UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institute de Génétique et de Biologie Moléculaire et Cellulaire

THÈSE présentée par :

Mónica PEÑA LUNA

soutenue le : **21 septembre 2021**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

Functional study of the Elongator complex from activity to diseases

Characterization of Kti12, a cofactor with a kinase domain

THÈSE dirigée par :

Dr. SÉRAPHIN Bertrand Directeur de recherche, Université de Strasbourg

RAPPORTEURS :

Dr. GOLINELLI-PIMPANEAU Beatrice Directeur de recherche, Collège de France

Prof. VAN TILBEURGH Herman Professeur des universités, Université Paris-Saclay

---

AUTRES MEMBRES DU JURY :

Dr. MARÉCHAL-DROUARD Laurence Directeur de recherche, Université de Strasbourg

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank the University of Strasbourg, particularly the Doctoral School ED414 'Sciences de la Vie et de la Santé', for the support, courses and activities organized that contributes to my development as a PhD candidate.

I am grateful to the 'Institut de Génétique et de Biologie Moléculaire et Cellulaire' (IGBMC) for providing me with a high-quality environment and internationally competitive place to perform my research

I want to thank the Mexican National Council of Science and Technology (CONACYT) for awarding me a fellowship to support the first three years of my PhD studies in France. Fellowship number: 471716, CVU: 330482.

I want also to thank the 'Fondation pour la Recherche Médicale' (FRM) for their fellowship to support my 4th year of PhD, which allowed me to conclude this project.

Special thanks to the PhD associations: ADDAL, SPB and BIOTechno for all the experience, self-development and good times.

I would like to thank Dr. Bertrand SÉRAPHIN and his lab ‘Protein networks and complexes regulating eukaryotic mRNA decay’ for the opportunity to do my PhD with them. I am very grateful to the whole team for these four years spent in the lab, time in which I acquired plentiful scientific knowledge and developed numerous transferable skills. I am especially grateful to Bertrand and Eric HUNTINGER for guiding me, for all their contributions to my project, for their time spent making drawings to explain me everything from basic concepts to experiment design; as well as for their patience, amiability and motivation during the last four years. You demanded a lot from me, but you gave me a lot more. I admire both of you so much!

I would also like to thank the jury members:

Dr. Beatrice GOLINELLI-PIMPANEAU

Dr. Laurence MARÉCHAL-DROUARD

Prof. Herman VAN TILBEURGH

for their time reading and improving my thesis and contributing to my scientific development.

I extend special thanks to the members of my ‘Comité de suivi de thèse’ (CST): Dr. Albert WEIXLBAUMER and Dr. Philippe GIEGÉ their time, support, valuable advices as well as for sharing their knowledge during my PhD project.

I am very grateful to all my lab ‘family’: Bertrand SÉRAPHIN, Claudine GAUDON-PLESSE, Eric HUNTZINGER, Fabienne MAUXION, Guillaume CAULIER, Hamza AMINE, Jeremy SCUTENAIRE, Jordan SINTEFF, Mélody MATELOT, Mélanie MAHE, Severine MANGOLD and Yasmine AMRANI for making me feel welcome since the first time I arrived, for being my teachers in the French culture, for teaching me French, for giving me a second family.

Thanks to all my friends who have made the last four years some of the best of my life

Also thanks to the Mexican team: Dana DÍAZ-JIMÉNEZ, Israel TORRES-CASTRO, Jessica Danielly MEDINA-SÁNCHEZ, José RIVERA-ÁLVAREZ and Roberto BAHENA-CERON for your friendship and for making me feel in Mexico while a thousand kilometers away. ¡Los quiero!

I would like to thank Alexey ROZOV, for his personal and scientific support, for inspiring me and accompanying me on this journey and for being generally useful.

Finally, I want to thank the two people I admire the most, my parents:

Lourdes LUNA-AMOZURRUTIA

Jaime PEÑA-VARGAS

For their love, motivation, guidance and invaluable support during all my life. For encouraging me to pursue my dream. This work symbolizes my gratitude for everything you have gave me; and thanks to you I became what I am and came to where I am.

ABSTRACT

The transcription of genetic information from DNA into messenger RNA (mRNA) and its translation into proteins lies at the heart of gene expression. Ribosomes 'read' the sequence of the mRNA and translate it into a protein sequence by sequential addition of amino acids carried by transfer RNAs (tRNAs) that recognize specific three-nucleotide 'codons' in the mRNA. Position 34 in tRNAs is known as the 'wobble' position because it allows several types of non-Watson-Crick base paring during codon recognition which is critical for the proper translation of the genetic code.

The tRNAs of all organisms are chemically modified by enzymes such as methylases, thiolases and hydroxylases among others. These modifications contribute to the stability, folding and function of the tRNA. Modifications of U residues at the wobble position (U_{34}) are particularly important because they contribute to the efficiency and fidelity of translation.

Elongator is a protein complex present in eukaryotes whose main function is the post-transcriptional modification of U_{34} in tRNAs to form 5-carboxylmethiluridine (cm^5U) and its derivatives. It is a dodecamer of six subunits, Elp1–Elp6, each of which is present in two copies. In addition, to perform its enzymatic activity, Elongator requires four proteins of the killer insensitive toxin (Kti) protein family (Kti11- Kti14). We previously demonstrated the roles of Kti11 and Kti13 as electron donors to the iron-sulfur (Fe-S) cluster in Elp3. Also, Kti12 was proposed to regulate Elp1 phosphorylation together with Kti14; however, its detailed molecular function remains ill-defined.

Alterations of post-transcriptional events and disruptions of translational control play a major role in human diseases, including neurological disorders and cancer. In humans, Elongator germline mutations are associated with neurological disorders such as Familial Dysautonomia resulting from mutation in the Elp1 homologue-coding gene, while alterations of the expression of several Elongator subunits promote metastasis of bladder and breast cancer cells, as well as the survival of melanoma cells.

In this study, I investigated the structure and function of Kti12 from the yeast *Saccharomyces cerevisiae*. I showed that Kti12 has sequence similarities with the *O*-phosphoseryl-$tRNA^{Sec}$ kinase (PSTK) and therefore that it is composed of two domains: the N-terminal domain (NTD) contains a kinase signature whereas the C-terminal domain (CTD) binds RNA with high affinity to tRNA. Electrophoretic mobility shift assays showed that the CTD bind tRNAs indiscriminately, independently of whether or not they are modified by Elongator. The $Kti12_{CTD}$ also showed a preference to bind tRNA over an equivalent DNA molecule.

S. cerevisiae strains lacking Kti12 showed phenotypes similar to those observed for cells lacking Elp3, the catalytic subunit of Elongator. The latter mutation prevents the Elongator-dependent post transcriptional modification of U_{34} in tRNAs. Despite the fact that both domains of Kti12 are required for efficient Elongator-dependent modifications, the NTD deletion has a

stronger phenotypic impact compared to the CTD deletion. The CTD domain seems to have an accessory role. Interestingly, substitution of the catalytic kinase residues of Kti12, did not eliminate its role in Elongator function, suggesting that $Kti12_{NTD}$ is important for Elongator function but not through a kinase activity.

Kinase assays revealed the presence of a kinase associated to Kti12 and Elongator responsible for the formation of 150 kDa phosphoprotein identified as Elp1. Level of Elp1 phosphorylation decreases when the $Kti12_{CTD}$ is deleted and is completely abolished when the $Kti12_{NTD}$ is deleted. Elp1 remained phosphorylated, albeit at reduced level, in strains carrying substitutions of the two of the catalytic kinase residues of Kti12, indicating that another kinase is responsible for Elp1 phosphorylation. Our experiments and published data suggest that this other kinase is Kti14.

Co-immunoprecipitation assays performed to assay interaction of Kti12 with Kti14 showed a strong reduction of levels of Kti14 co-precipitation in the absence of Kti12 or of its NTD. In contrast, in the absence of the $Kti12_{CTD}$, Kti14 still co-precipitates. Recombinant protein expression also provided evidence for a direct interaction between Kti12 and Kti14 mediated by the N-terminal domain of the former.

My observations support a mechanistic model in which Kti12 recruits Kti14 via its N-terminal domain in order to phosphorylate Elp1. However, additional assays are needed to fully understand the role of Kti12 in the Elongator dependent modifications.

RÉSUMÉ

La transcription de l'information génétique de l'ADN en ARN messager (ARNm) et sa traduction en protéines sont au cœur de l'expression des gènes. Les ribosomes « lisent » la séquence de l'ARNm et la traduisent en une séquence protéique par addition séquentielle d'acides aminés portés par des ARN de transfert (ARNt) qui reconnaissent des « codons » de trois nucléotides dans l'ARNm. La position 34 des ARNt est connue sous le nom de position « wobble », car elle permet plusieurs types d'appariement de bases non Watson-Crick lors de la reconnaissance des codons, ce qui est essentiel pour une traduction correcte du code génétique.

Les ARNt de tous les organismes sont modifiés chimiquement par des enzymes telles que des méthylases, des thiolases et des hydroxylases... Ces modifications contribuent à la stabilité, au repliement et à la fonction de l'ARNt. Les modifications des résidus U à la position wobble (U34) sont particulièrement importantes car elles contribuent à l'efficacité et à la fidélité de la traduction.

Elongator est un complexe protéique présent chez les eucaryotes dont la fonction principale est la modification post-transcriptionnelle de U_{34} dans les ARNt pour former la 5-carboxylméthiluridine (cm^5U) et ses dérivés. C'est un dodécamère de six sous-unités, Elp1-Elp6, dont chacune est présente en deux exemplaires. De plus, pour exercer son activité enzymatique, Elongator a besoin de quatre protéines de la famille des protéines « insensibles de la toxine tueuse » (« Killer-toxin insensitive » : Kti11-Kti14). Nous avons précédemment identifié les rôles de Kti11 et Kti13 en tant que donneurs d'électrons pour le cluster fer-soufre (Fe-S) de Elp3. En outre, Kti12 a été proposé réguler la phosphorylation d'Elp1 par Kti14 ; cependant, sa fonction moléculaire détaillée reste mal définie.

Les altérations des événements post-transcriptionnels et le contrôle de la traduction jouent un rôle majeur dans les maladies humaines, notamment pour les troubles neurologiques et le cancer. Chez l'homme, les mutations germinales d'Elongator sont associées à des troubles neurologiques tels que la dysautonomie familiale résultant de l'altération du gène codant l'homologue d'Elp1, tandis que l'altération de l'expression de plusieurs sous-unités d'Elongator favorise la métastase des cellules cancéreuses de la vessie et du sein, ainsi que la survie de cellules du mélanome.

Dans cette étude, j'ai étudié la structure et la fonction de Kti12 de la levure *Saccharomyces cerevisiae.* J'ai montré que Kti12 a des similarités de séquence avec la O-phosphoseryl-tRNASec kinase (PSTK) et donc qu'elle est composée de deux domaines : le domaine N-terminal (NTD) qui contient une signature kinase alors que le domaine C-terminal (CTD) se lie ARN avec une haute affinité pour les ARNt. Des tests de compétition analysés par retard électrophorétique ont montré que le domaine CTD se lie indistinctement aux ARNt qu'ils soient ou non modifiés par Elongator. Le CTD de Kti12 a également une préférence pour lier un ARNt à une molécule d'ADN équivalente.

Les souches de *S. cerevisiae* dépourvues de Kti12 ont montré des phénotypes similaires à ceux observés dans des cellules dépourvues d'Elp3, la sous-unité catalytique d'Elongator. Cette dernière mutation empêche la modification post-transcriptionnelle des U_{34} dans les ARNt catalysée par Elongator. Malgré le fait que les deux domaines de Kti12 soient nécessaires pour des modifications efficaces des ARNt par Elongator, la suppression du NTD de Kti12 a un impact phénotypique plus important que la suppression son CTD. Le domaine CTD semble donc avoir un rôle accessoire. Fait intéressant, la substitution des résidus catalytiques de kinase de Kti12 n'élimine pas son rôle dans la fonction Elongator, indiquant que $Kti12_{NTD}$ est important pour la fonction Elongator mais pas par une activité de kinase.

Un test d'activité kinase a révélé la présence d'une kinase associée à Kti12 et Elongator responsable de la formation d'une phosphoprotéine de 150 kDa identifiée comme Elp1. La phosphorylation d'Elp1 diminue lorsque le domaine CTD de Kti12 est délété mais le signal est aboli lorsque le domaine NTD de Kti12 est absent. Elp1 est reste phosphorylé, bien qu'à un niveau réduit, dans les souches portant la substitution des deux résidus catalytique du domaine kinase de Kti12, indiquant qu'une autre kinase est responsable de la phosphorylation d'Elp1. Nos données et des arguments de la littérature suggèrent que cette autre kinase est Kti14.

Des tests de co-immunoprécipitations effectués pour analyser l'interaction de Kti12 avec Kti14 ont montré une forte réduction de la co-précipitation de Kti14 en l'absence de Kti12 ou de son domaine NTD. En revanche, en l'absence du domaine CTD de Kti12, Kti14 coprécipite toujours. L'expression de la protéine recombinante fournit également des évidence indiquant une interaction directe de Kti12 avec Kti14 médiée par le domaine N-terminal du premier.

Mes observations argumentent pour un modèle mécanistique dans lequel Kti12 recrute Kti14 via son domaine N-terminal pour phosphoryler Elp1. D'autres tests seront nécessaires pour bien comprendre le rôle de Kti12 dans les modifications des ARNt effectuées par Elongator.

LIST OF FIGURES

Introduction

Materials and Methods

Results

Discussion

Appendix

LIST OF TABLES

Introduction

Materials and Methods

Results

LIST OF ABBREVIATIONS

Ab	antibody
AMP, ADP, ATP	adenosine 5'-mono, di-, and triphosphates
A	adenine residue
Arg	arginine residue (also R)
Asp	aspartic acid residue (also D)
ATPase	adenosine triphosphatase
bp	base-pair
C-terminal	carboxy-terminal
CoA	coenzyme A
cpm	counts per minute
CRISPR	clustered regularly interspaced short palindromic repeats
CSM	Complete Supplement Mixture
Cys	cysteine residue (also C)
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
G	guanine base
Glu	glutamic acid residue (also E)
GTP	guanosine 5'-triphosphate
h	hour
His	histidine residue (also H)
KD	dissociation constant
kDa or Da	kilodalton, dalton
Leu	leucine residue (also L)
Lys	lysine residue (also K
Met	methionine residue (also M)
min	minutes
mRNA	messenger RNA
MS	mass spectrometry/spectroscopy
N-terminal	amino-terminal
nm	nanometers
nt	nucleotides
ºC	degrees Celsius
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction

PEG	polyethylene glycol
pH	negative logarithm of hydrogen ion concentration
P_i	inorganic phosphate
Pro	proline residue (also P)
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
S	Svedberg sedimentation unit (10-18 s)
SDS	sodium dodecyl sulfate
SUP4	suppressor tRNA Tyr
T	thymine residue
tRNA	transfer RNA
U	uracil residue
WT	wild-type
YPDA	Yeast Peptone Dextrose Adenine
β-gal	β-galactosidase

I. INTRODUCTION

1.1 THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology describes the flow of genetic information of living organisms. In other words, it explains how the hereditary material inside the cell governs three main processes: replication, transcription and translation; involving three types of molecules: DNA, RNA and proteins.

This dogma was originally presented in 1957 by the Nobel Laureate Francis Crick at a scientific lecture held at University College London, and officially published one year after[1]. This new statement was revolutionary and generated controversy among scientists due to the lack of experimental evidence supporting Crick´s idea. At that time, it was not generally accepted that nucleic acids were involved in protein synthesis. Crick proposed four routes for information transfer:

- DNA → DNA: DNA replication
- DNA → RNA: transcription, first step of protein synthesis
- RNA → protein: translation, second step of protein synthesis
- RNA → RNA: RNA virus replication

There were two extra information transfer mechanisms for which there was no evidence, but that Crick did not discard:

- DNA → protein: this suggested that RNA was not a mandatory intermediate in protein synthesis
- RNA → DNA: structurally possible, but without known biological function at that time

And three steps that he considered impossible due to both lack of evidence and lack of biochemical mechanism:

- protein → protein
- protein → RNA
- protein → DNA

The two last points underlined an important feature of unidirectional transfer of information from nucleic acids to proteins, but not backwards.

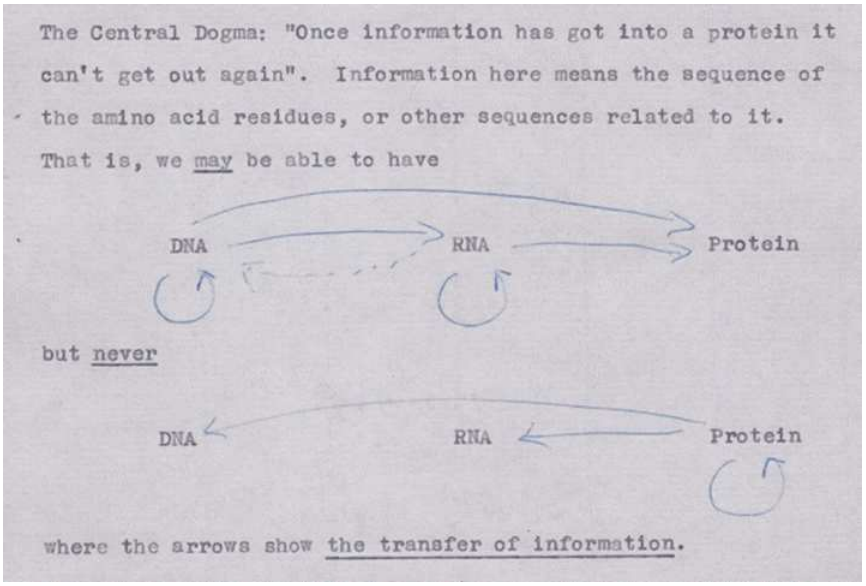


Figure 1. Crick's first outline of the central dogma
Drawing obtained from an unpublished note made in 1956. Taken from [2].

Nowadays, the central dogma of molecular biology is widely and erroneously presented as a linear flow which allows information to travel only one way:

DNA → RNA → protein

This concept builds on the ideas raised already in the 40s and 50s by authors such as Jean Brachet[3], André Boivin and Alexander Dounce and was popularized by James Watson in 1965 trough his book '*Molecular Biology of the Gene*'[4] *(reviewed in [2])*. It took until 1970, when the reverse transcriptase was discovered in the independent experiments of Satoshi Mizutani, Howard Temin, Renato Dulbecco and David Baltimore, to prove that information can also flow in the direction 'RNA → DNA' (reviewed in [5]). This discovery put Crick´s work into spotlight again, and lead to emergence of different scientific stances. Also in 1970, 12 years after his original publication, Crick published an article called Central Dogma of Molecular Biology where he restated that he never discarded the possibility of the RNA → DNA direction of the genetic information flow[6] (Figure 2A). This paper came out as a reply to an anonymous article published in Nature called Central Dogma Reversed which argued that the central dogma, as formulated by Crick, was likely to be a considerable over-simplification of the keystone process of molecular biology[7]. Nevertheless, further experiments demonstrated beyond doubt that the central point of the dogma is correct in living cells; however, some modifications to the basic scheme must be done (Figure 2B). In eukaryotes and prokaryotes, the information follows the primary route DNA → RNA → protein via transcription and translation. The DNA → DNA flow represents the way this molecule replicates. Some viruses utilize the RNA → RNA transfer bypassing the need for DNA template which is called RNA replication. Finally, the RNA → DNA transfer is used by a special kind of viruses called retroviruses. There is no evidence *in vivo* of protein going back to RNA or DNA molecules, as the translation step appears to be strictly unidirectional[8]; and the DNA → protein route has been demonstrated only under artificial conditions[9–11].

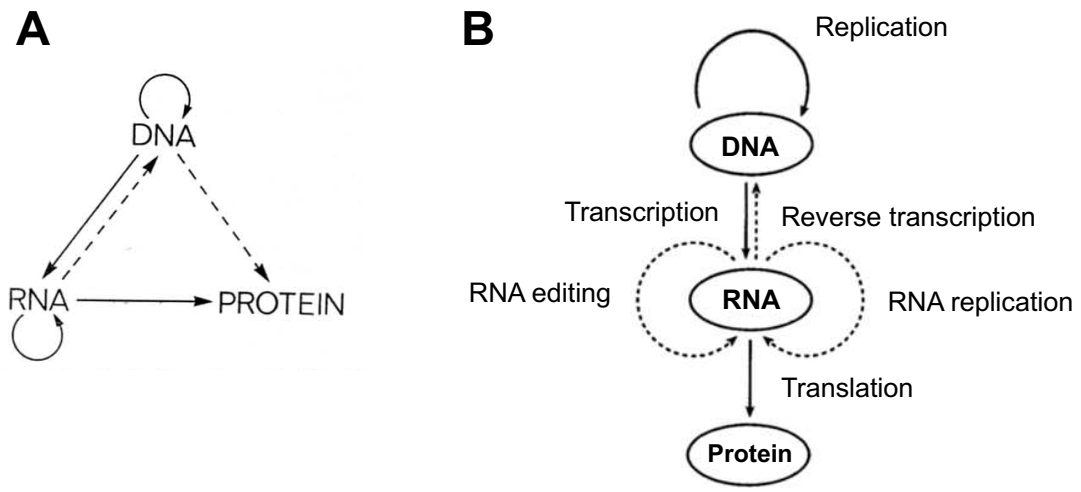


Figure 2. Central dogma of molecular biology

(A) Flow of genetic information proposed by Crick in 1958. Solid arrows represent probable transfers and the dotted arrows possible transfers[6]. **(B)** Flow of genetic information as is known nowadays. Solid arrows represent known wide-spread transfers in living organism. Dotted arrows represent exceptions to the dogma.

1.1.1 Deoxyribonucleic acid (DNA)

Deoxyribonucleic acid (DNA) is the molecule that contains the hereditary material of almost all organisms (including some viruses). In other words, it contains the instructions needed to reproduce. In eukaryotes, these instructions are mostly found inside the cell's nucleus and are passed down from parents to their children. DNA is also found in some cell organelles such as mitochondria and chloroplasts.

1.1.1.1 DNA structure

Biologically, DNA is primarily organized as two polynucleotidic chains that coil one around the other forming a double helix. Each strand is composed of numerous copies of monomeric subunits, named nucleotides. Every nucleotide consists of a nitrogen base, a five-carbon sugar and a phosphate group (Figure 3A). DNA takes its name from the 2'-deoxyribose pentose that forms its backbone. The sugars are linked via phosphate groups attached to the 5' and 3' positions on the pentose cycle by phosphodiester bonds forming the DNA backbone. The nitrogen base, which gives name to each nucleotide, is linked to the pentose by a glycosidic bond at the position N1 of pyrimidines or N9 of purines. The nitrogen bases are divided into two groups: purines and pyrimidines. The four most common types of nitrogen bases are purines adenine (A) and guanine (G) and pyrimidines thymine (T) and cytosine (C), however, their modified derivatives can be present as well (Figure 3B). For example, some viruses, such as the ones infecting photosynthetic bacteria, contain a fifth base, the 2-aminoadenine. This base was discovered in 1977 in the Soviet Union, but it resurfaced in the scientific community only recently. The 2-aminoadenine (dubbed Z) pairs with T via triple hydrogen bond resulting in a more stable pair than the ordinary A-T[12]. Nitrogen bases join the two DNA strands via complementary hydrogen bonds between purines of

one chain and pyrimidines of the other (A-T or G-C), creating the antiparallel structure of the molecule (Figure 3C).

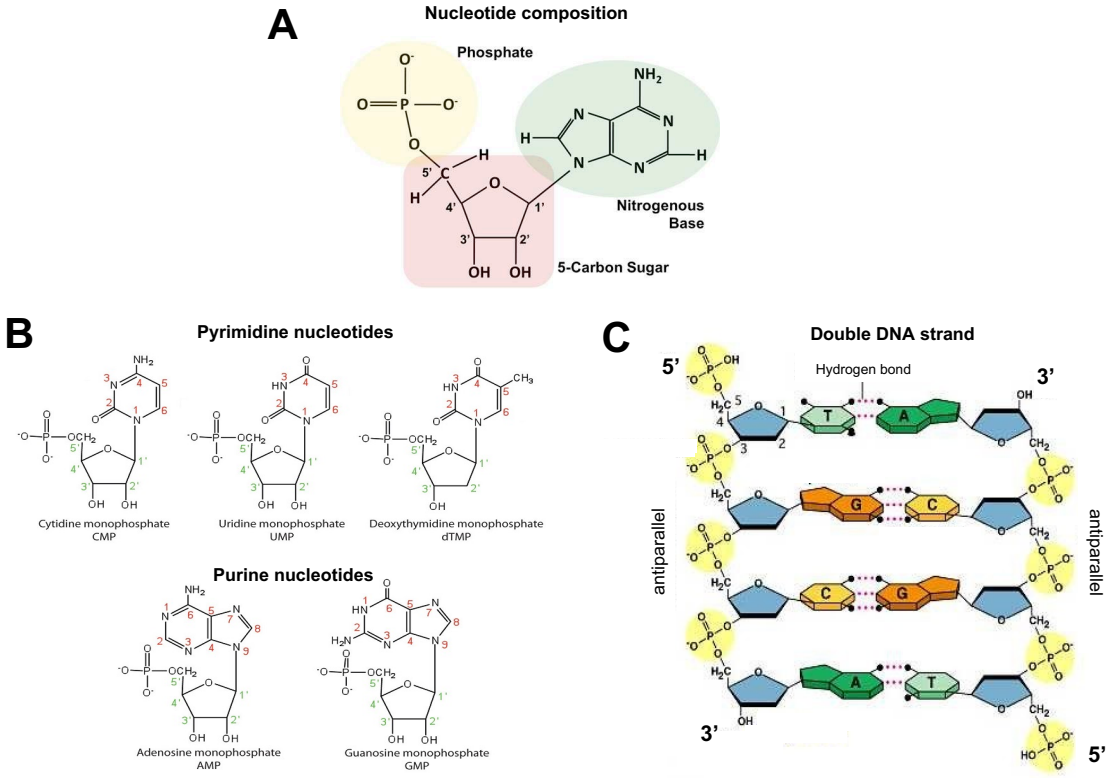


Figure 3. Deoxyribonucleic acid structure

(A) Nucleotide composition: phosphate group (yellow), sugar (pink), nitrogenous base (green). **(B)** Structure of nitrogenous bases found in DNA and RNA. **(C)** Structure of the DNA double helix. Two single strands are joined by hydrogen bonds between purines and pyrimidines. Taken from [13–15], respectively.

DNA molecules differ by the unique order of the nucleotides that compose them; so even in organisms of the same species, it constitutes a unique sequence of nucleotides which encodes the genetic information of the organism. In the DNA double helix, the terminal nucleotide of a DNA strand has a free 5' phosphate group; the terminal nucleotide at the other end has a free 3' OH group. Conventionally, the nucleic acid sequences are written in a 5' → 3' direction. DNA usually adopts a right-handed double helix structure that has distinct minor and major grooves.

Most eukaryotic DNA molecules are linear; nevertheless, some viral, prokaryotic and organelles genomes contain circular DNA molecules. In many cases DNAs are extremely long molecules; for example, every chromosome in animals and plants is formed by only two DNA molecules. Hence, in humans, the two DNA molecules forming chromosome 1 are 8.5 cm long each. Their accommodation in cells is possible due to the DNA ability to wrap around packaging proteins such as nucleosomes in eukaryotes. In humans, nuclear DNA contains about 3 billion bases, while in mitochondria, the organelles which generate the cell´s energy, DNA consists of 16,569 base pairs encoding 13 proteins. Genome sizes of other organisms are compared in table 1.

Table 1. Comparison of the genome and proteome of model organism[16].

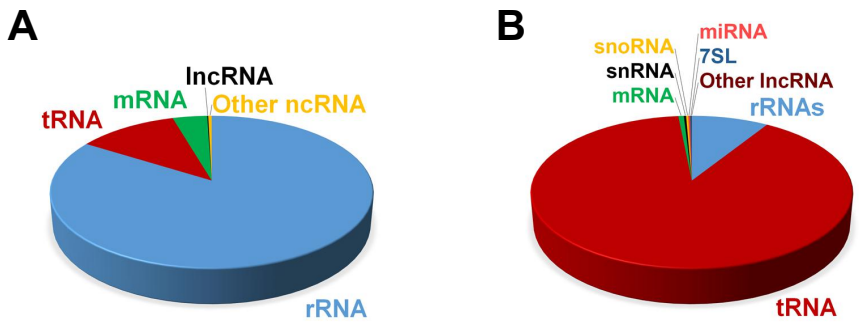
Species	Genome size (Mb)	Protein count
Homo sapiens (humans)	2,996.4	71,340
Mus musculus (mouse)	2,671.8	39,412
Danio rerio (zebrafish)	1,391.7	47,861
Gallus gallus (chicken)	1230.3	46,393
Drosophila melanogaster (fruit fly)	148.5	30,443
Caenorhabditis elegans (nematode)	100.7	28,026
Saccharomyces cerevisiae (yeast)	12.3	5,404
Escherichia coli (bacterium)	5.2	4,931

A crucial property of the double helix is its ability to separate the two strands without disrupting the covalent bonds (between phosphate and the sugar groups) of the backbone. This characteristic allows helix to dissociate and reform under specific physiological conditions. This property is crucial for DNA functions and is used every day by scientists to manipulate these molecules [17].

1.1.2 Ribonucleic acid (RNA)

The ribonucleic acid (RNA) is a molecule structurally similar to DNA; however, with the exception of some viruses, RNA does not carry the genetic material and does not act as a template for its own replication. The RNA molecule has three main differences compared to DNA. First, the RNA backbone is composed of ribose sugar instead of 2'-deoxyribose; it has a hydroxyl group (OH) in its 2' position. Second, the RNA contains uracil nitrogen base instead of thymine; uracil has the same structure as thymine except it lacks the methyl group in position C5. Third, RNA is generally a single-stranded molecule. As a consequence, RNA tends to fold back on itself to form short double helix regions with complementary sequences. If the 2 complementary sequences are close enough the RNA can adopt a stem-loop or hairpin structure. Base paring can also occur between more distant sequences to form pseudoknots and more complex structures. An important feature of RNA is its propensity to form a non-Watson-Crick base pairs (when guanine pairs with uracil (G:U) in its double-helical structures. Due to this and other non-canonical pairing as well as wide rotational freedom of its backbone, RNA is able to form various stable complex tertiary structures (reviewed in [4,17–19]). Such variety allows some types of RNA, known as ribozymes, to have catalytic activity similar to the protein enzymes (reviewed in [20,21]). Ribozymes can be divided into two families: large ribozymes which are typically larger that 200 nucleotides (e.g., introns group I, II and RNase P), and small nucleolytic ribozymes (hammerhead, hepatitis delta virus (HDV), hairpin Varkud satellite (VS)). Small ribozymes often perform a reaction in which the 2'-OH of the upstream nucleotide is activated for nucleophilic attack at the scissile phosphate to cleave their RNA substrate and generate 2',3'-cyclic phosphate and 5'-OH termini on the cleavage products. Therefore, only RNA molecules containing 2'-OH can be substrate of ribozymes and not the DNA which only contains 2'-H[22].

RNA is synthetized in cells using DNA as a template in a process called transcription. Resulting RNAs have several functions in cells and different RNA populations have been defined.



RNA by mass RNA by number of molecules

Figure 4. RNA estimation levels in a mammalian cell

(A) Proportion of different RNAs by mass. **(B)** Proportion of the numbers of different RNA molecules. Other ncRNA includes snRNA, snoRNA, and miRNA. Note that due to their relatively large sizes, rRNA, mRNA, and lncRNAs make up a larger proportion of the mass as compared to the overall number of molecules. Taken from [23].

In a general manner, RNA can be divided into two main groups: coding RNA (cRNA), the ones which can be translated into a protein; and non-coding RNA (ncRNA) which can have a wide range of functions such as regulating gene expression. Recent studies[24,25] have shown that some non-coding RNAs can encode small peptides named micropeptides up to 100 amino acids long which are involved in biological processes.

Up to now, more than 70 different RNA populations have been identified. The following tables summarizes some of the most well-known families according to their functions (modified from [26–28]).

