article

Visualisation of +1 frameshifting during tRNA accommodation.

Soumis à la revue "*Nature Structural & Molecular Biology*"

Visualisation of +1 frameshifting during tRNA accommodation

Matthieu Saguy[1,6], C. Alistair Siebert[2,6], Hélène Chommy[1], David Cornu[4], Olivier Namy[1,3,5] and Robert J C Gilbert[2,5]

[1] Université Paris-Sud, Institut de Génétique et Microbiologie, UMR8621, 91400 Orsay

[3] CNRS, 91400 Orsay

4 Institut de Chimie des Substances Naturelles, Centre de Recherche de Gif, CNRS, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette.

[2]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Univeristy of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

[5]Correspondence: olivier.namy@igmors.u-psub.fr and gilbert@strubi.ox.ac.uk

[6] These authors contributed equally to this paper.

Abstract

Ribosomes are regulated by secondary structures formed within mRNAs they translate. Such stem-loops and pseudoknots trigger, among other effects, -1 and +1 frameshifts. Here we describe the structure of a +1 frameshifting ribosome on the mammalian ornithine decarboxylase antizyme 1 mRNA. Two different ribosomal states are observed: an unratcheted state with the translation elongation factor eEF1a bound and a ratcheted state with a P-site tRNA bound (pre- and post-frameshift, respectively). We confirm the identity of the bound eEF1a using mass spectrometry and from the unratcheted state of the ribosome with which it is engaged conclude that +1 frameshifting occurs by the out-of-frame engagement of a tRNA at the A-site. This demonstrates that +1 frameshifting has evolved at the accommodation stage of translation elongation whereas we previously showed that -1 frameshifting occurs at the translocation stage.

Introduction

Translational accuracy at the ribosome is essential to ensure the expression of genetic information as protein structure and major advances have been made towards an atomic resolution understanding of this process in recent years [1]. From the initiating start codon, the sequence of amino acids is determined by consecutive triplet codons of nucleic acid bases on the messenger RNA (mRNA) read by complementary anticodon sequences on amino acid-bearing transfer RNAs (tRNAs). X-ray crystallography [2,3], cryo-electron microscopy (cryo-EM) [4-7] and single molecule fluorescence [8,9] and optical trap [10,11] approaches to understanding this process in mechanical terms have proved particularly insightful in recent years. Besides the normal translocational process, a variety of recoding events have also evolved which give added diversity and plasticity to the meaning of genetic sequences [12]. For example, a triggered shift out of the triplet reading frame has evolved -1 (towards the 5' end of the mRNA) or +1 (towards the 3' end), the trigger being either a stem-loop or a pseudoknot structure formed by the mRNA itself [12,13]. A large variety of cellular and viral genes' expression is regulated in this way and although pseudoknots are structures which commonly alter ribosome fidelity, -1 and +1 frameshifter pseudoknots are not interchangeable suggesting they work through different mechanisms [12]. We have previously shown that a pseudoknot stimulates the -1 programmed ribosomal frameshift (PRF) by interfering with the elongation cycle, stalling eEF2 in the ribosomal A-site [5,6,14]. Our work suggested that pseudoknot strength would correlate with frequency of frameshift and this hypothesis has been tested and confirmed [15,16]. -1 PRF was found to be a tension-dependent process in which a translocating tRNA undergoes a mechanical bending as the ribosome engages with the pseudoknot secondary structure [6]. This finding has also been found to be in agreement with single molecule studies of the ribosomal helicase unwinding secondary structures [10].

We now turn to the question of the mechanical basis of the +1 PRF used to regulate ornithine decarboxylase antizyme 1 (*OAZ1*) mRNA expression. OAZ1 is a protein controlling polyamine biosynthesis [17-19]. Despite differences in the identity of the P-site codon, the *OAZ1* stimulatory stop codon is universally conserved and the sequence 3' to it can fold to form a pseudoknot in all vertebrate orthologs of *OAZ1* [20,21]. This pseudoknot strongly stimulates frameshifting in a polyamine-independent manner [19]. +1 PRF is also stimulated by the presence of polyamines [18] which have recently been reported in a yeast system to act by binding directly to the nascent antizyme peptide, promoting the frameshift and thus completion of antizyme synthesis [17]. We programmed an *in vitro* translation reaction using the mammalian *OAZ1* and purified the frameshifting ribosomes by affinity chromatography. In addition to a pseudoknot the +1 frameshift trigger is often also enhanced by inefficient codons (a rare codon or a stop codon) in the ribosomal A site. Here we show that an mRNA pseudoknot-triggered +1 ribosomal frameshift mechanism that regulates metabolic synthesis of polyamines occurs during the tRNA accommodation stage of translation. Cryo-EM reconstructions of frameshifting and post-frameshift ribosomes reveal in the first case eEF1A engaged in out-of-frame accommodation of a tRNA into the ribosomal A-site of an unratcheted ribosome [22-24]. We confirm the identity of the bound eEF1A using mass spectrometry of the purified complexes. In the second case a P-site tRNA is bound within a tightly-ratcheted ribosome. This is the first direct observation of eEF1A-tRNA bound to a eukaryotic ribosome, obtained without additional fixation using an antibiotic, and shows that +1 frameshifting occurs during tRNA accommodation. Moreover, our results provide a mechanical explanation of the role of a pseudoknot during +1 frameshifting in a mechanism conserved throughout vertebrate species that directly links cellular information control and metabolism.

Results

Detection of a pause involved in OAZ1 +1 PRF

To gain a direct molecular insight into the role of a pseudoknot in stimulating +1 PRF we first demonstrated that the +1 pseudoknot induces a ribosomal pause as efficiently as the -1 pseudoknot in rabbit reticulocyte lysate (RRL) (Figure 1a). In this assay, the presence of stalling was monitored by [^{35}S]-methionine incorporation into a protein product, showing that the +1 PRF pseudoknot induces a pause of up to 15 minutes, with optimum paused product presence at 7.5-10 minutes. This is the first time a pause has been directly shown to be associated with a +1 PRF. We also tested the dependence of frameshifting on spermine and spermidine concentration (Figure 1b), as previously described [18], showing that the optimum spermidine concentration is 0.4 mM but that the ability of spermine to stimulate the frameshift depends directly on its concentration. On the basis of this benchmark we then purified stalled ribosomes in the process of decoding the *OAZ1* +1 frameshifting pseudoknot and performed a cryo-EM analysis of the stalled complex.

Cryo-EM reconstruction of +1 PRF ribosomes: overview

195 micrographs of the +1 PRF sample were selected as being images of good quality and signal:noise. Applying iterative reference-free image classification and alignment of ribosomes identified within them we derived a dataset of 63,372 images that represented an even sampling of ribosomal views. These we subjected to iterative reference-based alignment and reconstruction beginning with a stem-loop-stalled rabbit reticulocyte ribosome (RRL) [5] filtered to 80Å resolution to remove all possible conformation-specific model bias in our alignment and reconstruction procedure. After three iterations we had obtained a map with clear ribosomal features at a resolution of 16.7Å; calculation of a difference map at this stage with reference to an Apo ribosome reconstruction [6] indicated tRNA density in the P-site of the

ribosome. Following this we carried out subclassification of our data using a combination of multivariate statistical analysis-based and projection matching-based classification, full details of which are given in the Materials and Methods. Structural changes specifically associated with the paused ribosomes were determined by comparison with an Apo ribosome, a ribosome pausing at a -1 pseudoknot [5], and published crystal structures of the yeast 80S ribosome and of prokaryotic ribosomes bound with EF-Tu-tRNA and EF-G [2,5,22,25-27].

The +1 PRF population of ribosomes consists of two states, an elongation factor- (EF-) bound (model 1) state and a P-site tRNA-bound (model 2) state. The final reconstructions used 26,049 images in the EF-bound model 1 reconstructed at 15.6Å resolution and 37,323 images in the tRNA-bound model 2 reconstructed at 15.4Å resolution (see Supplementary Figure S1 for resolution estimation). In neither reconstruction have we been able to resolve convincingly density deriving from the pseudoknot; this may be due to its occupying variable positions, or being partly unwound and/or too closely associated with the ribosome to be distinguishable from it at the resolutions obtained. The +1 PRF *OAZ*1 pseudoknot within our programming construct is closer to the A site codon than our previously-reported pseudoknot-stalled reconstructions [5,6] and thus is very likely partially melted by the ribosome's helicase action [10,28], reducing its size and thus the likelihood it would be observed. The headline resolutions of our reconstructions are lower than those of several ribosomes published in recent years, most likely due to a combination of the relatively small size of our image dataset and the fact that our complexes have not been fixed by antibiotics but are stalled solely by the biological pseudoknot mechanism. The similar resolution of model 1 with 18% fewer images than model 2 suggests that the limited resolution is mostly a product of the sample itself. Nevertheless, the ribosomal structures are resolved with great clarity, giving an excellent basis for interpretation, with both the ligands being bound with full occupancy (Materials and methods)

(see Supplementary Movies 1-3). Overall our reconstructions reveal that the pseudoknot stalls the ribosome in the presence of an unaccommodated eEF1A-tRNA complex delivering the tRNA out-of register with the A site.

Detailed description of complexes

Orthogonal views of the two ribosomal complexes found in our +1 PRF reaction reveal the global conformational differences between them highlighting the positioning of their ligands (Figure 2). To assist in interpreting and validating our structures we fitted them with atomic models from X-ray crystallography of a representative eukaryotic ribosome, the yeast 80S complex [22]. It was noted in the paper describing these crystal structures that one of the ribosomes (copy 1 in the asymmetric unit) was in a ratcheted state and the other (copy 2) in a state correlating to some translocational intermediate but not ratcheted [22]. We found that the yeast copy 2 atomic model fitted very well to our EF-bound ribosome, while the copy 1 atomic model fitted similarly to the P-site tRNA-bound ribosome. This immediately indicates the EF-bound ribosome as in an unratcheted state and the P-site bound ribosome as in a ratcheted state. This can be further evidenced by consulting Supplementary Figure S2, in which the two states found in the +1 PRF population are shown, and Supplementary Movies 1-4 in which a morphing between the two states is shown. Moving from model 1 (EF-bound) to model 2 (tRNA-bound) involves a series of well-defined movements and breathing motions by rRNA helices and proteins. Among the most salient are those which result in a rotation of the 40S small subunit clockwise with respect to the 60S large subunit when viewed from its solvent face, and the 'nodding' of the 40S head in the same direction over a larger arc resulting in a closing of the mRNA entrance channel and an engagement with helix 18. In addition, as the 40S head rotates down towards the body a connection with the L1 stalk of the large subunit is broken and there is additional independent movement of the ribosome kinase

RACK1. Finally, the P0 stalk base beside the A-site region of the ribosome moves away from the intersubunit space; its protrusion into this space is associated with the EF-bound structure and is known to be involved in engagement of the GTPase activating centre of the ribosome with the GTPase region of elongation factors.

The diversity of conformations which the ribosome can adopt is increasingly appreciated [4] and conformational differences between the 80S crystal structures and the reconstructions of rabbit ribosomes to which they are fitted here underscores this fact. For example although there are many detailed similarities between the unratcheted and ratcheted crystal structures and models 1 and 2 respectively, in the P-site bound ribosome the L1 stalk is in an open state rather than closed against the small subunit as in the matching crystal structure [22].

Ligands found bound to the two ribosomal reconstructions

Calculating difference density maps for the two stalled complexes allows the shape of the ligands bound to be better defined by reference to an Apo structure. In Figures 3a and 3b we show close-up views of the EF and tRNA binding regions of the two reconstructions obtained from the +1 frameshifting ribosome image population (for global views see Supplementary Figures S3a-b). Superimposition with their respective difference maps clearly identifies distinct ligands as bound in the individual reconstructions. The unratcheted nature of the EF-bound ribosome structure strongly suggests that the EF bound is not eEF2, because eEF2 is only known to bind to ratcheted ribosomes, like its cognate in prokaryotes EF-G. eEF2/EF-G engagement follows after A-site tRNA accommodation, which is accompanied by the ratchet movement as the ribosome moves from a state in which P-site and E-site tRNAs are clasped tightly to one from which the E-site tRNA is released and a new, A-site tRNA is bound [1,29]. By contouring the maps at a high level we are able to show the regions which have the

greatest electron density, which will be regions correlating to RNA more than protein [30,31]. As shown in Figures 3a and 3b, the rRNA is then well highlighted and the ribosomal proteins are relatively uncovered by the map. A direct analysis of the reconstructed density shows that most of the EF is uncovered at the higher cutoff shown in Figure 3a but its tRNA region is still observable, particularly at its anticodon end, and is of comparable salience to the tRNA bound in the P-site. These two arguments clearly identify the extra-ribosomal density as an eEF1A-tRNA complex.

In order to confirm the identity of the bound elongation factor as eEF1A, we performed peptide mapping of proteins in the purified sample by mass spectrometry (see Materials and Methods). Among the non-integral ribosome proteins identified in the 50-100 kDa range were various initiation factors and their subunits and two components of eEF1A, eEF1A and eEF1G (Supplementary Figure S4; Table S1). This demonstrates that the bound EF is certainly eEF1A. In Figure 3c we show the fit of a homology model of rabbit eEF1A alongside a tRNA positioned as in the prokaryote homologue EF-Tu in its GTP-bound state (with GDPnP) [32]. The global fit is excellent, possible variability in the exact angle at which eEF1A presents its bound tRNA compared to EF-Tu notwithstanding. We show in Figure 3d a tRNA fitted into the density found within the P-site: there is an excellent agreement also between the density observed for the bound tRNA and a matching crystal structure, including the detailed envelope of the tRNA. The P-site tRNA footprint is adjacent to but not overlapping the footprint of eEF1A-tRNA, showing that they are entirely distinct ribosomal complexes (Supplementary Figure S5). The eEF1A-tRNA and P-site tRNA are stoichiometrically resolved in these two complexes: not only are they visible in the maps contoured at high density (Figures 3a and 3b) but their features are the most significant

aspects of their respective difference maps (Supplementary Figures S3c –f). This allows us to interrogate the position of eEF1A-tRNA in our complex in some detail.