Table 2. Major RNAs involved in protein synthesis

Name	Size (nucleotides)	Taxa distribution	Function
Messenger RNA (mRNA)	Up to ~100,000	All organisms	Code for proteins
Ribosomal RNA (rRNA)	Small subunit: 16S: 1,542 (*E. coli*), 18S: 1,869 (humans) Large subunit: 23S: 2,906 and 5S:120 (*E. coli*), 28S: 4,718, 5.8S: 1560 and 5S: 120 (humans)	All organisms	Translation, form the basic structure of the ribosome and catalyze protein synthesis
Transfer RNA (tRNA)	~76-90	All organisms	Adaptors between mRNA and amino acids. Transfer individual amino acids to the growing peptide chain during translation
Transfer-messenger RNA (tmRNA)	Most are 325-400	Bacteria	Rescues stalled ribosomes during translation
Signal recognition particle RNA (7SLRNA or SRP RNA)	~300	All organisms	Directs proteins to plasma membrane (prokaryotes) or endoplasmic reticulum (eukaryotes)

Table 3. Major RNAs involved in post-transcriptional modification or DNA replication

Name	Size (nucleotides)	Taxa distribution	Function
Small nuclear RNA (snRNA)	~150	Eukaryotes	Variety of nuclear processes, including splicing of pre-mRNA
Small nucleolar RNA (snoRNA)	Most around 70-120	Eukaryotes	Processing and chemical modification of rRNAs
SmY RNA	~70-90	Nematodes	Trans-splicing (the splicing of exons from two different transcripts)
Small Cajal body-specific RNA (scaRNA)	~130-300	Eukaryotes	Subset of snoRNAs involved in nucleotide modification of RNAs. Modify snoRNAs and snRNAs
Guide RNA (gRNA)	~20-50	Kinetoplastid protists	RNA editing of mitochondrial mRNAs
Ribonuclease P (RNase P)	341 'H1 RNA component' (*H. sapiens*)	Archaea, bacteria and most of eukaryotes	A ribozyme that catalyzes the maturation of tRNA
Ribonuclease MRP (RNA component of mitochondrial RNA processing endoribonuclease; RNase MRP)	277 (*H. sapiens*)	Eukaryotes	rRNA maturation
Y RNA	~85-112	Vertebrates, nematodes and bacteria	RNA processing, DNA replication
Telomerase RNA component (TERC)	541 (*H. sapiens*)	Most eukaryotes	Telomere maintenance during DNA replication
Spliced leader RNA (SL RNA)	100 (*C. elegans*)	Lower eukaryotes and nematodes	Trans-splicing and RNA processing

Table 4. Examples of regulatory RNAs

Name	Size (nucleotides)	Taxa distribution	Function
Antisense RNA (aRNA/asRNA)	Various	All organisms	Transcriptional attenuation / mRNA degradation / mRNA stabilization / Translation block
CRISPR RNA (crRNA)	~100	Bacteria and archaea	Resistance to infection by targeting pathogen DNA or RNA
Long noncoding RNA (lncRNA)	>200	All organisms	Regulatory functions typically in gene expression
MicroRNA (miRNA)	~22	Most eukaryotes	Gene regulation, typically by blocking translation of select mRNAs
Piwi-interacting RNA (piRNA)	~26-31	Most animals	Transposon defense, among other proposed functions
Enhancer RNA (eRNA)	~50-2,000	Mammals	Thought to regulate gene expression

Among the different RNAs, the three universal types found in every cells and organelles are messenger RNAs, ribosomal RNAs and transfer RNAs (reviewed in [4,17,18]).

1.1.2.1 Messenger RNA (mRNA)

The messenger RNA is the coding RNA and acts as an intermediate between DNA and protein in the gene expression process. Each mRNA contains the genetic information derived from one gene that is translated into a given protein. The mRNA represents 3-5 % of the total cellular RNA (by mass). mRNA´s length can vary depending on the size of the protein it encodes. After being transcribed the mRNA has to go through a maturation process in order to be translated into a protein. In eukaryotes, this maturation process includes capping and addition of a poly-A tail on the precursor pre-mRNA as well as splicing (see below).

1.1.2.2 Ribosomal RNA (rRNA)

The ribosomal RNA is a type of non-coding RNA that, similar to some protein enzymes, has the ability to catalyze specific biochemical reactions during the protein synthesis. Ribosomal RNA is the most abundant product of transcription, constituting 80 to 90% of the total RNA mass in prokaryotes and eukaryotes. In eukaryotes, the rRNA genes are contained in tandem clusters and rRNAs are synthetized in the nucleolus, a dense compartment of the nucleus that contains the corresponding genes. After an intricate assembly process with ribosomal proteins; rRNAs present in ribosome subunits, are exported to the cytoplasm where it catalyzes information translation of mRNA into proteins. Ribosomes are composed of around 60% rRNA and 40% ribosomal proteins. Ribosomes are organized into two subunits: a large ribosomal subunit (LSU) and a small ribosomal subunit (SSU). In eukaryotes, in a single cell can be present from 50 to 5,000 rRNA genes and more than 10 million ribosomes; while a prokaryotic cell can contain about 7 rRNA genes and 15,000 ribosomes[29].

1.1.2.3 Transfer RNA (tRNA)

The transfer RNA it is a type of non-coding RNA whose main function is to carry activated amino acids to the protein synthesis site. tRNAs are adaptor molecules 70 to 90 nucleotides long which act as physical link between the mRNA and the protein; because of this, they are indispensable for translation. Each tRNA carries a specific amino acid according to the triplet sequence in its anticodon which is complementary to the codon of the mRNA. In Eukaryotes, tRNAs are transcribed by the RNA polymerase III (reviewed in [30]). Many tRNAs are synthesized as precursors and post-transcriptionally processed through modification and/or splicing as well as generation of correct ends.

In all organisms tRNAs are overall well conserved. They are generally composed of an acceptor stem, dihydrouridine (D)-arm, anticodon stem-loop which includes the anticodon region, variable arm and a TΨC-arm. At the 3' end of the tRNA, a CCA trinucleotide is added as part of the tRNA maturation process in some organisms[31]. The corresponding amino acid is attached by a specific enzyme called aminoacyl-tRNA synthetase forming an ester bond between the 2' or 3' hydroxyl group of the terminal adenosine and the α-carboxyl group of the amino acid. There are 20 different synthetases in most of the living cells, and each one recognizes the unique structure

of its cognate tRNAs (reviewed in [32]). The secondary structure of a tRNA is typically represented in 2D as a cloverleaf which shows the base paring regions and the loops (Figure 5A). The tertiary structure determined by X-ray crystallography reveals hydrogen bonds that bend the cloverleaf into a 3D L-shaped structure (Figure 5B).

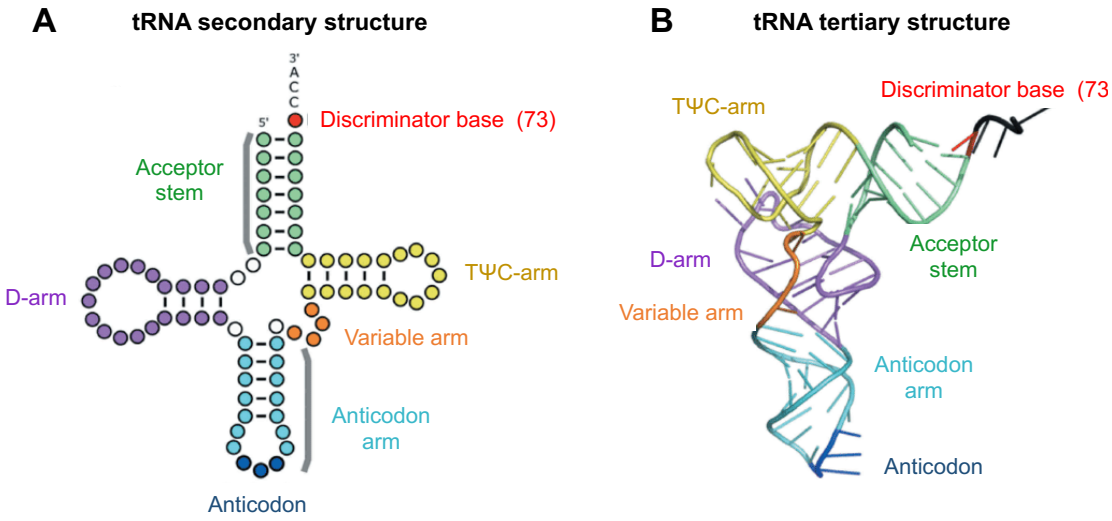


Figure 5. Transfer RNA structure

(A) tRNA secondary 2D structure represented as cloverleaf shape. **(B)** tRNA tertiary 3D structure in which the tRNA acquires an L-shape. Every stem is colored: acceptor stem (green), dihydrouridine (D)-arm (purple), anticodon stem (light blue), anticodon (bases 34, 35, 36 dark blue), variable arm (orange), TΨC-arm (yellow) and the discriminator base (red). Taken from [33].

In all organisms, tRNAs are the most abundant RNA molecules; therefore, in most cases a single gene copy could most often not satisfy the high demand; therefore, it is usual to find more than one tRNA that is able to carry one amino acid. According to the Human Genome Assembly Issue 19 (hg19), the human genome contains 613 tRNA genes, for 20 amino acids[34]. tRNAs that have the same anticodon but otherwise different sequences are named isodecoders[33,35]. Due to the degeneracy of the genetic code, there is more than one anticodon for every amino acid. Thus, distinct from the tRNAs isodecoders, the isoaceptors tRNAs are molecules with different sequences and anticodons but carrying the same amino acid. Therefore, it is important to have a clear nomenclature to distinguish different tRNAs (Figure 6). In bacteria, tRNAs are often clustered in polycistronic transcription units, while in eukaryotes they are usually not tightly clustered (reviewed in [30]).

tRNA nomenclature [18] :

- The specific tRNA that is able to carry a determined amino acid is indicated with a superscript on the right; e.g. $tRNA^{Glu}$.
- The specific anticodon for a tRNA is indicated with a subscript at the right or after the superscript preceded by a slash or between parentheses; e.g. $tRNA^{Glu}_{UUC}$, $tRNA^{Glu/UUC}$ or $tRNA^{Glu(UUC)}$.
- The amino acid linked to a charged tRNA is indicated with a prefix on the left; e.g. Glu-$tRNA^{Glu}$

- The right suffix is used to specify the organism from each the tRNA is derived; e.g. $tRNA_{yeast}$

Figure 6. Genomic tRNA database

Gene Symbol. **(1)** RNA type. **(2)** Isotype: three letters of the carrying amino acid. **(3)** Anticodon: nucleotide sequence of the tRNA anticodon. **(4)** Isodecoder: for a particular isotype and anticodon. **(5)** Gene copy: for genes that have multiple identical copies in different loci [36].

In addition to cytoplasmic tRNAs animal cells contain a specific set of mitochondrial tRNAs that are 59 to 75 nucleotides long. These tRNAs have smaller loops and arms and their structures are also less rigid than the cytoplasmic ones (reviewed in [37]).

Besides their canonical role in protein biosynthesis, some tRNAs are involved in other cellular functions in eukaryotes and prokaryotes including regulation of gene expression, lipid aminoacylation, post-translational protein labeling, bacteria conjugation and bacterial cell-wall biosynthesis [38].

1.1.3 Proteins

Proteins are cellular building blocks and the molecules which execute the functions programed in the genes. They are directly involved in chemical processes essential for life and constitute most of the dry mass of a cell. Proteins are linear polymers which contains ten to several thousand amino acid residues linked by peptide bonds; therefore, proteins are polypeptides. Every monomeric amino acid is composed of a central α-carbon atom (C_α) joined to four different groups: an amino group (NH_2) (except for proline), a carboxyl group (COOH), a hydrogen atom (H) and a variable group (R) also called side chain (Figure 7A). Due to the asymmetry of the α-carbon, amino acids, with the exception of glycine, can exist in two mirror-image conventionally called dextro (D) and levo (L) isomers. Although chemical properties of isomers are identical, their biological activities are different. With rare exceptions, proteins are formed from the L-amino acid isomers (Figure 7B).

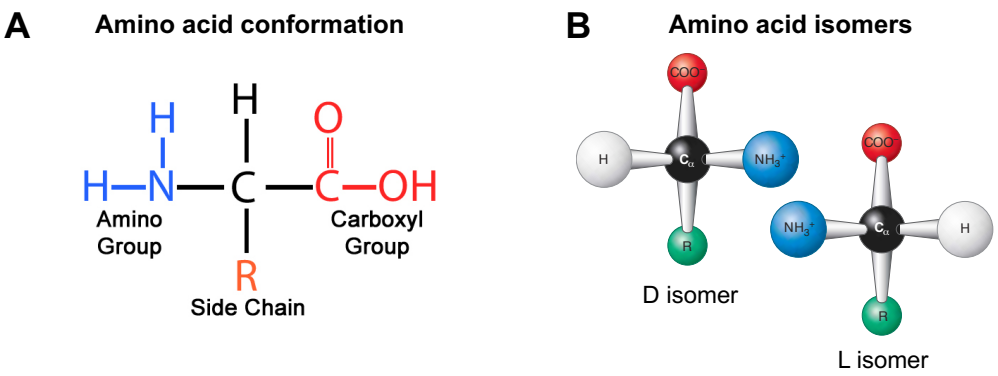


Figure 7. Common structures of amino acids

(A) Chemical groups conforming an amino acid, the central α-carbon (black) is bonded to four chemical groups: amino group (blue), hydrogen atom (gray), carboxyl group (read) and side chain or R variable group (green) is unique to each type of amino acid. (B) Dextro and levo amino acid isomers due to the C_α asymmetry. Taken from [39] and [19], respectively.

Except for the variable group, the rest of the atoms form the polypeptide backbone of the chain. Due to asymmetry of the peptide bond, the backbone presents directionality, i.e. the two ends of a chain are chemically different, one side carries a free amino group (NH_2) called amino terminal (N-terminus), while the other side has a free carboxyl group called carboxyl-terminal (C-terminus). Conventionally, the amino acid sequence of a protein is presented on an N → C direction, reading it form left to right (reviewed in [40]).

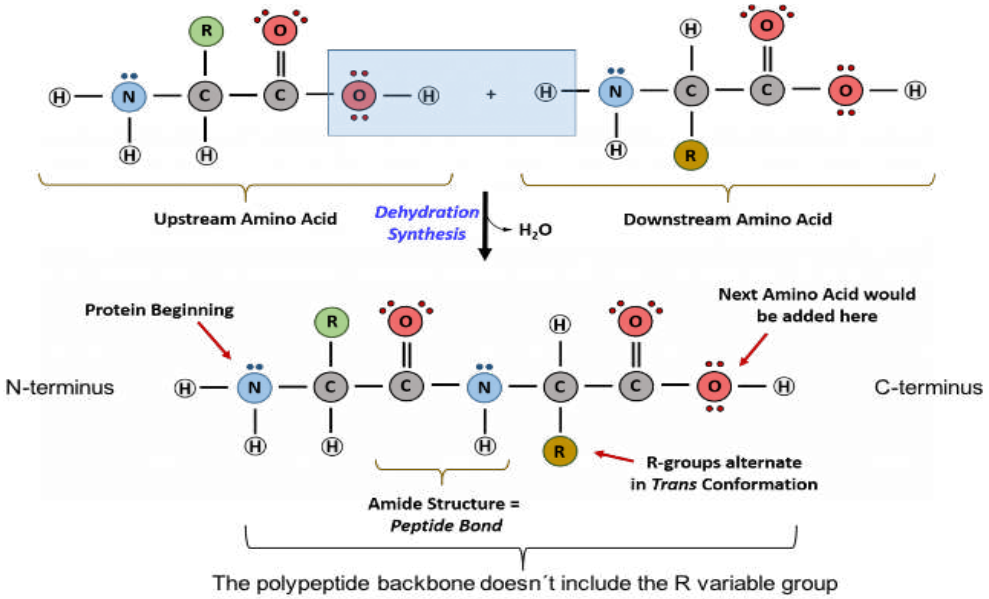


Figure 8. Formation of the peptide bond

The linked of two amino acids through a peptide bond requires dehydration synthesis. The R variable groups alternate in trans conformation. Taken from [40].

The flexibility of the polypeptide chain gives rise to a multitude of various protein structures. The wide array of possible structures allows proteins to perform their widely varying functions. Four levels of protein structure are recognized hierarchically: primary, secondary, tertiary and quaternary (reviewed in [19,41–44]).

The primary structure is the sequential arrangement of the amino acids (residues). A short chain of amino acids is named peptide (less than 20-30 amino acids), while a longer chain is called polypeptide. The term protein is used to refer to a polypeptide, or a complex of them, that have a well-defined 3D structure. The size of a protein is typically given in Daltons, one Dalton meaning 1 atomic mass unit.

The secondary structure arises from interactions between adjacent amino acid residues. At this level polypeptide chain can have defined structures or random unstructured coils. The most common secondary structures are alpha (α) helix, beta (β) strand or a short U-shaped turn. In an average protein, the α-helices and β-strands represent around 60% while the rest are coils and turns.

The tertiary structure is the level of three-dimensional organization of all the amino acids of a polypeptide chain which includes one or more secondary structure elements. The tertiary structure is stabilized by hydrophobic, polar interactions and hydrogen bonds between distant amino acid residues and peptide bonds. These interactions are weak: this is an important characteristic for the structural variations required for the protein function. Specific spatial combinations of secondary structure elements form motifs that very often have specific functions.

The quaternary structure is the number (stoichiometry) and relative positions of the separate subunits in multimeric proteins. Such complexes can be formed by any number of identical or different subunits, where each subunit can have different function or can cooperate with others. Some of these protein associations are huge macromolecular assemblies larger than 1 MDa. Some examples are viral capsids and enzymes.

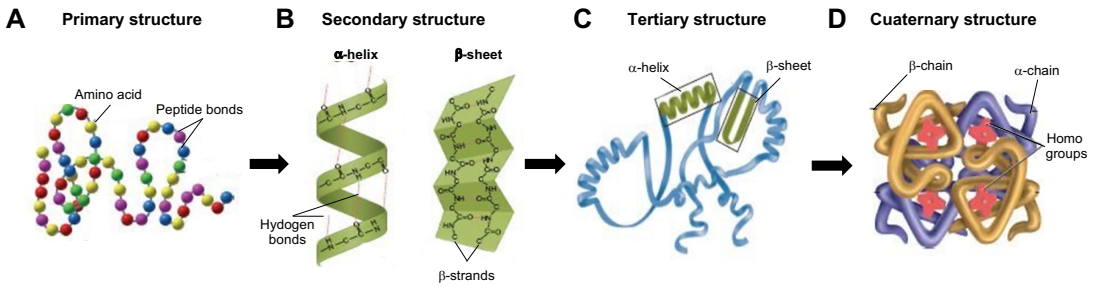


Figure 9. Different structures of a protein

(**A)** Primary structure: amino acid sequence linked by peptide bonds. **(B)** Secondary structure: α-helix, β-beta sheet formed by hydrogen bonds. **(C)** Tertiary structure: folding and coiling due to non-covalent interactions among R groups. **(D)** Quaternary structure: association of two or more polypeptide chains with each other[45].

1.1.4 Gene expression

Gene expression is the process by which the information contained in the DNA is converted into a functional polypeptide. Almost every cell of an organism contains the same DNA information (with exception of some cells in which genome rearrangements have occurred); however, not all of the genes are transcribed in each cell type. A gene is expressed when its product is synthesized in a cell. In eukaryotes, to express genes in a canonical way, the information coded in the DNA is

first copied to a mRNA molecule by a transcription process. Then, the mRNA is subjected to maturation and is exported to the cytoplasm where it is translated into a protein. Therefore, transcription and translation are the means by which cells express their genetic information, at a specific time point, to cover particular needs or to perform specific functions. Due to the time and space specificity, gene expression requires a strict regulation such as: control of chromatin state through histones, modulation of DNA transcription through epigenetic marks and transcription factors among others; and translation control via, for example, adjustment of mRNA stability.

In the next sections, I will describe the genetic code as well as the basic processes responsible for gene expression: transcription, RNA processing, post-transcriptional modifications and translation.

1.1.4.1 Genetic code

The genetic information is stored in the DNA molecules. The DNA regions that encode proteins are transcribed into RNA. In the latter, coding sequences consist of triplets of nucleotides also called codons, which correspond to the amino acid sequence of a protein from its N to C-terminus. The relationship between codons and amino acid residues is called genetic code[17]. Due to the fact that each of the 4 possible nucleotides (A,T,C,G) can take each of the three positions of a triplet (4^3), there are 64 possible nucleotides combination or codons, from which, in the standard code, 61 code for an amino acid and 3 cause the termination of the protein synthesis[46] (Figure 10). Some organisms use modified codes derived from the standard one.

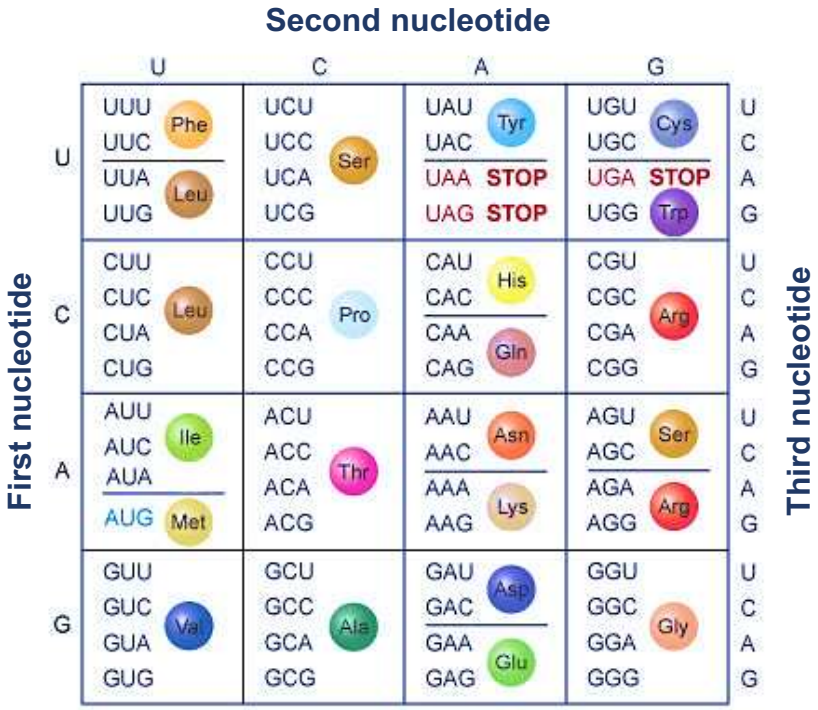


Figure 10. Representation of the standard genetic code

The 20 amino acids (represented in circles) are coded by different codons. Multiple codons can code for the same amino acid. Codons are written 5'→3' as in the mRNA. AUG is an initiation codon while UAA, UAG and UGA are termination/stop codons. Taken from [47].

1.1.4.2 Transcription

The transcription is the first of a set of biological processes to transform DNA into proteins. A gene is a DNA sequence that contains the information to produce a functional RNA and in many cases a protein. However, in some organisms, an RNA transcript can contain information from more than one gene.

In all organisms, transcription starts with the binding of the RNA polymerase (RNAP) complex, also called holoenzyme, to a specific region at the extremity of the DNA sequence named promoter. The activation of the holoenzyme leads to the transcription initiation, that is followed by the transcript elongation and finally the termination. Due to the fact that the elongation of the transcript enables the clearing of the promoter, a new transcript can be synthetized after the previous one. Although transcription is a common process for prokaryotes and eukaryotes, the mechanisms behind the eukaryotic process are more complex.

1.1.4.2.1 Transcription in prokaryotes

In prokaryotes, transcription is performed by a single type of RNA polymerase. In *Escherichia coli* for example, this RNAP is composed of 4 subunits α, α, β, and β′ forming the catalytic core plus a regulatory subunit called σ-factor, which is necessary for a subset of genes[48]. In most cases the promoter regions are located upstream of the transcription start site (TSS). In bacteria, there are two promoter elements, the first one is normally located 10 nucleotides upstream the TSS, this promoter is called TATA box because of its sequence (TATAAT); the second one is located at position -35 (TTGACA) and binds the σ-factor (reviewed in [49]). During the elongation step, the RNAP moves along the template "unzipping" the double DNA helix in a direction 5' to 3' (5'→3') adding nucleotides to the synthetized RNA chain with a speed of about 40 nucleotides per second. Termination can be dependent or independent of the Rho terminator that destabilizes the DNA-RNA interaction and disassociating the RNAP complex; otherwise, a C-G-rich region close to the end of the transcript have a similar effect (reviewed in [50–52]).

1.1.4.2.2 Transcription in eukaryotes

In eukaryotes, the transcription is highly complex and strictly regulated. This process occurs by multiple rounds of initiation-elongation-termination cyclic patterns. Different form prokaryotes, the complexity of the eukaryotic transcription begins with the genome organization into chromatin. A complex of DNA and proteins which include histones and non-histones proteins pack the genome in a highly compacted structure that is localized in the nucleus of the cell; this packing serves as a first regulatory mechanism of gene expression. This regulation is mediated mostly by epigenetic marks as acetylation, methylation and phosphorylation of histones such as H3K9, H3K4, H3K14 among others. Histone acetylation and phosphorylation 'relax' chromatin structure exposing sites needed to activate gene transcription; whereas DNA methylation marks recruit proteins involved in gene repression and inhibit binding of transcription factors (reviewed in [53]).

In bacteria, such as *E.coli,* there is only one RNA polymerase (RNAP) composed of five subunits including two copies of subunit α and one copy of subunits β, β' and ω (reviewed in [54]). In eukaryotes there are at least three different nuclear RNA polymerases (reviewed in [55,56]):

- RNA polymerase I (RNAPI) synthetize the main ribosomal RNAs (rRNAs), it is formed by 14 protein subunits.
- RNA polymerase II (RNAPII) transcribes messenger RNAs (mRNAs) and the majority of the non-coding regulatory RNAs. It contains a C-terminal domain (CTD) that contains phosphorylation marks as Serine 2 (Ser2) that are critical for transcription and RNA processing events. It is formed by 12 subunits.
- RNA polymerase III (RNAPIII) produces transfer RNA (tRNA) and the ribosomal 5S rRNA. It consists of 17 subunits.

Additional RNA polymerases have been identified in plants and participate, for example, in the production of small interfering RNA (siRNA) involved in RNA methylation[57].

For mRNAs, transcription starts when an activator protein binds a specific enhancer DNA sequence that initiates the recruitment of general transcription factors (TFIIB, TFIID, TFIIE, TFIIF and TFIIH) and the RNAPII onto the promoter of the gene forming the pre-initiation complex (PIC) (Figure 11). A central DNA region is melted leading to the formation of the open promoter complex (reviewed in [58]). After the formation of PIC, the RNAP releases transcription initiation factors and recruits new ones that help its displacement along the DNA template strand in the 3'→5' direction catalyzing the RNA synthesis by adding nucleotides to the 3' end of the nascent RNA strand.

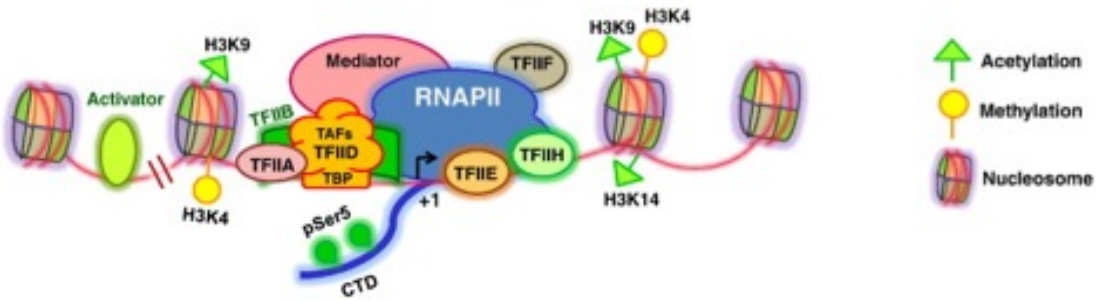


Figure 11. Pre-initiation complex (PIC) assembly

Binding of activators to its cognate enhance and recruitment of general transcription factors. Transcription starts at site +1. Taken from [59].

Transcription termination mechanisms are different for the main three eukaryotic RNAPs. In humans, genes transcribed by the RNAPI contain an 11 bp sequence that is recognized by a transcription terminator factor (TTF-1) blocking transcription and releasing the polymerase from the newly-synthesized RNA. The RNAPII does not follow a specific sequence for termination, it can continue the transcription after the end of the gene; however, at the end of the gene the transcript is cleaved. This cleavage is considered as the end of the gene. There are two such pathways, the poly (A)-dependent pathway and the Nrd1-Nab3-Sen1-dependent (reviewed in [59]). The RNAPIII transcripts contain a sequence of 4-7 uridines (Us) at their 3' end triggering the

formation of pre-termination complex and subsequent termination controlled by sequence specific elements[60].

For coding genes, there are several critical steps that convert the newly synthesized pre-mRNA into a fully functional mature messenger RNA. This process, called pre-mRNA processing or maturation, is discussed in the next section.

1.1.4.3 RNA processing

In bacteria, RNA transcripts are ready to work as messengers in protein translation even before they finish the transcription process. However, in eukaryotes freshly transcribed RNAs, also called pre-mRNAs (for the coding RNAs), require a set of modifications to prepare them for the translation process; the RNA processing is a collection of events which leads to the maturation of the primary transcript to a functional form. There are three main steps for the mRNA processing:

- 5' cap addition
- Splicing
- Polyadenylation

1.1.4.3.1 5' cap addition

To protect the newly transcribed mRNA from degradation, the pre-mRNA is modified at its 5' end by adding a modified guanine to form a N7-methyl guanosine (m7G) nucleotide. Capping is the first event of the mRNA processing of eukaryotes and viruses. It occurs during the early transcription after the RNAPII have transcribed first 25-30 nucleotides. It plays a key role in mRNA biogenesis, stability, splicing, export and translation. The addition of the cap is performed by a set of 3 enzymes: RNA triphosphatase, guanylyltransferase and 7-methyltransferase (reviewed in [61]). Even though the term 'capping' often refers to the addition of m^7G nucleotide, several other different caps exist. For example, there is a hypermethylated form of m^7G-cap, in which there is another methylation at 2'-OH of the ribose of the first nucleotide. The trimethylguanosine (TMG) is another type of cap that can be found in small nuclear RNAs (snRNAs) such as U1, U2, U4 and U5, small nucleolar RNAs (snoRNAs) as well as in the telomerase RNA TLC1. Other small RNAs can also be capped with a *γ*-methylphosphate (reviewed in [37,55,61]).

1.1.4.3.2 Splicing

In eukaryotes, the DNA sequence of a coding gene often contains nucleotides that must be removed to generate a mature mRNA; these non-coding regions are called introns, whereas the coding nucleotides are grouped in exons. Splicing is the step of gene expression in which introns are removed from the pre-mRNA and exons are ligated together. Similar to the capping process, the splicing occurs co-transcriptionally when the RNAPII is still in the elongation process [62,63]. Depending on the organism, intron structure or RNA type the splicing mechanism can change.

1.1.4.3.2.1 Removal of introns from coding genes

In mammals and yeast, this reaction is mainly catalyzed by a large set of RNAs and proteins that assemble in a dynamic complex called spliceosome. In most eukaryotes there are two main types of spliceosome: the major or U2-dependent and the minor or U12-type. The major spliceosome catalyzes the majority of splicing reactions (99%); it involves 5 different RNP subunits (U1, U2, U4, U5 and U6) and around 100 different proteins[64]. Introns are removed by transesterification reaction, branching and exon ligation catalyzed in one active site (reviewed in [65]). The intron excision occurs in 2 chemical steps (Figure 12): a 5' splice site cleavage and lariat formation that is followed by a 3' splice site cleavage and exon ligation. First, the U1 snRNP selects the 5' splice site (5'SS) and forms the so-called early complex (E-complex). The U2 snRNP selects a specific adenosine in the branch point (BP); the interaction between U1 and U2 results in formation of a pre-spliceosome, called Complex A. After, the pre-spliceosome associates with the preassembled triple-snRNP (U4/U6-U5) to form the pre-catalytic spliceosome also called as Complex B. Then, U1 is released and the U5 shifts from exon to intron while U6 binds at the 5' SS in the so-called Complex B*; which carries out the first catalytic step of splicing and releases U4 snRNP generating the catalytic spliceosome or Complex C. Complex C is able to perform the second catalytic step producing the lariat intron and spliced exons. Finally, U2/U5/U6 snRNPs are released from the complex and recycled to repeat the process (reviewed in [65–68]).

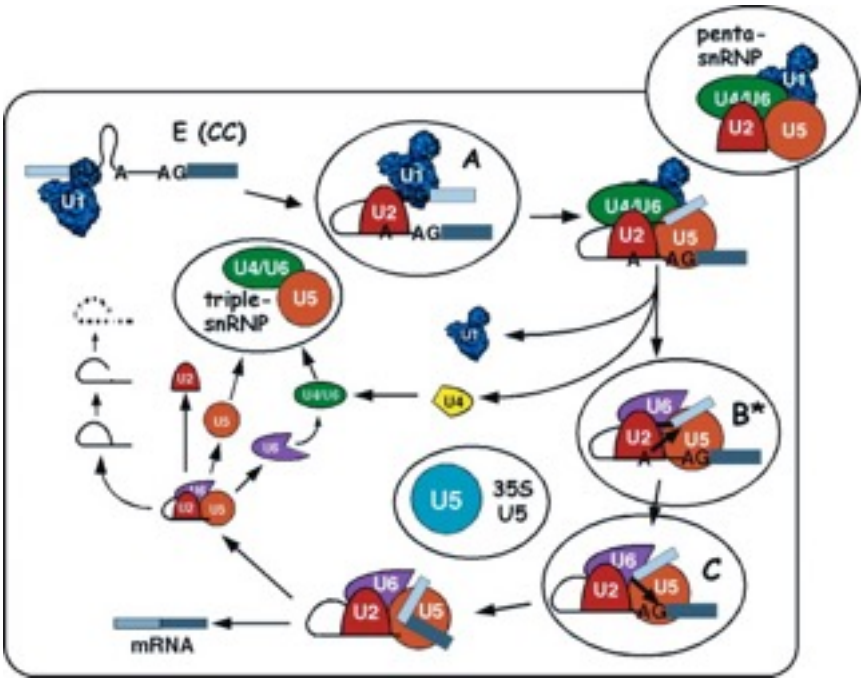


Figure 12. The spliceosome assembly cycle
Taken from [67].

The minor spliceosome or U12 type is similar in function to the major; however, there are some crucial differences. The minor spliceosome removes less than 1% of the introns and it involves U11, U12, U4atac/U6atac and U5 snRNAs. The 5'SS, 3'SS and BP sequences are

different between the two spliceosomes[69]. The minor spliceosome rarely produces alternative isoforms and it may be involved in expression control during proliferation as it has been shown to be active during mitosis[70] (reviewed in [71]).

1.1.4.3.2.2 Alternative splicing

It is the process in which a single pre-mRNA molecule is alternatively processed to generate different mature mRNAs and therefore different proteins isoforms, increasing the complexity and diversity of the proteome. In this context, the spliceosome "chooses" which exons to include or to exclude during splicing. There are four main types of alternative splicing such as exon skipping, intron retention, mutually exclusive exon usage and alternative splice site selection (reviewed in [72–74]). In humans, more than the 90% of the genes are alternatively spliced[75].

1.1.4.3.2.3 Self-splicing

Some RNA molecules such as ribosomal RNA (rRNA), phage and fungal mitochondrial RNA can form ribozyme structures have the ability to splice themselves without the intervention of any protein. There are at least three different types of self-splicing mechanism according to the groups in which introns are classified. This self-splicing is mediated through a phosphoester transfer mechanism in which a 2'-OH (or 3'-OH depending on the type of intron) or a nucleotide cofactor (GMP, GDP, GTP) attacks the 5' exon splice site, followed by a join of the two exons by transesterification (reviewed in [76]).

1.1.4.3.2.4 tRNA splicing

Transfer RNAs containing introns have been detected in the three domains of life. Introns of tRNAs are normally small in size and, despite the fact that there are exceptions, the canonical location for tRNA introns is: one nucleotide downstream of the anticodon loop of the pre-tRNA. The tRNA splicing is carried out by specific protein complexes. In *S. cerevisiae* this protein complex is called SEN (splicing endonuclease), whereas in human is called TSEN (tRNA splicing endonuclease). The TSEN complex is formed of proteins TSEN2, TSEN15, TSEN34, TSEN54, pre-mRNA processing factor and the RNA kinase (CLP1). TSEN15 and TSEN54 have structural function; TSEN2 performs 5'-exon/intron cleavage and TSEN34 performs intron/3'-exon cleavage. These cleavage reactions generate non-canonical RNA ends: free 5'-OH on both intron and 3'-exon and 2',3'-cyclic phosphate on both 5'-exon and intron. After some phosphorylation steps, the tRNA exon halves are ligated by the Rlg1/Trl1 (reviewed in [77]).

1.1.4.3.3 Polyadenylation

The polyadenylation is a post transcriptional modification in which a long (100-300) chain of adenine nucleotides is added to the 3' end of messenger RNA (mRNA) immediately after transcription. In eukaryotes, the poly(A) tail function is to protect the mRNA from degradation by increasing the stability which contributes to the export process and translation efficiency. In prokaryotes, oligo(A) tails have shown to be mainly involved in quality control and RNA degradation (reviewed in[78]).

The poly(A) tail addition can be decomposed into two processes: cleavage and polyadenylation. Polyadenylation starts by cleavage of the 3'-end of pre-mRNA by the Cleavage/Polyadenylation Specificity Factor (CPSF) which is assisted by cleavage factor I (CFI) and RNAPII to detect the cleavage site; normally, RNA is cleaved before the transcription termination. Once the cleavage is done, the poly(A) polymerase (PAP) adds the adenine nucleotides chain until the PAP cannot longer bind the CFP [79–81]. Similar to the alternative splicing, the alternative cleavage and polyadenylation (APA), occurs when a gene has multiple polyadenylation sites, resulting in multiple RNA transcripts. In humans, around 70% of genes are prone to APA[82].

1.1.4.3.4 RNA export

Due to compartmentalization in eukaryotic cells gene expression requires constant transport of RNA species; thus, export of coding and non-coding RNA molecules from the nucleus to the cytoplasm becomes one of the key steps. The cell nucleus is surrounded by the nuclear envelope (NE). Nuclear pore complexes (NPCs) perforate the NE and are the main gateways to connect nucleus and cytoplasm. The NPCs are formed from nucleoporins, a set of around 30 proteins which form a hydrophobic network that acts as barrier limiting the exchange of molecules (reviewed in[83]).

1.1.4.3.5 tRNA processing

Different from mRNA processing, tRNA biogenesis occurs post-transcriptionally at different subcellular localizations. In *S. cerevisiae* (Figure 13), tRNA biogenesis starts with the transcription of the primary tRNA (pre-tRNA) by the RNA Polymerase III, in the nucleolus. While still in nucleolus, the pre-tRNA is subjected to 5' maturation, catalyzed by the RNase P, and 3'-end processing. Some of the tRNA modifications (such as pseudouridine modification) or CCA addition, are performed in the nucleoplasm or inner nuclear membrane. The pre-tRNA is then exported to the cytoplasm via Los, Mex67-Mtr2 or by Crm1 in the so called tRNA primary nuclear export. tRNA which contain introns, are subjected to a splicing process on the surface of mitochondria by the SEN complex and ligated by Rlg1/Trl1. Spliced tRNAs undergo a second step named retrograde nuclear import, which is mediated by Ssa2 and possibly Mtr10. Once tRNAs are back to the nucleus, spliced tRNAs can be modified by enzymes, such as Trm5 which catalyzes m^1G_{37} modification, which only recognize intron-free tRNAs. tRNAs are then re-exported to the cytoplasm. Some tRNAs, such as $tRNA^{Phe}$, receive then new modifications such as wybutosine (yW) (reviewed in [84]).

tRNA biogenesis

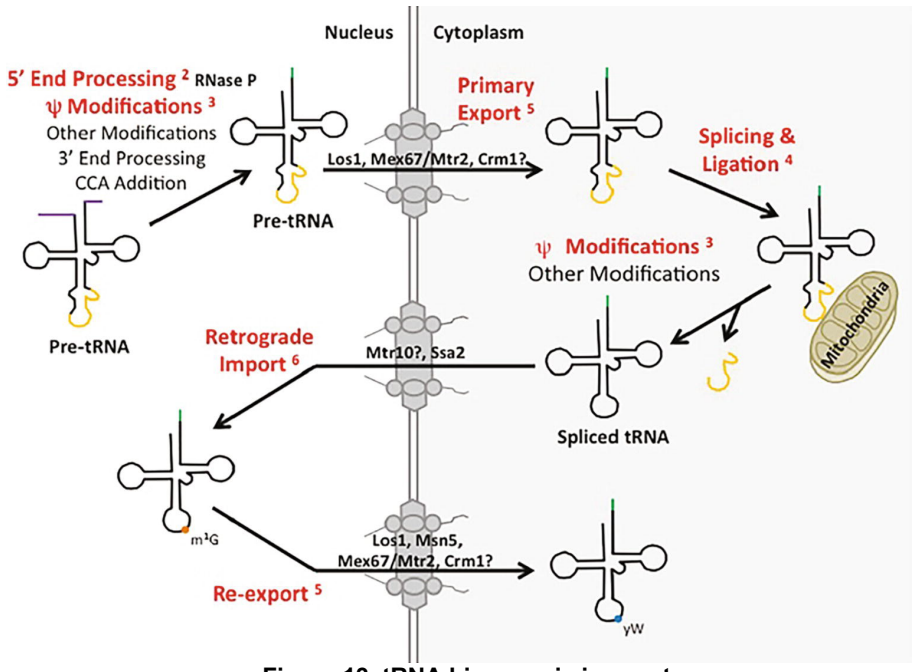


Figure 13. tRNA biogenesis in yeast
Taken from [84].

1.1.4.4 Translation

Translation is the second major step of gene expression in which mRNA is interpreted according the genetic code producing a chain of amino acids called polypeptide, that will finally fold into a 3D protein structure. In all organisms, translation is carried out by the ribosomes; in prokaryotes, this process occurs while the transcript is still transcribed and attached to RNAP. In eukaryotes, the mature mRNA has to travel to the cytoplasm where the ribosomes are located. Ribosomes are large asymmetric RNA-protein complexes formed of two subunits: a large subunit (LSU) and a small subunit (SSU) (reviewed in [85,86]) (Figure 14). Ribosomal subunits are commonly formed of one to three ribosomal RNA (rRNA) molecules and several ribosomal proteins (r-proteins). Sizes of ribosomes and their subunits are usually characterized in Svedberg (S) units; which are a measure of sedimentation velocity. Comparison between the high-resolution X-ray and cryo-electron microscopy of structures from eukaryotic, bacterial and archaeal ribosomes, reveals a universally conserved structure and function throughout the three domains of life[87,88] (reviewed in [89]); however, their rRNA and r-proteins and therefore their S-unit counts change depending on the organism and on the subcellular localization. In eukaryotes for example, ribosomes differ between the cytosolic[90], mitochondrial[90,91] or chloroplast[92] compartments. Table 5 shows the ribosomes composition of the model organism *E. coli* (prokaryote) and the cytoplasmic ribosomes of *S. cerevisiae* (eukaryote).

Table 5. Ribosome composition of model organisms.

Prokaryotic (*E. coli*) and the cytoplasmic eukaryotic (*S. cerevisiae*).

Organism	Ribosome	Subunit	rRNAs	r-proteins
Escherichia coli	70S	50S (large)	23S (2,904 nt)	31
			5S (120 nt)	
		30S (small)	16S (1542 nt)	21
Saccharomyces cerevisiae	80S	60S (large)	25S (3,396 nt)	46
			5,8S (158 nt)	
			5S (121 nt)	
		40S (small)	18S (1,800 nt)	33

The LSU contains a peptidyl-transferase center (PTC) that catalyzes the peptide bond formation in the growing polypeptide. It also contains three specialized tRNA binding sites: the aminoacyl (A) site which accepts aminoacyl-tRNAs, the peptidyl (P) site which holds the tRNA together with the nascent polypeptide, and the exit (E) site which holds the deacylated tRNA before it leaves the ribosome structure. The SSU contains a decoding site which mediates the interaction of mRNA codons and the tRNA anticodons. Coordinated conformational changes between the two subunits, allow translocation of tRNA/mRNA pair through the ribosome during translation elongation (reviewed in [86,93]).

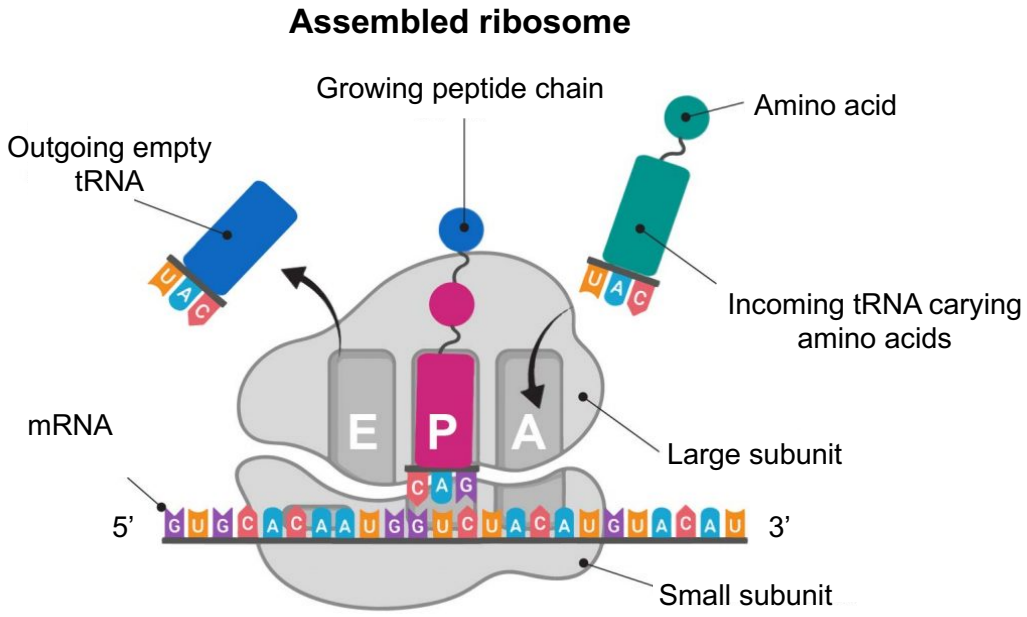


Figure 14. Diagram of an assembled ribosome

The LSU and SSU are assembled together in an elongation-competent subunit, able to bind an aminoacyl-tRNA in the A-site, a peptidyl-tRNA in the P-site and a deacylated-tRNA in the E-site. Taken from [94].

Translation of a mRNA molecule by ribosomes occur in three stages: initiation, elongation and termination.

1.1.4.4.1 mRNA translation initiation

Initiation of mRNA translation is a key checkpoint for regulation and protein synthesis fidelity. Translation initiation is the process in which small and large subunits assemble together to generate an elongation-competent ribosome. In archaea and bacteria, translation starts with the binding of the ribosome to the Shine-Dalgarno sequence of the mRNA which is followed by the elongation step; in eukaryotes however, the process is more complex.

1.1.4.4.1.1 Eukaryotic translation initiation by scanning mechanism

The translation (Figure 15) starts with the formation of a ternary complex (TC) from eIF2, Met-tRNA^{i} and GTP [1]. The TC is then recruited to the 40S subunit in a reaction promoted by eIF1, eIF1A, eIF5, and the multi-subunit eIF3 to form the 43S preinitiation complex (PIC) [2]. The 43S PIC then binds to the 5' end of pre-activated mRNA to form the 48S PIC. Pre-activation is made by association of eIF4 (eIF4: m^{7}G cap-binding protein eIF4E, the DEAD-box helicase eIF4A and the scaffold subunit eIF4G) with the cap and poly-A tail [3]. When the mRNA enters to the P-site, the 5' untranslated region (5'UTR) is scanned for complementarity to the Met-tRNA^{i} anticodon [4]. AUG recognition triggers the hydrolysis of GTP in the TC and release of eIF2-GDP from Met-tRNA^{i} [5], that is followed by the joining of the 60S LSU, that is stimulated by eIF5B, to form the 80S initiation complex [6], completing the translation initiation [7] (reviewed in [95,96]).

Translation initiation in eukaryotes

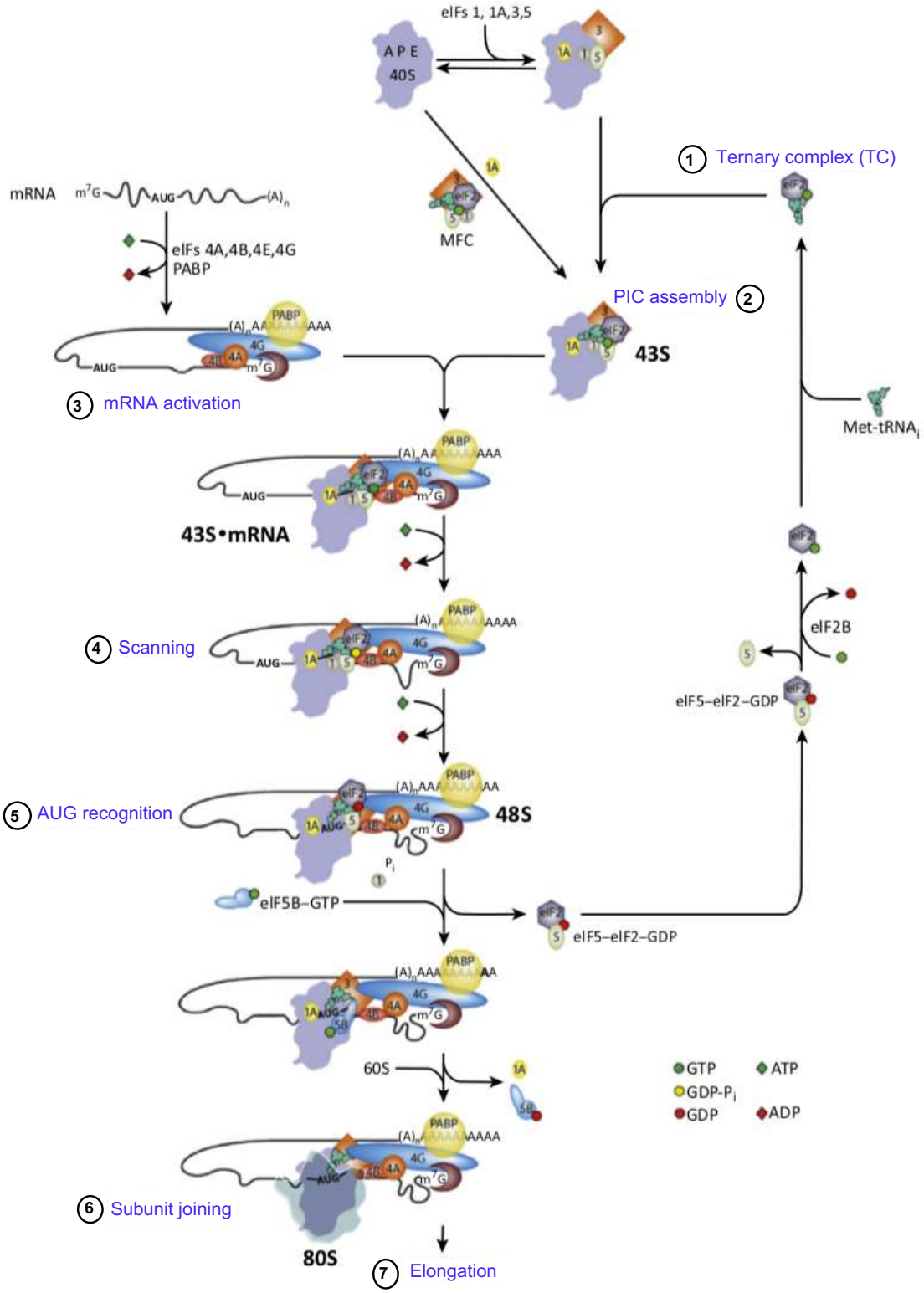


Figure 15. Model of scanning mechanism of eukaryotic translation initiation

The pathway is shown as a series of 6 steps that lead to the elongation. Taken from [95].

1.1.4.4.2 Translation elongation

The cyclic reaction in which new amino acids are added to the growing polypeptide chain during the protein translation is called elongation. This repeated multistep process consists in the selection of the proper aminoacyl-tRNA (aa-tRNA), the peptide bond formation and translocation. Different from the initiation and termination steps, elongation is very well conserved between prokaryotes and eukaryotes. After the initiation and the assembly of the 80S ribosome (70S in prokaryotes), the Met-tRNA^{i} is located in the P-site and base paired with the mRNA initiation codon (Figure 16A); the elongation continues with the incorporation of the aa-tRNA into the ribosome (Figure 16B). First, aa-tRNA is recruited to the ribosome as ternary complex with GTP and the GTPase eukaryotic elongation factor 1A (eEF1A) (or EF-Tu in prokaryotes) and occupies A-T state. When the complementary base pairing between the mRNA codon and the tRNA anticodon matches, the eEF1A is activated and triggers a GTP hydrolysis; and therefore, the release of eEF1A*GDP+Pi complex from the ribosome. Subsequently, the growing peptide chain is transferred from the peptidyl-tRNA in the P-site to the amino group of the aminoacyl-tRNA in the A-site to form a new peptidyl-tRNA (Figure 16C). The pre-translocated ribosome now contains a peptidyl-tRNA in the A-site and a deacylated tRNA in the P-site. Further, ribosome subunit ratcheting shifts the tRNAs into 'hybrid' states (A/P and P/E, respectively). These states are called hybrid since the anticodon stem loops (ASL) of tRNA are located in the A and P sites of the small subunit (SSU) of ribosome whereas the acceptor arms are accommodated in the P and E sites of the large subunit (Figure 16D). Translocation of the tRNAs to the canonical E and P sites is finalized by the GTPase eEF2 (EF-G in prokaryotes) (Figure 16E) (reviewed in [97,98]).

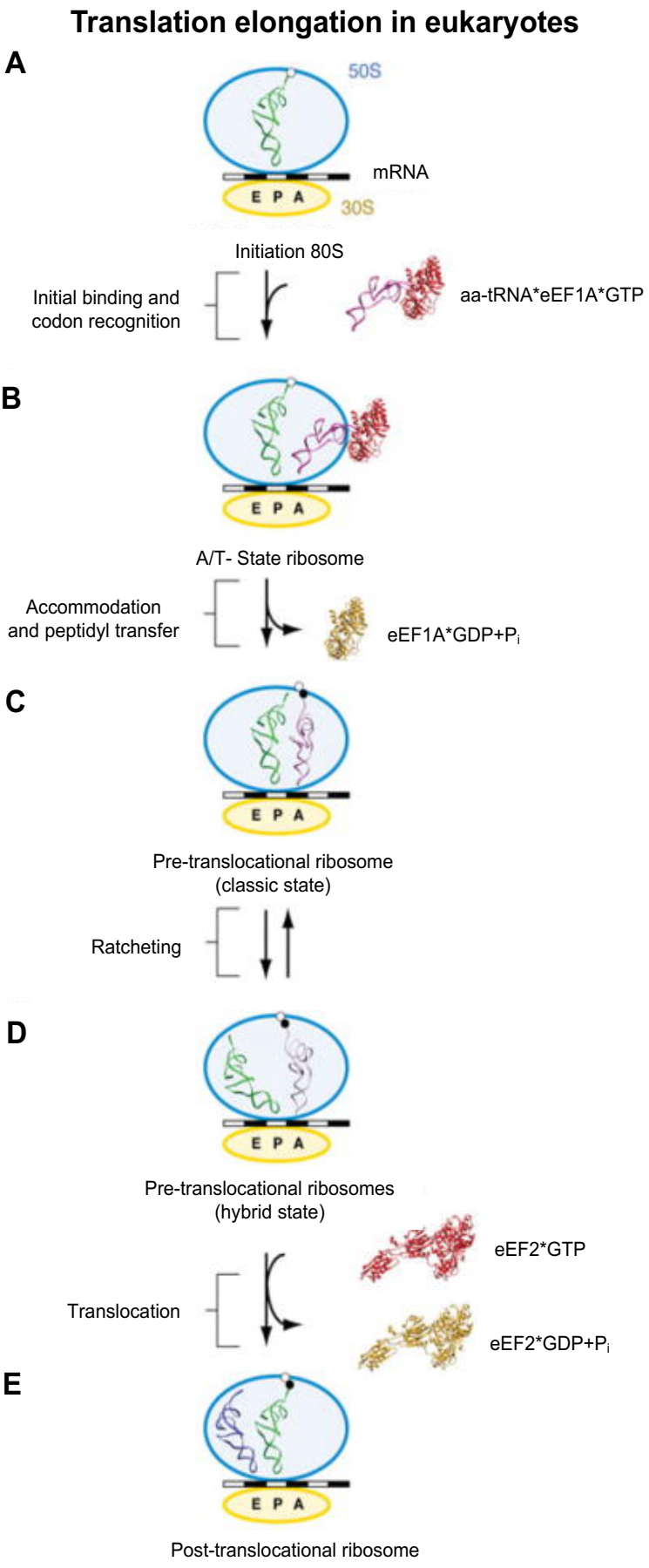


Figure 16. Overview of the translation elongation cycle **(A)** At the end of the initiation, the ribosome contains Met-tRNAi in the P-site. **(B)** First aa-tRNA is incorporated. **(C)** Matching pairing of codon-anticodon triggers GTP hydrolysis on eEF1A resulting in peptide bond transfer eEF1A-GDP release. **(D)** Pre-translocational ribosome hybrid A/P and P/E states. **(E)** eEF2*GTP catalyzes the complete translocation of the mRNA-tRNA complex. GTP-containing complexes are depicted in red; GDP-containing - in golden. Modified from [98].

1.1.4.4.3 Translation termination

The translation termination occurs when a stop codon (UAA, UGA or UAG) enters in the A-site of the ribosome (reviewed in[97,99]). Most of the translation termination mechanisms were first elucidated in prokaryotes, where this process is mediated by three released factors: RF1, RF2 and RF3[100]. RF1 and RF2 recognize stop codons in the A site and catalyze the hydrolysis of the peptidyl-tRNA; while RF3, an GTP-binding protein, promotes the dissociation of RF1 and RF2 from the post-termination complex. Later, ribosome recycling is triggered by the ribosome recycling factor (RRF) together with the initiation factor 3 (IF3) and EF-G that segregate the ribosome subunits to start a new round of translation (reviewed in [101,102]). In eukaryotes, translation termination is catalyzed by two eukaryotic released factors, eRF1 and eRF3[103,104]. eRF1 contains a NIKS (Asn-Ile-Lys-Ser) motif, that, together with other conserved motifs, recognize the stop codons with high specificity[105]. eRF1 also contains a GGQ (Gly-Gly-Gln) motif that promotes the release of the nascent polypeptide chain from the PTC of the ribosome[106]. eRF1 requires eRF3, a specific GTPase to bind the ribosome. Therefore, when the stop codon is localized at the A-site of the ribosome, the eRF1-eRF3-GTP complex binds to it (Figure 17A). The codon-anticodon recognition of the termination signal triggers the eRF3 GTPase activity (Figure 17B); and therefore, its release from the complex in a GDP form (Figure 17C). Next, Rli1, also known as ABCE1, an ATPase of the ABC-family that converts chemical energy into mechanical, is recruited to the ribosome. First, Rli1-ABCE1-ATP complex binding pushes eRF1 GGQ motif into the PTC of the LSU (Figure 17D). This step leads to ATP-independent hydrolysis of the bond between peptidyl chain and tRNA and release of the nascent protein from the ribosome (Figure 17E). Another mechanism of translation termination, besides the canonical eRF1-eRF3-dependent one, involves the Dom34-Hbs1 (these proteins are homologous to eRF1 and eRF3 respectively) heterodimer which induces termination in a codon-independent way, and promotes dissociation of empty ribosomes[72,107]

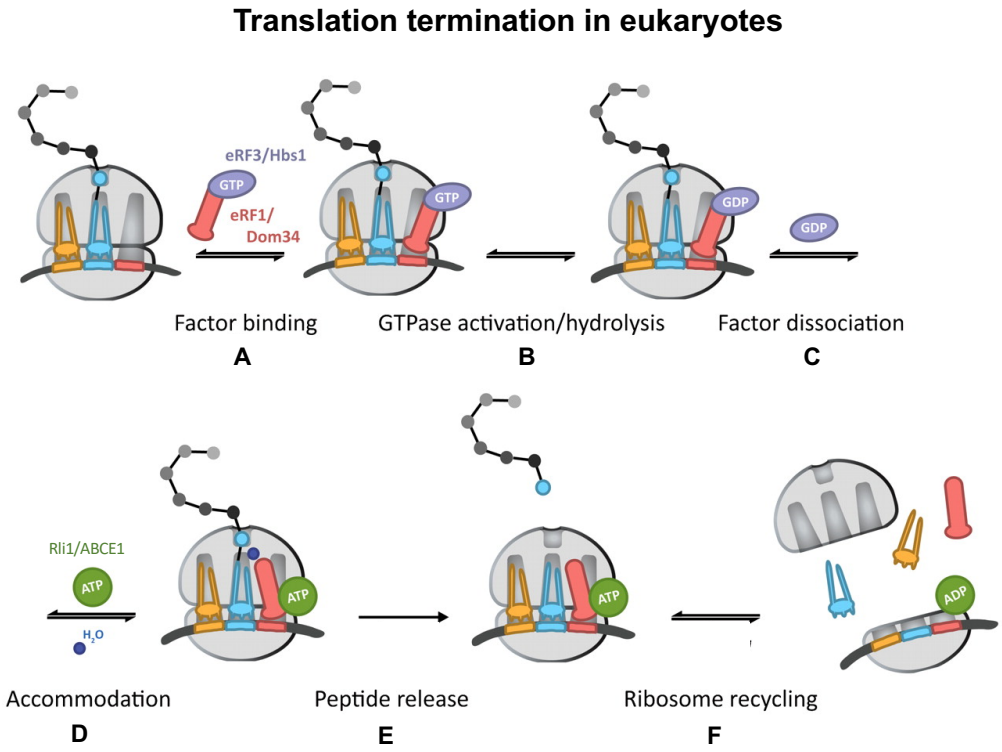


Figure 17. Model of the translation termination

(A) eRF1/Dom34-eRF3/Hbs1 complex binding to the ribosome A-site after stop codon recognition. **(B)** GTPase activity activation of eRF3/Hbs1 after codon-anticodon pairing. **(C)** GDP released form the complex. **(D)** Mechanical accommodation of eRF1/Dom34 by Rli1/ABCE1 in ATP-independent way. **(E)** Peptide is released by eRF1/Dom34. **(F)** Ribosome dissociation and recycling promoted by Rli1. Modified from [107].

1.1.4.4.4 Ribosome recycling

The final step of translation, the recycling, is necessary for the disassembly of the ribosome complex. After translation termination, the fully assembled 80S ribosome containing deacylated tRNA in its P-site, mRNA and eRF1 must be recycled to be used again in another translation cycle. Rli1/ABCE1 protein initiates ATP-dependent ribosome dissociation into subunits (Figure 17F)[107,108]. mRNA and tRNA are subsequently removed from the 40S subunit by the eIF2D factor (reviewed in[109,110]).

1.1.4.4.5 Suppressor transfer RNAs

As discussed above, gene expression is regulated at different stages by a variety of mechanisms and factors. For example, translation termination provides a regulatory mechanism of gene expression that is used in a natural way by the cell, and also offers to scientists a robust tool for genetic analyses.

Mutations are changes in the DNA sequence that can arise fortuitously *de novo* or can be inherited. Depending on specific factors such as: mutation type and affected gene, they can confer advantages or disadvantages to the organism. Considering a mutation as a primary event, suppression will be a secondary event that will camouflage the effects of the mutation; in other words, will phenotypically revert the mutation impact to a (near) wild-type state.

Suppression can be caused by changes in external conditions or result from alterations at the genome level. Changes in conditions mediated by environmental factors are transient and the mutant phenotype is restored once these environmental factors are removed. Temperature, ultraviolet radiation[111], salt concentration, carbon source among others are examples of such factors (reviewed in [112]). Genotypic suppression is caused by a secondary mutation in the genome which is able to suppress the effects of the first one while it is still present. This suppressor mutations can be located either in the same gene that carries the first mutation (intragenic suppressor), or in a different gene (extragenic suppressor) (reviewed in [112]). Suppressor genes were first identified in revertant strains of a specific mutant phenotype in the early 1960´s[111,113]. One example of such genes are suppressor transfer RNAs, which code mutant tRNAs able to insert a suitable amino acid at a mutated codon of the mRNA (reviewed in [112]). Two of the main types of tRNA suppressors: frameshift and nonsense suppressors, are described in more detail below.

1.1.4.4.5.1 Frameshift suppressor tRNAs

A frameshift occurs when a mutation produces the insertion or deletion of one or more nucleotides (but not a multiple of 3) changing the normal open reading frame (ORF) of a gene, and in consequence, the amino acid sequence of the polypeptide. Frameshift mutations often lead to the generation of premature stop codons and truncated protein products. Frameshift suppressor tRNAs are mutant tRNAs, which generally contain four bases in the anticodon loop, and are able to suppress +1 frameshift mutation reverting the mutation effect. As an example, the *SUF16-1* suppressor gene, encodes an altered glycine tRNA containing the four-base anticodon 3'-CCCG-5' in place of the wild-type anticodon 3'-CCG-5'[114] (reviewed in [115]). Besides suppressor tRNAs, frameshifts may be suppressed by random or programmed error of translating ribosomes. For example, in some viruses, controlled ribosome frameshifting permits production of different proteins from the same DNA sequence, allowing viruses to keep small DNA or RNA genome which are perfectly suitable to viral replication in a host cell and packaging. In eukaryotes, frameshifting also plays a regulatory role of gene expression[116].

1.1.4.4.5.2 Nonsense suppressor tRNAs

In the majority of organisms and organelles, the stop codons UAG (amber), UAA (ochre) and UGA (opal) are nonsense codons for which cognate tRNAs do not exist. Therefore, the release factors bind to them promoting the end of the peptide chain and the ribosome disassembly (reviewed in [117]). This statement has some exceptions e.g., in some *trypanosomatid* protists UGA has been reassigned to code for tryptophan, whereas UAG and UAA code for glutamate (reviewed in [118]). Also, in mammals and some other organisms, selenocysteine amino acid can be incorporated at some UGA codons[119].

The nonsense termination codons can be read by nonsense suppressor tRNAs as sense codons, thus skipping termination and continuing elongation. In *S. cerevisiae*, the first evidence

of nonsense suppressors dates to the early 1970´s and comes from the study of the iso-1-cytochrome *c* mutants[120,121] .