Although there are no other structures of eEF1A bound to the eukaryotic ribosome, there are crystal structures of both EF-Tu-tRNA and EF-G bound to the prokaryotic ribosome. These were compared to the position of the tRNA and eEF1A in our reconstruction to investigate any unique features of our 80S-eEF1A-tRNA complex. The structural alignment was effected by fitting the small subunit atomic model from each crystal structure into the small subunit of our reconstructed density. In Supplementary Figure S6a we show a view of the unratcheted 80S crystal structure with the density we find in our reconstruction of eEF1A-tRNA, fitted with atomic models for its components as shown also in Figure 3. By showing the eEF1A-tRNA difference density at two contour levels (1σ and 1.5σ) we can underscore the expected high density of the tRNA region. In Supplementary Figure S6b we show the difference density and the EF-Tu-tRNA crystal structure [25] as fitted in the 40S region of the reconstruction: the EF-Tu-bound tRNA is more deeply inserted into its binding pocket than the eEF1A-tRNA. However, the core rRNA of the yeast 40S and prokaryotic 30S structures superimposes well suggesting no major difference in rRNA conformation in our structure compared to the prokaryotic crystal structure. In Supplementary Figure S7 we repeat this analysis for the EF-G-bound crystal structure [27] and find again that the EF-G is positioned more deeply inserted into the A-site than eEF1A-tRNA in our reconstruction. In the inset to Supplementary Figure S7 we show the EF-Tu and EF-G bound prokaryotic structures superimposed – the footprints of EF-G and of EF-Tu's tRNA are essentially identical. Thus, the tRNA in our eEF1A-tRNA-bound ribosomal complex seems to be engaged improperly at the A-site: this is exactly consistent with the structure being unratcheted. If the A-site tRNA were properly engaged the ribosome should, by all accounts, be ratcheted. The absence of a P-

site tRNA from the eEF1A-tRNA bound structure is most likely due to a combination of at least three factors: the short (8 residue) amino acid nascent chain attached to its aminoacyl arm is too small to interact with the exit channel and stabilize the tRNA [33], the near-cognate nature of the P-site tRNA probably will participate in tRNA destabilization, and the open conformation of the ribosome facilitates P-site tRNA drop-off.

Discussion

The structures we have resolved from a +1 PRF population of ribosomes consist of an eEF1A-tRNA bound complex in which the tRNA has not engaged properly but is instead held stably out of register with the A-site codon, and a ratcheted P-site tRNA-bound complex. These are pre- and post-frameshifting assemblies respectively in the +1 frameshift. Successful binding and accommodation of a tRNA into the A-site is known to result in ratcheting of the small subunit towards the A site, clasping the newly engaged tRNA [23,24]. This means that eEF1A-tRNA, whose identity we have confirmed; by density analysis, because eEF2 only binds to ratcheted ribosomes, and by mass spectrometry of the purified samples, is not properly engaged. It is, however, tightly enough bound that it is well resolved in our reconstruction. This suggests that the pseudoknot stalling the ribosome to promote the +1 PRF has induced a novel conformation of the small subunit facilitating table binding by its engagement with the solvent face helicase centre of the ribosome [10,28], a hypothesis that requires further investigation. Single-molecule studies on ribosomes have made plain the complex nature of the engagement they make with mRNAs, with variable pause times between translocation [11] and two helicase mechanisms at once – a tension-based mechanism and a mechanism sensitive to the stereochemistry of the stem junction being opened [10]. We have observed a novel effect here, the induction of an out-of-frame tRNA binding, which may be of relevance not only to +1 PRF but also to other recoding effects [12]. Our previous study of

-1 PRF [5,6,14] provided clear hypotheses and insights which the results of others have validated [10,11,15,16]. This showed the power of cryo-EM studies to frame experiments establishing the molecular bases for biological phenomena. Turning now to +1 PRF we have been able to suggest similarly the mechanical basis of a specific recoding event.

Drawing on currently-available insights we can hypothesize how the *OAZ1* pseudoknot and polyamines interplay to regulate antizyme synthesis. Ribosome stalling at the *OAZ1* mRNA pseudoknot will either result in a +1 frameshift or more likely termination at the stop codon intervening at the end of the first open reading frame. Rising polyamine concentrations will result in an increased frameshift frequency, either by the nascent peptide-mediated effect observed in yeast [17] or by decreasing stop codon recognition by translation termination complex as observed in several organisms [19,34]. The synergistic effect of the pseudoknot and polyamines will lead to more antizyme synthesis and the enhanced degradation of ODC, reducing the polyamine concentration. When polyamine levels fall again then the frequency of frameshifting will reduce and so will the synthesis of antizyme, allowing ODC levels to rise and more polyamines to be made.

In the pre-frameshifting complex we observe the mechanism by which the +1 move would occur (Figure 4): by inducing the engagement of a tRNA out of register with the cognate codon in the A-site the ribosome will avoid termination at the preceding stop codon and instead continue to synthesise antizyme. After accommodation out-of-frame, for which the pseudoknot-induced pause may be critical as it allows time for an otherwise unfavourable event to occur – the leader peptide for antizyme will undergo peptidyltransfer to the amino acid on the +1 PRFed eEF1-delivered tRNA, allowing continuation of antizyme synthesis. For future testing, this indicates that (like the -1 PRF [15,16]), the stability of a pseudoknot will

correlate directly with the frequency of +1 PRF. In our model 1 with eEF1-tRNA bound there is no P-site tRNA because it has been lost from this pre-frameshift state due to possession of only a near-cognate anticodon and of a short nascent chain of 8 residues. The alternative is that translation termination occurs *via* binding of release factor. If the +1 PRF occurs, the ribosomes will continue to translate full-length antizyme; if it does not they will proceed to termination and subunit disassembly. This explains how polyamine biosynthesis is regulated by the pseudoknot-stimulated *OAZ1* +1 PRF. The mechanism is quite distinct from -1 PRF: +1 PRF is shown to occur at tRNA accommodation, while -1 PRF occurred at tRNA translocation. Furthermore, +1 PRF is found to be based on delay in translation (the pausing shown in Figure 1) and improper binding during this time of eEF1a-tRNA; -1 PRF was found to be based on mechanical tension [6,10,15]. Primary information control by genetics and metabolic control are most usually separately studied and it is very intriguing how far-downstream metabolism can regulate gene expression. The deeply conserved *OAZ1* mechanism of +1 frameshifting provides a very simple, elegant and powerful metabolic control mechanism by which this is achieved which, having evolved, has remained largely unimproved throughout eukaryotic evolution.

Figure legends

Figure 1

a Time course of translational pausing: mRNA with either *OAZ1* pseudoknot or no PK was translated in RRL at 27°C in the presence of [^{35}S]methionine for 5 min prior to addition of edeine to a final concentration of 6μM, inhibiting translation initiation. Translation was stopped at various times and the protein concentration determined by autoradiography. M: run-off translation of the mRNA from XhoI linearized plasmid to mark the expected position of a pseudoknot-induced ribosomal pause product. The pause product is indicated by an arrow.

b The effect of spermine and spermidine concentration on the *OAZ1* frameshifting efficiency. Mammalian *OAZ1* pseudoknot is inserted in the PB1 gene of pPS0 plasmid as previously done to test frameshifting efficiency *in vitro* [35]. This mRNA is obtained by SP6 transcription, then translated in RRL as in **a** and in the presence of a range of polyamine concentrations . The frameshifting efficiency indicated is the median value of at least three independent measurements, taking into account the number of methionines in each product (i.e frameshifted versus non-frameshifted).

Figure 2

a Orthogonal views of the model 1 complex in which the ribosome is unratcheted and bound with an elongation factor (EF). EF difference density is shown in red, and the reconstruction is fitted with the small and large subunits of asymmetric unit copy 2 of the 80S yeast crystal structure [22]. 40S rRNA blue, 60S rRNA yellow, proteins pink.

b As **a** for the model 2 complex in which the ribosome is ratcheted and bound with P-site tRNA. tRNA difference density is shown in green, and the reconstruction is fitted with the small and large subunits of asymmetric unit copy 1 of the 80S yeast crystal structure [22]

Figure 3

a A close-up of the intersubunit space of the EF-bound structure fitted as in Figure 1. The reconstruction is shown at a high contour level (2.7σ) to emphasise high density features and the difference density for the EF is shown as a red mesh (1σ). For a global view see Supplementary Figure S6a.

b As **a** for the P-site tRNA-bound structure (at 3.8σ), with difference density in green mesh (1.25σ). For a global view see Supplementary Figure S6b.

c Fit of a tRNA and eEF1A homology model to the EF difference density (at 1σ); tRNA blue ribbon, eEF1A green ribbon. Two views are shown – from the side looking between the subunits (as in **a** and **b**) and from above, as indicated by the thumbnails.

d Fit of a tRNA to its difference density (at 1.25σ), with its backbone a red-coloured ribbon.

Figure 4

A schematic of the +1 PRF mechanism revealed by the reconstructions described in this paper. The ribosome begins in a P/E tRNA-occupied post-translocational state, stalled at the pseudoknot. At this point eEF1-tRNA can bind, or a release factor (RF). If the RF binds then the ribosome will terminate and the P-site tRNA will be released and the subunits recycled (left hand path). If eEF1-tRNA binds it will do so out of register inducing a +1 PRF and proceeding to tRNA accommodation and subsequent translocation into the P-site (right hand path). Full antizyme synthesis will then ensue.

Materials and Methods

Plasmid construction.

The *OAZ1* frameshift pseudoknot or a destabilized pseudoknot (no PK) was inserted in the pPS0 plasmid [6,35]. This plasmid is transcribed by SP6 RNA polymerase, to generate an mRNA translating a long enough protein to be detectable *in vitro* to quantify either frameshifting efficiency or kinetic pausing. Digestion at the AvaII site of the parental plasmid is used as a size marker for translational pause products.

Cryo-EM complex: The *OAZ1* pseudoknot from *O. cuniculus* (CCCTCACCCACCCCAGAAGATCCCAGGTGGGCGAGGGGATAGCCAGAGGGATC) was inserted in the pEMC plasmid [6] to generate a short mRNA dedicated to ribosome pausing purification.

mRNA transcription

Plasmids were transcribed by the SP6 RNA polymerase (NEB) in SP6 buffer 1X (NEB) in the presence of 10mM DTT, 0.5mM RNA CAP analog (NEB) and low mix GTP (2mM ATP, 2mM CTP, 2mM UTP, 0.1mM GTP). This solution was incubated at 40°C for 30min, then 2mM GTP was added, and incubated at 40°C for a further 30min. The resulting mRNA was loaded on a Quick spin RNA column (Roche) then Phenol/chloroform extracted and precipitated before use.

Translational pause test

This test was essentially performed as described by Kontos at al [35]. A translation reaction (25µl) is followed by addition of 8µCi of ^{35}S-Met in 22µl of nuclease-treated rabbit reticulocyte lysate (RRL) (Promega) at 25°C for 5min. Initiation was inhibited by adding 6µM edeine. 1.5µl of the reaction was stopped at 0, 2.5, 5, 7.5, 10, 12.5, 15, 20min after

addition of edeine. The sample was RNase A treated for 20 min at room temperature and loaded on a 4-12% gradient pre-casted SDS-PAGE gel (Invitrogen). After migration, the gel was dried, then exposed to X-ray film over night. The XhoI digested pPS0 plasmid was translated in RRL and used as a size marker for the translational pausing band.

Cryo-EM complex purification.

A short synthetic mRNA containing the *OAZ1* pseudoknot was incubated at 68°C for 5min, then 5min at 37°C and finally 10min on ice, in the presence of a biotinylated RNA oligonucleotide in Mg^{2+} buffer. The translational solution (1ml of RRL (Promega) + 0.2ml H_20 + 3mM $MgCl_2$) was pre-heated at 27°C for 10min. The complex was added to the *in vitro* translational assay and incubated at 27°C. 1mM of cycloheximide was added to stop the reaction. Paused ribosomes were purified as previously described [6].

Frameshifting quantification

In vitro transcription was performed with SP6 RNA polymerase including the synthetic cap structure 7meGPPG (NewEngland Biolabs) to generate capped mRNA. Translation products were analysed on 12% polyacrylamide gels. The relative abundance of non-frameshifted and frameshifted products was determined by quantification of [^{35}S]methionine incorporation using a Typhoon analyser (GE Healthcare). The frameshifting efficiency indicated is the median value of at least three independent measurements, taking into account the number of methionines in each product. We tested a range of concentrations for spermidine and spermine, according to a previously described range of concentrations [18]. Endogenous *in vitro* concentration of polyamines in RRL is variable depending on batch. It was quantified by HPLC to be 0.68mM spermidine and 0.0038mM spermine (S. Matsufuji personal communication).

Cryo-EM methods

The +1 PRF ribosome population was flash frozen on glow discharged lacey carbon grids (Agar) and imaged in low-dose mode on a Tecnai F30 cryo-electron microscope. Micrographs with no apparent drift or astigmatism were scanned on a Zeiss SCAI scanner and then individual particles were picked using Boxer from the EMAN suite [36]. CTF correction was by CTFIT, also from the EMAN suite [36]. Particle images were triaged by iterative reference-free classification and alignment using EMAN routines operated through the ProjectViewer GUI component of the WellMap software (Flanagan, J.F., IV and RJCG, unpublished). Images classifying as crowded or containing ice were discarded at this stage. The final dataset was initially reconstructed using a stem-loop stalled RRL ribosome [5] filtered to 80Å resolution using SPIDER routines again operated through the ProjectViewer GUI. Difference density was computed between reconstructions Fourier scaled to the reference Apo map [6] using WellMap. This difference density showed a tRNA present in the P-site and was used in a multivariate statistical analysis- (MSA-) based two-way classification of the data in which a soft mask in the region of the identified tRNA was used to separate the data into those projection images with high density in the equivalent region *versus* those with low density in that region. This resulted in two maps: one had the P-site tRNA, the other had density adjacent to the A site in the region where an elongation factor might bind. A difference map for the likely elongation factor density was then computed against the Apo ribosome and used to generate a mask used in a new MSA-based classification. Again, two models were obtained, one with tRNA in the P site and one with likely elongation-factor density. This showed that whether the images were triaged based on the density in the P site or the density in the EF-binding region, the same classification results: one structure with a tRNA alone, the other with an elongation factor. Next, a three-way reference-based classification was carried

out in which the alignment models were the P-site tRNA-containing ribosome, the EF-containing ribosome, and a simulated structure made from the two in which both the tRNA and EF were present. Although three models were used only two reconstructions resulted, and these were with the tRNA *or* the EF but not both. On this basis, a two-way reference-based classification was carried out in which selection of particles was made using a soft-edged mask covering both the P-site tRNA and the EF; in this case, although the mask covered both regions, the resulting models were again distinct having one or the other ligand but not both. Finally, a three-way alignment involving (i) the P-site tRNA containing ribosome, (ii) the EF-containing ribosome and (iii) an Apo ribosome was carried out in which a soft mask focused the particle selection to the region covering both binding sites. Again, only two structures resulted, with one ligand or the other but not both, and there appeared to be no empty ribosomes. Thus there appear to be no empty ribosomes detectable in the sample, but all ribosomes appear to be either EF-bound (model 1) or P-site tRNA bound (model 2). The final reconstructions consisted of 26,049 images in the EF-bound model 1 reconstructed at 15.6Å resolution and 37,323 images in the tRNA-bound model 2 reconstructed at 15.4Å resolution (see Supplementary Figure S1 for resolution estimation, by 0.5-cutoff Fourier Shell Correlation). Perhaps the most convincingly validating feature of the resulting maps is that although, both in the MSA-based and projection matching-based classifications, the masks covered only ligand binding sites (stretching from the P-site out beyond the A-site) the classification resolved dramatically different conformations for the whole ribosomes (see main text).