In yeast, most suppressors are denoted by three letters followed by a locus designation, e.g., *SUF1*, *SUP4*, *sup35, sup46,* etc. Frameshift suppressors are denoted *SUF* (or *suf*) while other metabolic suppressors are denoted with a variety of symbols as *srn* (suppressor of *rna1-1*), and *suh* (suppressor of *his2-1*). By convention upper case names are used to indicate dominant and lower cases denote recessive suppressors. In some cases, the ochre (UAA) and amber (UAG) suppressors are designated **o** (ochre) or **a** (amber) following the locus, e.g., *SUP4-**o**, SUP4-**a*** (reviewed in [115]). For example, in *S. cerevisiae*, nonsense suppressors (SUP) insert tyrosine, serine or leucine into the corresponding positions. Suppressor tRNAs can be used in combination with a selective marker which carries a premature stop codon[122]. Changes in the expression of the marker gene can be used to monitor changes in the functionality of the suppressor tRNA.

Since the discovery of sup-tRNAs, scientists have focused their efforts on adapting them for various therapeutic approaches; for example, to rescue protein function loss due to a premature stop codon (reviewed in [123]). In 1982, Temple *et al.*, restored the β-globin chain synthesis, protein that is deficient in the β0-thalassemia, by microinjection of a human amber (UAG) sup-$tRNA^{Lys}$ in *Xenopus laevis* oocytes[124]. In 2014, Bordeira-Carrrico *et al.*, rescued the expression of the human associated-cancer gene E-cadherin, by restoring the CDH1 gene from a premature stop codon hereditary mutation that causes gastric cancer using the sup-$tRNA^{Arg}$ suppressor[125].

1.1.4.4.5.3 SUP4 suppressor

The *SUP4 gene* is localized in X chromosome and encodes a tyrosine tRNA ($tRNA^{Tyr}_{G\psi A}$). Seven other genes encoding tyrosine tRNAs ($tRNA^{Tyr}_{G\psi A}$) are disseminated throughout the yeast genome. When SUP4 is mutated into a suppressor, its anticodon becomes cognate for ochre or amber stop codons and thus inserts a tyrosine residue at UAA (*SUP4-**o***) or UAG (*SUP4-**a***) positions[126]. The presence of multiple tRNAs is important as the other copies still allow normal incorporation of tyrosine at cognate codons. *SUP4-**o*** is particularly interesting as it encodes an altered variant of tyrosine tRNA ($tRNA^{Tyr}_{U\psi A}$) which carries a G to U substitution at the wobble base (U_{34}) which is modified by the Elongator complex. *SUP4* suppressor is used in a wide range of genetic screens, frequently combined with the mutant genes *ADE1*, *ADE 2* and *CAN1*. In yeast, the *ADE1* and *ADE2* genes encode two enzymes involved in the biosynthesis of the adenine: the phosphoribosylamino-imidazole-succinocarbozamide synthetase and the phosphoribosylamino-imidazole-carboxylase, respectively. Specific mutant forms of these genes (*ade1* and *ade2*) carry mutations that generate a premature UAA stop codon, preventing synthesis of full-length active enzyme and blocking adenine production. Moreover, these mutants accumulate intermediates of the adenine biosynthetic pathway and form polymerized products of phosphoribosylamino-imidazole (AIR) or phosphoribosylamino-imidazolecarboxylate (CAIR) intermediates. These compounds generate a red pigment in the yeast vacuoles that can be visually detected (Figure 18); therefore, the *ade1* and *ade2* genes can be used as markers[127].

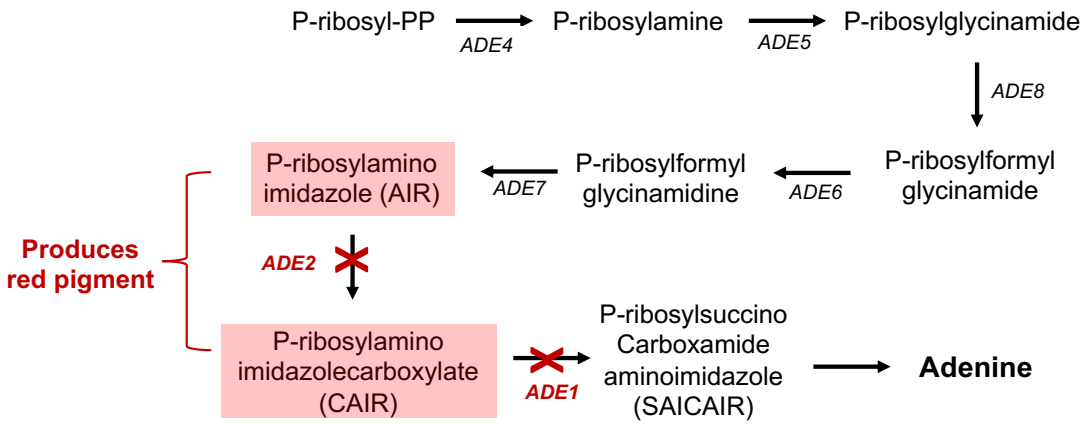


Figure 18. Mutations in *ADE1* or *ADE2*.
They cause accumulation of the intermediate phosphoribosylamino-imidazole (AIR) or phosphoribosylamino-imidazolecarboxylate (CAIR) which polymerize to form the red pigments. Modified from [128].

Suppression of the red pigment formation by *SUP4-**o***, has allowed scientists to use it in a variety of techniques not only to select transformants and conditional mutants but also to test the activity of enzymes involved in the pathway of mutant genes, e.g. the Elongator complex which catalyzes the formation of cm^5U_{34} derivates in tRNAs (reviewed in [115]). In a yeast strain containing the *ade2-1* UAA allele and the *SUP4-**o***, a functional Elongator complex will modify the mutant U_{34} position of $tRNA^{Tyr}_{U\psi A}$ which will allow at a certain frequency the recognition of the premature stop codon (UAA) as a cognate codon and permit the synthesis of the functional Ade2 enzyme. Sufficient levels of the latter can be produced to generate a viable amount of CAIR intermediate. Therefore, yeast cells will be able to grow in medium lacking adenine **(**Figure 19A). In contrast, in cells with a dysfunctional Elongator complex, unmodified $tRNA^{Tyr}_{U\psi A}$ will not be able to recognize UAA at a sufficiently high frequency and cells will not be able to grow in medium lacking adenine. In parallel, with suppression of adenine auxotrophy, *ade2-1* cells with dysfunctional Elongator accumulate the red pigment whereas cells with a functional Elongator remains white.

Another example of selectable marker used as a mutation reporter is the *can1* gene[129]. L-canavanine is a toxic analog of the arginine amino acid which can be imported into the yeast cells by the arginine transporter Can1[130]. This selectable marker is based in the fact that wild-type yeast cells growing in a media lacking arginine will import canavanine and die while *can1* mutants will be able to proliferate. In a host strain containing the *can1* UAA mutant, colony formation is inhibited by fully active *SUP4-**o***. Therefore, in media containing canavanine, cells carrying a defective Elongator will be able to grow as they are unable to synthesize the arginine transporter and thus to incorporate the toxic amino acid. In contrast, cells carrying a functional Elongator, will incorporate canavanine and die (Figure 19B).

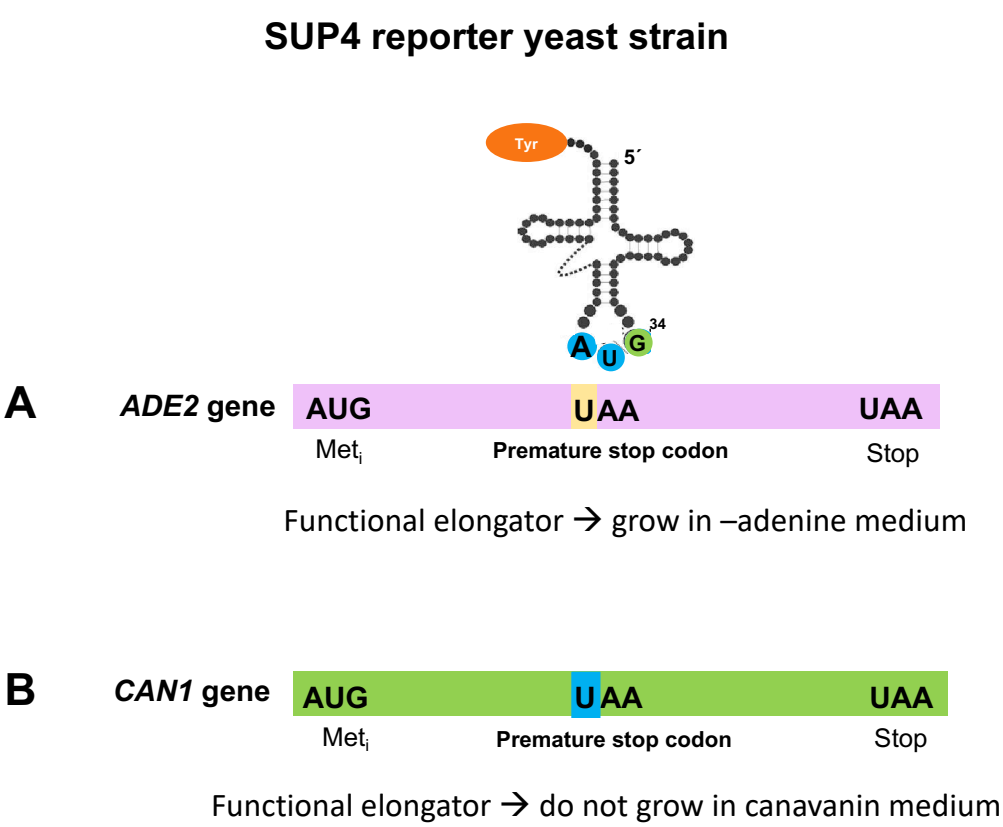


Figure 19. SUP4 reporter strain

(A) Yeast cells carrying a mutation in the *ADE2* gene, which leads to a premature UAA stop codon are able to grow in medium lacking adenine if the Elongator complex modifies the *SUP4-**o***. These cells do not accumulate high level of adenine biosynthesis intermediate and thus are white (or pink). **(B)** Yeast cells carrying a UAA mutation in the *CAN1* gene do not grow in canavanine-containing media if they carry a functional Elongator complex, since the *SUP4-**o*** will suppress the premature stop codon and ultimately allow the incorporation of the toxic canavanine amino acid.

1.1.4.5 RNA Decay

RNA decay is the process by which RNA molecules are degraded. This process plays a key role in regulating both quantity and quality of RNAs present in cells. RNA degradation is performed by enzymatic systems that cleave phosphodiester bonds on the polynucleotides chains to generally generate mononucleotides. This process is performed not only to degrade aberrant transcripts but also to maintain a homeostasis in the temporal and spatial requirements for RNA. The RNA decay process is caried out by various enzymes with exquisite specificities (Table 6).

Table 6. Some key enzymes and factors involved in eukaryotic mRNA decay. Modified from[131].

Protein name	Full name	Function
Deadenylation		
PAN2	Poly(A)-nuclease deadenylation subunit 2	Enzymatically active subunit of the PAN2/3 complex that is responsible for the initial step of poly(A) trimming

Protein name	Full name	Function
PAN3	Poly(A)-nuclease deadenylation subunit 3	Cofactor and regulatory subunit of the PAN2/3 complex that is responsible for the initial step of poly(A) trimming
CNOT6/CNOT6L (CCR4)	CCR4-NOT complex subunit 6	3′ to 5′ exonucleases that constitute one of the catalytically active subunit of the CCR4-NOT complex that performs most of mRNA deadenylation its function is somewhat redundant to CNOT7/CNOT8 (CCR4)
CNOT1 (NOT1)	CCR4-NOT complex subunit 1	Scaffolding component of the CCR4-NOT deadenylation complex
CNOT7/CNOT8 (CAF1/POP2)	CCR4-NOT complex subunit 7	3′ to 5′ exonucleases that constitue one of the catalytically active component of the CCR4-NOT complex that performs most of mRNA deadenylation; its function is somewhat redundant to CNOT6/CNOT6L (CCR4)
PARN	Poly(A)-specific ribonuclease	3′ to 5′ exoribonuclease that interacts with the 5′ cap and preferentially degrades poly(A) tails
Decapping		
DCP2	Decapping protein 2	Catalytically active component of the decapping complex that removes the 5′ mRNA cap through hydrolysis reaction
DCP1	Decapping protein 1	Primary cofactor for DCP2; it enhances hydrolysis activity of DCP2 and interacts with other proteins to recruit DCP1/2 complex to mRNA substrates
PATL1 (Pat1)	PAT1 homolog protein 1	Scaffolding protein in deadenylation-dependent decapping; it connects decapping and the deadenylated 3’ end of mRNAs
Lsm1-7	Sm-like protein complex	Identifies and associates with deadenylated transcripts; it interacts with PATL1 to recruit the DCP1/2 complex
DDX6 (Dhh1)	DEAD-box helicase 6	RNA helicase enzyme that interacts with other deadenylation and decapping factors and is thought to remodel the transcript to promote efficient decapping complex assembly
LSM14 (Scd6)	Sm-like protein 14	Serve as a scaffold for activation of the DCP1/2 complex
EDC1,2,3	Enhancer of mRNA-decapping proteins	Serve as a scaffold for activation of the DCP1/2 complex
5′ to 3′ decay		
XRN1	5′-3′ exoribonuclease	Major 5′ to 3′ exoribonuclease that preferentially degrades RNAs with a 5′ monophosphate
3′ to 5′ decay		
Exo9	Exosome complex (9-subunit)	The core of the exosome that consists of nine different protein subunits
DIS3/RRP44	Exosome complex exonuclease RRP44	Catalytically active subunit of the exosome with 3’-5’ exonuclease and endonuclease activity
SKI7	Superkiller protein 7	GTP-binding protein that recruits the exosome to nonpoly(A) mRNAs; it is involved in several activities, including NSD, NMD, and antiviral
DCPS	Decapping scavenger enzyme	Catalyzes the cleavage of the residual 5′ cap from dinucleotide (or short oligonucleotide) following 3′ to 5′ decay

Protein name	Full name	Function
DIS3L2	DIS3-like exonuclease 2	3′ to 5′ exoribonuclease that specifically recognizes polyuridylated RNAs and mediates their degradation
TUT4/7	Terminal uridyltransferase 4/7	Uridyltransferases that mediate terminal uridylation of mRNAs with short poly(A) tails, thus facilitating their degradation
Endonucleolytic decay		
MCPIP1	Regnase 1/endoribonuclease ZC3H12A	Endonuclease that facilitates mRNA decay and is linked to various biological functions, including immune and inflammatory responses
HRSP12	Heat-responsive protein 12	Protein with endonuclease activity that is involved in activity and formation of the GMD complex
IRE-1	Inositol-requiring enzyme 1	Endonuclease that targets a series of mRNAs connected to several cellular processes
RNase L	Ribonuclease L or ribonuclease 4	Inducible endoribonuclease associated with antiviral response

1.1.4.5.1 Main pathways for mRNA turnover in eukaryotes

Deadenylation (shortening of the poly(A) tail), is the initial step of traditional exonucleolytic mRNA decay pathways. A variety of RNA-binding proteins (RBPs) recruit deadenylases to the substrate RNAs (reviewed in[131]). The first of the two main deadenylation complexes in eukaryotes is the poly(A) nuclease (PAN) PAN2/PAN3 complex, in which PAN2 has the catalytic activity. The PAN2/PAN3 complex is thought to perform the initial RNA trimming; then, the actual deadenylation is performed by the CCR4-NOT complex, with CCR4 and CAF1 performing catalytic function. The second step is the decapping (removal of the 5' m^7Gppp-cap to generate a 5' monophosphate and m^7GDP). The decapping process can be dependent on prior deadenylation or independent when the RNA degradation is initiated by alternative mechanisms. The decapping is mainly carried out by the DCP2/DCP1 complex, and involves many other cofactors including Lsm1-7, PATL1 (Pat1), DDX6 (Dhh1), LSM14 (Scd6), and EDC proteins, but can also be mediated by other factors such as DXO proteins. The third step of the main pathway is the 5' to 3' exoribonucleolytic decay in cytoplasm; in this step XRN1 acts as the primary 5' to 3' cytoplasmic exonuclease[132]. It degrades RNA containing free 5' termini formed by decapping. Another option after deadenylation is 3' to 5' exonucleolytic decay in cytoplasm which is performed by the multi-subunit complex named exosome. Deadenylation generates free 3' mRNA ends which the exosome uses to initiate the decay. 3' ends substrate of exosome are also generated by the RNA-induced silencing complex (RISC). The later recruits the exosome to degrade the 5' product of Ago2-mediated cleavage. After the 3'-5' decay, the m7G cap is removed from the remaining oligo-RNAs by the scavenger decapping DCPS or by the FHIT/Aph1 (Figure 20) (reviewed in [133,134]).

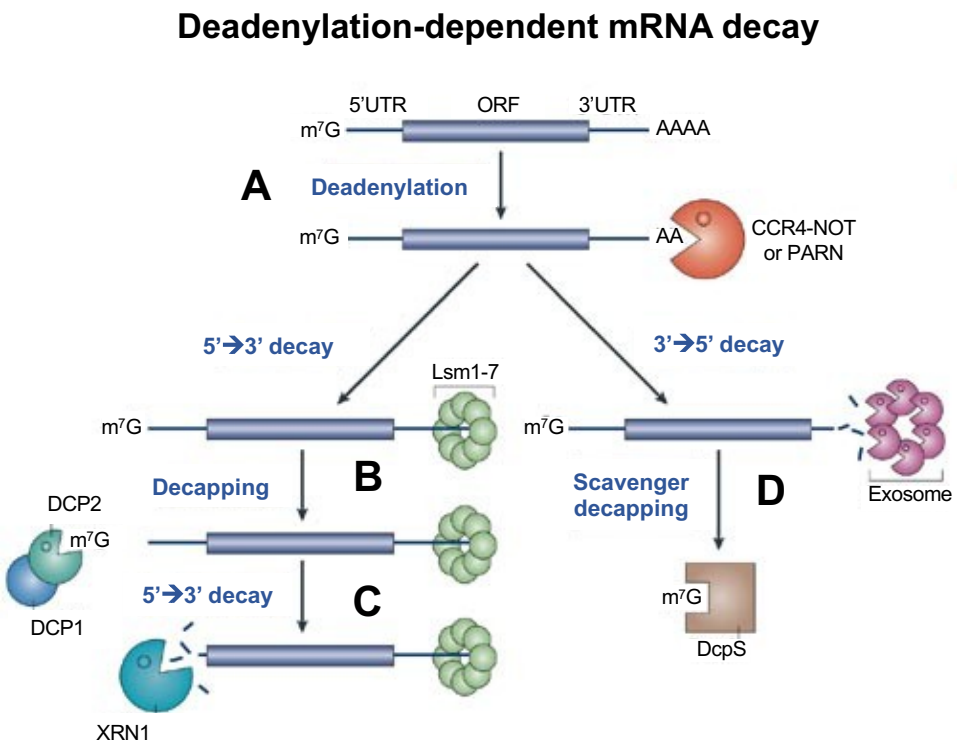


Figure 20. Deadenylation-dependent mRNA decay

(A) Step one: deadenylation, removal of the poly(A) tail. **(B)** Step two (option 1): decapping, removal of the 5′ m^{7}Gppp modification to generate a 5′-monophosphate. **(C)** Step three (option 1): 5′ to 3′ exoribonucleolytic decay in the cytoplasm. **(D)** Step two (option 2): 3′ to 5′ exonucleolytic decay in the cytoplasm. Taken from [135].

1.1.4.5.2 mRNA modification and turnover: m^{6}A

Rates of decay of individual mRNAs can vary in time and space according cellular needs. The study of *N*^{6}-methyladenosine (m^{6}A) modification (see below) as well as the fat mass and obesity associated protein (FTO), an RNA demethylase, have provided useful insights into how the m^{6}A modification contributes to the processing, localization, translation and turnover of eukaryotic mRNAs[44,47]. Some scenarios on how the m^{6}A modification affects the stability of a transcript have been proposed, e.g., mRNA modified with m^{6}A may be subject to binding 'reader' proteins near to the translation termination codon promoting deadenylation of the target mRNA. Another scenario is that m^{6}A modifications could control local RNA secondary structure; then, the m^{6}A modification of a base-paired adenylate leads to exposure of a single-stranded RNA motif, allowing its binding to a cognate RNA-binding protein (RBP); and therefore, leading to stabilization or destabilization of the transcript depending on nature of the RBP (reviewed in [139]).

1.1.4.5.3 mRNA quality control linked to translation

Despite the high fidelity of the transcription and translation processes, there is always the possibility to produce defective transcripts or proteins which will not have a proper function. Therefore, cells have developed quality control pathways to eliminate defective mRNA preventing production of defective proteins. In eukaryotes, there are three main surveillance pathways that

ensure the mRNA quality control. The no-go decay (NGD) pathway degrades mRNAs that are stalled during translation. The non-stop decay (NSD) pathway prevents translation of mRNAs lacking natural stop codons. Finally, the nonsense-mediated (NMD) pathway recognizes and eliminates mRNAs containing premature stop codons (reviewed in [140]).

The NMD pathway operates in the cytoplasm and is closely linked to translation termination. This pathway is not exclusive for mRNAs as long non-coding RNAs (lncRNAs) can also be substrates of NMD. A key characteristic of NMD is the ability to distinguish between normal and premature terminator codons; once a codon is recognized as premature, it recruits and activate enzymes to destroy the transcript (reviewed in [141]). Obviously, tRNA suppressors may counteract NMD by preventing premature stop codon recognition. Thus, suppressor tRNA increase production of genes carrying a stop codon mutation both by stabilizing the corresponding transcript (if subject to NMD) and by allowing translation through the stop codon.

1.1.4.5.4 tRNA quality control

At least two pathways are responsible for quality control at different stages of the tRNA maturation. The first one, which acts at a premature tRNA level, was discovered when Kadaba *et al.* found that pre-$tRNA_i^{Met}$ lacking m^1A_{58} due to a mutation on *GCD10* and *GCD14* are degraded by the exosome and Rrp6 after polyadenylation by Trf4[142,143], a component of the Trf4/Air2/Mtr4p polyadenylation complex (TRAMP) complex[144,145]. In the second pathway, mature tRNAs lacking nonessential post-transcriptional modifications are degraded by the rapid tRNA decay (RTD)[146,147]. For example, yeast strains lacking m^7G_{46} and m^5C_{49} modifications on $tRNA^{Val(AAC)}$ as a result of the deletion of the *TRM8* and *TRM4* genes, are temperature-sensitive and have growth defects at 37ºC; in these strains the $tRNA^{Val(AAC)}$ is deacylated and degraded via the RTD pathway independently from the *TRF4*/*RRP6*[148].

1.2 POST-TRANSCRIPTIONAL MODIFICATIONS

RNA molecules are frequently subject to chemical modifications. Since the discovery of the first chemically modified nucleoside, pseudouridine more than 70 years ago[149], the presence of modified ribonucleosides has been recognized. Modifications carried by various RNA molecules have been attributed a particular importance by scientists due to their biological impact, being indispensable for a wide range of cellular processes. Therefore, understanding their distribution, function and regulation will shed light on the processes in RNA biology.

Post-transcriptional modifications of RNA are defined as a set of biochemical steps in which an RNA transcript is altered after been transcribed (reviewed in [150]). These modifications include the addition, removal or substitution of chemical groups of the four canonical nucleosides (adenosine, guanosine, uridine, cytidine), for example: methylation (guanosine → 7-methylguanosine), deamination (adenine → inosine), sulfur substitution (uridine → 4-thiouridine); or structural rearrangements (uridine → pseudouridine) and the saturation of existing double bonds (uridine → dihydrouridine) (reviewed in [151–153]). Among the broad range of functions regulated by RNA modifications are: translation initiation, RNA processing, splicing, polyadenylation, structure, localization and stability (reviewed in [154,155]).

RNA modifications occur in organisms of all three domains of life. Different types of RNA including, such as messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear RNA (snRNA) can be modified. Initially, studies were limited to the more abundant modifications: N^6-methyladenosine (m^6A), 5-methylcytidine (m^5C) and 2′O methylation (2′OMe) in mRNA and viral RNA; and pseudouridine (Ψ), N^1-methyladenosine (m^1A) and 2′-O-methylation (2′OMe) in tRNA and rRNA (reviewed in [156]). However, technological advances of the last decades have allowed more complex analysis of both less frequent modifications and less abundant RNA species [150].

According to MODOMICS database[157], 173 different modifications have been identified to date in the so called epitranscriptome[158], a collection of biochemical modifications of the transcriptome in coding and non-coding RNAs, with the majority reported in tRNAs and rRNAs[157]. Modifications cover a wide range of chemical structures, from the simple addition of a methyl group to the very complex multistep alterations. These chemical marks can be unique to a specific type of RNA or can be shared among several (Figure 21).

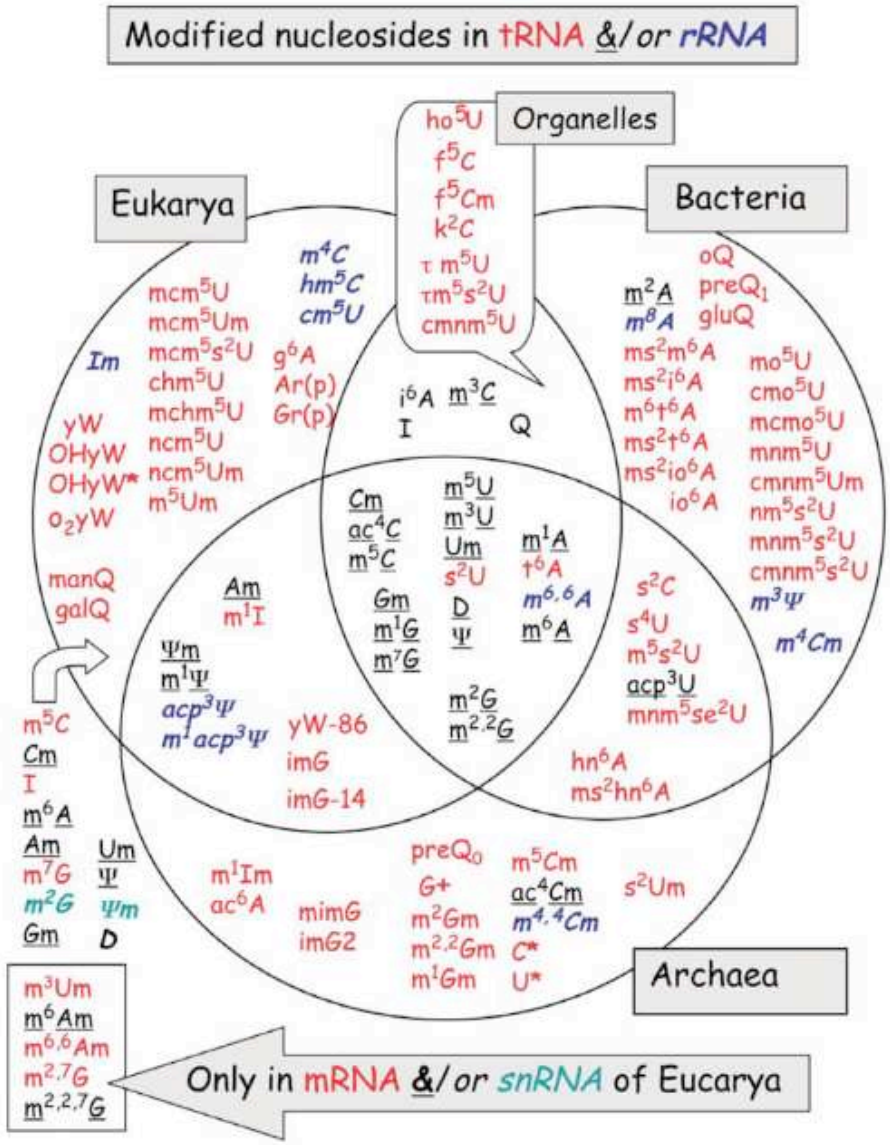


Figure 21. Phylogenetic distribution of some identified RNA modifications presents the three domains of life

Diagram shows 107 out of the 173 identified nucleosides in Archaea, Bacteria and Eukarya. Color code represents those that are belonging to a particular type of RNA or shared between types. Inside the circles: tRNA (red), rRNA (blue), both (black). Modifications presented in the organelles section, refer to mitochondrial and chloroplastic modifications which, while they are present in eukaryotic cells have an ancient bacterial origin. Outside the circles are presented modifications belonging to the Eukarya domain: mRNA (red), snRNA (green), both (black). Inside a box are present the modifications of the eukaryotic group that are unique to mRNAs and/or snRNAs, while the others are also shared by tRNAs and/or rRNAs. Taken from [159].

Post transcriptional modifications in figure 21 are presented as systemic abbreviations; a short name which describes its chemical structure. In table 7 some of the most common chemical modifications present in RNA are listed in alphabetical order.

Table 7. Abbreviations of the chemical groups present in the most common RNA modifications.

Abbreviation	Modification	Abbreviation	Modification
ac	acetyl	man	mannosyl
acp	aminocarboxypropyl	mchm	carboxyhydroxymethyl methylester
chm	carboxyhydroxymethyl	mcm	methoxycarbonylmethyl
cmo	oxyacetic acid	mcmo	oxyacetic acid methyl ester
cmnm	carboxymethylaminomethy	mnm	methylaminomethyl
f	formyl	mo	methoxy
g	glycinyl	ncm	carbamoylmethyl
gal	galactosyl	nm	aminomethyl
hn	hydroxynorvalylcarbamoyl	r(p)	5-O-phosphono-b-D-ribofuranosyl
ho	hydroxy	s	thio
i	isopentenyl	se	seleno
inm	isopentenylaminomethyl	t	threonylcarbamoy
io	cis-hydroxyisopentenyl	tm	taurinomethyl
m	methyl	2′OMe	2′-O-methylation
cm	carboxymethyl		

The systemic abbreviation of RNA modification is built using combinations of the abbreviations according the chemical moieties present in the nucleotides. Figure 22 depicts schematization of the systemic abbreviations. In the top part the examples of purine and pyrimidine nucleosides are shown; and in the bottom two examples of RNA modifications: N^2,N^2,2′-O-trimethylguanosine ($m^2{}_2$Gm) on the left and 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U) on the right. Abbreviations before the nitrogen base letter (A,U,C,G) represent a modification in purine or pyrimidine rings, while abbreviations after the nitrogen base describes an alteration in the ribose. Superscripted numbers precises the modified position of the base and subscripted numbers refer to the number of identical moieties at the same position.

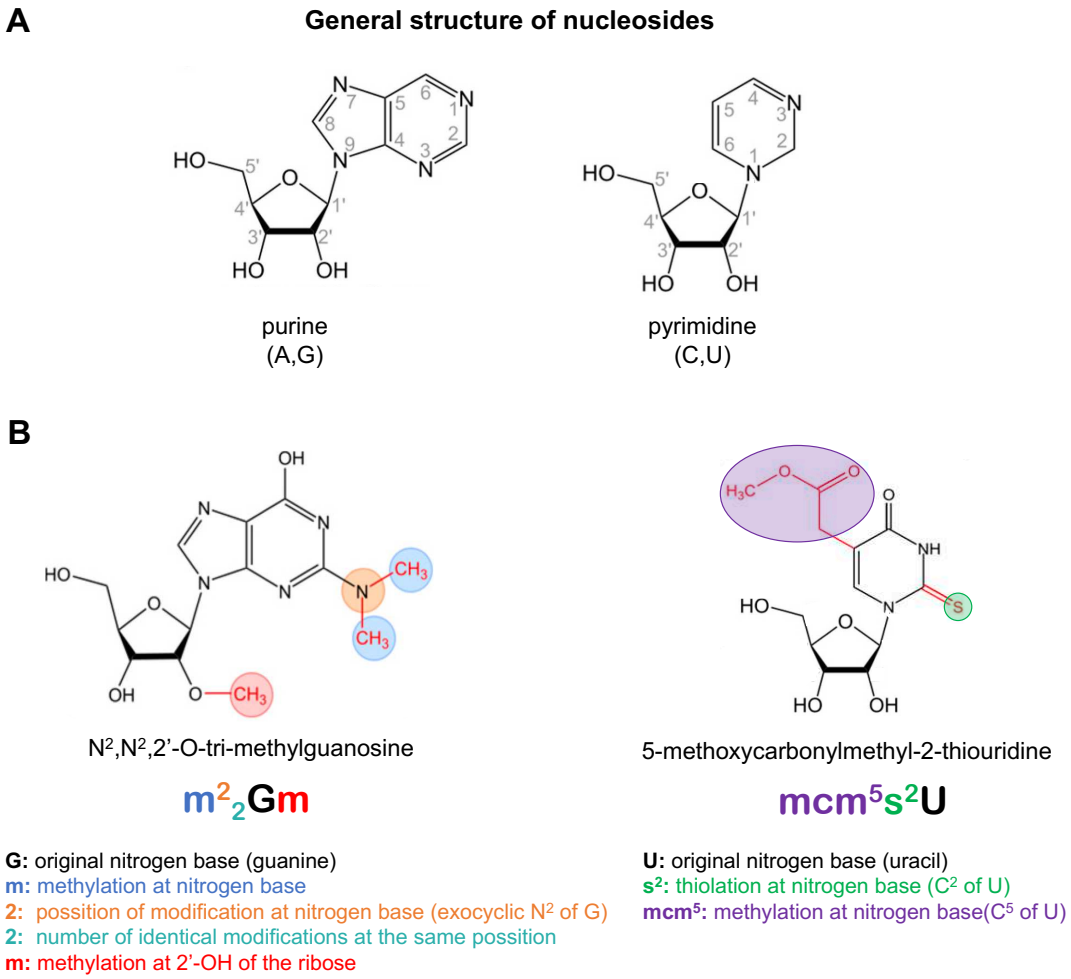


Figure 22. Representation of the systemic abbreviation of RNA modification according its addition of its chemical groups
(A) General nucleoside structure. **(B)** Examples of systemic abbreviation usage.

The emergence of next generation sequencing technologies coupled to RNA immunoprecipitation methods such as RIP, RIP-seq, RIP-chip and derivatives have led to the identification of specific enzymes that interact with RNA fragments that contain modified nucleosides such as m^6A (reviewed in [150,158]). Enzymes interacting with modifications, also called RNA modifying proteins (RMPs) are normally multiprotein complexes which have been classified in three principal groups "writers" which add the chemical group to the RNA, "readers" that carry out the function of the modification and "erasers" which are able to remove the mark. The reversibility of RNA modifications highlights the dynamic nature of the RNA marks and suggest their use as "switches" to control gene expression in response to external or internal stimuli[160,161]. However, not all the modifications are dynamic; some of them are unidirectional and nonreversible, such as pseudouridine, adenosine-to-inosine (A to I), cytidine to uridine (C to U) and ribose 2′-O-methylation, which are among the most common examples of eukaryotic RNA modifications.

Aberrant modifications as well as defects of enzymes catalyzing RNA modifications are directly linked to a variety of human diseases including neurological and metabolic disorders, genetic birth defects, infertility and cancer (reviewed in [162,163]). For example, mutation of the

KEOPS complex, responsible for the N6-threonyl-carbamoylation of adenosine 37 of ANN-type tRNAs (t^6A) results in early-onset steroid-resistant nephrotic syndrome and microcephaly[164,165]. If a tRNA molecule is lacking a chemical group, the protein translation can be interrupted at many sites, resulting in an increased amount of protein aggregates[166].

In the last decades, advances of clinical oncology have been compromised by the patient acquisition of cancer drug resistance, which is often consequence of the metabolic plasticity of cancer cells provoked by defects on the post-transcriptional modifications pathways (reviewed in [167,168]). Table 8 shows some examples of human diseases attributed to a dysregulation of certain RNA modifications. These dysregulations are usually a consequence of defects in proteins creating such modifications.

Table 8. RNA modifications associated to some human diseases. Modified from [169]

Affected modification	Mutation	Gene	Disease
m^6A	Polymorphism	*FTO*	Obesity, type 2 diabetes
m^6A	Inactivating mutation	*FTO*	Autosomal-recessive lethal syndrome. Overcoming drug resistance in cancer
Ψ	Mutation	*DKC1*	X-linked form of dyskeratosis congenita
Ψ	Mutation	*DKC1*	Hoyeraal-Hreidarsson syndrome
m^5C	Inactivating mutation	*NSUN2*	Intellectual disability and Dubowitz-like syndrome
m^5C, f^5C	Inactivating mutation	*NSUN3*	Mitochondrial disease
mcm^5s^2U	Inactivating mutation	*IKBKAP*	Familial dysautonomia (recessive neurodegenerative disorder)
m^7G	Inactivating mutation	*WDR4*	Microcephalic primordial dwarfism
2′-O-methylation	Mutations and copy-number variation	*FTSJ1*	Non-syndromic X-linked intellectual disability
i^6A	Inactivating mutation	*TRIT1*	Encephalopathy and myoclonic epilepsy
I	Mutation	*ADAT3*	Intellectual disability and strabismus
m^1G	Mutation	*TRMT10A*	Young onset diabetes, short stature and microcephaly with intellectual disability
m^2G and $m^2{}_2G$	Frameshift alteration	*TRMT1*	Intellectual disability
t^6A	Mutation	*YRDC, GON7*	Galloway-Mowat syndrome

5-methylcytidine (m^5C) as an example is a common modification that can be found in the three domains of life and it is present in a wide range of RNA types. Particularly in tRNA, it influences translation rates. Blanco et al. [170] demonstrated that in humans and mice, the lack of NSun2 protein leads to a complete abolishment of m^5C in a broad set of tRNAs. The loss of m^5C increases the angiogenin-mediated endonucleolytic cleavage of tRNAs. Angiogenin is a ribonuclease that promotes formation of blood vessels, and this cleavage leads to the accumulation of 5' tRNA-derived small RNA fragments which activate stress pathways and reduce protein translation rates, mechanisms that increase apoptosis of cortical, hippocampal and neuronal cells[170]. Whereas mRNA hypermethylation with m^5C can be found in cancer-related pathways such as the Mitogen-Activated Protein Kinases (MAPK) or the Phosphatidylinositol 3-Kinase (AKT) and oncogenes like Heparin Binding Growth Factor (*HDGF*) in cancer tissue samples compared with normal tissues[171].

1.2.1 mRNA and rRNA post-transcriptional modifications

In addition to the terminal modifications of mRNA discussed in section 1.1.4.3 that play a key role in RNA stability, nuclear export and translation, mRNA undergoes other types of post transcriptional modifications. The implementation of poly(A) tail-based mRNA purification techniques in the early 1970s allowed the obtention of pure mRNA preparations, shedding light on the existence of a wide group of post-transcriptional modifications. The first modification to be identified in mRNA was the N^6-methyladenosine (m^6A) which is the major form of internal methylation and is estimated to be present in 25% mammalian mRNA[44,47]. As the most abundant internal modification in eukaryotic mRNA, m^6A also has been the most extensively studied. New analyses have revealed many other mRNA chemical modifications including: N^6,2′-O-dimethyladenosineinosine (m^6Am), N^1-methyladenosine (m^1A), 5-methylcytosine (m^5C), 5-hydroxymethylcytosine (hm^5C), N^4-acetylcytidine (ac4C), pseudouridine (Ψ), N^7-methylguanosine (m^7G), and 2′-O-methylated nucleotides (Nm).

mRNA modifications are not randomly distributed, they occur in specific locations at certain sequence motifs and their position correlates with their molecular function. m^6A for example, is primarily localized at the end of an ORF suggesting a potential role in regulating translation termination and RNA decay, while m^1A and m^6Am are enriched at the 5'UTR of a mRNA (Figure 23).

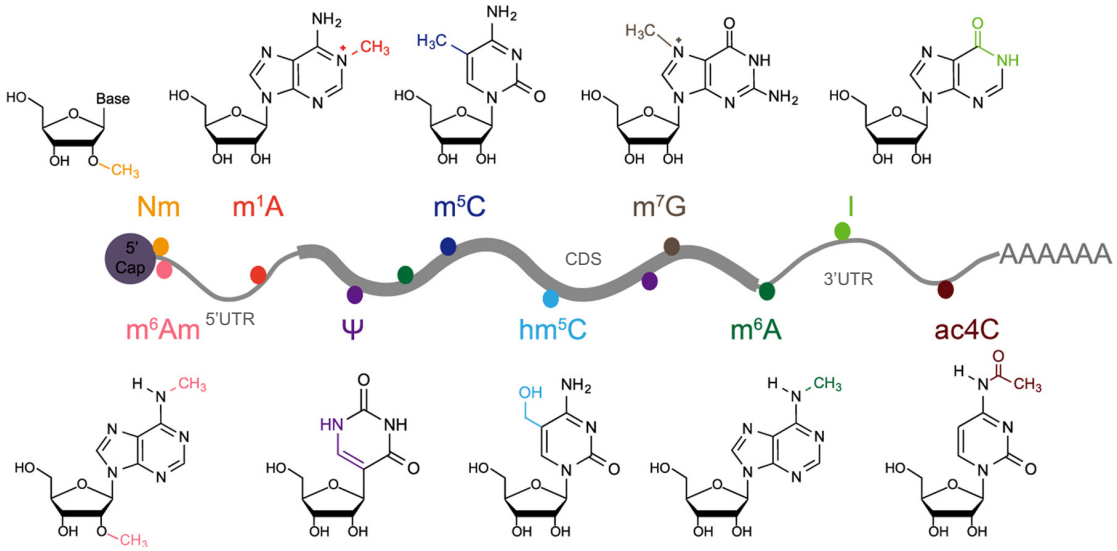


Figure 23. Structure and frequent localization of mRNA modifications
Predominant locations of specific modifications in transcripts. mRNA is divided into 5' untranslated region (5'UTR), coding sequence (CDS) and 3'UTR. Taken from[172].

In ribosomal RNA (rRNA) methylation of nucleotides at both nucleobase and ribose are the predominant type of modification. The majority of the methylation sites are clustered near the ribosome functional centers: decoding and peptidyl transferase centers. Their function is not totally yet elucidated; however, it can be associated with the structural stabilization of RNA substrate as well as the modulation of ribosome function including maintenance and translational fidelity (reviewed in [173,174]).

1.2.2 tRNA post-transcriptional modifications

In the early 1960's, the sequencing of the first biological RNA, the yeast alanine transfer RNA, identified 10 modified nucleosides including pseudouridine (Ψ)[175]. Over 170 various modifications have since been discovered in RNA, and around 100 are exclusively found in transfer RNAs[157]. This makes tRNA the molecule with the largest number of modified nucleotides from all RNA species. Depending on the organism, tRNAs can have up to 23% of their nucleotides modified (~14 modifications per tRNA), while demonstrating also the highest chemical complexity and variety of modifications[157]. These modifications vary in frequency of their occurrence, some of them are present in almost all the tRNAs while others are unique to isoaceptors tRNAs or limited to specific organisms. The RNA database MODOMICS[157] shows the distribution of modifications in tRNA. The highest average level of post-transcriptionally modified residues is found in the cytosol of eukaryotic single cell organisms, fungi and metazoa (16%), while gram-negative and gram-positive bacteria contain the lowest fraction of modifications with an average of 9 and 6 percent respectively (Table 9). In general, eukaryotic tRNAs contain a higher number of modifications that their bacteria homologues.

Table 9. Minimal, average and maximal percentage of modified residues per tRNA molecule in 7 organism groups. Taken from[176].

Group	Minimum	Average	Maximum
Gram-negative bacteria	4%	9.55%	16.88%
Gram-positive bacteria	2.6%	6.53%	10.53%
Cytosol of eukaryotic single cell organisms, Fungi and Metazoa	3.53%	16.53%	23.08%
Mitochondria	1.67%	8.58%	17.86%
Plastids	6.49%	10.22%	16.88%

The large diversity of tRNA modifications can be grouped in three main categories:

- In the first category are the specific modifications which occupy conserved positions in the majority of the tRNA species, for example 2′-*O*-methylguanosine (Gm) and dihydrouridine (D) which belong the D-loop, or pseudouridine (Ψ) and 5-methyluridine (m^5U) present in the TΨC-loop.
- The second category groups modifications which are present in the same positions but are not necessarily chemically identical. These modifications include nucleotides at positions 34 and 37 of the anticodon loop.
- In the final category are all other modifications that are not included in the first two categories[176].

More than half of the tRNA modifications are specific to one domain of Life and only one fifth is shared by the three domains and another one fifth is in the overlapping regions between domains (Figure 24). Simple modifications like pseudouridine (Ψ), inosine (I), 1-methyladenosine (m^1A) and 5-methylcytidine (m^5C) are found in the three domains. Pseudouridine is an isomer

form of uridine arising from the substitution of the canonical carbon-nitrogen glycosidic bond (C1'-N1) with a carbon-carbon bond (C1'-C5) between the uracil and the ribose. This isomerization reaction is catalyzed by the pseudouridine synthases (PUS) family (reviewed in [177]). Inosine is a modification that can be found at three different localizations of tRNAs: position 34, 37 and 57. It is produced as result of a deamination reaction of adenine catalyzed by the adenosine deaminases acting on tRNAs (ADATs)[178].

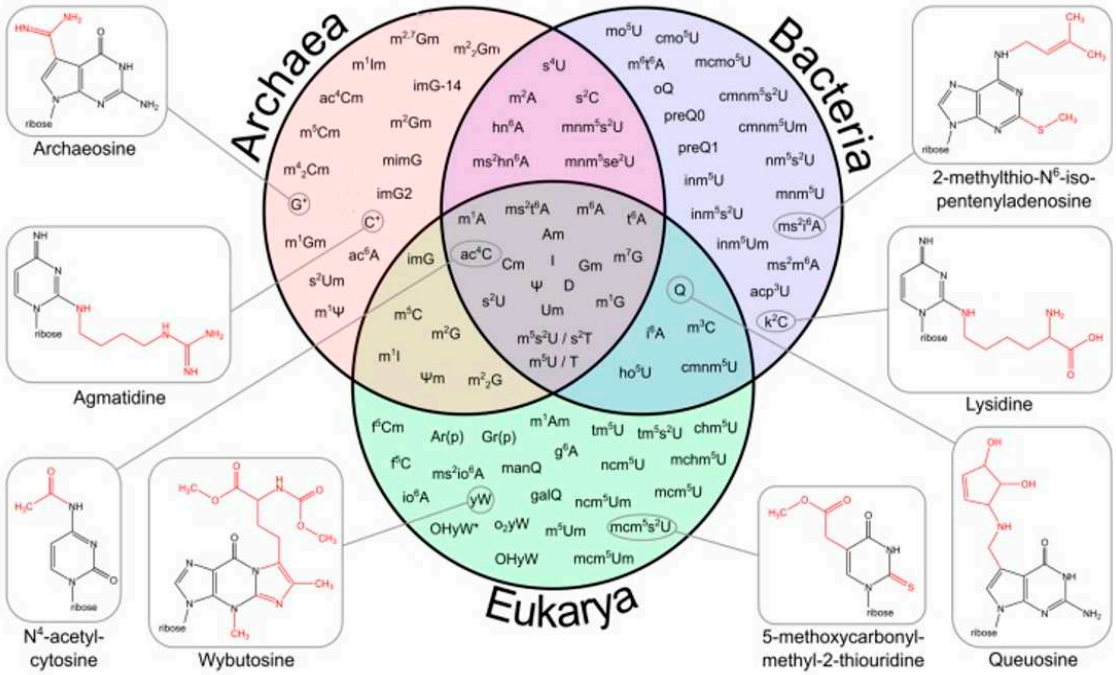


Figure 24. Distribution of tRNA modifications into the three domains of life.
The Venn diagram shows the spread of tRNA modifications between three domains of life Archaea (pink) Bacteria (purple) and Eukarya (green). The chemical structures of select examples are also shown[153].

Modifications are present throughout all the tRNA structure (Figure 25). Depending on the particular position, these modifications can contribute to specific functions of the tRNA[179]. For example, modifications present on the tRNA are important for establishing tertiary contacts between the D-loop and the TΨC-loop; this interaction contributes to the structural stabilization of the tRNA secondary L-shape. In bacteria and archaea, tRNA modifications in the D-loop and the TΨC-loop also maintain structural rigidity in a wide temperature ranges; whereby, fully modified tRNA transcripts are densely folded and thermostable[153]. On the other hand, modified residues at the anticodon loop influence efficiency and fidelity of translation as well as for a proper aminoacylation process (reviewed in [151]).

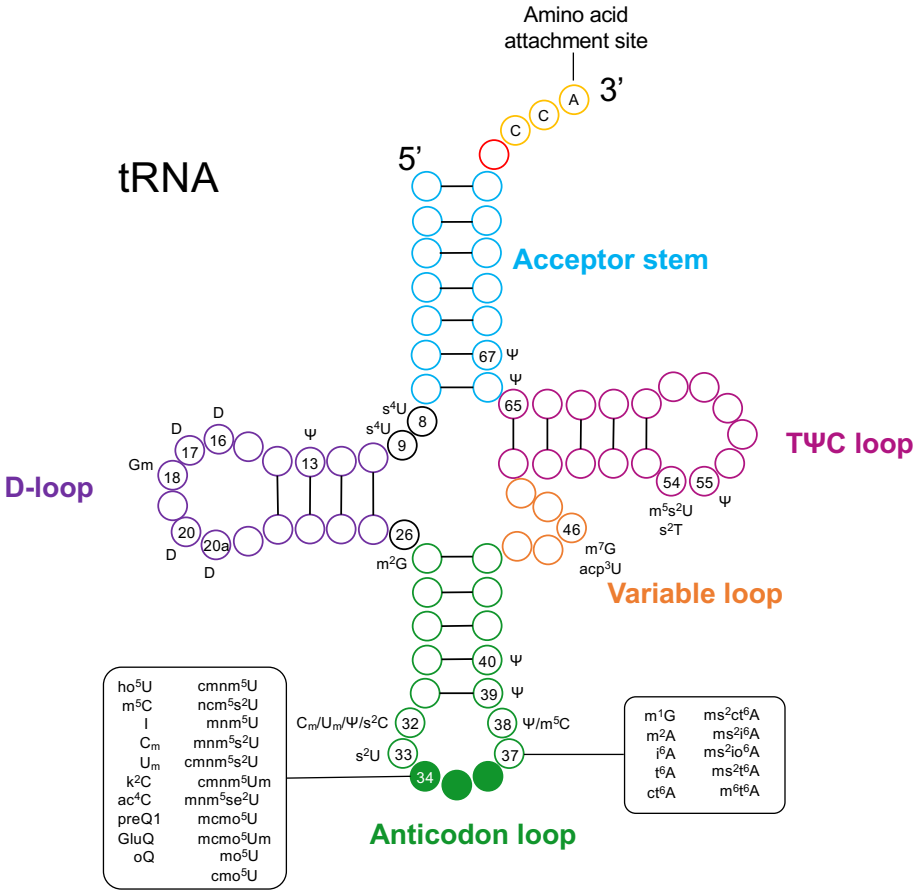


Figure 25. Most common tRNA post-transcriptional modifications
The acceptor stem is marked blue, the D-loop - purple, the TΨC-loop - magenta and the anticodon loop - green. Positions 34 and 37 can carry the widest variety of modifications. Modified from [180].

Translation of a mRNA codon requires pairing interaction with the tRNA anticodon of its cognate tRNA. Due to the genetic code degeneracy multiple codons can pair with one anticodon, each interaction having different stability and efficiency. Because of this, translation efficiency rests on three main parameters, the first one being the biased codon usage, the second being abundance of various types of tRNAs, and the third being the post-transcriptional chemical modifications of tRNA anticodons or adjacent nucleotides. The latter is dynamic and flexible, and it can be adapted to specific time and stress conditions. Actually, some of the tRNA modifications are stress-induced and confer the ability to regulate cellular adaptation (reviewed in [181]).

Two crucial positions for tRNA post-transcriptional modifications occur at residue number 34 also known as wobble position and at residue 37 which is the 3' adjacent site to the anticodon. These positions play a key role in translation accuracy and reading-frame maintenance as they demonstrate strong effect on codon-anticodon pairing interactions[182].

1.2.2.1 U_{34} (wobble) tRNA modification

According to the universal genetic code, there are 61 codons that are decoded by complementary tRNA anticodon sequences and three codons that are recognized as stop codons. However, the canonical Watson-Crick pairing does not explain how 61 amino acid codons can be decoded by far fewer tRNAs, e.g., 42 in yeast (reviewed in [183]).

In the Wobble Hypothesis[184], Francis Crick proposed that the first two base pairs between codon and anticodon always follow a canonical pairing between purines and pyrimidines (A-U, U-A, C-G or G-C). However, the third base could also follow or not canonical pairing and instead pair in the following way: first base G of tRNA can recognize U or C, first base U can recognize A or G, and first base inosine (I) can recognize U, C or A (Figure 26). Therefore, additional wobble base pairs are guanine-uracil (G-U), inosine-adenine (I-A), inosine-uracil (I-U) and inosine-cytosine (I-C). A wobble base pair is then, a pair between two RNA nucleotides that uses expanded set of Watson-Crick base pair rules.

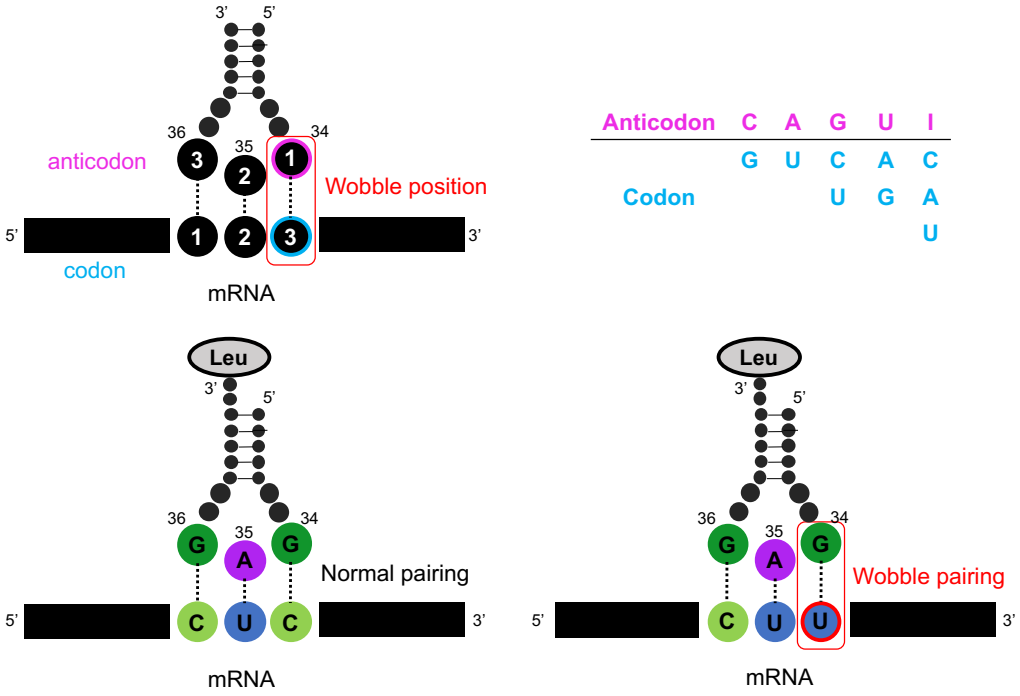


Figure 26. Crick´s wobble hypothesis

Non-canonical pairing between the first anticodon nucleotide and the third codon nucleotide. The two first nucleotides of codon and last two of anticodon are more critical for selection of the correct tRNA.

In 1991, a Modified Wobble Hypothesis[185] was suggested stating the wobble pairing was not restricted to only four additional wobble base pairs (G-U, I-A, I-U, I-C), but that it was also mediated by the presence of post-transcriptional modifications (other than inosine), which change the architecture of nucleotide at position 34 and its base pairing capability (reviewed in [183]).

From broad range of post-transcriptional modifications found in tRNAs, the modifications of uridine 34 are well conserved among eukaryotes. This evolutionary maintenance suggests its biological importance; however, the absence of U_{34} modifications does not cause lethality in

several yeasts as well as in *C. elegans*[186], but it does in a specific *S. cerevisiae*[187,188] strains and in mice[189] .

In yeast, from the 64 total codons in the genetic code, 16 codons ending with an A are potentially complementary to tRNAs carrying a uridine at wobble position. Of these UAA and UGA are stop codons recognized by Release Factors (e.g., eRF1 in eukaryotes), thus no natural tRNA recognizes them. For the remaining 14 anticodons, $tRNA^{Ile(UAU)}$ carries a pseudouridine, $tRNA^{Leu(UAG)}$ is not modified and $tRNA^{Arg(UCG)}$ does not exist in yeast. The remaining 11 tRNA species contain 4 different types of modifications: 5-methoxycarbonyl-methyluridine (mcm^5U), 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U), 5-carbamoylmethyl-2'-O-methyluridine (ncm^5Um) and 5-carbamoylmethyluridine (ncm^5U) (Figure 27).

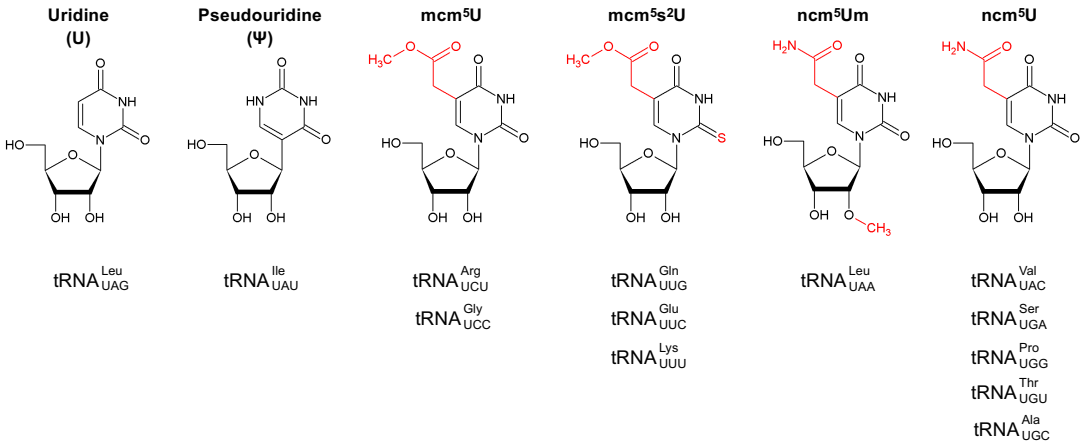


Figure 27. Modifications of tRNA wobble uridines (U_{34})
Abbreviation and structure of uridine derivatives found in yeast cytoplasmic tRNAs. Chemical modifications are highlighted in red. Abbreviations: mcm^5U: 5-methoxycarbonyl-methyluridine; mcm^5s^2U: 5-methoxycarbonylmethyl-2-thiouridine; ncm^5Um: 5-carbamoylmethyl-2'-O-methyluridine and ncm^5U: 5-carbamoylmethyluridine. Modified from [190].

Sulfur-containing 2-thio (s^2) modifications are commonly found on uridine at wobble position (U_{34}) of tRNA molecules coding for glutamate, glutamine and lysine. The exchange of the uridine C2-oxygen atom for a sulfur atom improves the s^2U stacking interaction and RNA stability; therefore, s^2U_{34} enhances stacking interaction between anticodon loop nucleotides and stabilizes codon-anticodon pairing compared to unmodified U_{34}[191]. While the role of s^2U_{34} modification is to restrict codon conformation, the mcm/ncm^5U function is to stabilize U-G wobble pairing[192]. Indeed, the principal activity of the mcm^5 group is to modify the electron structure of the ring and to shift the keto-enol equilibrium towards enol form; therefore, facilitating U_{34}-G wobble pairing (reviewed in [190]). Loss of the $mcm^5s^2U_{34}$ modification in $tRNA^{Lys}$ reduces growth and disbalances protein homeostasis in *S. cerevisiae* by affecting recognition of cognate codons during yeast translation. In contrast, in bacteria U_{34} modification prevents missense errors in decoding (reviewed in [180]). Modifications in anticodon loop guarantee mRNA decoding fidelity; X-ray structural data from Rozov A. and co-workers demonstrated that in the case of $tRNA^{Lys(UUU)}$ the modified 5-methylaminomethyl-2-thiouridine ($mnm^5s^2U_{34}$) nucleotide influences base-pairing interactions in

the wobble position, thus, helping to discriminate between cognate and near-cognate codons and emphasizing the powerful role of tRNA modifications in translation[193].

1.2.3 tRNA Elongator-dependent modifications

Behind all RNA modifications, there is a set of enzymes which catalyze their formation. The synthesis of modified derivatives of uridine at position 34 (U_{34}) of tRNA, such as mcm^5U_{34} and $mcm^5s^2U_{34}$ is highly complex and requires at least eleven enzymes for the s^2 group formation and 15 for the mcm^5 addition (reviewed in [166,194]). In eukaryotes, eleven out of thirteen tRNAs carrying a uridine at wobble position contain a 5-carboxymethyl (cm^5) group which is added by the Elongator complex (Elp1-Elp6) in the presence of acetyl-CoA, S-adenosylmethionine, an electron donor (Kti11-Kti13 heterodimer) with cofactors (Kti12 and Kti14), the suppressor of transcription initiation (Sit4) phosphatase, and its associated partners Sap185 and Sap190 (Figure 28). The yeast cytochrome *b_5* reductase (Cbr1) was identified as a NADH-dependent reductase cofactoring Kti11 in the electron transfer for radical SAM enzymes[195]. Other reductases such as Mcr1 and Ncp1 are also considered as possible actors in Kti11 redox reaction (reviewed in [196]). Interestingly, in this reaction SAM is not used as a methyl group donor but as substrate for radical formation.

The cm^5U is an intermediate which can either be converted into a 5-carbamoylmethyl (ncm^5) group by an unknown process; or used by the methyltransferase complex Trm9/Trm112 as a substrate for the formation of mcm^5U_{34} on $tRNA^{Arg(UCU)}$, $tRNA^{Gly(UCC)}$, $tRNA^{Gln(UUG)}$, $tRNA^{Glu(UUC)}$ and $tRNA^{Lys(UUU)}$, using S-adenosyl methionine (SAM) as methyl donor and producing S-adenosyl homocysteine (SAH) as by product[197] (Figure 28). In three of these tRNAs, $tRNA^{Gln(UUG)}$, $tRNA^{Glu(UUC)}$ and $tRNA^{Lys(UUU)}$ the position 34 is further modified by addition of 2-thio (s^2) group resulting in formation of $mcm^5s^2U_{34}$ nucleotide (reviewed in [190,194,196,198]). The s^2 group addition requires the ubiquitin-related modifier 1 (Urm1), its activator protein Uba4, as well as other associated partners such as the cysteine desulfurase (Nfs1), the thiouridine modification protein (Tum1), the iron-sulfur cluster binding proteins (Isu1, Isu2, Cfd1, Cia1, Nbp35) and the s^2U ligases Ncs2 and Ncs6. Cysteine is used as the source of sulfur (reviewed in[194,196]) (Figure 28).

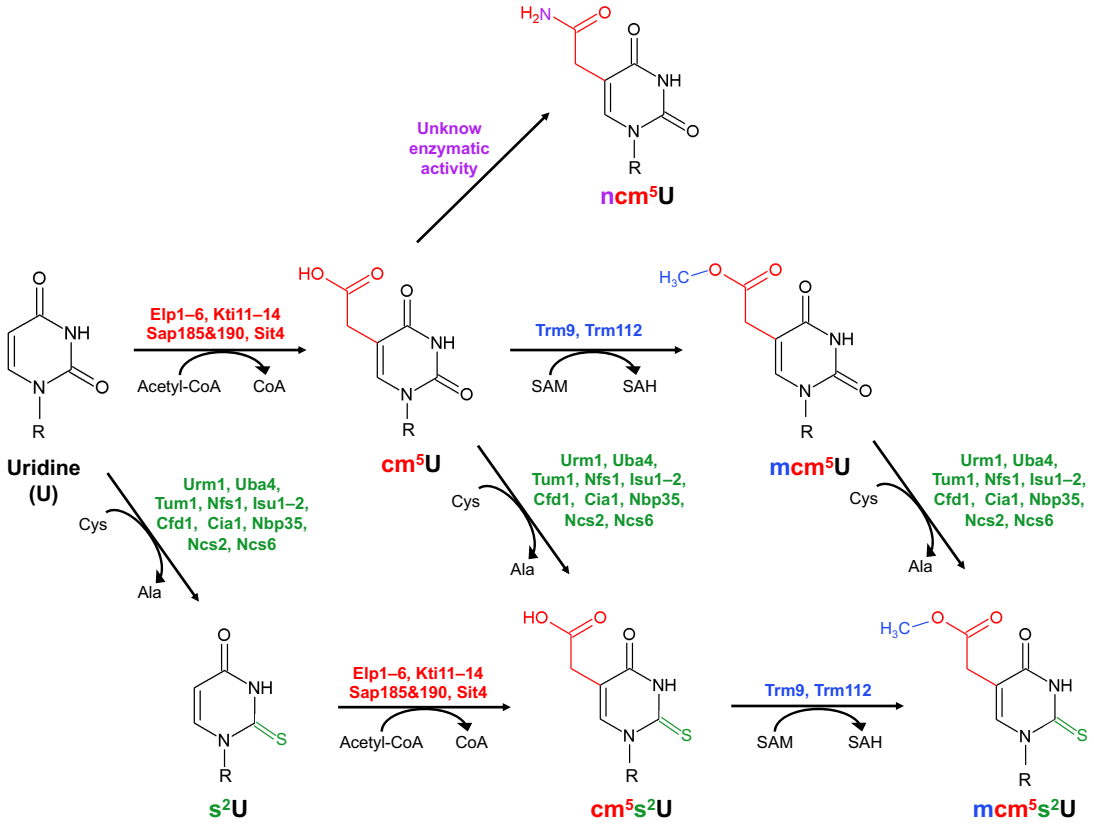


Figure 28.Elongator dependent tRNA modification pathway in yeast

Different pathways for the modification of uridine at position 34 (U_{34}) of tRNA are shown. In red, is shown the addition of cm^5 mediated by Elongator complex and its associated partners using acetyl-CoA as substrate and SAM as cofactor. In green, is shown the s^2 substitution by Urm1, Uba4 and partners using cysteine as sulfur donor. In blue, is indicated the methylation by Trm9 and Trm112 using SAM as methyl donor. The reaction involved in the formation of ncm^5 remains unknown. All possible reactions are shown but those may, or may not, follow an obligatory order. Modified from [196].