Mass spectrometry

Purified ribosomes were run on an SDS-PAGE and silver stained. Four bands corresponding to proteins with a molecular weight between 50 and 100kDa were cut from the gel. In-gel

enzymatic digestion was automatically performed using standard conditions. After overnight tryptic digestion, peptides were extracted with 60% acetonitrile and 0.1% (v/v) formic acid. Acetonitrile was removed under vacuum and peptides were resuspended in 0.1% (v/v) formic acid prior to LC-MS/MS mass spectrometry analyses.

LC-MS/MS analyses were performed with the LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) coupled to the EASY nanoLC HPLC system (Proxeon). Peptides were eluted from a reverse phase C18 column with a 5-40% (v/v) acetonitrile gradient. Raw data were processed and analysed with the Proteome Discoverer 1.3 software using the MASCOT algorithm and the Swissprot database (Mammalian taxonomy). Protein identification searches were performed considering carbamidomethylation of cysteines and oxidation of methionines as complete and variable modifications respectively. Protein identifications from the 4 bands were merged to a single list and only proteins identified with at least two peptides (MS/MS ion score >20) and a total score >300 were retained. For sake of ease, while many ribosomal proteins were identified, only ribosomal associated factors are shown in Table S1. The elongation factor eEF1A was identified in the band at 50 kDa (predicted MW 50.1 kDa).

Acknowledgments

MS and HC were founded by ANR and FRM grants respectively from ON. This work is supported by ANR and the PRES UniverSud. This work has benefited from the facilities and expertise of the SICaPS platform of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr). C.A.S. was funded by the BBSRC. R.J.C.G. is a Royal Society University Research Fellow. We gratefully acknowledge financial support from INSTRUCT (http://www.structuralbiology.eu/).

References

1. Schmeing, T.M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461, 1234-42 (2009).
2. Schmeing, T.M., Voorhees, R.M., Kelley, A.C. & Ramakrishnan, V. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* 18, 432-6 (2011).
3. Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935-42 (2006).
4. Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466, 329-33 (2010).
5. Flanagan, J.F.t., Namy, O., Brierley, I. & Gilbert, R.J. Direct observation of distinct A/P hybrid-state tRNAs in translocating ribosomes. *Structure* 18, 257-64 (2010).
6. Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441, 244-7 (2006).
7. Fu, J., Munro, J.B., Blanchard, S.C. & Frank, J. Cryoelectron microscopy structures of the ribosome complex in intermediate states during tRNA translocation. *Proc Natl Acad Sci U S A* 108, 4817-21 (2011).
8. Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M. & Blanchard, S.C. Navigating the ribosome's metastable energy landscape. *Trends Biochem Sci* 34, 390-400 (2009).
9. Munro, J.B., Wasserman, M.R., Altman, R.B., Wang, L. & Blanchard, S.C. Correlated conformational events in EF-G and the ribosome regulate translocation. *Nat Struct Mol Biol* 17, 1470-7 (2010).
10. Qu, X. et al. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* 475, 118-21 (2011).
11. Wen, J.D. et al. Following translation by single ribosomes one codon at a time. *Nature* 452, 598-603 (2008).
12. Namy, O., Rousset, J.P., Napthine, S. & Brierley, I. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* 13, 157-68 (2004).
13. Brierley, I., Gilbert, R.J. & Pennell, S. RNA pseudoknots and the regulation of protein synthesis. *Biochem Soc Trans* 36, 684-9 (2008).
14. Moran, S.J. et al. The mechanics of translocation: a molecular "spring-and-ratchet" system. *Structure* 16, 664-72 (2008).
15. Hansen, T.M., Reihani, S.N., Oddershede, L.B. & Sorensen, M.A. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc Natl Acad Sci U S A* 104, 5830-5 (2007).
16. Chen, G., Chang, K.Y., Chou, M.Y., Bustamante, C. & Tinoco, I., Jr. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A* 106, 12706-11 (2009).
17. Kurian, L., Palanimurugan, R., Godderz, D. & Dohmen, R.J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* 477, 490-4 (2011).
18. Matsufuji, S. et al. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* 80, 51-60 (1995).
19. Petros, L.M., Howard, M.T., Gesteland, R.F. & Atkins, J.F. Polyamine sensing during antizyme mRNA programmed frameshifting. *Biochem Biophys Res Commun* 338, 1478-89 (2005).
20. Ivanov, I.P., Atkins, J.F. & Michael, A.J. A profusion of upstream open reading frame

mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res* 38, 353-9 (2009).
21. Ivanov, I.P., Gesteland, R.F. & Atkins, J.F. Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. *Nucleic Acids Res* 28, 3185-96 (2000).
22. Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* 330, 1203-9 (2011).
23. Frank, J. & Agrawal, R.K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406, 318-22 (2000).
24. Valle, M. et al. Locking and unlocking of ribosomal motions. *Cell* 114, 123-34 (2003).
25. Schuette, J.C. et al. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *Embo J* 28, 755-65 (2009).
26. Villa, E. et al. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A* 106, 1063-8 (2009).
27. Gao, Y.G. et al. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* 326, 694-9 (2009).
28. Takyar, S., Hickerson, R.P. & Noller, H.F. mRNA helicase activity of the ribosome. *Cell* 120, 49-58 (2005).
29. Ogle, J.M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292, 897-902 (2001).
30. Gilbert, R.J. et al. Three-dimensional structures of translating ribosomes by Cryo-EM. *Mol Cell* 14, 57-66 (2004).
31. Spahn, C.M., Penczek, P.A., Leith, A. & Frank, J. A method for differentiating proteins from nucleic acids in intermediate-resolution density maps: cryo-electron microscopy defines the quaternary structure of the Escherichia coli 70S ribosome. *Structure* 8, 937-48 (2000).
32. Nissen, P. et al. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* 270, 1464-72 (1995).
33. Huang, Y. & Sprinzl, M. Peptide bond formation on the ribosome: the role of the 2'-OH group on the terminal adenosine of peptidyl-tRNA and of the length of nascent peptide chain. *Angew Chem Int Ed Engl* 50, 7287-9 (2011).
34. Namy, O. et al. Epigenetic control of polyamines by the prion [PSI+]. *Nat Cell Biol* 10, 1069-75 (2008).
35. Kontos, H., Napthine, S. & Brierley, I. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* 21, 8657-70 (2001).
36. Ludtke, S.J., Baldwin, P.R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128, 82-97 (1999).

a
M
0
2.5
5
7.5
10
12.5
15
20
min
OAZ1 PK
No PK

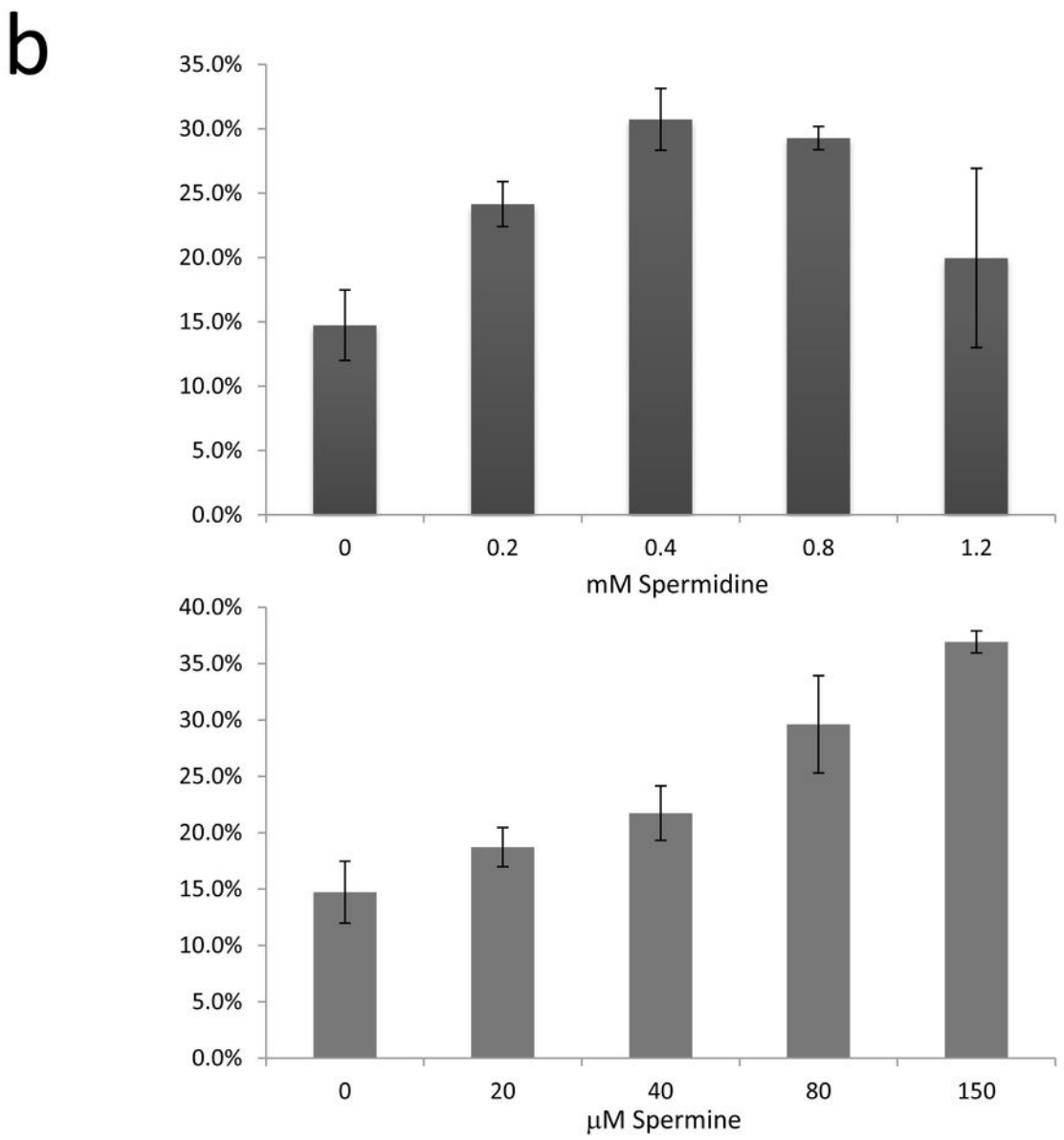

b
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0
0.2
0.4
0.8
1.2
mM Spermidine
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0
20
40
80
150
µM Spermine

Figure 1

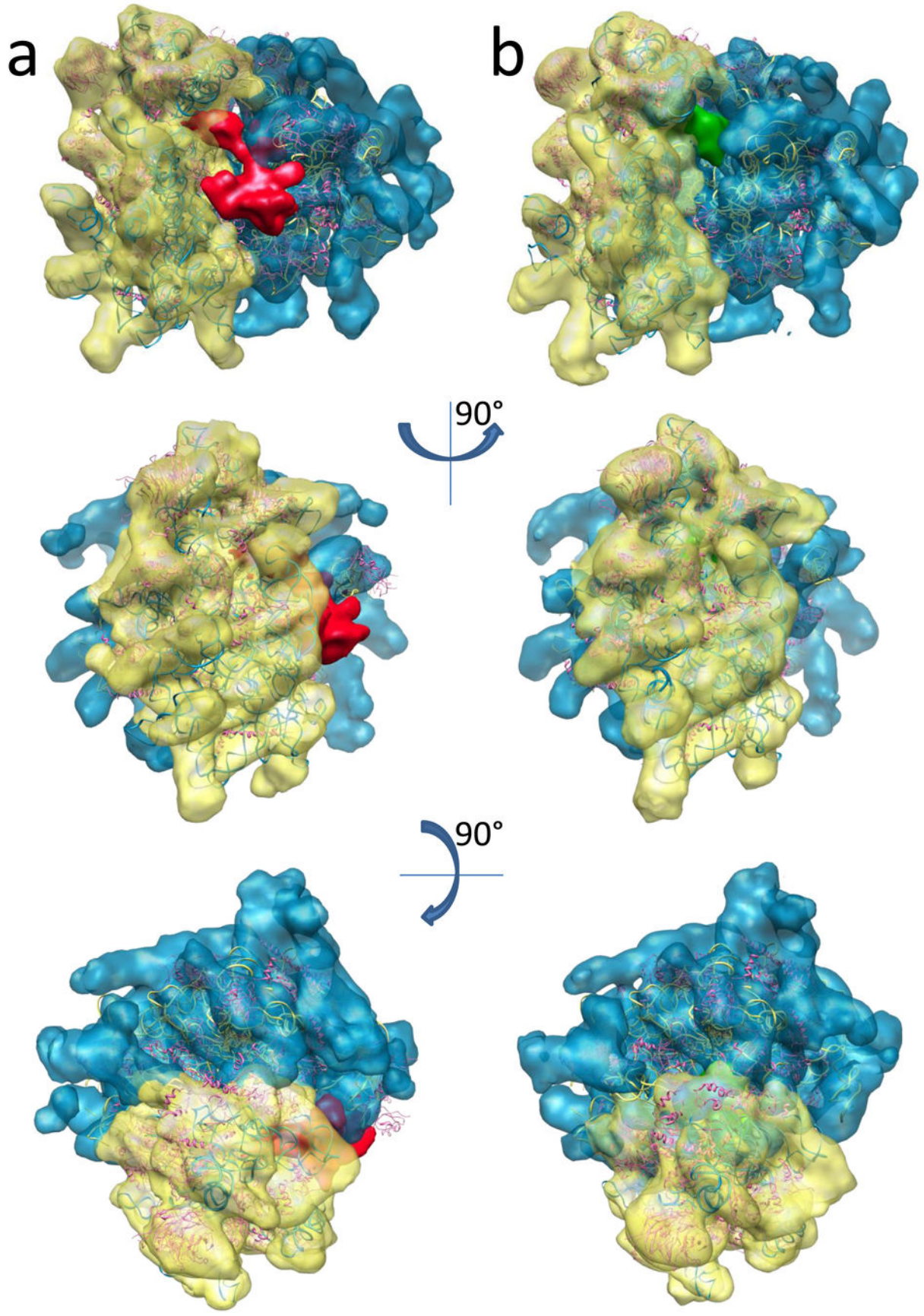


Figure 2

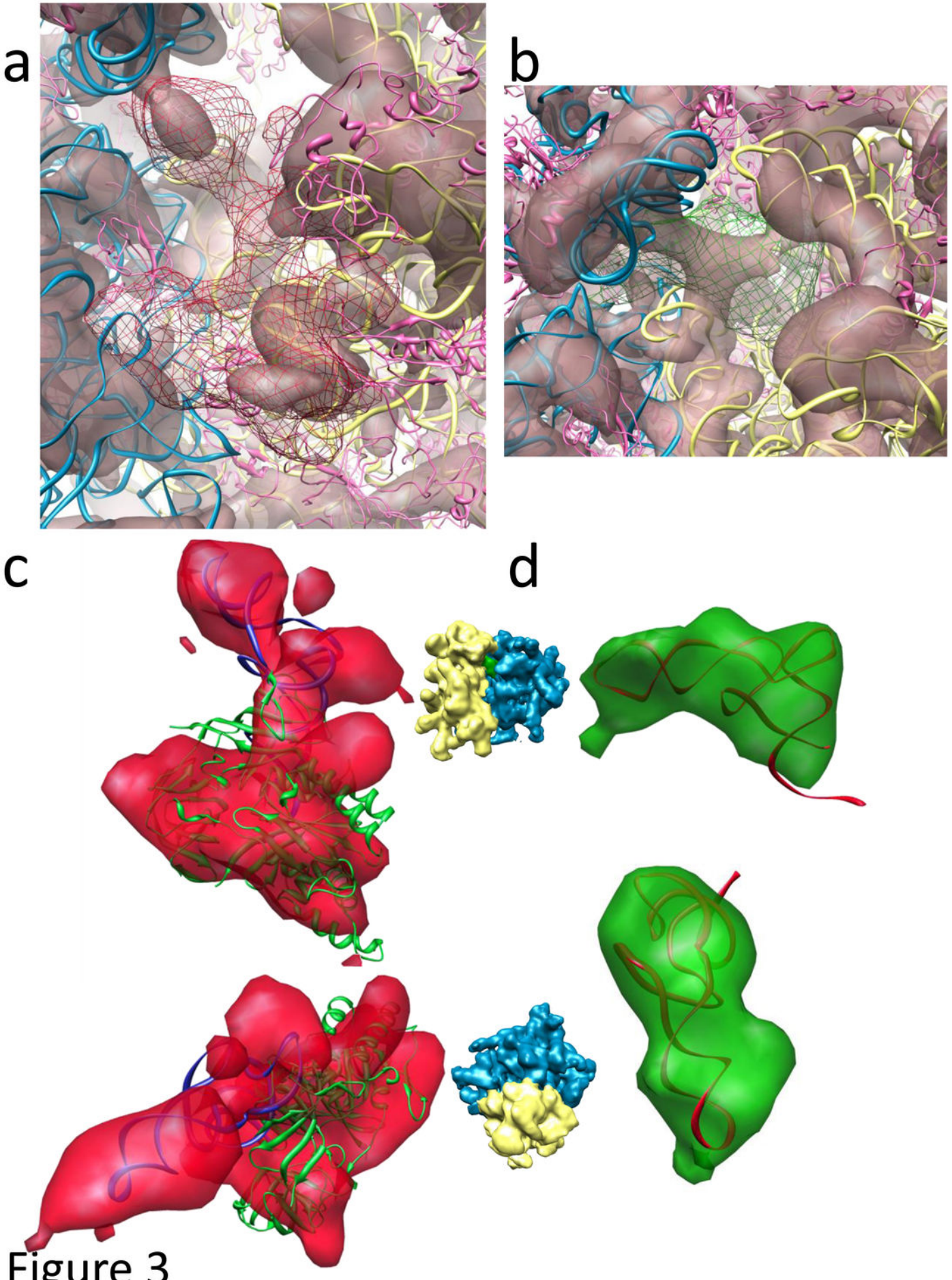


Figure 3

Figure 4

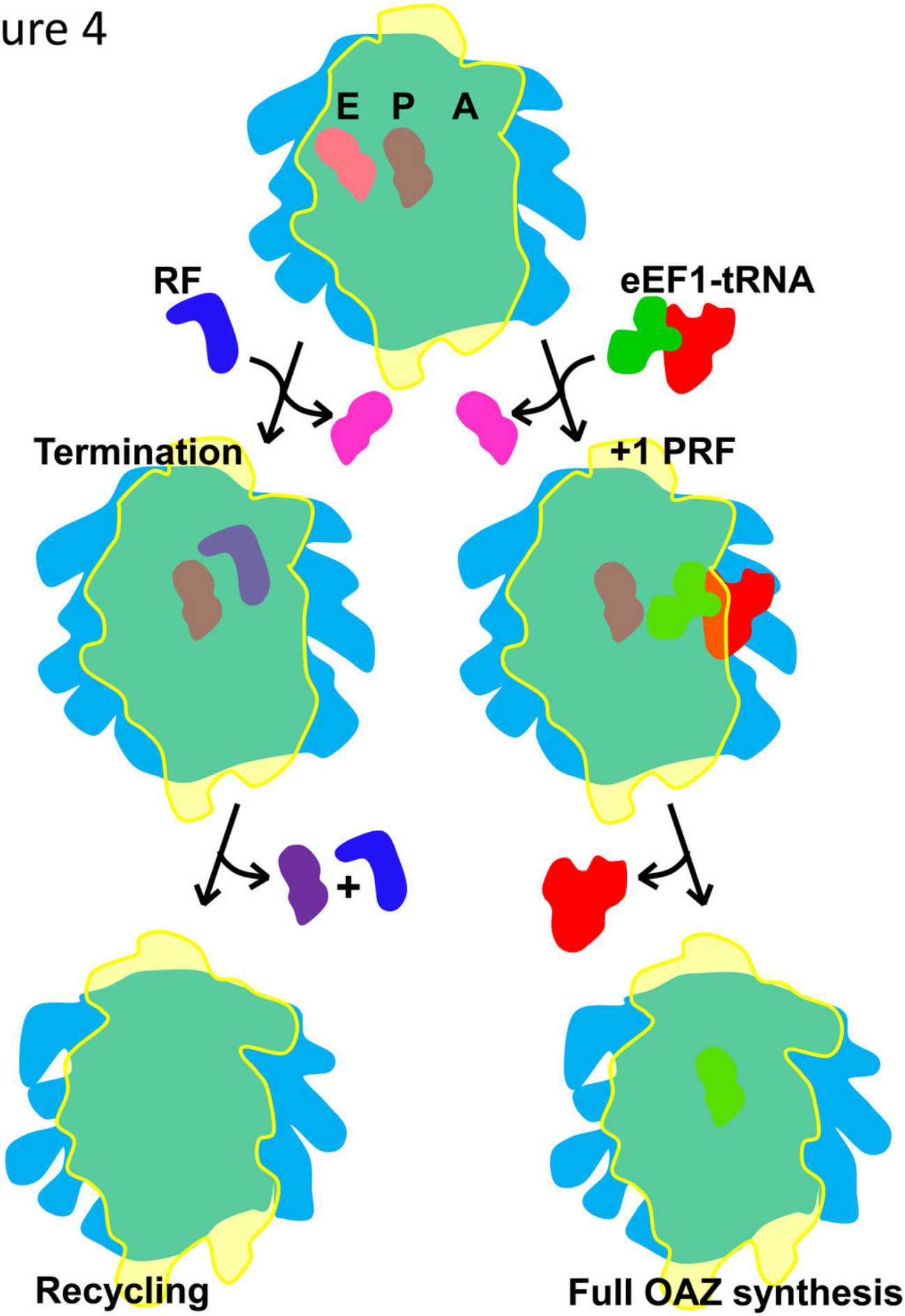

E P A
RF
eEF1-tRNA
Termination
+1 PRF
+
Recycling
Full OAZ synthesis

Table S1: Results of fragment mapping of gel bands using mass spectrometry.

Swiss-Prot AC.	Gene Name	N° of unique peptides	Protein Coverage (%)
P60228	eIF3E	18	41
A5A6M4	eIF3L	17	30
Q3T122	eIF3D	14	26
P55010	eIF5	14	31
P81795	eIf2s3	12	31
O00303	eIF3F	10	32
P41035	eIF2S2	7	24
Q3SZ65	eIF4A2	6	20
A2Q0Z0	**eEF1A1**	**6**	**16**
P29694	eEF1G	5	13
Q5RK09	eIF3g	4	13

Supplementary Figure Legends

Figure S1: Fourier shell correlation of the two models: EF-bound (model 1) in grey, 15.6Å resolution, P-site tRNA-bound (model 2) in red, 15.4Å resolution.

Figure S2: Two views of the ratchet visualized by superposition of the EF-bound model 1 and P-site tRNA-bound model 2. Model 1 is shown in blue and yellow mesh for large and small subunits respectively, model 2 in blue and yellow semitransparent surface. Both difference densities, for the EF and for the P-site tRNA, are shown (in red and green, respectively).

Figure S3:

a High contour level (2.7σ) rendering of the EF-bound model 1 reconstruction fitted with the unratcheted (copy 2) 80S yeast crystal structure: small subunit rRNA aquamarine, large subunit rRNA yellow, proteins pink. A difference map calculated between this reconstruction and an Apo map is shown in red for the EF. The fit of eEF1A (in green) and its tRNA (in blue) is shown also.

b As **a** for the tRNA-bound model 2 reconstruction (at 3.8σ) fitted with the ratcheted (copy 1) 80S yeast crystal structure. The difference map shown in green is again with respect to an Apo map and is for the P-site tRNA. The fit of a tRNA atomic model (in red) is shown also.

c The copy 2 80S crystal structure with the whole difference map for model 1 shown in red at an intermediate contour level (1.25σ), with atomic models as in **a**.

d As **c** for the copy 1 80S crystal structure and the model 2 difference map (at 1.75σ), in green, with atomic models as in **b**.

e As **c** for a higher contour level (1.7σ): the EF-bound density is the most significant difference of all compared to an Apo state.

f As **d** for a higher contour level (2.5σ): the P-site tRNA density is the most significant difference of all.

Figure S4: The silver-stained gel of purified +1 PRF ribosomes showing standard markers (left) and bands used for mass spectrometry finger printing (right). The asterisked bands were extracted for analysis.

Figure S5: Views of both the EF and P-site tRNA difference densities, with their respective atomic fits, in two views: one looking down the intersubunit space and one from above, as indicated by the whole-ribosome views to the right.

Figure S6:

a *Left* a side-view of the small subunit of the model 1 eEF1-bound ribosome fitted with the copy 2 unratcheted crystal structure [1], with difference density in red mesh and the eEF1A (green) and tRNA (blue) fitted. *Right* a view from the intersubunit face of the small subunit

with just the crystal structures and fitted eEF1-tRNA and difference density at two contour levels (1σ and 1.5σ) to emphasise the most intense regions.

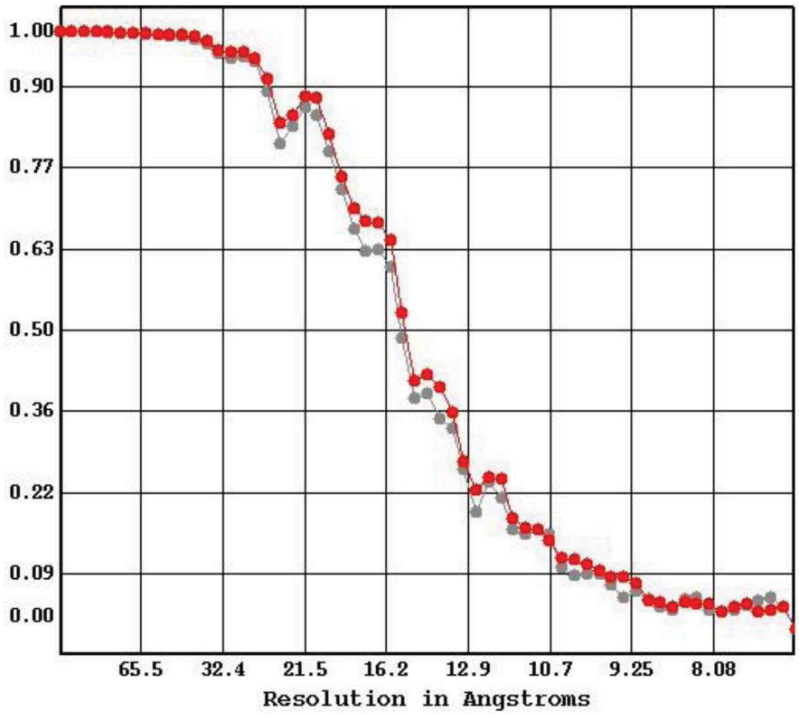
b *Left* and *centre* as **a** for a fit of the EF-Tu-tRNA-bound small prokaryotic subunit structure [2]. The rRNA is blue, the ribosomal proteins pink, EF-Tu is green and its tRNA is blue. *Right* a close-up of the yeast 40S crystal structure (grey proteins and black rRNA) and the 30S-EF-Tu-tRNA crystal structure (coloured as in the *centre*).

Figure S7:

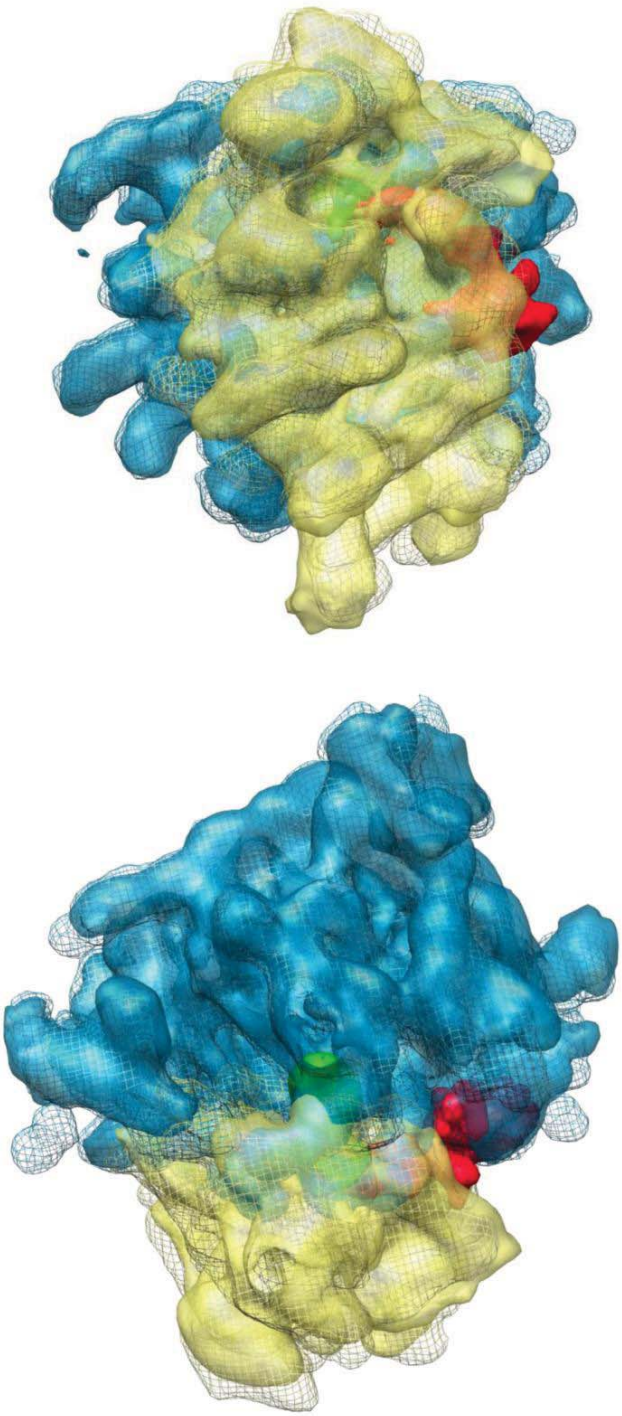
Left, *centre* and *right* as in Figure S6b for a fit of the EF-G-bound small prokaryotic subunit structure [3] with the EF-G in green. Inset (red box) is the EF-Tu and EF-G crystal structures superimposed by common alignment to the 40S envelope.