The reactions to form cm^5U_{34} derivatives may not follow an obligatory order. In 2011 yeast strains lacking Trm9 and Trm112 were analyzed by Chen *et al.*[199]. They observed an accumulation of ncm^5U and ncm^5s^2U nucleosides in total tRNA instead of the expected cm^5U and cm^5s^2U. This led to the proposal that cm^5U is formed before ncm^5U and mcm^5U or that there is an additional pathway to generate cm^5U from ncm^5U by an unknown enzymatic activity.

1.3 ELONGATOR COMPLEX

Elongator is a multi-subunit protein complex highly conserved among eukaryotes. It is composed of two copies of six proteins named Elp1-6 which are arranged into two subcomplexes: Elp123 and Elp456. The whole complex is then composed of twelve proteins with total molecular weight of ~850 kDa[200]. Elongator homologs are found among eukaryotic organisms including but not limited to yeast, nematodes, insects and mammals. Table 10 shows some of the different aliases of yeast Elongator subunits.

Table 10. Different aliases for the yeast Elongator subunits

Systematic name	Elongator protein	Killer Toxin Insensitive	Insensitive to Killer toxin	Zymocin target complex
YLR384C	Elp1	KTI7	IKI3	TOT1
YGR200C	Elp2	KTI3	—	TOT2
YPL086C	Elp3	KTI8	—	TOT3
YPL101W	Elp4	KTI9	—	TOT7
YHR187W	Elp5	—	IKI1	TOT5
YMR312W	Elp6	KTI4	—	TOT6

1.3.1 Discovery of the Elongator complex

The Elongator complex was first identified in the *Saccharomyces cerevisiae* yeast as part of the RNA Polymerase II (RNAPII) complex by Svejstrup´s team in 1999[201]. They observed that during gel filtration as the last step of the RNAPII purification, the eluted species were significantly bigger in size (5-600 kDa) than the expected RNAPII core. Accordingly, additional polypeptides were observed coeluting with subunits of the RNAPII. A peptide mass fingerprinting using MALDI-TOF mass spectrometry identified a polypeptide of 150 kDa encoded by the YLR384C ORF localized on yeast chromosome XII. Sensitive database searches failed to find homology to any known protein and the team named the gene *ELP1* for Elongator protein 1. Further purifications of Elp1 showed that the protein coeluted with two other factors of 90 (Elp2) and 60 kDa (Elp3), suggesting these 3 proteins as subunits of a complex. These three novel proteins also precipitated with the RNA Polymerase II holoenzyme, so the team named the complex “Elongator” to indicate their association with the transcription elongation form of the RNAPII. Phenotypic assays in yeast lacking the Elp1 gene shown a variety of phenotypes including slow growth adaptation under new conditions, slow gene activation, salt and temperature sensitivity.

Later in the same year, the Svejstrup´s team identified Elp3, the 60 kDa protein, as a highly conserved protein with a histone acetyltransferase (HAT) signature. The 60 kDa protein was analyzed with peptide mass fingerprinting and the gene was named *ELP3* for Elongator protein 3. Database mining revealed that the C-terminal domain (CTD) of Elp3 contains Gcn5-related *N*-acetyltransferase (GNATs) conserved motifs, a family of acetyltransferases which acetylate lysine residues on histones[202,203]. To explore if Elp3 contains HAT activity, an in-gel assay was used utilizing protein overexpressed in insect cells. Elp3 was fractionated in denaturing gels containing

histones, renatured and acetylation activity tested by incubating the gel with radioactively labeled acetyl-CoA. After removing excess substrate by washing, a 60 kDa band was detected by exposing to an autoradiographic film. The authors concluded that Elp3 was able to acetylate the amino terminal tails of the four histones, or synthetic peptides derived thereof. No product was detected when the histones were replaced by BSA. While suggestive, it is noteworthy that these experiments do not demonstrate that histones were modified. The same observation may indicate strong attachment of acetyl-CoA to Elp3 occurring selectively in the presence of basic histones. Further experiments identified Elongator´s HAT domain to acetylate specifically histone H3 and H4 through a physical interaction with DNA and nucleosomes[204]. ChIP experiments also detected Elongator binding to genes whose expression was not affected by Elongator subunits depletion, showing that Elongator interacts with the coding region but not with the promoter of these genes[205]. These findings supported the idea that Elongator was involved in transcription assisting RNAPII during elongation by catalyzing histone H3 acetylation (reviewed in [206]). Isolation, purification and characterization of Elongator subunits together with RNA polymerase II in human cells additionally supported the proposal that Elongator is part of the RNAPII holoenzyme[207,208].

Elongator was also associated to exocytosis. This hypothesis emerged when deletion of Elp1 was shown to rescue the yeast phenotypes caused by *Sec2* depletion, an essential GTPase involved in exocytosis[209]. The hypothesis was supported by a two-hybrid assay in which Elp1 fragment interacted with Sec2. However, evidence suggest that the interactions detected were due to an unspecific binding exacerbated by the use of a yeast strain *sec2* allele which harbors a premature stop codon which is likely to be suppressed by a posttranscriptional mechanism linked to tRNA modification defects (reviewed in [206]).

Elongator has been associated with a wide range of metabolic processes including tubulin acetylation[210], DNA repair, transcriptional and telomeric silencing[211] (reviewed[196,206,212]). In *Caenorhabditis elegans*[186] and *Drosophila melanogaster*[213] dysfunction of Elongator subunits lead to defects in embryonic development; while in humans and mice, several neural degenerative diseases[214] and cancer[2154] have been associated with mutations and altered expression of different Elongator proteins (reviewed in [166]).

Despite the numerous and diverse roles in which the Elongator complex has been involved, many if not all of the associated phenotypes may be a consequence of altered translation. For example, embryonic and neurological disorders could result from translational defects, as neural and embryonic cells are highly sensitive to translation perturbations due to their high rate of protein synthesis. Telomeric gene silencing and DNA damage response defects observed in Elongator mutants have been suppressed by elevated levels of some tRNAs normally harboring the mcm/ncm$^5U_{34}$ modification[216]. A similar effect was also described for Elongator mutants phenotype attributed to RNA polymerase II transcription and exocytosis[217] . The observation that overexpression of hypomodified tRNA correct most if not all phenotypes associated with Elongator inactivation, together with Elongator-dependent tRNA modification, strongly argue that impaired translation is the unique primary defect resulting from inactivation of this factor.

Nowadays, there is a consensus in the scientific community that the main function of Elongator complex is the posttranscriptional modification of the U_{34} (wobble) nucleotide of tRNA[218,219], leading to the formation of 5-methoxycarbonylmethyluridine (mcm^5U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm^5s^2U) and 5-carbamoylmethyluridine (ncm^5U) groups (reviewed in[196,218]).

1.3.2 Elongator function among different domains

1.3.2.1 Elp3 in archaea and bacteria

Elongator is highly conserved among eukaryotes. However, in other domains of life, the holoElongator complex does not seem to exist. Available evidence suggests that Elongator subunit 3 (Elp3) is responsible for the catalytic activity of the complex and the formation of the cm^5U derivative on tRNA. Bioinformatic research of Elp3 protein revealed its homologs in almost all archaea, a small number of bacteria and two viruses (Figure 29)[220].

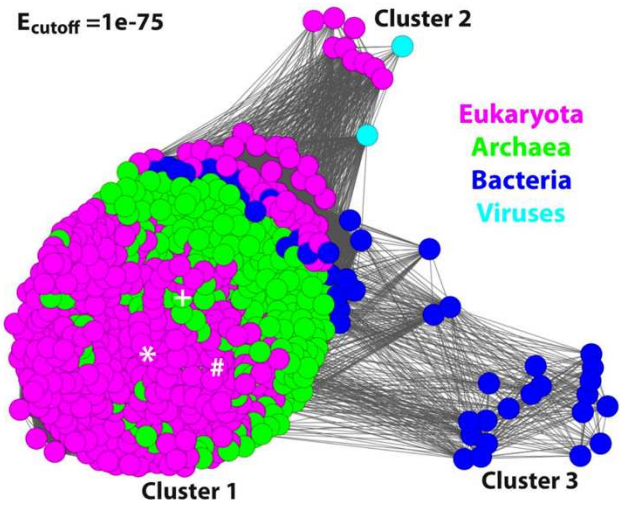


Figure 29. Visualization of Elp3 presence distributed among the different domains

Sequence similarity network of pairwise protein BLAST of Elp3 sequences using Cytoscape. Each node, colored as a circle, represents an individual Elp3. Lines represent connections between the nodes. Taken from[220].

Therefore, Elp3 is the only highly conserved Elongator subunit that can be found among the three different domains of life. Structurally, bacterial Elp3 shares 50% sequence similarities with its archaeal homolog, which harbors an extended N-terminal region of 75 amino acids present in most archaeal and eukaryotic Elp3 sequences[221]. Bacterial Elp3 also shares 41.4% identity with its *S. cerevisiae* homologue and 42.2% with the human Elp3[222]. Different from the archaeal protein which is exclusively monomeric, the bacterial DmcElp3 was shown to form monomers and dimers in solution[221,222].

Elp3 also shares sequence similarities with the dual-specificity RNA methyltransferase (RlmN) which catalyzes methylation of both rRNA and tRNA[223]. In *E. coli*, RlmN is responsible for the 2-methlyadenosine (m^2A) synthesis at position 37 of tRNA. Remarkably, archaeal and

bacterial organisms containing Elp3-like proteins do not contain the *MnmE* nor *MnmG* genes which, in bacteria and archaea, are responsible for or implicated in in the tRNA modification of c_5U_{34}[224] respectively. This fact supports the hypothesis that the main function of the Elongator complex is the post-transcriptional modification of tRNA.

1.3.2.2 Elongator among eukaryotes

As mentioned before, Elongator main function is the formation of cm^5-derivatives of uridine 34 of tRNA. Studies in yeast have shown that lack of mcm^5 or s^2 modification on U_{34} of $tRNA^{Glu(UUC)}$, $tRNA^{Gln(UUG)}$ and $tRNA^{Lys(UUU)}$, arising from deletion or inactivation of Elongator subunits or partners, causes ribosome pausing at GAA, CAA and AAA codons respectively[225], affecting mRNA decoding and tRNA–mRNA translocation during the elongation step[226]. Reciprocally, mcm^5 and s^2 modifications of U_{34} modifications were shown to enhance binding of $tRNA^{Lys(UUU)}$ to the ribosomal A-site[227].

In *S. cerevisiae* for example, Elongator mutants present a very characteristic phenotype. This includes slow growth, delayed response to stress stimuli and zymocin resistance[201]. Zymocin is a toxin excreted by killer strains of *Kluyveromyces lactis*. This compound arrests sensitive *S. cerevisiae* strains in the G1 phase blocking cell division (see section 1.3.4 Elongator cofactors for more details). Additionally, in *S. cerevisiae,* deletion or functional mutations of the Elongator subunits or its cofactors as well as enzymes involved in the tRNA U_{34} thiolation pathway (Ncs2/Ncs6, Urm1 or Uba4) are individually tolerated. However, combinations of deletions or inactivating mutations of both pathways have been reported to have lethal effects. *S. cerevisiae* strains lacking *elp3* and *ncs6* genes, resulting in the simultaneous disruption of mcm^5U_{34} and s^2U_{34} formation, have been reported to be synthetic lethal and inviable using a plasmid shuffling strategy[187]. This growth defect can be partially restored by the overexpression of $tRNA^{Gln}$ and $tRNA^{Lys}$[187]. In contrast to *S. cerevisiae*, combined loss of uridine 34 mcm^5 modification and thiolation is tolerated in *Caenorhabditis elegans*[186] and in fission yeast[228], leading to the conclusion that mcm^5s^2U is not absolutely required for life in eukaryotes. Consistently, tests conducted in our lab for synthetic lethality between Elongator subunits and enzymes involved in the tRNA U_{34} thiolation pathway by yeast crossing demonstrated that they were not synthetic lethal (C. Faux and B. Séraphin, unpublished).

Dictyostelium discoideum, an amoeba used as model organism of the evolutionary supergroup of Amoebozoa, is an interesting organism to study Elongator complex. It contains one of the most AT-rich genomes among eukaryotes. Due to its particular characteristic of low CG-content, *D. discoideum* has a strongly overrepresented A-ending codon usage compared to yeasts *S. cerevisiae* or *S. pombe*. Therefore, codons being decoded by tRNAs with $mcm^5s^2U_{34}$ modification such as Glu (GAA), Gln (CAA) and Lys (AAA), appear more often than their counterparts ending with G such as Glu (GAG), Gln (CAG) and Lys (AAG)[229]. *D. discoideum* Elp3 shares 70% identity with its *S. cerevisiae* homolog, making it the most conserved protein among the complex, followed by Elp1 and Elp2 with less than 30% identity and finally the subcomplex

Elp456 which shares less than 20% identity. Other proteins involved in the U34 modification pathways such as Trm9, Trm112, Ctu1 and Ctu2 (homologs of Ncs2 and Ncs6 respectively) have some sequence similarities between the two species. A double mutant *elp3* and *ctu1* also shows that the s^2U, cm^5U and ncm^5s^2U modifications do not follow a linear order[229].

Toxoplasma gondii, is a protozoan parasite that can cause a life-threatening opportunistic infectious disease. It contains an Elp3 homolog, which, different from other organisms, is localized in the outer mitochondrial membrane (OMM) via a tail-anchored mechanism[230]. The exact reason for this localization is unknown. However, a possibility is that specific tRNAs may need to be modified in order to be imported to the *T. gondii* mitochondrion, since this parasite completely lacks tRNA genes in its mitochondrial genome[230]. Previous reports have suggested that tRNA modifications may determine which tRNAs remain in the cytosol and which are imported into the mitochondria. In *Leishmania tarentolae* for example, tRNAs encoded in the nucleus that contain the mcm^5 modification are found in mitochondria while tRNAs contain the mcm^5s^2 moiety are primarily found in cytoplasm, indicating that thiolation (s^2) may have an inhibitory function in mitochondrial tRNA import[231].

In *Arabidopsis thaliana*, genetic analysis of leaf development using large-scale mutant screening for abnormal leaves, revealed a phenotypic class called *elongata* (*elo*). *elo* mutants are characterized by decreased cell division rate resulting in narrow leaves and reduced root growth [232]. Point mutations of the *A. thaliana* Elongator orthologs elo1(*elp4),* elo2 (*elp1)* and elo3 (*elp3)* were shown to be responsible for these delays as well as slow germination and seedling growth[233]. Disruption of Elongator in *A. thaliana* also leads to additional phenotypes including resistance to oxidative stress, altered cell progression, hypersensitivity to abscisic acid, aberrant auxin phenotypes and abnormal root development (reviewed in [234]). Additionally, the *A. thaliana* Elp2 (AtElp2) ortholog and the elo3 (*elp3)* were shown to regulate kinetics of defense gene induction and the establishment of full basal and effector-triggered immunity (ETI)[235]. Elongator is also required for the induction of the jasmonic acid (JA)/ethylene (ET) defense pathway and for resistance to the necrotrophic fungal pathogens *Alternaria brassicicola* and *Botrytis cinerea*[236].

Plants are continually exposed to pathogens; however, just a few of them have the ability to infect a particular plant species. Sometimes non-adapted pathogens attempt to colonize plant outside of their normal host range, triggering the nonhost plant resistance (reviewed in [234]). A study showed that Elongator is required for *A. thaliana* nonhost resistance against *P. syringae* pv. *phaseolicola* (*Psp*) and *Xanthomonas citri* subsp. *citri* (*Xcc*), two pathogenic bacteria which cause halo blight of common bean and citrus canker respectively[234]. Based on the pathogenic resistance properties conferred by Elongator; an interesting study attempted to improve the resistance of tomato (Solanum lycopersicum) to the disease “bacterial speck” caused by the Pseudomonas syringae pv. tomato strain J4 (Pst J4), using transgenic overexpression of *A. thaliana* genes encoding Elongator homologs AtElp3 (elo3) and AtElp4 (elo1). Results showed that both Elongator homologs overexpression improved tomato resistance. This phenomenon was also

seen in strawberry where AtElp3 and AtElp4 overexpression significantly increased resistance to several fungal and bacterial pathogens[237].

In *Drosophila melanogaster*, mutations in *elp3* homolog result in larval lethality at pupal stage[213]. Additionally, larval growth is dramatically impaired during early development, with progressions significantly delayed. Microarray analysis of late-stage Elp3 larvae mutants have shown the upregulation of genes associated with oxidative stress response[213]; while whole-genome expression profile revealed the induction of many genes involved in the activation of larval immune system[238]. Mutations in *poly*, an *elp6* homolog, affect the insulin receptor–target of rapamycin (InR–TOR) pathway and have shown a specific phenotype including an extreme larval longevity and the appearance of melanotic masses at multiple locations of larval body[239]. A recent study also showed that Elongator is required in adults' neurons for long-term behavioral memory and long-term memory (LTM) trace in *D. melanogaster*; since *elp1* and *elp3* RNAi knockdown impairs LTM and co-expression of *elp1* and *elp3* cDNA reverse the phenotype[240]

In *Caenorhabditis elegans,* Elongator acts in neurological and development processes by modulating translation; this activity is mediated by Elongator homologues *elp1*(*elpc-1*) and *elp3* (*elpc-3*) which have been described as responsible for the ncm^5U and mcm^5U formation. Elongator mutants have been associated with a loss in translation efficiency[186]. Mutants combining defects in Elongator as well as in the *tuc-1* gene*,* which is responsible for the s^2 formation, cause development defects. This phenomena could be linked to the reduction of neuropeptides production and excretion which leads to an inefficient communication among neurones[186]. Elongator was also implicated in microtubule acetylation by regulating α-tubulin function in early development stages, as well as in α-tubulin turnover and vesicle transport[241]. In another study, mutations on *elpc-1*, *elpc-2,* and *elpc-3* were found to cause shortened locomotor healthspan[242].

In mouse, inactivation of the *elp1* homolog gene (*ikbkap*) as well as elp3, results in embryonic lethality with severe growth retardation, caused by an aberrant cell cycle progression due to an inefficient translation[189,243]; while reduced expression of *ikbkap* causes peripheral neuropathies[189]. A recent study reported that mice embryonic lethality caused by homozygous deletion of the mouse *elp1 (elp1$^{-/-}$)* gene, can be rescued in a dose-dependent manner by introducing different copy numbers of the human elp1[244]. Additionally, an *elp6* point mutation, which destabilizes the complex and causes protein misfolding, was reported to cause Purkinje neuron degeneration and ataxia-like phenotype[214].

In humans, defects in Elongator subunits and Kti cofactors are related to neurological disorders and cancer, more information is presented in the section 1.3.5 Elongator in human diseases.

1.3.3 Elongator subunits

1.3.3.1 Elongator protein 1 (Elp1)

The Elongator protein 1 (Elp1), also named Insensitive to Killer toxin protein 3 (IKI3), is the largest protein of the six Elongator subunits. In yeast, it has a molecular weight of 152.2 kDa and is encoded by a gene located in the chromosome XII. Elp1 was first identified and characterized by Hishinuma team in 1997. They isolated, sequenced and characterized the *IKI3* gene. Their experimental data showed *IKI3* as a non-essential gene; however, its disruption in yeast results in a slow growth rate and a killer-insensitive phenotype[245]. Two years after, another group independently identified the same protein copurifying with the RNA polymerase II, naming it Elp1[201].

Elp1 contains two WD40 β-propeller domains at its N-terminus and one tetratrico-peptide repeat (TPR) domain at its C-terminus (Figure 30). WD40 repeats are composed of repetitions of short structural motifs (44-66 residues) that contain tryptophan-aspartic acid (W-D) dipeptide often presented in tandem forming stranded antiparallel β-strands. The latter assemble as blades which in turn form a β-propeller. Commonly WD40 domains serve as scaffolds mediating protein-protein or protein-RNA/DNA interactions. WD40 domain-containing proteins participate in a wide range of cellular processes; however, no enzymatic activity has been attributed to these domains (reviewed in [246]). TPR domains are a structural motif containing 3-16 repeats of 34 amino acids, arranged into two anti-parallel α helices separated by a turn (helix-turn-helix). This structural arrangement generates a large area for a ligand binding facilitating protein-protein interaction and assembly of multiprotein complexes (reviewed in [247,248]). The presence in Elp1 of domains that are involved in protein-protein interactions suggests that Elp1 acts as a scaffold for the other Elongator subunits, possibly also mediating interaction with other Elongator binding partners.

The Elp1 C-terminus contains a region enriched in arginine and lysine that resembles a nuclear localization signal (NLS) sequence[249], an amino acid sequence that tag proteins to be imported into the nucleus[250]. Despite this, this part of Elp1 does not function as an NLS. It has been shown that this region binds tRNA and it is essential for tRNA U_{34} modification, since mutations of the putative NLS lead to an *elp* phenotype indicative of defective Elongator dependent tRNAs modification[249].

In 2015 two X-ray crystal structures of dimers of Elp1 C-terminal domains were solved by Xu *et al.*: the yeast Elp1 (Elp1-DD-$CT_{919\text{-}1349aa}$) at 2.7 Å resolution (PDB ID 5CQS, figure 30 and the human Elp1 (hElp1-$CT_{715\text{-}1332aa}$) at 3 Å resolution (PDB ID 5CQR). The two proteins show a very similar dimeric horseshoe shaped conformation despite their low sequence identity and contain an evolutionary conserved dimerization domain which is necessary for Elongator assembly[251].

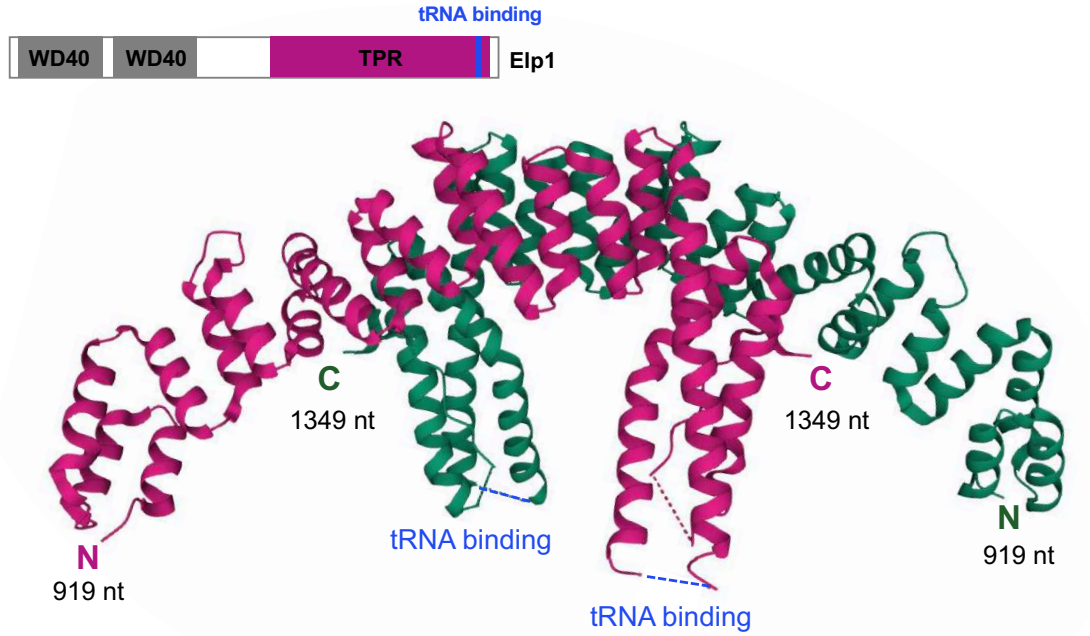


Figure 30. Crystal structure of the homodimeric C-terminal region of yeast Elongator Elp1

The ribbon diagram shows the Elp1 C-terminal domain, residues 919–1349. The two monomers are colored magenta and green with its tRNA binding site indicated in blue. N and C termini of each monomer are labeled.

The Elp1 C-terminus also contains several phosphorylation sites (Ser1198, Ser1202,Ser1205/Thr1206 and Ser1209) required for tRNA modification[252]. Several lines of evidence have demonstrated that Kti14, also called Hrr25, directly phosphorylates Elp1 (Ser1198 and Ser1202)[252]. Opposite to the Kti14 effect, Sit4 is a phosphatase which is implicated in Elp1 dephosphorylation, yeast cells lacking Sit4 or its associated proteins Sap185 and Sap190 show hyperphosphorylation of Elp1; in contrast, hypophosphorylation is seen in strains lacking Kti14[253]. Excess of Sit4-Sap190 counteracts the Kti14 phosphorylation effect and reestablishes zymocin sensitivity, suggesting competition between these proteins[253].

1.3.3.2 Elongator protein 2 (Elp2)

The Elongator protein 2 (Elp2) is the second largest subunit of the Elongator complex with a molecular weight of ~90kDa in yeast. In 2000, Svejstrup´s team, predicted that Elp2 contains eight WD40 repeats[254]. In 2015 and 2017, two crystal structures of *S. cerevisiae* Elp2 were solved independently by two teams at 3.2 Å (PDB ID 4XFV)[255] and 2.8 Å (PDB ID 5M2N)[256] resolution. The structures revealed that Elp2 is composed of two seven-bladed WD40 β propellers slightly twisted with respect to each other (Figure 31A). The N-terminal β propeller (propeller I) is formed by blades 1-7 containing amino acids 20-380, the C-terminal β propeller (propeller II) is formed by blades 8-14 containing amino acids 1-10 and 380-788 (Figure 31B).

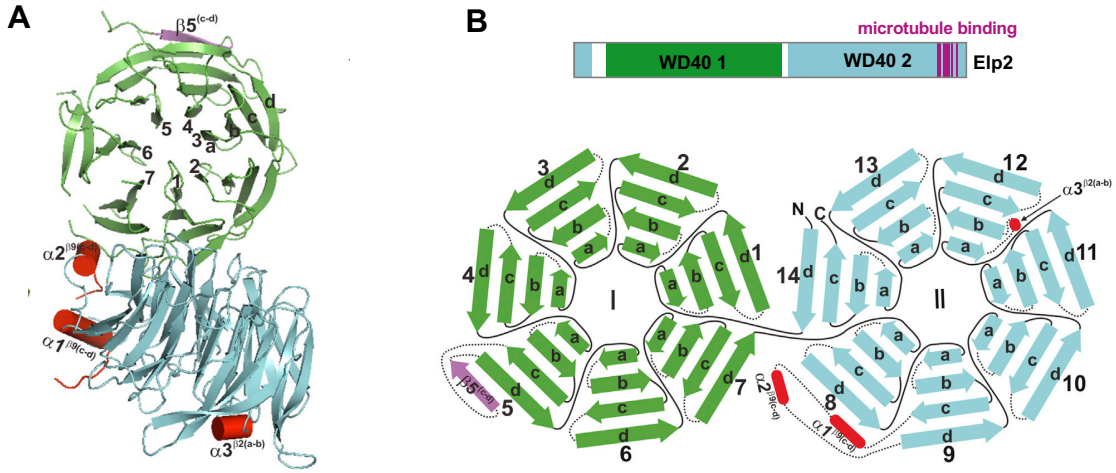


Figure 31. Crystal structure of Elongator protein 2

(A) Ribbon diagram of Elp2 view down the axis of the N-terminal β propeller I (green) and side view of the C-terminal β propeller II (cyan). **(B)** Schematic diagram of the secondary structure of Elp2 WD40 domains. Solid lines indicate interblade connection strands (d-a and b-c) of the top surface loop of each propeller while dashed lines interblade connection strands (a-b and c-d) of the bottom surface loop of each propeller. In purple is marked the extra **β**-sheet in the blade 5 WD1 and in red are present the 3 α-helix of blades 9 and 12 of the WD2. Modified from [255].

The 14 blades composing Elp2 are similar, however blade 5 of the propeller I contains an additional β strand ($\beta5^{(c-d)}$), while in propeller II there are 3 extra α-helices: two at blade 9 ($\alpha1^{\beta9(c-d)}$ and $\alpha2^{\beta9(c-d)}$) and one at blade 12 ($\alpha3^{\beta2(a-b)}$). Both β propeller domains are similar in size, their main difference is that propeller I is formed by a continuous string of residues while the propeller II blade 14 includes the 10 first N-terminal residues of the protein. This structure is commonly known as a 'molecular velcro' and has been observed in other WD domains. Supposedly it contributes to the stability of WD40 domain by keeping the β propeller tightly folded.

Mutant yeast strains lacking the first 14 residues of Elp2 located in the βd14 of the β-propeller II have shown complete loss of interaction with Elp1 and Elp3[28]. *In vivo* studies have demonstrated that mutant strains lacking Elp2 gene (*elp2*Δ) display the typical *'elp'* phenotype[20,27] that can be rescued by expression of the wild type gene but not with the Elp2-βd14Δ strain[255]. The same team also showed that Elp2 binds α-tubulin of microtubules (MTs). These data suggested that Elongator complex may recognize the cytoskeleton of the cell through the Elp2 binding to MTs via charged-charged interactions altering the integrity of the functional Elongator complex[255]. This was linked to altered α-tubulin acetylation in Elongator mutant. However, the latter phenomenon is probably controlled indirectly by Elongator tRNA modifications rather than by direct α-tubulin acetylation.

Human Elp2 is also known as Stat3-interacting protein 1 (StIP1) and was reported to interact with members of the signal transducer and activator of transcription (STAT) family, having preference for non-phosphorylated proteins as Stat1, Stat3 and Stat5, as well as with the Janus kinases (JAKs) family (reviewed in [196,258]). It has also been shown that StIP1 may act as a negative

regulator of Stat3, as its overexpression blocks Stat3 activation in human hepatocellular carcinoma cell line HepG[259] and it is involved in down-regulation of Stat3 transcript levels in the leukemia cell line K562[260].

1.3.3.3 Elongator protein 3 (Elp3)

The Elongator protein 3 (Elp3) is the third largest subunit of the Elongator complex with a molecular weight of 62.26 kDa in yeast. Elp3 is responsible for the catalytic activity of the whole Elongator and it is evolutionally the most ancient protein of the complex. Indeed, Elp3-like proteins can be found in some archaea, bacteria and viruses. At least in some bacteria and archaea, this single Elp3-homolog protein is able to catalyze the same tRNA cm^5U_{34} modification that the Elongator complex performs in eukaryotes[220]. The Elp3 protein has the same organization in all the domains of life, it contains a radical S-adenosyl-L-methionine (rSAM) binding domain and lysine acetyltransferase (KAT) domain (reviewed in [261]).

Elp3 was first identified by Svejstrup´s team in 1999 and predicted to contain a histone acetyltransferase (HAT) domain[203]. Crystal structure of full length Elp3 from the bacteria *Dehalococcoides mccartyi* (DmcElp3) was first solved in 2016 by Glatt *et al.*, at 2.15 Å resolution (PDB ID 5L7J)[222]. The structure revealed that DmcElp3 harbors an N-terminal rSAM and a C-terminal KAT domain which are connected by a zinc-binding motif. The rSAM domain of Elp3 is highly similar to the bacterial RlmN, which methylates rRNAs and tRNAs in a reaction involving the Fe-S cluster of its radical SAM center[223]. Also similar to other radical SAM enzymes, Elp3 contains a [4Fe-4S] cluster which is easily degraded to [2Fe-2S] cluster under aerobic conditions. Particularly for DmcElp3, crystallization of the Elp3 dimers has shown that Cys27 and Cys30 coordinate the [2Fe-2S] cluster[222], suggesting that Elp3 dimerization and coordination of the [Fe-S] cluster are essential for Elongator catalytic activity. The presence of a [4Fe-4S] cluster in a tRNA modification enzyme is rare but not unique; for example, this iron-sulfur cluster is present in the thermophilic bacteria and hyperthermophilic archaea enzyme tRNA thiouridine synthetases (TtuA) which mediates tRNA thiolation[262].

The Elp3 KAT domain is highly similar to the GCN5-like acetyl transferase superfamily which acetylates lysins by acetyl-CoA hydrolysis activity[263,264]; however, Elp3 lacks two helices and contains an additional β-strand (reviewed in [261]). Interestingly, the canonical KAT peptide-binding site, needed for acetylate lysine residues, is fully blocked by the radical SAM domain arguing against the hypothesis that recombinant Elp3 is able to acetylate proteins. The two domains together form a cleft containing highly conserved residues that accommodates the tRNA anticodon stem loop (ASL), allowing the wobble base to access the active site[221].

In 2019, Glatt *et al.*,[221] reported the crystal structure of the archaea *Methanocaldococcus infernus* Elp3 (MinElp3) with different versions of its extended N-terminus at 1.9 Å, 1.9Å and 2.05 Å resolution (PDB ID 61A8, 6IAZ, 6IAD) respectively. The group also obtained the structure of DmcElp3 complex with desulfo-CoA (DmcElp3-DCA), an acetyl-CoA analogue at 2.7 Å resolution (PDB ID 6IA6)[221] using an DmcElp3 blocking loop mutant ($390–406_{(GSGSG)}$) with an additional

active site mutation (E386A). Structure of DmcElp3-DCA closely matches the structure of DmcElp3, which indicates that the ligand binding does not have a strong impact on the structural arrangement of Elp3 (Figure 32).

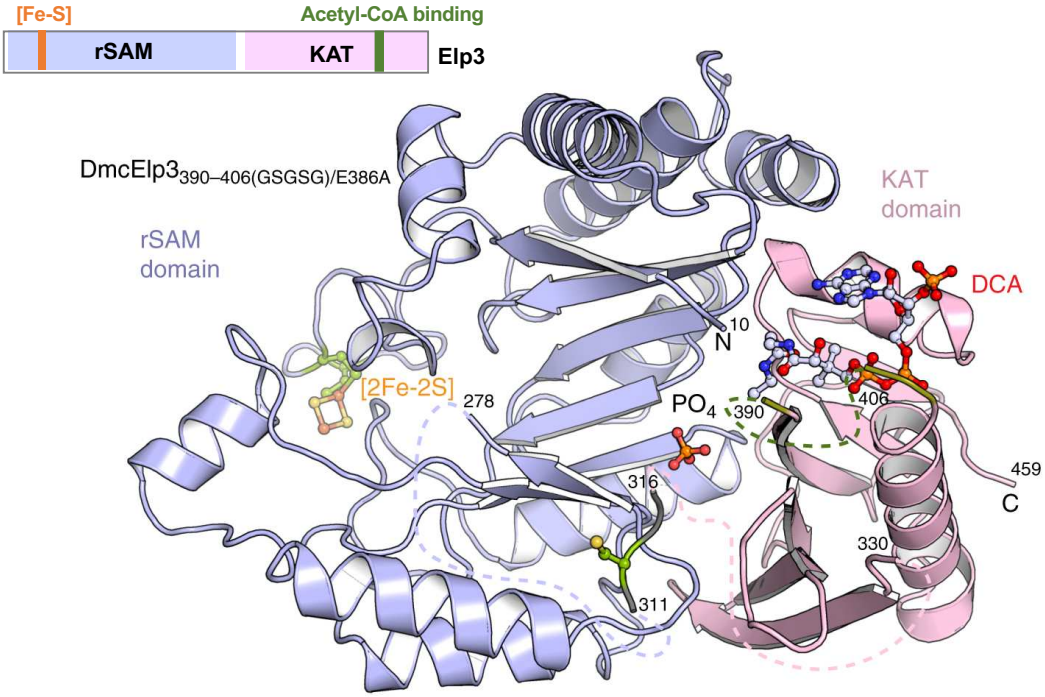


Figure 32. Crystal structure of *Dmc*Elp3

Blocking loop mutant with active site mutation ($DmcElp3_{390–406(GSGSG)/E386A}$) bound to DCA. Numbers indicate the residue position, together with its main components: rSAM domain (blue), KAT domain (pink), zinc-binding loop (gray), iron-sulfur cluster (orange and yellow), PO4 (orange and red spheres) and acetyl-CoA blocking loop (olive). Fe-S cluster binding and zinc binding residues are highlighted in green. Dotted lines indicate disordered loops[221].

The current model of addition of cm^5 modifications mediated by Elp3 proposes the following pathway[221,222]. Figure 33 shows a scheme of the Elp3 reaction:

1. Elp3 binds modifiable and non-modifiable tRNA with similar affinity
    a. The radical SAM domain detects the presence of 3'CCA to guide the ASL into the central Elp3 cleft
    b. tRNA binding shifts the acetyl-CoA blocking loop stimulating local structural rearrangements to allow acetyl-CoA binding
2. Elp3 hydrolyzes acetyl-CoA in its KAT domain
    a. Acetyl-CoA binding is secured, and hydrolysis starts only when the correct modifiable tRNA is in place (analogous to canonical acetyl transferase)
    b. The released acetyl group is transported to the rSAM domain
3. Elp3 recruits S-adenosyl-L-methionine (SAM) which is cleaved to 5'deoxyadenosine (5'-dA) radical by its radical SAM domain.
    a. The Fe-S cluster generates a 5'-dA radical
    b. An acetyl radical is produced by the contact of the acetyl group with the 5'-dA radical

4. The acetyl radical is transferred to the C5 position of the tRNA U_{34}
    a. The modified tRNA leaves the Elongator complex

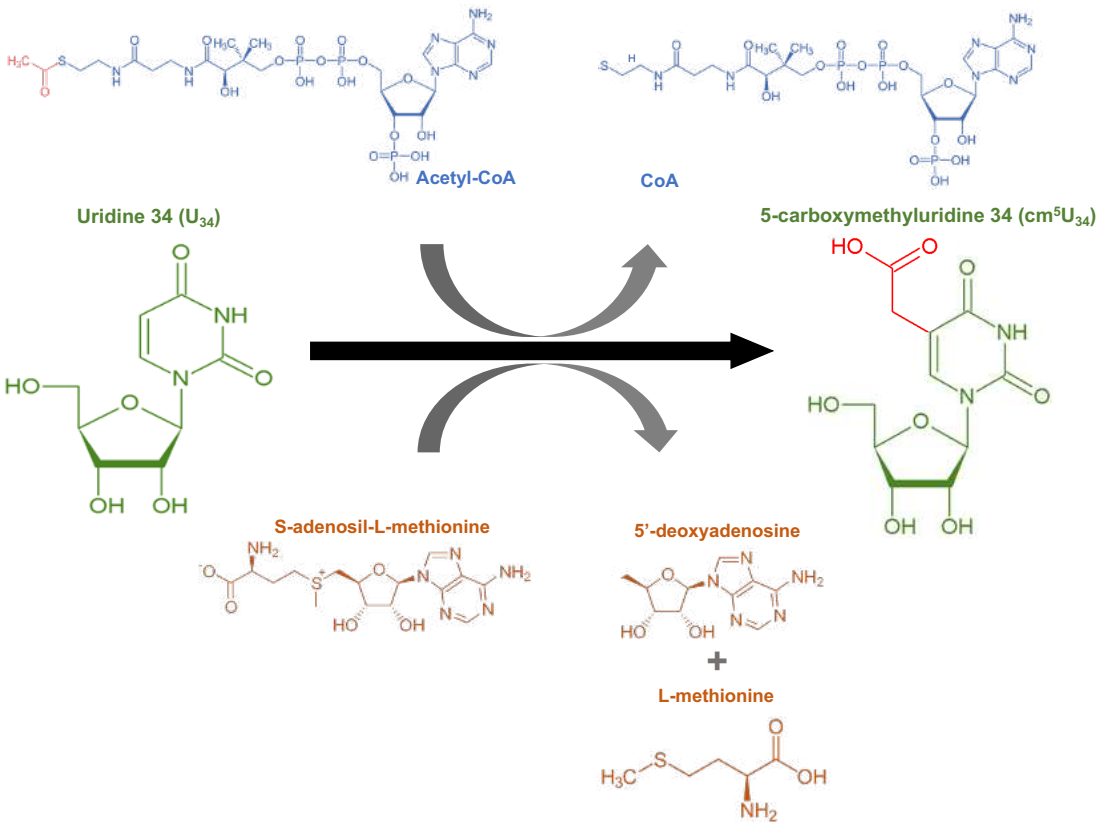


Figure 33. Scheme of Elp3 chemical reaction to modified tRNA U_{34}
Acetyl-Co-A acts as a donor for carboxymethylation of uridine (cm^5U_{34}) modification generating CoA as subproduct while SAM is used to create a 5'-dA radical generating L-methionine as subproduct.

1.3.3.4 Subcomplex Elp123

The structure of the subcomplex Elp123 free and bound to $tRNA^{Ala}$ was solved by Dauden *et al.*[265], at 3.3 and 4.4 Å using cryo-electron microscopy in 2019 (Figure 34). This publication revealed details of the Elongator active site and how the tRNA binding induces conformational changes in the subcomplex and stimulates the acetyl-CoA hydrolysis. The iron-sulfur (4Fe-4S) cluster of Elp3 appears to provide the electron for reductive cleavage of SAM, the first step of carboxymethylation of U_{34} (Figure 33), generating the 5'deoxyadenosine radical in close proximity to the active site, thus minimizing the risk of unspecific acetylation[265].

The Elp123 subcomplex dimer was shown to bind one or two molecules of tRNA[256]. tRNA binds to the Elp123 subcomplex with higher affinity than to the holoElongator, suggesting that the tRNA binds first to the Elp123 subcomplex alone. In addition to the lower affinity of tRNA to the holoElongator, a lower level of tRNA-induced acetyl-CoA hydrolysis by the full complex was also detected[265]. Fitting the model of the Elp123-tRNA complex into the negative-stain density of the fully assembled Elongator (Dauden *et al.* [256]) allowed the authors to assess the position of Elp123-bound tRNA in relation to the Elp456 subcomplex within Elongator: contacts of the Elp456 ring with two anchoring points of the tRNA in the active site of Elp3 and the CTD-loop of Elp1

suggested that the role of the Elp456 subcomplex is to release of the tRNA after the cm^5 -tRNA modification[265].

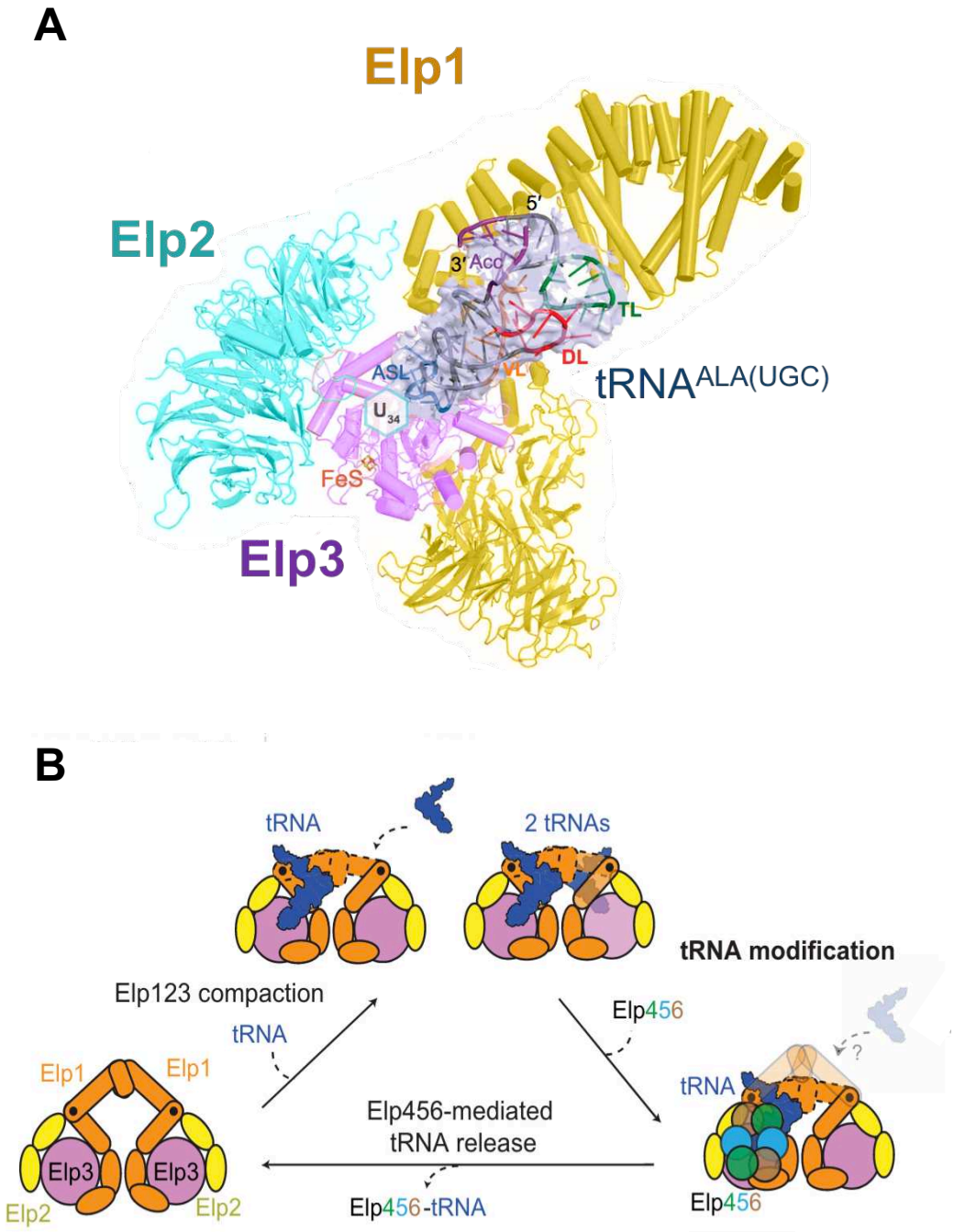


Figure 34. Structure of Elp123 complex bound to $tRNA^{Ala}$

(A) Atomic model of the Elp123 subcomplex bound to $tRNA^{Ala(UGC)}$: Elp1 (yellow), Elp2 (aqua), Elp3 (purple) and electron density of $tRNA^{Ala}$ at 4.4 Å resolution. **(B)** Illustration of Elp123 structural rearrangements upon tRNA binding and Elp456 interaction. Taken from[265].

1.3.3.5 Elongator proteins 456 (Elp456)

The Elp456 subcomplex is an heterohexameric ring formed by two copies of Elp4 (50 kDa), Elp5 (35 kDa) and Elp6 (30 kDa) with a molecular weight of ~230 kDa in yeast. In 2012 Zhijie *et al.* and Glatt *et al.*, independently solved the crystal structure of the truncated Elp456 subcomplex ($Elp4_{66-426}$, $Elp5_{1-270}$, $Elp6_{1-273}$) at 2.6 Å (PDB ID 4EJS)[266] and 2.1 Å (PDB ID 4A8J)[200] resolution respectively (Figure 35). Both structures were highly similar and revealed that the three subunits adopt identical RecA-ATPase-like folds despite the lack of obvious sequence similarity to RecA. This RecA-like domain contains a central parallel and twisted β-sheet flanked by α-helices but lacks the key sequence signature of ATPases, specifically the Walker A motif (P-loop)[200,266]; however, an intrinsic basal ATPase activity appears to be retained, and it is not nucleoside triphosphate specific[200].

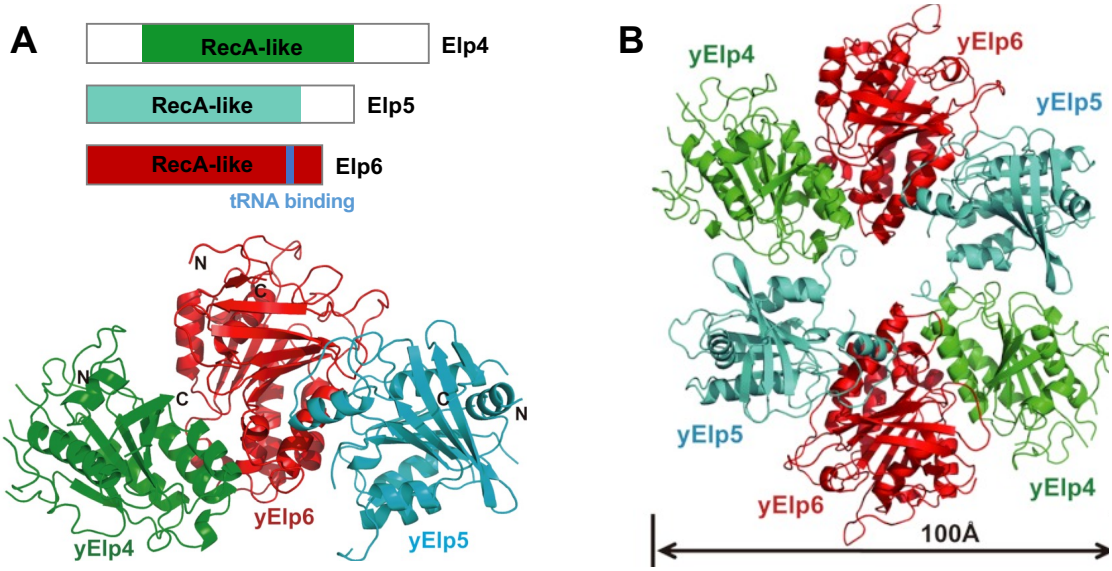


Figure 35. Crystal structure of the yeast Elp456 subcomplex
(A) Schematic and ribbon representation of Elp4 residues 67-372 (green), Elp5 residues 1-238 (cyan) and Elp6 residues 1-273 (red). The N and C termini are labeled. **(B)** Ribbon representation of the heterohexameric ring of the Elp456 subcomplex. Modified from[266].

RecA-like proteins are known to bind, rearrange and translocate nucleic acids[267]. Electrophoretic mobility shift assay (EMSA) revealed that Elp456 hexamer binds tRNA but not single-stranded DNA (ssDNA) neither oligo(U) RNA. Similarly to other Elongator subunits, tRNA binding is not dependent on whether the tRNA is modified or not by Elongator[200]. Different from other RecA-like proteins, Elp456-tRNA interaction is not mediated by positively charged binding loops. Instead, tRNA appears to bind a central L2 loop of Elp6, since mutations in this area showed to lower tRNA binding but did not affect complex stability.

Unlike other RecA-like proteins whose activity increases in the presence of a substrate, Elp456 ATPase activity was not stimulated by tRNA or anticodon loop binding. Interestingly, ATP reduces Elp456-tRNA affinity in a concentration dependent manner, suggesting that ATP binding and its hydrolysis inhibit tRNA binding or promote tRNA dissociation from Elp456[200].

Originally, the proximity of Elp456 to the Elp3 catalytic site suggested a possible role of this subcomplex in tRNA recruitment and release[256]. As mentioned in section 1.3.3.4, a recent publication[265] showed that Elp456 and the tRNA contact the same anchoring points on Elp3 and Elp1, suggesting that both molecules cannot be attached at the same time and proposing Elp456 subcomplex to be implicated in the tRNA release[265]. However more studies are required to confirm or disprove this model.

1.3.3.6 The holoElongator

In 2017, two groups independently determined the three-dimensional structure of the fully assembled yeast Elongator by using single-particle[268] and negative-stain[256] electron microscopy (EM), as well as the structure of the Elp123 subcomplex[265]. The structure of Dauden *et al.*[265] revealed that the holoElongator adopts a moth-like bilobal conformation in which two copies of Elp1, Elp2 and Elp3 subunits form a symmetric assembly which is linked together by an arch formed by the C-termini of the two Elp1 subunits. In each lobe Elp1 acts as a central platform holding Elp3 centrally and Elp2 in a peripheric position[269]. Despite the fact that previous studies showed that bacterial Elp3 homologue is capable of forming dimers in solution, there was no evidence of contact between the two Elp3 subunits in the holoElongator[269].

The hexameric Elp456 subcomplex binds to one of the Elp123 lobes forming the holoElongator complex (Figure 36).

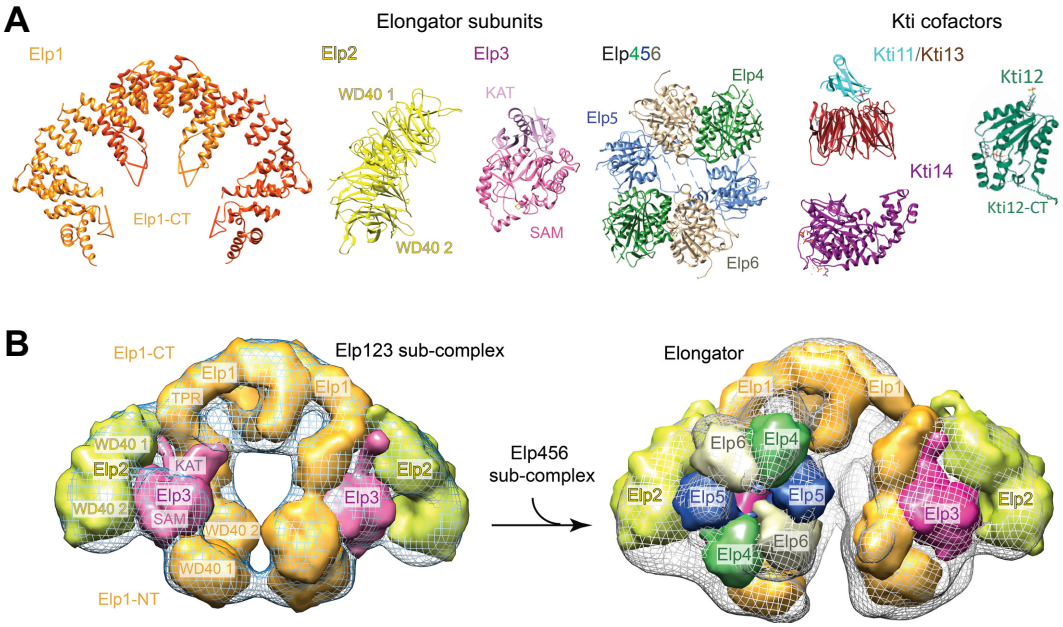


Figure 36. Overview of the eukaryotic holoElongator and cofactors

(A) Crystal structure of the Elongator subunits and the Kti cofactors family. Dimer of human Elp1 C-terminal PDB ID 5CQR (orange). *S. cerevisiae* Elp2 PDB ID 5M2N (yellow). *D. mccartyi* Elp3 PDB ID 5L7J (pink). *S. cerevisiae* Elp456 heteroexamer PDB ID 4A8J (green, blue and brown). *S. cerevisiae* Kti11/Kti13 heterodimer PDB ID 4X33 (cyan and red). *C. glabrata* Kti12 PDB ID 4XHH (purple). *C. thermophilum* Kti12 C-terminal PDB ID 6QP0 (dark green). **(B)** Topological models of the Elp123 subcomplex alone and attached to Elp456 hexameric ring. Modified from [269].

The Elp3 active site is localized in a conserved cavity formed by Elp3, Elp2 WD40 domain and the α-solenoid domain of Elp1 C-terminus. Elp3 is anchored to Elp1 by its KAT domain. The proximity of Elp3 to the Elp1 tRNA binding region (121-1259) suggests that both subunits might bind the same tRNA molecule simultaneously[256]. The data also showed that Elp123 undergoes a conformational rearrangement when Elp456 binds; this might be related to a possible activation mechanism of the complex[256]. The Elp456 hexamer is localized on the top of the cleft containing the active site of Elp3 and interacts with the Elp123 subcomplex via flexible N-terminus of Elp4 and the WD40 of Elp1.

However, some evidence has suggested that Elongator is able to bind two Elp456 hexameric rings, one on each lobe, implicating that the reported asymmetric architecture may be due to the lower levels of the Elp456 subcomplex compared to the Elp123 within the yeast cells. This hypothesis arose from *in vitro* experiments that demonstrated that purified yeast Elongator can bind two recombinant Elp456 hexameric rings, upon incubation with an excess of Elp456[268]. This results open the question on which of the two conformations, single or double Elp456 hexamer occurs *in vivo* (reviewed in [196,258]). Moreover, despite the complex asymmetry, all of the subunits have been shown to be equally important for Elongator function.

1.3.4 Elongator cofactors

In nature, limitation of resources is a factor that contributes to specific evolutionary characteristics that confer advantages against competitors. In microorganisms for example, one of these tactics is the production of toxic compounds that will kill or stop the growth of other species. *Kluyveromyces lactis* is a yeast that excretes zymocin, a toxin which inhibits growth of some other yeasts including *Saccharomyces cerevisiae* by arresting the G1 phase of the cell cycle (reviewed in [270]).

Zymocin is a heterotrimeric protein toxin complex formed by three subunits alpha, beta and gamma (αβγ) of 99, 30 and 28 kDa respectively (reviewed in [271]). The β-subunit binds chitin via its hydrophobic C-terminal domain, a polysaccharide found as main component of cell walls in fungi; while the α-subunit is an exochitinase which cleaves chitin allowing the toxin to enter the cell. Therefore, chitin binding and chitin hydrolysis by α and β-subunits is crucial for zymocin docking causing also cell wall damage, whereas toxicity hinges on the γ-subunit since its intracellular expression leads to cell lethality. Further studies have demonstrated that zymocin γ-subunit is an endonuclease that specifically recognizes $mcm^5s^2U_{34}$ tRNA modifications and cleaves the tRNA between the bases 34 and 35, leading to the tRNA degradation (Figure 37)[271,272].

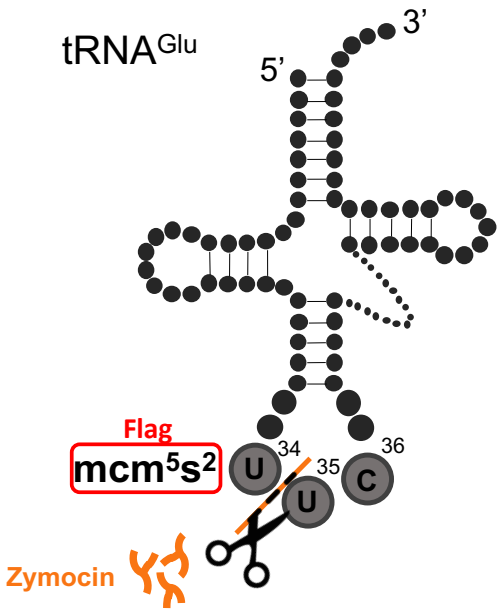


Figure 37. $mcm^5s^2U_{34}$ mediates zymocin cleavage of tRNAs
Zymocin endonuclease γ-subunit cleaves tRNA containing $mcm^5s^2U_{34}$ ($tRNA^{Glu}$, $tRNA^{Gln}$ and $tRNA^{Lys}$) between the positions 34 and 35 of the anticodon. Modified from [271].

A genetic screen from a collection of 4826 homozygous diploid *S. cerevisiae* strains each containing the deletion of a non-essential gene, identified 63 mutants which were resistant to zymocin[273]. These mutants have received different names as they were discovered independently in different labs.

The most common terms are:

- *skt* (sensitive to *K. lactis* toxin)[274]
- *iki* (insensitive to killer) [275]
- *tot* (target of γ-toxin)
- *kti* (*K. lactis* toxin insensitive or killer toxin insensitive) [276]

Zymocin-resistant mutant genes can be divided into two main classes.

- Class I: mutants resistant to extracellular zymocin but sensitive to intracellular expression of the γ-toxin, therefore, defective in the zymocin docking or uptake.
- Class II: mutants resistant to extracellular and intracellular expression of γ-toxin, therefore, suggested as probable zymocin targets

Table 11 displays some of the genes related to the zymocin-resistance phenotype. Among class II genes we can find several proteins required for the tRNA U_{34} modification[276,277]. Besides the six Elongator subunits we can find there four KTI proteins (Kti11, Kti12, Kti13 and Kti14) which act as Elongator cofactors and are indispensable for its activity (reviewed in [196]). Therefore, to decipher the complete details of Elongator function it is important to understand their contribution.

Table 11. *S. cerevisiae* genes whose deletion causes zymocin resistance. Modified from[271,278]

Name/aliases	Function	Zymocin-relation
Mutants class I		
CHS3/KTI2	Chitin synthase III	Docking
CHS4/SKT5	Chitin synthase III activator	Docking
CHS5	Chs6 interactor	Docking
CHS6	Chitin synthase III regulator	Docking
CHS7	Chitin synthase III regulator	Docking
ISR1	Putative kinase	Docking
UGP1	UGPase	Docking
GRX3	Glutaredoxin	Post-docking?
KTI6/IPT1	Sphingolipid $M(IP)_2C$ synthesis	γ-toxin import?
KTI10/PMA1	H^+-ATPase	γ-toxin activation?
PTK2	Pma1 activator	γ-toxin activation?
Mutants class II		
Elp1/TOT1/KTI7	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Elp2/TOT2/KTI3	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Elp3/TOT3/KTI8	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Elp4/TOT7/KTI9	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Elp5/TOT5	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Elp6/TOT6/KTI14	Elongator subunit	γ-toxin target / cm^5U_{34} formation
Kti11	Elongator cofactor	γ-toxin target / cm^5U_{34} formation
Kti12/TOT4	Elongator cofactor	γ-toxin target / cm^5U_{34} formation
Kti13/ATS1	Elongator cofactor	γ-toxin target / cm^5U_{34} formation
Kti14/HRR25	Elongator cofactor/casein kinase 1	γ-toxin target / Elp1 phosphorylation
SIT4	Phosphatase	γ-toxin target / Elp1 phosphorbalance
SAP185,SAP190	Sit4 subunits	γ-toxin target / Elp1 phosphobalance
SAP155	Sit4 associated protein	γ-toxin target / Elp1 phosphobalance
URM1	Urmylation (ubiquitin-like)	U_{34} thiolation
UBA4/NCS3	Urm1 activator	U_{34} thiolation
NCS2	Cla4 interactor	U_{34} thiolation
TRM9/Kti1	tRNA methylase	U_{34} methylation
$tRNA^{Glu,Gln,Lys}$	tRNAs	γ-toxin target

1.3.4.1 Kti11/Kti13 heterodimer

Kti11, also known as Dph3, is the smallest protein of the KTI cofactors family with a molecular weight of 9.33 kDa in *S. cerevisiae*. Kti11 is highly conserved among eukaryotes. Kti11 belongs to the CSL zinc-finger family; and it has been demonstrated that it copurifies with the Elongator subcomplex Elp123[279] and co-precipitates with the Elp2 and Elp4 subunits[280]. Kti11 appears to be involved in the modification of tRNA uridine 34, since Kti11 deletion leads to the characteristic yeast *elp* phenotype including zymocin resistance.

Kti11/Dph3 is also implicated in the biosynthesis of diphthamide, a post-translationally modified histidine found in the translation elongation factor 2 (EF-2) of eukaryotic and archaeal cells. Kti11 was shown to interact with the diphthamide synthesis factors 1 and 2 (Dph1 and Dph2). Similar to mcm^5s^2U modified tRNA and zymocin, diphthamide poses as a target of diphtheria toxin which inactivates EF-2 through ADP-ribosylation of the diphthamide residue leading to cell death. The diphthamide pathway requires seven proteins (Dph1-Dph7) and depends on the iron-sulfur cluster of Dph1 and Dph2 (reviewed in [281,282]).

Kti11 also interacts with Kti13, another of the four KTI Elongator cofactors. Kti13, also known as α-tubulin suppressor 1 (Ats1), has a molecular weight of 36.46 kDa in yeast and was identified as a class 2 suppressor of α-tubulin mutation. Sequence similarities show that Kti13 contains a RCC1 domain, characteristic for Ran guanine exchange factor-like (GEF-like) proteins. Deletion of Kti13 confers zymocin resistance, but unlike deletions of Kti11 and Elongator subunits, reduces but does not completely abolish wobble uridine modification[219]. Kti13 was also implicated in the diphthamide synthesis pathway alongside Kti11[283].

Kti11 and Kti13 were shown to co-immunoprecipitate together in yeast forming a stable heterodimer even in the absence of the Elongator complex[284,285]. In 2005, the structure of Kti11 was solved by Sun *et al.*[286] using NMR (PDB ID 1YOP), and in 2015 two crystal structures of the Kti11/Kti13 heterodimer were published independently by our laboratory[285] (1.45 Å resolution PDB 4X33) (Figure 38A), and by Glatt´s team[283] (2.9 Å resolution PDB 4D4O). The structure of Kti13 alone was also solved by Glatt *et al.*[283], at 2.4 Å resolution (PDB 4D4Q).

Results showed that Kti11 folds into a closed globular structure comprising 3.10 helix, two turns, two α-helices (H1 and H2), and two β-sheets and coordinates a metal ion, often interpreted as zinc. Interestingly, despite strong structural similarity to a C_4 Zn-finger, and initial attribution as such, Kti11 “Zn-binding domain” only binds iron[285]. The iron is coordinated by four conserved cysteines (C25, C27, C47 and C50) with a distorted tetrahedral geometry (Figure 38B). Remarkably, iron binding occurs preferentially when Kti11 is in heterodimer with Kti13. Consistent with its iron binding ability, Kti11 is a redox-active protein[285], which acts as carrier in electron transfer to the [4Fe-4S] cluster of Dph2[287] initiating the diphthamide synthesis pathway by an unusual SAM radical reaction.

Kti13 contains a WD40 domain (Figure 38C), which adopts a seven-bladed β-propeller fold characteristic of the RCC1-like domains (RLDs). Blades 1, 2, 3, 6 and 7 contain four β-strands

while blades 4 and 5 contain only three β-strands complemented with a short helix and an extended loop respectively. A β-strand end stabilizes blade 7, forming a molecular Velcro between the C-terminal strand (330-334aa) and the N-terminal strand (5-10aa) which complete the last blade, in a similar way to Elp2[283,285]. Kti13 is present as monomer in solution [283].