References

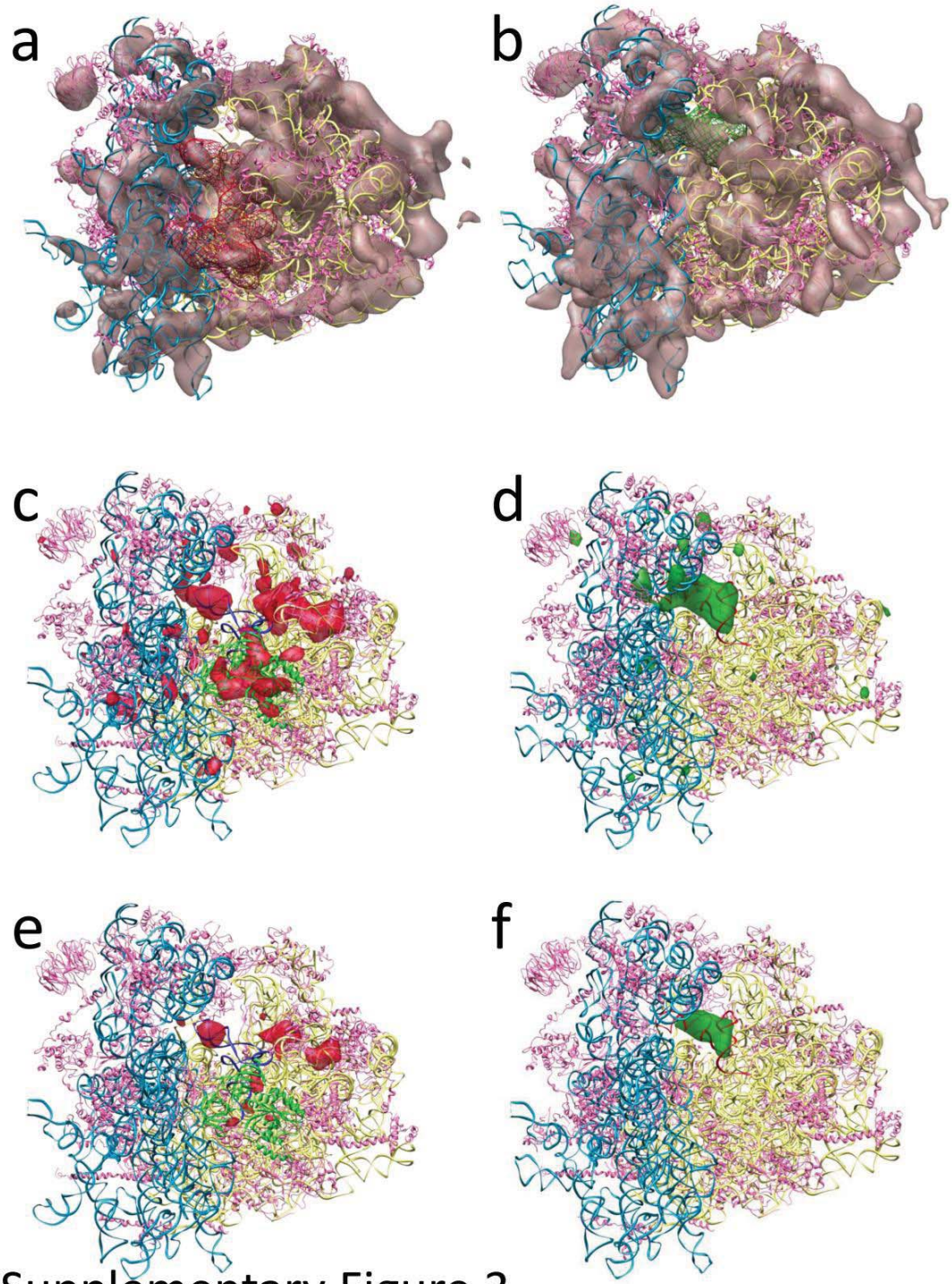
1. Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203-9 (2011).
2. Schmeing, T.M., Voorhees, R.M., Kelley, A.C. & Ramakrishnan, V. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* **18**, 432-6 (2011).
3. Gao, Y.G. et al. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694-9 (2009).



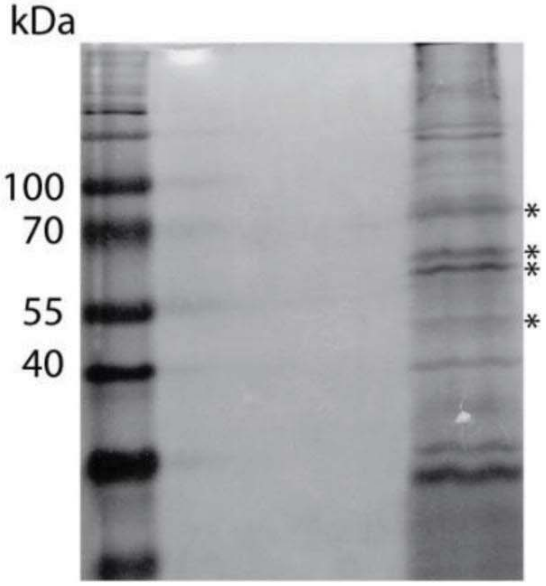
Supplementary Figure 1



Supplementary Figure 2



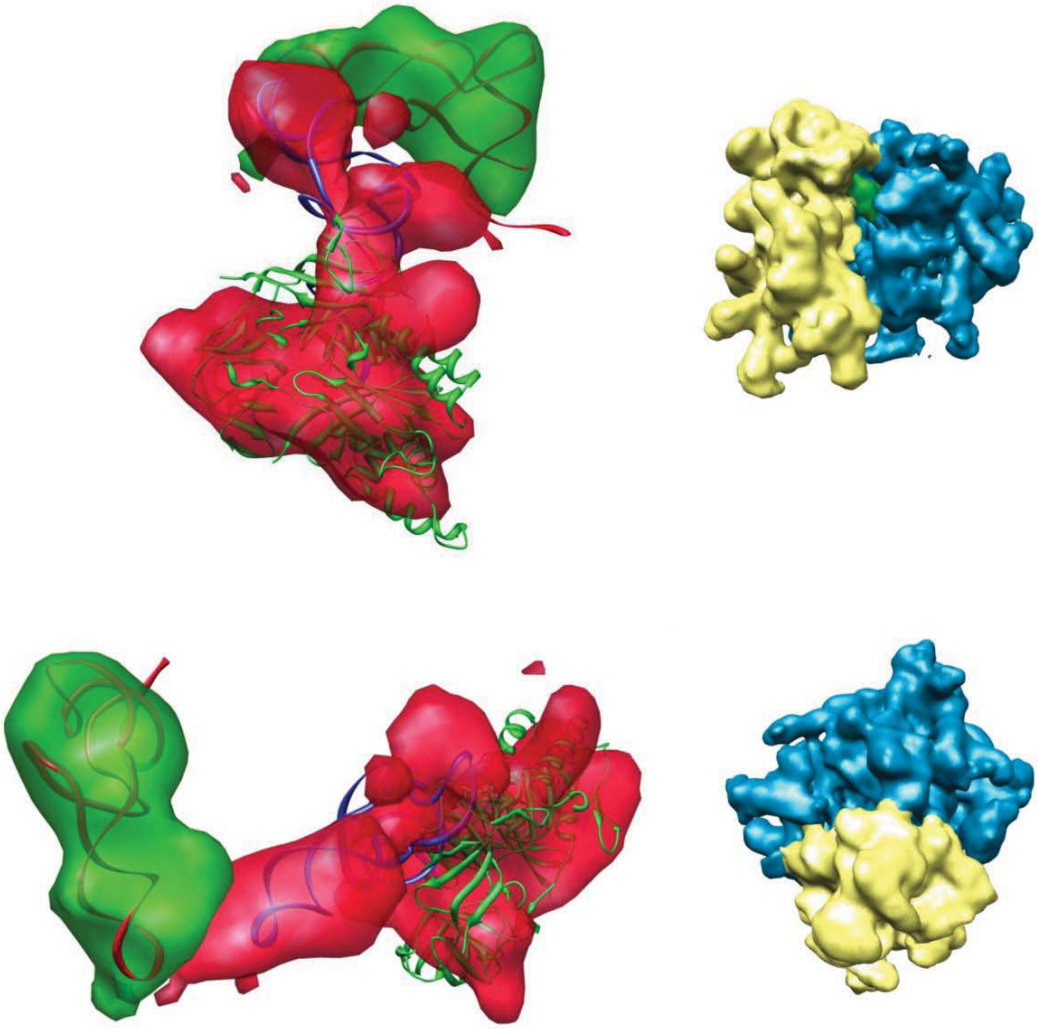
Supplementary Figure 3



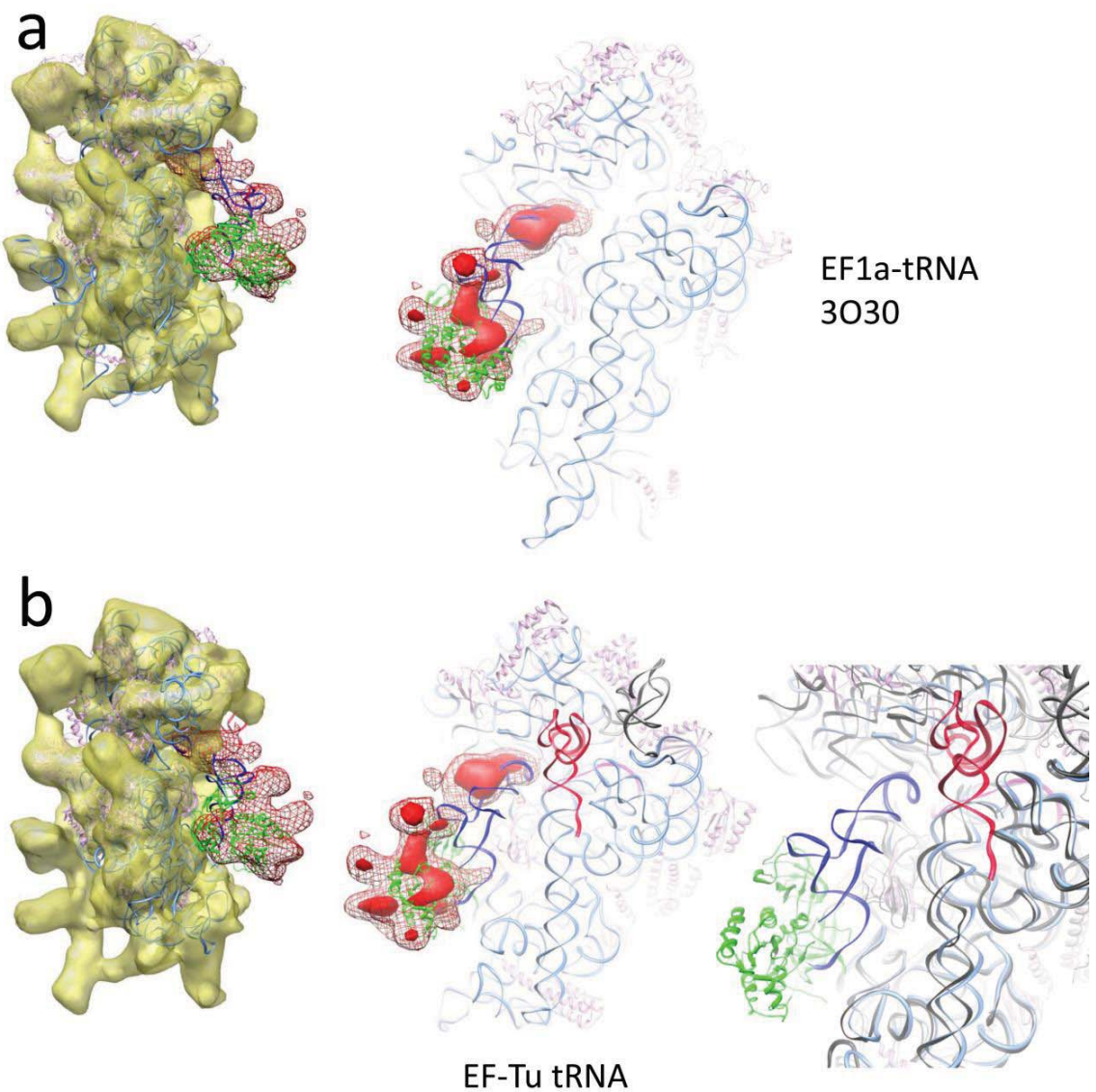
EEF1A1

MGKEKTHINI VVIGHVDSGK **STTTGHLIYK** CGGIDKRTIE KFEKEAAEMG KGSFKYAWVL DKLKAERERG ITIDISLWKF ETSK**YYVTII DAPGHR**DFIK NMITGTSQAD CAVLIVAAGV GEFEAGISKN GQTREHALLA YTLGVKQLIV GVNKMDSTEP PYSQKRYEEI VKEVSTYIKK IGYNPDTVAF VPISGWNGDN MLEPSANMPW FKGWKVTRKD GNASGTTLLE ALDCILPPTR PTDKPLR**LPL QDVYKIGGIG TVPVGRVETG VLKPGMVVTF APVNVTTEVK** SVEMHHEALS EALPGDNVGF NVKNVSVKDV RRGNVAGDSK NDPPMEAAGF TAQVIILNHP GQISAGYAPV PDCHTAHIAC KFAELKEKID RRSGKKLEDG PKFLKSGDAA IVDMVPGKPM CVESFSDYPP LGRFAVRDMR **QTVAVGVIK**A VDKKAAGAGK VTKSAQKAQK AK

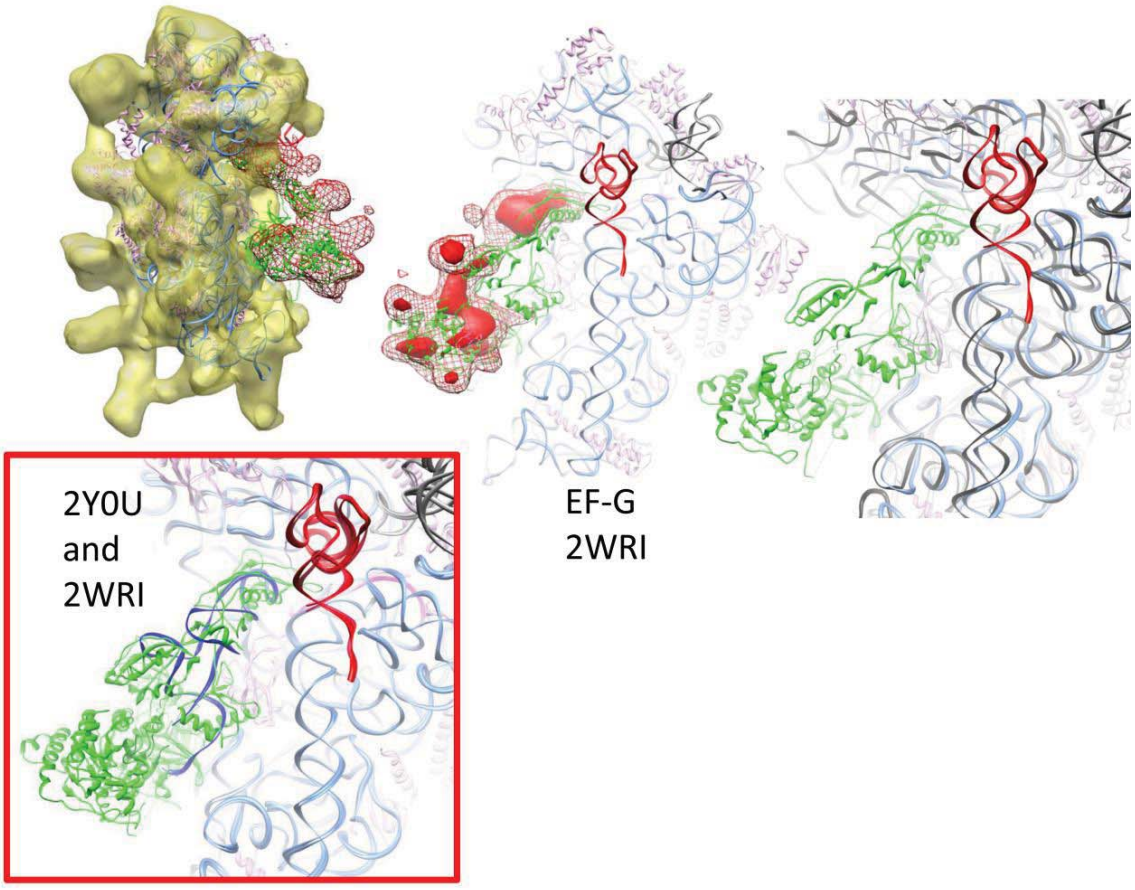
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7

Matériels et méthodes.

Sommaire

I. Chapitre 1 : Etude du rôle d'eEF2 dans le maintien du cadre de lecture..... i

1. Les plasmides..... i
2. Les souches.....ii
3. Mutagénèse aléatoire par PCR.....iv

II. Chapitre 2 : Etude de l'élongation en molécule unique.....v

1. Les plasmides.....v
2. Les oligonucléotides et fluorophores.....vi
3. Préparation des sous-unités ribosomiques à partir de RRL.....vi
4. Transcription *in vitro*.....vii
5. Expériences de retard sur gel.....viii
6. Traduction sur billes magnétiques.....viii
7. Montage microscopique.....x
8. Chimie de surface.....x
9. Traduction en cellule microfluidique.....x
10. Purification des sous-unités 40S et 60S à partir de *S. cerevisiae*.....xi

I. Chapitre 1 : Etude du rôle d'eEF2 dans le maintien du cadre de lecture.

1. Les plasmides

La liste et la description des plasmides utilisés sont indiquées dans le tableau suivant :

Plasmides	Description
pFL38	Origine de réplication centromérique ; Amp^R ; *URA3*
pFL38-EFT2	pFL38 ; *EFT2*
pUC19	Amp^R
pUC19-EFT2	pUC19 ; *EFT2*
pUC19-eft2(H699N)	pUC19-EFT2 ; eft2(H699N)
pUC19-PGK	pUC19 ; Promoteur / terminateur PGK1
pUC19-PGK-EFT2	pUC19-PGK ; $Prom_{PGK1}$-*EFT2*-Ter_{PGK1}
pFL61	Origine de réplication 2µm ; Amp^R ; *URA3 ;* Promoteur / terminateur PGK1
pFL44H	Origine de réplication 2µm ; Amp^R ; *HIS3*
pFL44H-(NotI)	pFL44H
pFL44H-PGK	pFL44H-(NotI) ; Promoteur / terminateur PGK1
pFL44H-PGK-EFT2	pFL44H-PGK ; $Prom_{PGK1}$-*EFT2*-Ter_{PGK1}
pFL44H-PGK-eft2(H699N)	pFL44H-PGK ; $Prom_{PGK1}$-eft2(H699N)-Ter_{PGK1}

Le plasmide pFL38-EFT2 a été obtenu en clonant *EFT2* dans le plasmide d'expression pFL38 au site de restriction EcoIcRI. *EFT2* a été amplifié à partir d'ADNg de cellules HFY en incluant sa région promotrice (990bp) et sa région terminatrice (572pb). Les amorces utilisées étaient EFT2-990.w et EFT2+575.c. La construction a été vérifiée par séquençage.

Le plasmide pUC19-EFT2 a été obtenu en clonant le gène EFT2 au site de restriction SmaI du pUC19. Le fragment *EFT2* a été amplifié par PCR à partir d'ADNg de cellules HFY en incluant 2 codons en 5' du codon d'initiation et codon stop additionnel après le codon stop d'*EFT2*. Les amorces utilisées étaient EFT2atg.w et EFT2stop.c. La construction a été vérifiée par séquençage, les codons stops manquaient.

Le plasmide pUC19-eft2(H699N) a été obtenu en réalisant une mutagénèse dirigée par PCR du plasmide pUC19-EFT2. Le codon CAC n°699 d'EFT2 a été en AAC grâce amorces H699N.w et H699N.c. La construction a été vérifiée par séquençage, les codons stops manquaient à cause d'une synthèse défectueuse des oligonucléotides.

Le plasmide pUC19-PGK a été obtenu en digérant le plasmide pFL61 par les enzymes SmaI et HindIII. Le fragment d'ADN contenant le promoteur et le terminateur PGK1 a été inséré aux sites de restriction SmaI et HindIII du pUC19.

Le plasmide pUC19-PGK-EFT2 a été obtenu en clonant le gène *EFT2* au site de restriction MscI du pUC19-PGK entre le promoteur et le terminateur PGK1. Le fragment *EFT2* a été amplifié par PCR à

partir d'ADNg de cellules HFY en incluant 2 codons en 5' du codon d'initiation et un codon stop additionnel après le codon stop d'*EFT2*. Les amorces utilisées étaient EFT2atg.w et EFT2stop.c. La construction a été vérifiée par séquençage.

Le plasmide pFL44H-(NotI) a été obtenu en cassant le site NotI du plasmide pFL44H. Après digestion par NotI, les extrémités 3' sortantes ont été remplies par polymérisation avec la polymérase Vent. Cette enzyme a été choisie car elle n'ajoute pas un A en 3' après la fin du duplex d'ADN. Le plasmide a été ensuite ligaturé sur lui-même.

Le plasmide pFL44H-PGK a été obtenu en clonant le promoteur et le terminateur PGK1 au site de restriction EcoIcRI du plasmide pFL44H-(NotI). Le fragment contenant le promoteur et le terminateur PGK1 a été amplifié par PCR à partir du plasmide pUC19-PGK avec les amorces PGK5'prom.w et PR. La construction a été vérifiée par des profils de digestion.

Le plasmide pFL44H-PGK-EFT2 a été obtenu en clonant le gène EFT2 entre le promoteur et le terminateur PGK1 du plasmide pFL44H-PGK. Le fragment EFT2 a été amplifié par PCR à partir d'ADNg de cellules 74-D694 en utilisant les amorces EFT2atg-EcoRI.w et EFT2stop-MscI.c. Le fragment de PCR a été cloné entre les 2 sites de restriction NotI du pFL44H-PGK qui ont été remplis au préalable. La construction a été vérifiée par séquençage.

Le plasmide pFL44H-PGK-eft2(H699N) a été obtenu en clonant l'allèle eft2(H699N) entre le promoteur et le terminateur PGK1 du plasmide pFL44H-PGK. Le fragment a été amplifié par PCR à partir du plasmide pUC19-eft2(H699N) en utilisant les amorces EFT2atg-EcoRI.w et EFT2stop-MscI.c. Le fragment de PCR a été cloné entre les 2 sites de restriction NotI du pFL44H-PGK qui ont été remplis au préalable. La construction a été vérifiée par séquençage.

Les séquences des amorces utilisées pour construire les plasmides sont détaillées dans le tableau suivant :

Amorces	Séquences	
EFT2-990.w	ATTGGTGTTTCCATTAGGCC	*EFT2* avec son promoteur et terminateur
EFT2+575.c	CGGGAGATATGTGGTGAATTTTGC	
EFT2atg.w	TGCAAG**ATG**GTTGCTTTCACTGTTGAC (**codon initiation**)	*EFT2*
EFT2stop.c	**TTATTA**CAATTTGTCGTAATATTCTTGCC (**2 stops**)	
H699N.w	CATGCCGATGCTATC**A**ACAGAGGTGGTGGTC (**mutation**)	Mutagénèse H699N *d'EFT2*
H699N.c	GACCACCACCTCTG**T**TGATAGCATCGGCATG (**mutation**)	
PGK5'prom.w	AGATTCCTGACTTCAACTCAAGACGCACAG	Promoteur et terminateur PGK1
PR	AGCGGATAACAATTTCACACAGG	
EFT2atg-EcoRI.w	G**GAATTC**TGCAAG*ATG*GTTGCTTTCACTGTTGAC (**EcoRI**) (*codon d'initiation*)	(EcoRI)-*EFT2*-(MscI)
EFT2stop-MscI.c	T**TGGCCA***TTATTA*CAATTTGTCGTAATATTCTTGCC (**MscI**) (*2 stops*)	

2. Les souches

Les souches utilisées sont répertoriées avec leur génotype dans le tableau suivant :

Souches	Génotypes	Références

HFY1200	MATa ; ade2-1 ; trp1-1 ; can1-100 ; leu2-3, 112 ; his3-11,15 ; ura3-1	(He and Jacobson 1995)
FS1	MATα ; ade2-592 ; lys2Δ201; his3Δ200; leu2-3,112 ; ura3-52	(Hatin et al. 2009)
74-D694	MATa ; ade1-14 ; trp1-289 ; his3Δ200; leu2-3,112 ; ura3-52; [psi-]	(Chernoff et al. 1995)
FY1679 -06C	MATα ; ura3-52 ; trp1Δ63 ; his3Δ200 ; leu2Δ1 ;	(Winston, Dollard and Ricupero-Hovasse 1995)
FY IBV(G3U3G)	FY1679-06C, *ade2-IBV(G3U3G)*	A. Baudin Baillieu
HC06	FYG3U3G ; eft2Δ::Kan^R,	H. Chommy
HC07	FYG3U3G ; eft2Δ::Kan^R ; eft1Δ::*TRP1-KL* ; pFL38-EFT2	
BY4742-Y10000	MATα ; his3Δ1 ; leu2Δ0 ; ura3Δ0 ; lys2Δ0	(Brachmann et al. 1998)
BY4742-Y12389	BY4742-Y10000 ; eft1Δ::Kan^R	
BY4742-Y04221	MATa ; his3Δ1 ; leu2Δ0 ; ura3Δ0 ; met15Δ0 ; eft2Δ::Kan^R	
TKY826	MATa ; ade2 ; leu2 ; ura3 ; his3 ; leu2 ; trp1 ; eft1Δ::*HIS3* ; eft2Δ::*TRP1* ; peft2(I698A)-6xHis (*CEN* ;*LEU2*)	(Ortiz et al. 2006)

La souche HC06 a été construite par délétion du gène *EFT2* avec la cassette de résistance Kan^R issue du plasmide pFA6a-kanMX6. Cette cassette confère, chez les eucaryotes, une résistance au G418. La cassette a été amplifiée par join-PCR avec des amorces contenant les régions flanquantes d'*EFT2* : de -927 à -400 en 5' de l'ATG (soit 526pb); de +52 à +494 en 3' du stop (soit 423pb). Ces régions ont été précédemment amplifiées par PCR à partir d'ADNg. Des cellules FY ont été transformées avec le produit de PCR contenant la cassette Kan^R. Les transformants sont sélectionnés sur milieu YPG contenant du G418 à 200µg/ml. L'intégration de la cassette par recombinaison homologue au locus d'*EFT2* a été vérifiée par PCR sur colonie avec des amorces de part et d'autre de la cassette et d'*EFT2* dans les zones de recombinaison. La discrimination se fait par la taille du produit de PCR. Dans le cas d'une souche sauvage le produit de PCR a une taille attendue de 3130pb, et dans le cas d'une délétion *eft2Δ::Kan^R* le produit a une taille attendue de 1516pb.

La souche HC07 a été construite à partir de la souche HC06. Les cellules HC06 sont d'abord transformées avec le plasmide pFL38-EFT2. Les transformants sont sélectionnés sur milieu YNB CSM-Ura. Ces cellules sont utilisées pour la délétion du gène *EFT1* avec la cassette de sélection TRP1-KL issue du plasmide pBS1479. La cassette a été amplifiée par join-PCR avec des amorces contenant les régions flanquantes d'*EFT1* : de -700 à -41 en 5' de l'ATG (soit 660pb); de +19 à +532 en 3' du stop (soit 513pb). Ces régions ont été précédemment amplifiées par PCR à partir d'ADNg. Le produit de PCR contenant la cassette *TRP1-KL* a été utilisé pour transformer des cellules HC06 pFL38-EFT2. Les transformants sont sélectionnés sur milieu YNB CSM-Trp. L'intégration de la cassette *TRP1-KL* par recombinaison homologue au locus d'*EFT1* a été vérifiée pas PCR sur colonie avec des amorces de part et d'autre de la cassette et d'*EFT1* dans les zones de recombinaison. La discrimination se fait par la taille du produit de PCR. Dans le cas d'une souche sauvage, le produit de PCR a une taille attendue de 3277pb ; et dans le cas d'une délétion eft1Δ::TRP1-KL, le produit a une taille attendue de 1748pb. Les colonies ont aussi été testées sur milieu YNB CSM-Ura contenant du 0,1% 5-FOA.

Les séquences des amorces utilisées pour construire les souches sont détaillées dans le tableau suivant :

Amorces	Séquences

EFT2 74.w	ATGCTTCTTCCTAACGCTATTG	Région d'homologie en 5' d'*EFT2*.
EFT2 620.c	**CTGGGCCTCCATGTCGCTGG**TACGTTCCAGGTATTTCGCACC **(homologie avec la cassette KanR)**	
EFT2 3581.w	**GCGTCAATCGTATGTGAATG**GTATTGGGCACCTTTTATGTCC **(homologie avec la cassette KanR)**	Région d'homologie en 3' d'*EFT2*.
EFT2 4023.c	TAACTTTCTTTAGATCTGTAGG	
EFT2.w	TGCGAAATACCTGGAACGTAAG	Vérification de la délétion *d'EFT2*
EFT2.c	AAGCAAGATAAATCATAGTAAC	
EFT1 301.w	GACCCTTAAGCTTCAACGAATCACGTCC	Région d'homologie en 5' d'*EFT1*.
EFT1 trp5'.c	**CCCTATAGTGAGTCGTA**TCTTTCTGGCAAATGGGGTTAAGTAAAAAAACCAGTGAAGCG **(homologie avec la cassette TRP1-KL)**	
EFT1 trp3'.w	**GCTTCAGGTTGACTTCCC**GGTGGTTCTGTAAGAGCAAACCTTACCGCCTTATGATC **(homologie avec la cassette TRP1-KL)**	Région d'homologie en 3' d'*EFT1*.
EFT1 4060.c	GAGTTTTAGCCCGATGAGGC	
EFT1.w	GGCATGTCCACTAAATAATACG	Vérification de la délétion *d'EFT1*
EFT1.c	ATAGAAGCACCAACATACGGCG	

3. Mutagénèse aléatoire par PCR

La mutagénèse aléatoire a été réalisée dans 5 réactions de 50µl comprenant : 1x Tampon AmpliTaq (Applied Biosystems) ; 2,5mM $MgCl_2$ (ce qui, avec le tampon, porte la concentration finale totale à 4mM) ; 50µM de chaque dNTP (dATP, dCTP, dGTP et dTTP) ; 2 µM de chaque amorce (PU : AGGGTTTTCCCAGTCACGAC et PR : AGCGGATAACAATTTCACACAGG) ; 0,2ng/µl de pUC19-PGK-EFT2 et 0,1U/µl d'ADN polymérase AmpliTaq. La concentration finale en $MnCl_2$ retenue est 50µM.

Le programme d'amplification est : 1min à 94°C ; (30s à 94°C ; 15s à 55°C et 4min à 72°C) répété 30 fois et 5min à 72°C. La taille du produit de PCR est analysée par électrophorèse.

II. Chapitre 2 : Etude de l'élongation en molécule unique

1. Les plasmides

La liste et la description des plasmides utilisés sont indiquées dans le tableau suivant :

Plasmides	Description	Masse molaire de l'ARNm (g/mol)	Taille ARNm[1] (linéarisation)
pUC57 RNAopt	Construction de base avec le promoteur SP6, l'IRES CRPV, le site d'insertion HpaI et le pseudoneud de l'IBV et 2 stops.		
pUC57 RNAopt PPS0	Pour la traduction, séquence du PPS0 insérée au site HpaI	335.10^3 g/mol	1027nt (KpnI)
pUC57 RNAopt ATTO	Pour la microscopie, l'oligo H12.3A.c s'apparie à partir du 5ème codon en aval du codon d'initiation CCU	95.10^3 g/mol	293nt (XhoI)
pUC57 RNAopt ATTObis	pUC57 RNAopt ATTO mais avec le site de l'oligo H12.3A.c décalé de 9nt en aval.	335.10^3 g/mol	302nt (XhoI)
pUC57 RNAopt ATTO-PK	pUC57 RNAopt ATTO avec en aval de l'oligo H12.3A.c : 2 codons stops puis le site d'appariement pour l'oligo J12.3A.c.	121.10^3 g/mol	375nt (Ecl136II)
pUC57 RNAopt QF	pUC57 RNA opt, séquence QF inséré au site HpaI	143.10^3 g/mol	441nt (MscI)

Le plasmide pUC57 RNAopt a été synthétisé par la société Eurogentec.

Le plasmide pUC57 RNAopt PPS0 a été obtenu en insérant la phase codante du gène PPS0 (sans ATG ni codon stop) dans le plasmide pUC57 RNAopt au site de restriction HpaI.

Le plasmide pUC57 RNAopt ATTO a été obtenu en insérant la séquence TCTACTGCTGAACTCGCT dans le plasmide pUC57 RNAopt au site de restriction HpaI.

Le plasmide pUC57 RNAopt ATTObis a été obtenu en insérant la séquence AGCGCTGGA dans le plasmide pUC57 RNAopt ATTO par une mutagénèse pontuelle par PCR. Les amorces utilisées sont :
ATTObis.w : GCTTTCAAAGTTTCTAGCGCTGGAACTGCTGAACTCGCT
ATTObis.c : AGCGAGTTCAGCAGTTCCAGCGCTAGAAACTTTGAAAGC

Le plasmide pUC57 RNAopt ATTO-PK a été obtenu en ôtant le pseudonoeud de l'IBV (PK) dans le plasmide pUC57 RNAoptATTO. Le plasmide a été digéré par l'enzyme de restriction XhoI. La bande d'intérêt a été purifiée et l'ADN a été refermé sur lui même.

[1] La transcription commence au dernier G du promoteur SP6 et se termine au site de clivage du plasmide.

Le plasmide pUC57 RNAopt QF a été obtenu en insérant la séquence CTCGCTGA**AGCT**TCTACTAACAAGACGAATGCTTCTGCTCTC (**AluI**) dans le plasmide pUC57 RNAopt au site de restriction HpaI.

Les constructions ont toutes été vérifiées par séquençage.

2. Les oligonucléotides et fluorophores

Amorces	Séquences	
RNOB	CAGAUCUAUUAAGAGCGGUCGG	ARN 2'-O-allyl ; 3 biotines couplées en 5' et en 3'
H12.3A.c	AAAGAGUUCAGCAGU	ARN couplé avec un ATTO647N en 5' (λabs 644nm ; λem 669nm)
J12.3A.c	AAAUCGCGAAUGCAU	ARN méthylé et couplé avec un ATTO550 en 5' (λabs 554nm ; λem 576nm)
DNAopt D.c	GAGAGCAGAAGCATTCGTC	ADN couplé avec un ATTO565 en 3'
DNAopt A.c	AGTAGAAGCTTCAGCGAG	ADN couplé avec un ATTO647N en 5'

3. Préparation des sous-unités ribosomiques à partir de RRL

Le tampon A (20mM HEPES pH 7,8 ; 150mM KAc ; 6mM MgAc ; 2mM DTT[2]) et la solution pour le coussin de sucrose (20% sucrose dans du tampon A) doivent être filtrés avant d'être utilisés. La solution mère de puromycine est préparée dans de l'eau MilliQ à une concentration de 10mg/ml soit 18mM. Elle est aliquotée et conservée à -20°C.

Mélanger 1ml de RRL (TB127, Promega) avec 3mL de tampon A et 222µl de puromycine. Incuber dans la glace pendant 10min puis à 37°C pendant 10min.

Préparer 4 tubes (#362305 ; Beckmann) avec 3ml de solution avec 20% de sucrose.

Répartir le RRL délicatement sur chacun des coussins et équilibrer les tubes 2 à 2 avec du tampon A. Centrifuger dans un rotor TLA 110 à 44 000rpm (~ 80 000g) et 4°C pendant 4h.

Un culot doit être visible dans chaque tube. Oter le surnageant avec une pipette de 5ml en prenant soin de ne pas contaminer le culot avec les parties contenant du RRL. Le culot est transparent et collant donc bien attaché au fond du tube.

Rincer la surface de chaque culot avec 500µl de tampon A avant de les reprendre **DELICATEMENT** avec 100µl de tampon A par un va et vient avec une pipette. Cette étape est très importante car le culot est très dense et visqueux et il est important que les sous-unités se détachent les unes des autres.
Astuces : Indiquer le culot avec une marque au feutre. Pour rester à 4°C, poser le fond du tube sur la glace avec le culot vers le haut. Régler la pipette (cônes à filtres) sur 75µl afin de ne pas faire de bulles. Ejecter le tampon sur le culot et reprendre.

Récupérer les extraits dans un même tube 1,5ml à 4°C. Laver le fond des 4 tubes avec les mêmes 100µl de tampon A. Ajouter ces 100µL à l'extrait. Centrifuger à 13200rpm pendant une minute et

[2] Toujours utiliser une solution fraiche de DTT. La solution stock peut être aliquotée et congelée.

garder le surnageant. Cette étape a pour but d'éliminer d'éventuels contaminants colorés présents dans le RRL ou des morceaux de culots. Aliquoter les extraits par 12,5µl (à faire dans la glace) et garder à -70°C.

Mesurer la densité optique/ml (DO/ml) de l'extrait à 260nm et 280nm au Nanodrop dans le mode ARN. **ATTENTION**, la DO est directement donnée par ml, il n'y a pas de conversion à faire. Le ratio $\frac{DO\ 260nm/ml}{DO\ 280nm/ml}$ doit être proche de 2 car dans un ribosome, il y a environ deux fois plus d'ARNr que de protéines.

La règle de conversion pour calculer la molarité en sous-unités ribosomiques est :
1 unité de DO 260nm/ml équivaut à 25nM de sous-unité 60S et 50nm de sous-unité 40S. Pour le calcul de la molarité en ribosomes dans l'extrait, le ribosome 80S est approximé à la sous-unité 60S. La concentration sera donc surévaluée même si l'ordre de grandeur est bon.
Exemple : Pour une mesure de 12,047 DO 260nm/ml au Nanodrop, la concentration en ribosome est 12 x 25.10^{-9} M = 3.10^{-7} M soit 300nM.

4. Transcription *in vitro*

Important, les minipreps doivent être réalisées avec des solutions sans RNAses. Par exemple, le tampon de resuspension du kit « Wizard® Plus SV Minipreps DNA Purification System » (Promega) contient des RNAses.

Linéarisation du plasmide et purification de l'ADN
Mélanger : 25 µl H_2O + 15 µl plasmide + 5 µl de tampon 10x + 3 µl d'enzyme de restriction.
Incuber pendent 1h à 37°C et ajouter 2µl d'enzyme avant de continuer pour une heure de plus.

Important : Afin d'ôter tous les sels qui pourraient inhiber la transcription, purifier le plasmide sur une colonne NucleoSpin Extract II (Macherey-Nagel). Eluer avec 15µl d'H_2O

Transcription (elle commence au dernier G du promoteur SP6)
Mélanger : 46µl H_2O DEPC + 20µl de chaque NTP à 10mM + 10µl de DTT à 0,1M + 10µl de tampon 10x SP6 RNA polymérase + 10 µl de plasmide linéarisé + 4 µl de polymérase SP6 (NEB).

Aliquoter 50µl par tube et incuber à 40°C pendant 2h30. Vérifier la présence d'ARN en analysant 2µl d'échantillon par électrophorèse. (Bien laver la cuve avant)

Traitement à la DNaseI et purification
Ce traitement a pour but de dégrader la matrice ADN qui pourrait titrer les oligonucléotides lors de l'appariement.
Ajouter 1µL de DNaseI (NEB) pour 50µl d'échantillon et incuber à 37°C pendant 1h. Vérifier la digestion par électrophorèse.
Pour purifier l'ARNm, utiliser une colonne « Mini quick RNA spin column » (Roche).

Pour la microscopie, mesurer la DO à 260nm et adapter la concentration pour avoir environ 1µM d'ARNm. Pour la traduction sur bille magnétique, concentrer l'ARNm en le précipitant avec une solution de NaAc

Précipitation d'ARN

Pour 1 volume d'échantillon ajouter 1/1 de volume de NaAC à 3M pH5,2 et 3 volumes d'EtOH pur. Garder à -20°C sur la nuit. Centrifuger à 13 200rpm et 4°C pendant 30min. Oter le surnageant et laver le culot avec 100µl d70% EtOH. Centrifuger encore pendant 15min et ôter le surnageant. Reprendre le culot avec 15µl d'H_2O.

5. Expériences de retard sur gel

Solutions :

Tampon B(NaAc) : 20mM Hepes pH7,8 ; 250mM NaAc ; 100mM KAc ; 2,5mM MgAc ; 2mM DDT et 0,25mM spermidine.

Tampon migration : 66mM Hepes ; 34mM Tris pH7,5 et 2,5mM MgCl2

Gel d'agarose : 0,8% agarose dans tampon de migration. Ne pas faire rebouillir sinon des saletés apparaissent dans le gel et contaminent le gel lors de l'analyse de la fluorescence.

1µM **ARNm opt ATTO (XhoI)** ; 10µM **H12.3A.c** ; 300µM **sous-unités ribosomiques**

Appariement des oligonucléotides avec l'ARNm

Mélanger : 1µl ARNm opt ATTO + 3,5µL H2O + 1 µM H12.3A.c + 0,5µL tampon 10X T4 DNA ligase (Promega) soit 0,5µM ARNm, 0,1µM H12.3A.c. Incuber les 10µl à 68°C pendant 5min, 37°C pendant 5min et garder dans la glace.

Faire une dilution au 1/20 dans du tampon B (NaAc). (25nM ARNm, 5nM H12.3A.c)

Accroche des sous-unités ribosomiques sur les ARNm

Mélanger dans 25µl final: 17,5µL tampon B (NaAc) + 5µL d'appariement ARNm/oligos + 2,5µL 300nM sous-unités ribosomiques. Incuber 10min à 37°C et ajouter 5µL de glycérol 80%

Solutions mères des expériences contrôles : 250nM ARNt[3], 250nM ARNm opt QF.

Attention ne pas utiliser de bleu de bromophénol dans les échantillons car les pigments émettent dans le rouge.

Electrophorèse

Déposer un TOUT PETIT PEU de bleu de charge dans un puits éloigné pour avoir un repère de migration. Migrer dans le tampon de migration pendant 35min à 135V.

Scanner le gel immédiatement après la migration. Réglage du Typhoon : Excitation rouge (633m), Emission 670BP30, Sensitivity High, PMT 600V, focal plane +3mm

6. Traduction sur billes magnétiques

Le tampon A (20mM HEPES pH 7,8 ; 150mM KAc ; 6mM MgAc ; 2mM DTT) et le tampon de fixation 5x (100mM HEPES pH 7,8 ; 500mM KAc ; 25mM MgAc ; 5mM DTT) doivent être filtrés avant utilisation.

RRL (#TM232, Promega)

[3] Masse molaire des ARNt : 25.10^3 g/mol

LYS(BODIPY-FL)-ARNt^{LYS} (FluoroTect^{TM} GreenLYS, Promega). La solution stock est estimée à 30µM par Promega, mais ce n'est pas certifié. BODIPY-FL : λabs 502nm, λem 510nm.
400µM **RNOB**, 2,6 µM **ARNm opt PPS0 (KpnI)** et 25nM de **sous-unités ribosomiques** purifiées à partir de RRL

Préparation des billes magnétiques recouvertes de streptavidine
Utiliser un tube de « Streptavidin MagneSphere® Paramagnetic Particles » (Promega) par échantillon. Laver les billes aimantées 3 fois avec 600µl de tampon A.

Appariement de l'ARNm avec l'oligonucléotide RNOB
Mélanger 10µl de ARNmPPS0 + 2,5µl RNOB + 1µl de tampon 10x T4 DNA ligase (Promega) + 6,5µl H_20. Incuber 5min à 68°C, 5min à 37°C et garder dans la glace.

(Concentrations finales : 1,3µM ARNm PPS0 ; 50µM RNOB)

Accroche du ribosome sur l'ARNm/RNOB
Mélanger les solutions suivantes pour obtenir un volume final de 25µl.

Ribosomes (fmol)	0	10	50	250
ARNm/RNOB (pmol)	6,25 pmol ARNm ; 250 pmol RNOB			
ARNm/RNOB (µl)	5	5	5	5
Tp fixation 5x (µl)	5	5	5	5
Sous-unités (µl)	0	0,4	2	10
Tampon A (µl)	15	14,6	13	5

Incuber à 30°C pendant 15min. Ajouter 75µl de tampon A et incuber à température ambiante pendant 15min.

Fixation sur les billes
Après avoir ôté le dernier tampon de rinçage des billes, ajouter les complexes ARNm/RNOB/ribosome sur les billes aimantées pour les fixer à la streptavidine. Laisser incuber à température ambiante pendant 15min puis rincer les billes 2 fois avec 500µl de tampon A.

Traduction avec des LYS(BODIPY-FL)-ARNt^{LYS}
Ajouter sur les billes 47µl de RRL supplémenté avec le mélange d'acides aminés sans méthionine + 1µl L-méthionine (50mM) + 4µl KOAc (1M) + 3µl LYS(BODIPY-FL)-ARNt^{LYS}. Incuber à 37°C pendant 1h30 sous agitation.

Conserver les surnageants et laver les billes 3 fois avec 100µl de tampon A. Reprendre les billes dans 25µl de tampon A.

Analyser en faisant migrer les échantillons dans un gel SDS-PAGE 12%. Avant le dépôt des échantillons, chauffer les extraits seulement 2min à 70°C car les BODIPY-FL sont sensibles à la température. Scanner le gel (sans les plaques de verre car les UV traversent mal le verre) avec le Typhoon (GE Healthcare). Réglages : Excitation bleu (488m), Emission 520BP40, Sensitivity High, PMT 600V.

7. Montage microscopique

Les expériences de microscopie ont été réalisées avec un microscope inversé couplé avec un dispositif de TIRF monté au laboratoire. Le faisceau d'excitation rouge est généré par une diode laser à 638nm (LBX-638-100, Oxxius, 100mW) ayant une puissance maximale de 100mW en continu. La fluorescence est collectée avec un objectif à immersion Apoplan (Olympus) ayant une ouverture numérique de 1,45 et un grossissement x 60. L'huile à immersion utilisée a un indice de réfraction 1,512 (Fluka10976). Les images sont collectées par une caméra EM-CCD Andor (Ixon) à l'aide du logiciel Solis fourni par le constructeur. Tout le système (excitation laser, acquisition des images, auto-focus) est piloté par le programme LabView 8.2 (National Instrument). Les programmes d'analyses d'images ont été codés sous IgorPro 6.20.

8. Chimie de surface

L'amino-silane est du 3-(aminopropyl) triethoxysilane (APTES). Le PEG-NHS est du methyl-PEG-succinimidyl Glutarate, MW 5000 Daltons (Laysan Bio). Le Biotine PEG-NHS est du Biotin-PEGsuccinimidylcarbonate, MW 5000 Daltons (Laysan Bio)

20g PEG-NHS et 5g de PEG-NHS biotinylé sont dissouts dans 370µL d'une solution de tampon borate, pH=8,8. Cette solution est ensuite immédiatement déposée sur les lamelles et incubée pendant 4 heures avant d'être rincée avec du tampon borate puis du tampon Tris.

9. Traduction en cellule microfluidique

Solutions :

Tampon B(NaAc) : 20mM Hepes pH7,8 ; 250mM NaAc ; 100mM KAc ; 2,5mM MgAc ; 2mM DDT et 0,25mM spermidine.

1µM **ARNm opt ATTO (XhoI)** ; 10µM **RNOB** et 10µM **H12.3A.c** ; 280µM **Sous-unités ribosomiques**

Solution mère de **cycloheximide** (500x): 50mg/ml dans EtOH soit 178mM. La solution 10x est préparée dans du tampon B (NaAc).

Appariement des oligonucléotides avec l'ARNm

Mélanger : 5µl ARNm + 2µL H2O + 1,25µl RNOB + 1,25µM H12.3A.c + 0,5µL tampon 10X T4 DNA ligase (Promega) soit 0,5µM ARNm, 1,25µM RNOB et 1,25µM H12.3A.c. Incuber les 10µl à 68°C pendant 5min, 37°C pendant 5min et garder dans la glace.

Faire une dilution au 1/100 dans du tampon B (NaAc) : 98µl de tampon + 1µl hybridation + 1µl SuperaseIn (20u/µL, NEB). (5nM ARNm, 12,5nM RNOB et 12,5nM H12.3A.c)

Accroche des sous-unités ribosomiques sur les ARNm

Mélanger : 17,5µL tampon B (NaAc) + 5µL d'appariement ARNm/oligos + 2,5µL sous-unités ribosomiques. Incuber 10min à 37°C et injecter dans la cellule microfluidique. Laisser incuber 10min avant de rincer avec du tampon B(NaAc) et d'injecter le RRL. (1nM ARNm, 2,5nM RNOB, 2,5nM H12.3A.c et 28nM ribosomes)

Dans le cas du contrôle sans ribosome, mélanger 20µL tampon B (NaAc) + 5µL d'appariement ARNm/oligos.

Dans le cas du contrôle avec cycloheximide, mélanger : 15µL tampon B (NaAc) + 5µL d'appariement ARNm/oligos + 2,5µL sous-unités ribosomiques + 2,5µL 10x cycloheximide. Le rinçage de la cellule microfluidique est réalisé avec du tampon B (NaAC) 1x cycloheximide.

Préparation du RRL

Mélanger : 17,5µL de RRL (Promega) + 6µL H2O + 1µL SuperaseIn (20u/µL) + 0,5µL acide aminé-Met (Promega). Incuber 10min à 30°C avant de l'injecter dans la cellule microfluidique.

Dans le cas du contrôle avec cycloheximide, mélanger : 17,5µL de RRL (Promega) + 3,5µL H2O + 1µL SuperaseIn (20u/µL) + 0,5µL acide aminé-Met (Promega) + 2,5µL 10x cycloheximide.

10. Purification des sous-unités 40S et 60S à partir de *S. cerevisiae*

Protocole adapté de (Acker et al. 2007)

Solutions : Toutes doivent être filtrées sur membrane 0,2µm.

Solution 10x tampon HKM : 200mM Hepes(KOH) pH7,4 ; 1M KAc et 25mM MgAc

Héparine (#H3393-1MU, Sigma): solution mère à 100mg/ml dans 1x tampon HKM. Aliquoter par 1ml et garder à -20°C.

Tampon de lyse : 1x tampon HKM, 2mMDTT, inhibiteurs de protéases (Roche) et 1mg/ml d'héparine ajoutée extemporanément.

Solution coussin de sucrose 34% : 1x tampon HKM, 500mM Cl, 1M sucrose, 2mM DTT

Tampon de séparation des sous-unités : 50mM Hepes(KOH) pH7,4, 500 mM KCl, 2mM $MgCl_2$, 2mM DTT et 1mg /ml d'héparine ajoutée extemporanément.

Puromycine (10µg/µl) : solution mère 18,3mM dans H_2O MilliQ

Solutions gradient de sucrose : une avec 40% sucrose et l'autre sans. 50mM Hepes(KOH) pH7,4, 500mM KCl, 2,5mM MgCl2 et 2mM DTT.

Solution de stockage : 1x tampon HKM, 250mM sucrose, 2 mM DTT.

Cultures des cellules

Le matin, mettre en pré-culture la souche 74-D694 [psi-] dans du YPG. Le soir, ensemencer 6 cultures de 500ml de manière à avoir une concentration de 1 DO 600nm/ml le lendemain matin. Il n'est pas utile d'aller à une concentration supérieure à 1 car même si la quantité de cellules augmente le rendement d'extraction diminue.

Cassage des cellules

Culoter les 3L de cultures et congeler le culot de cellules (~ 8ml) dans l'azote liquide. Broyer ensuite le culot au pilon et au mortier jusqu'à obtenir une poudre aussi fine que du fard. Astuces : le broyage est optimal quand la quantité d'azote liquide permet d'obtenir une consistance de sable mouillé. Récupérer la poudre dans un tube 50ml et ajouter 40ml de tampon de lyse glacé contenant 1mg/ml d'héparine. L'héparine permet d'inactiver les RNAses. Secouer vigoureusement pour bien dissoudre l'extrait cellulaire. Afin d'ôter les débris, centrifuger à 13 000rpm et 4°C pendant 25min. Récupérer le surnageant (~ 36ml). Garder l'extrait dans la glace.

Isolation des ribosomes sur coussin de sucrose

Préparer 2 tubes (Beckmann #355 618) en déposant 2,5ml de coussin de sucrose glacé au fond. Ajouter les extraits au-dessus DELICATEMENT. Equilibrer les tubes avec du tampon de lyse contenant de l'héparine. Centrifuger à 4°C et 60 000rpm pendant 2h dans un rotor 70Ti (Beckmann-Coulter).

Le culot est visible au fond du tube, il est translucide et bien collé au fond du tube. Oter le surnageant avec une pipette 10ml en prenant soin de ne pas contaminer le culot avec la couche de lipides.

Laver rapidement chaque culot avec 200µl de tampon de séparation contenant 1mg/ml d'héparine. Reprendre chaque culot avec 250µl de tampon de séparation contenant de l'héparine. Cette étape est très importante pour bien séparer chaque molécule. Il faut éviter les bulles et rester sur de la glace. Transférer les extraits dans un tube 1ml. Laver les tubes avec 250µL de tampon et les ajouter à l'extrait. Centrifuger l'extrait 1min à 13 200rpm pour faire tomber d'éventuels débris et garder le surnageant.

Traitement à la puromycine pour séparer les sous-unités

Mesurer la DO/ml à 260 et 280nm au Nanodrop. Pour la mesure, diluer l'échantillon au 1/1000. Le rapport des concentrations 260nm/280nm doit être proche de 2.

Diluer l'échantillon avec du tampon de séparation contenant de l'héparine afin d'obtenir une concentration finale de 150 DO_{260nm} /ml.

Ajouter de la puromycine à l'extrait de manière à obtenir une concentration finale de 1mM. Laisser agir dans la glace pendant 20min puis incuber à 37°C pendant 15min. Garder l'échantillon dans la glace

Isolation des sous-unités ribosomiques sur un gradient de sucrose

Préparer le gradient de sucrose 10-30% dans 12 tubes (Beckman #326 823). Chaque couche a un volume de 5ml. Du fond du tube vers le haut : 30%, 27%, 23%, 20%, 17%, 13% et 10%. Déposer 300µl d'extrait sur chaque gradient. Réserver 6 préparations à 4°C et centrifuger les autres à 4°C et 29 000rpm pendant 7h30 dans un rotor SW32 (Beckman). Le lendemain matin, faire la deuxième centrifugation pendant l'analyse des 6 premiers tubes.

Collecte des sous-unités ribosomiques.

Utiliser le collecteur de fraction avec la tige entrant par le haut du tube (ISCO). Utiliser une solution de sucrose 60% pour pousser le gradient dans le capillaire et des tubes 2ml pour collecter les fractions. Réglages : délai 12s, sensibilité 2, collecte 30S, vitesse 100 x10 et défilement du papier 60cm/h. Collecter séparément les fractions correspondant aux sous-unités 40S et 60S.

Dialyse et concentration des échantillons

Utiliser des colonnes Amicon ultra (Millipore) ayant un seuil d'exclusion 100kD. Laver l'extrait avec 3V de tampon de stockage. Reprendre le culot avec le plus petit volume possible (500µl). Attention, il faut bien laver la membrane car les protéines précipitent dessus.

Mesurer la DO/ml à 260 et 280nm dans une **cuve de quartz**, en diluer 1µl d'extrait dans 1ml d'H_2O. Conversion 1DO/ml à 260nm correspond à 25nM de sous-unités 60S et 50nM de sous-unités 40S.

Fidélité de la traduction chez les eucaryotes : de la molécule au génome.

Résumé :

Ce travail porte sur l'étude de la fidélité de la traduction chez les eucaryotes d'un point de vue mécanistique et génomique. Au cours de ma thèse j'ai développé trois approches :

Le premier projet porte sur l'étude du rôle du facteur de l'élongation eEF2 dans le maintien du cadre de lecture. La stratégie associe une mutagénèse aléatoire du gène *EFT2* à un criblage phénotypique, elle permet d'isoler des mutants capables d'augmenter ou diminuer l'efficacité de recodage d'une séquence de décalage du cadre de lecture en -1.

Le second projet décrit la mise au point d'un système de traduction en molécule unique qui permet d'étudier le ribosome eucaryote. La traduction est initiée grâce à l'IRES CrPV qui a pour caractéristique d'être totalement indépendante des facteurs d'initiation et de l'ARNt initiateur. L'élongation de la traduction est détectée grâce au départ d'un oligonucléotide fluorescent qui est décroché par l'activité hélicase du ribosome. Les résultats de ces expériences constituent une preuve de principe démontrant que l'étude de la traduction eucaryote en molécule unique est possible.

Le troisième projet est une étude de génomique comparative qui permet de rechercher des événements de recodage ainsi que d'autres événements non-conventionnels de la traduction dans le génome de la levure *Saccharomyces cerevisiae*. L'approche est basée sur une recherche d'organisations génomiques conservées au sein de 19 génomes de levures. Les gènes candidats sont testés *in vivo* grâce à un vecteur double rapporteur. Cette étude a permis de mettre en évidence le gène VOA1 qui a été ensuite caractérisé plus en détails.

Mots clés : Fidélité de la traduction, Recodage, Ribosome, Eucaryotes, eEF2, Molécule unique, *Saccharomyces cerevisiae*, *VOA1*.

Translational fidelity in eukaryotes: from the molecule to the genome.

Abstract :

This report describes a study of translation fidelity in Eukarya. Two aspects are tested: the molecular mechanism of recoding and the research of recoding events at the genomic level. During my PhD I have developed three projects:

The first project deals with the role of the elongation factor eEF2 in reading frame maintenance. The strategy is based on a random mutagenesis of EFT2 and a phenotypic screening in order to isolate mutants increasing or reducing -1 frameshifting efficiency.

The second project describes the development of a single molecule translational system to study the eukaryotic ribosome. Translation initiation is mediated by the CrPV IRES which is initiation factor and initiation tRNA independent. Elongation is monitored with the dissociation of a fluorescent oligonucleotide by the helicase activity of the ribosome. This work is a proof of principle that studying eukaryotic ribosome with single molecule techniques is now feasible.

The third project is a comparative genomic approach to search for recoding and unconventional translational events in the genome of yeast *Saccharomyces cerevisiae*. The approach is based on the detection of conserved genomic organization among 19 Fungi genomes. The candidate genes are then tested in vivo with a dual reporter system. This study allowed the characterization of VOA1 which was further analysed.

Keywords : Translation fidelity, recoding, Ribosome, Eukaryotes, eEF2, Single molecule, *Saccharomyces cerevisiae*, *VOA1*.