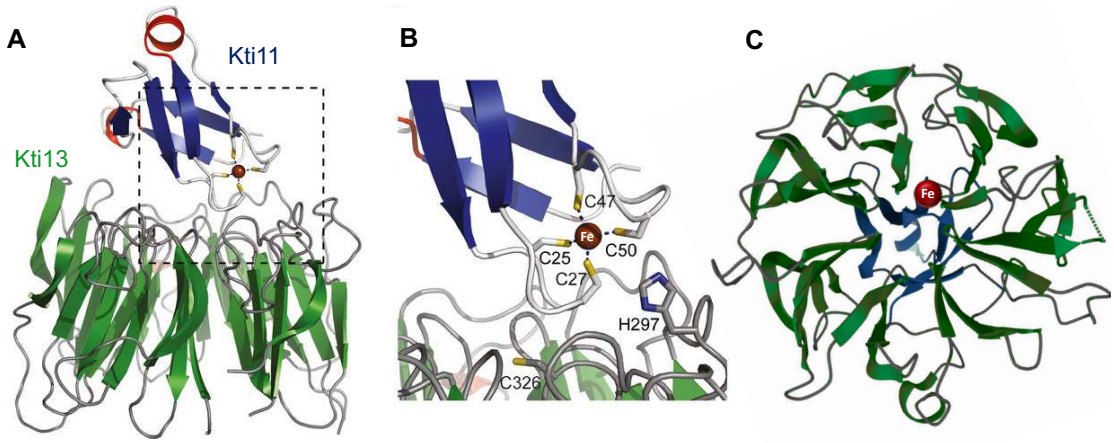


Figure 38. Crystal structure of the yeast Kti11-Kti13 Elongator cofactor heterodimer.

S. cerevisiae complex of Kti11$_{(2-57)}$ (blue) and Kti13 (green). The redox-active iron atom is shown as a red sphere. **(A)** Side and **(C)** bottom views of the heterodimer PDB ID 4X33. **(B)** Zoom of dashed area indicated in 'A'. Kti11 cysteine residues C25, C27, C49 and C50 coordinating iron as well as Kti13 C326 tunnel-closing are highlighted in yellow. Kti13 histidine residue H297 pointing towards iron atom is highlighted in blue. Modified from[285].

Formation of Kti11-Kti13 heterodimer is relevant for Elongator mediated tRNA modifications as well as for diphthamide synthesis. Structural analysis has suggested that Kti13 may act as chaperone of Kti11, as it orients Kti11 and protects the exposed iron atom preventing electron transfer to random redox active molecules (reviewed in [196]). This idea is supported by electron transfer assays which demonstrate that Kti13 modulates Kti11 electron transfer abilities. More specifically, Kti13 contains a highly conserved histidine (H297) which faces the Kti11 iron atom which is believed to influence the redox potential of Kti11[285]. Due to similarities between Elp3 and Dph2 it is likely that Kti11 acts as electron donor to the [Fe-S] cluster of Elp3 as well[283] (Figure 39).

In *S. cerevisiae* the cytochrome b_5 reductase (Cbr1) was identified as a NADH-dependent reductase for Kti11 in the electron transfer to radical SAM enzymes[195]. However, since the deletion of *cbr1* did not abolish the diphthamide synthesis, other reductases such as Mcr1 and Ncp1 arose as putative factors responsible for Kti11 redox reaction (reviewed in [196]).

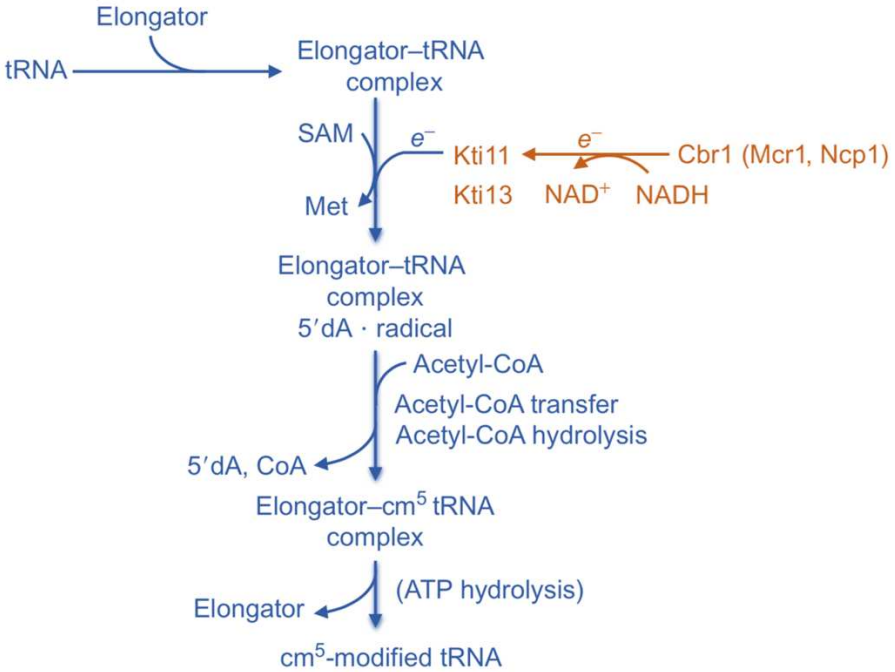


Figure 39. Proposed pathway for cm^{5}U tRNA Elongator dependent modification and the Kti11-Kti13 heterodimer interaction

Different steps of tRNA modification are presented in blue. Electron transfer mediated by Kti11-Kti13 and Cbr1 is shown in orange. Note that the pathway remains hypothetical. Modified from[196].

1.3.4.2 Killer Toxin Insensitive 12 (Kti12)

The *S. cerevisiae* Kti12 is a protein with a molecular weight of 35.32 kDa. Similar to other proteins of the KTI family, Kti12 mediates zymocin resistance, and its deletion in yeast confers the 'elp' phenotype. Different studies also revealed that Kti12 and Kti14 were required for phosphorylation of Elp1, since mutant yeast strains lacking Kti12 or Kti14 (ΔKti12, ΔKti14) showed hypo-phosphorylation of Elp1 while mutant strains overproducing Kti12 showed hyper-phosphorylation[253]. Remarkably, deletion of Kti12 in yeast does not perturb Elongator integrity, suggesting that phosphorylation of Elp1 impacts Elongator function but not its architecture or stability. Kti12 also copurifies with Kti14 using TAP-purification[288].

The contribution of Kti12 to the Elongator function was unknown at the beginning of this project. Two independent publications reported that Kti12 has sequence similarities to the *O*-phosphoseryl-tRNA^{Sec} kinase (PSTK), an enzyme involved in the biosynthesis of the selenocysteine amino acid. First (in 2008), Sherrer *et al.*[289] observed a homology between the kinase domain of Kti12 and PSTK proteins of archaea and eukaryotes (Figure 40).

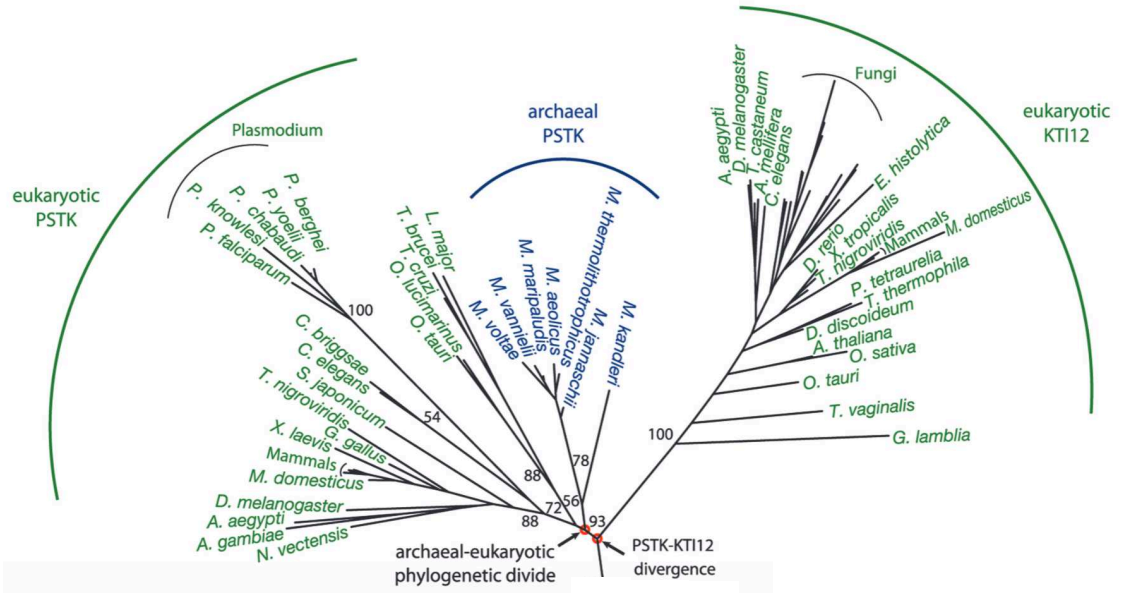


Figure 40. Phylogenetic tree of the kinase domains of PSTK and Kti12
Name of organisms are colored as follows: Archaea in blue and Eukarya in green. Taken from [289].

A second study discussing similarities between Kti12 and PSTK was published in 2017. In this article, Mehlgarten *et al.*[290] described a sequence alignment using the *Methanocaldococcus jannaschii* PSTK and fourteen Kti12 sequences from different organisms. The alignment identified three conserved elements between these proteins: the first was a P-loop near to the N-terminus of the proteins; the second and third elements were two putative calmodulin binding regions localized in the center and close to the C-terminus, respectively. Since then, the biological significance of the latter has not been confirmed. In the same study, the authors also used the PSTK crystal structure to build a structural model of the yeast *Saccharomyces cerevisiae* Kti12 protein and its *Arabidopsis thaliana* ortholog named Elo4. These models served as a basis for analysis of structural and functional conservation between elements shared by PSTK and Kti12. Overall, these early studies pointed out two important facts: (i) Kti12 and its homologs are only found in eukaryotes and (ii) Kti12 is related to eukaryotic and archaeal PSTK.

During the development of my project Krutyholowa *et al.*[291] solved the structure of the *Chaetomium thermophilum* Kti12 N-terminal domain (*Ct*Kti12$_{NTD}$) residues 2-252 in the presence of magnesium, ADP and aluminum fluoride (ADP-AlF_3) at 2.4 Å resolution. The structure showed that Kti12 contains a globular ATPase domain which consists of eight α-helices and five parallel β-sheets surrounding the hydrophobic core of the domain[291]. The crystal structure demonstrated that similarities between Kti12 and *Methanocaldococcus* jannaschii PSTK (*Mj*PSTK) extend from sequence to the 3D structure (Figure 41). The structure also confirmed existence of the previously described P-loop motif located between the first strand (β1) and the first helix (α1)[292].

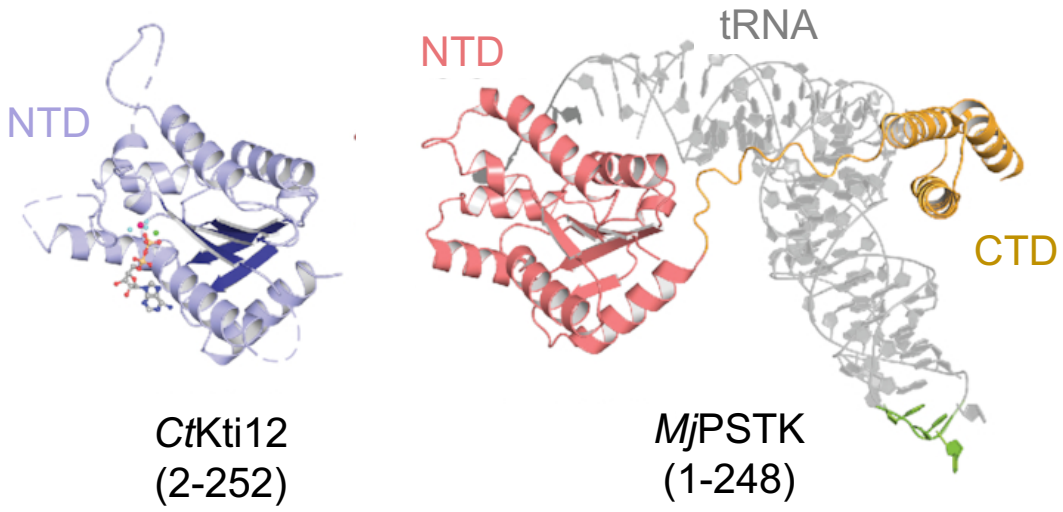


Figure 41. Kti12 has structural similarity to PSTK
Structural similarity between the N-terminal domains of *C. thermophilum* Kti12 (blue) and *Methanocaldococcus jannaschii* PSTK (red)

In this publication, the authors characterized Kti12 biochemically and functionally[291]. An in depth comparison between my results and the results published by Krutyholowa *et al.*[291] can be found in the discussion section.

In 2020 Krutyholowa et al.[293], also analyzed the linker of different Kti12 proteins. They found a subset of proteins with uncommon long linkers characterized by repeated substitutions of lysine (K) for leucine (L) in the Walker A motif which was thought to be invariable. The authors showed that the K14L mutation reduces the affinity of the protein to ATP, but not its ATPase catalytic activity[293].

Another recent publication comparing Kti12 and PSTK in humans, showed that Kti12 interacts with the protein Mindy-3[294] but more assays are required to establish a stronger functional connection between Kti12 and Mindy-3. Mindy-3 belong to a family of deubiquitinases (DUBs) which can remove ubiquitin from modified proteins.

1.3.4.2.1 O-phosphoseryl-tRNA^{Sec} kinase (PSTK)

The O-phosphoseryl tRNA^{Sec} kinase (PSTK) is an enzyme involved in the selenocysteine (Sec) synthesis in archaea and eukarya. Selenocysteine (Sec) is commonly known as the 21st amino acid, which is incorporated into proteins. It is widely present among the three domains of life (but not in *S. cerevisiae*) and its biosynthesis and incorporation into proteins differ from the canonical mechanisms. Sec is the only amino acid containing selenium, an essential micronutrient, and is part of the so called selenoproteins. Structurally selenocysteine is similar to cysteine (Cys) with selenium instead of sulfur; since the selenol (-SeH) group of Sec is more nucleophilic than the Cys thiol (-SH) group, its main function is to act as active oxidation-reduction center in a specific group of enzymes (reviewed in [295,296]).

Among the 22 amino acids inserted in proteins during translation, selenocysteine is the only one that constantly requires a specific multistep tRNA-dependent synthesis, ribosome delivery, and insertion. Sec is encoded by the UGA codon, which is a canonical a stop codon in most organisms. However, for Sec incorporation some UGA codons are read as coding codons (Figure 42). Selenocysteine is synthesized after aminoacylation of the cognate $tRNA^{Sec}$, by replacement of the hydroxy (OH) group of Serine-$tRNA^{Sec}$ by a selenol group (SeH). In mammals, a stem-loop structure named SECIS, located in the 3´-UTR of the mRNA, is recognized by SBP2 protein enabling the recruitment of Sec-$tRNA^{Sec}$ to the ribosome and decoding of the stop codon to incorporate Sec in the newly synthetized polypeptide. During Sec synthesis, discrimination between $tRNA^{Ser}$ and $tRNA^{Sec}$ plays a key role since the amount of the latter is much lower in the cell (reviewed in [294,295,297]). PSTK appears to be the major discriminator between Ser-$tRNA^{Ser}$ and Ser-$tRNA^{Sec}$ as it exclusively phosphorylates the latter[297].

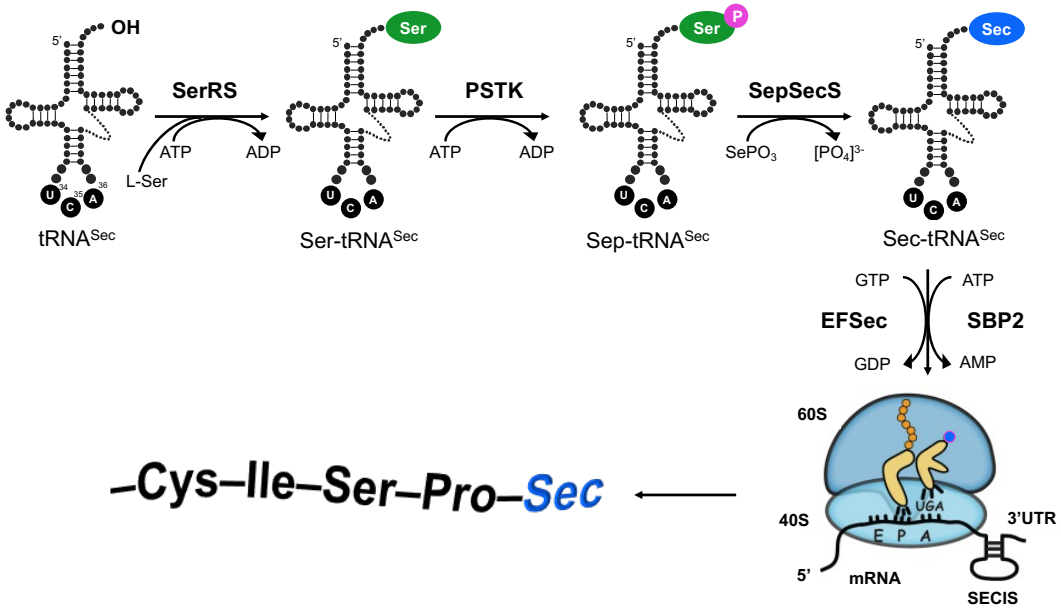


Figure 42. Scheme of selenocysteine synthesis in mammals

First, the seryl-tRNA synthetase (SerRS) attaches L-serine (L-Ser) to a non-cognate $tRNA^{Sec}$. Then the O-phosphoseryl-$tRNA^{Sec}$ kinase (PSTK) phosphorylates the seryl group of Ser to a phosphoseryl (Sep)-$tRNA^{Sec}$ intermediate. Next, the O-phosphoseryl-$tRNA^{Sec}$:selenocysteinyl-$tRNA^{Sec}$ synthase (SepSecS) catalyzes the conversion of Sep-$tRNA^{Sec}$ into selenocysteinyl (Sec)-$tRNA^{Sec}$ using selenophosphate ($SePO_3$). Finally, Sec-$tRNA^{Sec}$ is delivered to the ribosome by a specialized elongation factor (EFSec) and the auxiliary protein factor SECIS-binding protein 2 (SBP2). The Sec Insertion Sequence (SECIS), an in-cis element in the selenoproteins mRNA located in the 3′-UTR, forms a stem loop structure and is required for decoding of the selenocysteine UGA codon. Modified from [295].

Based on the solved structures of archaeal PSTKs, the protein is formed of two domains: an N-terminal domain with kinase activity and a C-terminal domain which binds tRNA. Both domains are joined by a flexible linker[297] (Figure 43).

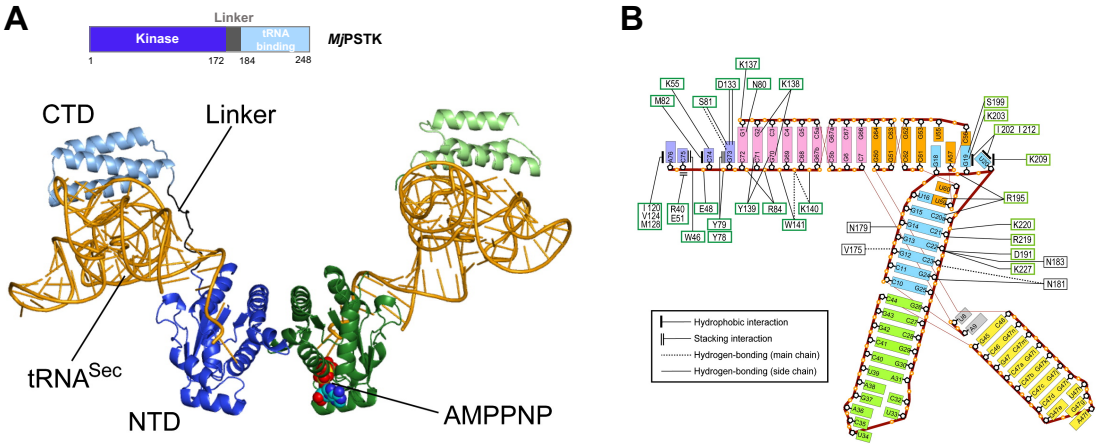


Figure 43. Crystal structure of PSTK-tRNA^{Sec} complex

(A) Crystal structure of complex of *M. jannaschii* O-phosphoseryl tRNA^{Sec} kinase (PSTK) dimer with with two molecules of *M. kandleri* tRNA^{Sec} and AMPPNP (non-hydrolysable ATP analog) (PDB ID 3ADB). **(B)** PSTK-tRNA^{Sec} interactions. Different tRNA parts are colored as follows: acceptor stem (pink), AD linker (grey), D-arm (light blue), anticodon arm (green), variable loop (yellow), TΨC-arm (orange), discriminator base and CCA-end (purple). Taken from [297].

The *M. jannaschii* PSTK (*Mj*PSTK) is known to form homodimers[298]; and similar to Kti12, it is localized in the cytoplasm in yeast and humans[294]. Interestingly, PSTK does not interact with any of the Elongator subunits and neither has any strong interaction partners in common[294].

Based on this information and on multiple sequence alignments, we made the hypothesis that Kti12 contains related domains and therefore similar activities: an N-terminal domain (NTD) with kinase signature and a C-terminal domain (CTD) with RNA binding capacity.

1.3.4.3 Killer Toxin Insensitive 14 (Kti14)

Kti14 also known as Hrr25 is a yeast protein encoded by the gene *hrr25* and has a molecular weight of 57.35 kDa. The Kti14 human homolog is the casein kinase 1δ (CK1δ), a well-studied protein belonging to the CK1-family of kinases. *S. cerevisiae* contains four members of the CK1-family, Yck1, Yck2, Yck3 and Kti14, the first three proteins (Yck1, Yck2, Yck3) are prenylated and form part of the yeast membrane[299] while Kti14 is the only soluble member. Kti14 has been implicated in a large number of functionally diverse processes including ribosome biogenesis, DNA repair, autophagy, vesicle trafficking and microtubule assembly among others (reviewed in [300]). Its participation in such diverse pathways makes Kti14 an essential gene as its deletion in yeast is lethal. *S. cerevisiae* Kti14 contains three domains, a kinase N-terminal domain (residues 1-290), a central domain which can only be found in *S. cerevisiae* and other closely related yeasts, and a disordered C-terminal domain rich in proline and glutamine (P/Q-rich) which is extensively and heterogeneously auto-phosphorylated[300,301]. Phosphorylation sites are quite commonly enriched in disordered regions[302]. Kti14 disordered domain has been shown to cause aggregation of the protein; therefore, only *S. cerevisiae* Kti14 construct without C-terminal domain can be expressed and purified from *E. coli*[301].

Besides many roles that Kti14 plays in diverse pathways, Kti14 is also a cofactor of Elongator and is involved in phosphorylation of Elp1. Abdel-Fattah *et al.*[252] identified 9 phosphorylation sites on Elp1 (Figure 44). Phosphorylated sites were mutated to a non-phosphorylable alanine. Results showed that from the 9 phosphorylation site yeast mutants, the S1209A mutation was the only that conferred zymocin resistance and therefore, impaired Elongator function. Combinations of phosphorylation site mutations were also tested; among those, S1198A together with S1202A also conferred zymocin resistance, showing an apparent redundancy in the function of phosphorylation sites.

In principle, as Elp1 is the central platform of the whole complex, its phosphorylation state could influence Elongator assembly. However, results of co-immunoprecipitation from mutant strains (including S1209A) revealed that phosphorylation (or lack thereof) of Elp1 did not prevent its interaction with other subunits and therefore Elongator assembly was not impacted; but interaction between Elp1 subunit and the Elongator cofactors Kti12 and Kti14 was affected[252]. As mentioned in section 1.3.3.1 Kti14 was also shown to coprecipitate with Kti12 and Elp1, revealing a physical interaction between Kti12 and Kti14. This co-precipitation requires the presence of all of the Elongator subunits [288].

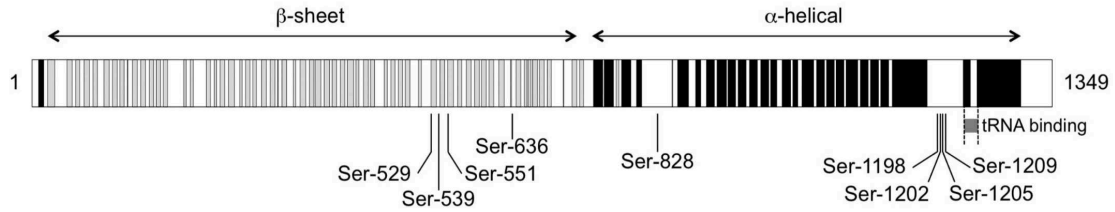


Figure 44. Elp1 phosphorylation sites

Mapped phosphorylation sites are labeled, as well as structural motifs of Elp1: α-helices (black) and β-sheets (gray).

Crystal structures of Kti14 from *Saccharomyces cerevisiae* and *Candida glabrata* were solved in 2016 by Ye *et al.*, at 3.01 Å (PDB ID 4XHL) and at 1.99 Å (PDB ID 4XH0) resolution respectively[300]. Protein constructs without the disordered P/Q-rich C-terminal domain were crystallized in the presence of the CK1 inhibitor CK1-7 which constrains the conformational freedom of the N-terminal domain. The *S. cerevisiae* Kti14 structure (Figure 45) also contained a substitution of lysine 38 by arginine (K38R) which turns off the kinase activity of Kti14 and prevents its autophosphorylation facilitating the formation of well-ordered crystal formation. Kinase domain of Kti14 is almost identical to other CK1 enzymes[300]; it is composed of two lobes with a cleft in between to which ATP binds. However, Kti14 also contains a well-ordered central domain that the other CK1 enzymes do not possess[303]. The central domain folds into a cluster of one extended loop and five α-helixes[300].

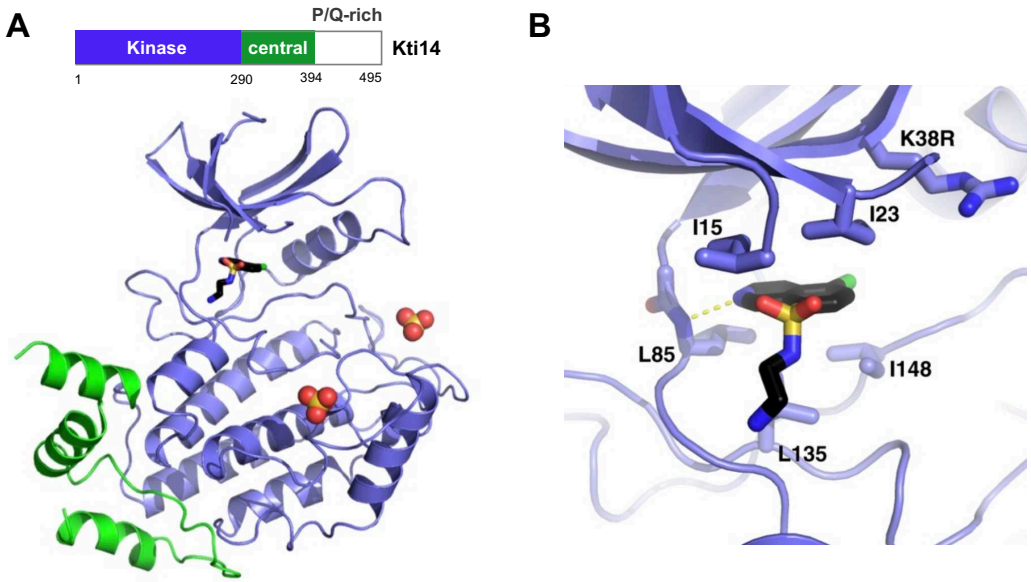


Figure 45. Crystal structure of the S. cerevisiae Kti14$_{1-394}$ bound to CK1-7
(A) The kinase N-terminal domain is depicted in blue and the Central domain - in green. **(B)** Enlarged view of CK1-7 binding to the to Kti14, conserved residues are labeled. Modified from[300].

The serine/threonine-protein phosphatase Sit4, was shown to be antagonist to Kti14 activity, Deletion of *sit4* gene is associated to zymocin resistance phenotype[273]. Yeast strains lacking *sit4* leads to an hyperphosphorylation of Elp1, double-mutant yeast deleting *sit4* and inactivating the catalytic activity of *hrr35*, reversed zymocin resistance and restored the wildtype-like phosphorylation balance of Elp1[288].

1.3.5 Elongator in human diseases

As we previously discussed, in most organisms and organelles, tRNAs are heavily modified after transcription. These chemical modifications are crucial for tRNA structure, stability and function. Consequently, alterations of post-transcriptional modification of tRNA and their impact on translational control play a major role in human diseases[304,305]. Even though it could be expected that the lack of modifications affects protein synthesis in cells globally, evidence has shown that certain tRNA modifications preferentially affect the translation of a specific subset of transcripts enriched in certain codons. Hence, a variety of human diseases were linked to mutations in genes involved in post-transcriptional modifications of cytoplasmic and mitochondrial tRNAs. Interestingly, many of the modifications implicated in human diseases are localized in the tRNA anticodon loop, and more precisely at the wobble position; this is not surprising as such modifications are particularly important because they modulate accuracy, efficiency and fidelity of translation[219]. Dysregulations of the wobble uridine modifications in cytosolic and mitochondrial tRNAs are linked to respiratory malfunctions, cognitive disorders and cancer (Table 12). For example, the Alkylated DNA repair protein alkB homolog 8 (ALKBH8) is the human homolog of the yeast methylase Trm9 (see Table 13). It participates in the formation of the mcm^{5}s^{2}U, (S)-mchm^{5}U, (R)-mchm^{5}U, mcm^{5}U, and mcm^{5}Um at the wobble position of specific tRNAs (see Table

X). Homozygous truncated mutants of ALKBH8 have been identified to cause of intellectual disability. Analysis of tRNAs from patients carrying the mutations showed a complete absence of the previously described U modifications[306].

Table 12. Wobble tRNA modifications linked to human diseases

Category	Disease	*Gene*	Modification
Neurological	Familial dysautonomia[307308]	*IKBKAP*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Ataxias and Purkinje neuron degeneration[214]	*Elp6*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Amyotrophic lateral sclerosis[309,310]	*ELP3, ELP4*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Intellectual disability[311–313]	*ELP2*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Autism spectrum disorder[313]	*ELP2, ELP4*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Rolandic epilepsy[314]	*ELP4*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	MERF (myoclonus epilepsy with ragged-red fibers)	*Mt-tRNA$^{Lys(UUU)}$), MTO1, GTPBP3, MTU1*	tm^{5}s^{2}U$_{34}$ (mitochondria)
Cancer	Invasive breast cancer[315]	*ELP3, CTU1/2*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Melanoma[316–318]	*ELP1,3,5,6 CTU1/2*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Sonic Hedgehog Medulloblastoma[215]	*IKBKAP*	mcm^{5}s^{2}U$_{34}$, ncm^{5}U$_{34}$, ncm^{5}Um
	Hepatocellular carcinoma[319,320]	*ELP3, ELP4*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
	Bladder cancer and Intellectual dysability[306,321]	*ALKBH8*	mcm^{5}s^{2}U
Respiratory	Bronchial asthma[322]	*IKBKAP*	mcm^{5}s^{2}U, ncm^{5}U, ncm^{5}Um
Blood	MLASA (Mitochondrial Myopathy and Sideroblastic Anemia)	*PUS1*	$\Psi_{multiple}$ (mitochondria)

Table 13. Human orthologs of the *S. cerevisiae* genes involved in the U_{34} modification pathways

Moieties in which individual genes are involved are underlined. Modified from [166].

Yeast Gene	Human Orthologs	Modifications
Elp1	*Elp1/IKBKAP*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Elp2	*Elp2*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Elp3	*Elp3*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Elp4	*Elp4*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Elp5	*Elp5*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Elp6	*Elp6*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Kti11	*Dph3*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Kti12	*Kti12*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Kti13	*Ats1/DelGEF*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Kti14	*CK1δ*	$\underline{mcm}^5s^2U$, $\underline{mcm}^5U$, $\underline{ncm}^5U$, $\underline{ncm}^5Um$
Trm9	*Alkbh8*	$\underline{m}cm^5s^2U$, $\underline{m}cm^5U$
Trm112	*Trmt112*	$\underline{m}cm^5s^2U$, $\underline{m}cm^5U$
Nfs1	*Nfs1*	$mcm^5\underline{s}^2U$
Tum1	*Tum1*	$mcm^5\underline{s}^2U$
Urm1	*Urm1*	$mcm^5\underline{s}^2U$
Uba4	*Uba4*	$mcm^5\underline{s}^2U$
Ncs2	*Ctu1*	$mcm^5\underline{s}^2U$
Ncs6	*Ctu2*	$mcm^5\underline{s}^2U$

Among all the cells in the body, neurons are particularly sensitive to translation defects, due to their high demands of synaptic plasticity and polarized morphology; even slight reductions in tRNA modification levels cause dramatic cellular malfunction while stronger alterations are lethal (reviewed in [323]). Absence of wobble modifications at tRNA U_{34} has also been connected with accumulation of protein aggregates in yeast, nematodes, mice and human cells (reviewed in [166]). In humans, protein aggregation is considered as a neurodegenerative hallmark with deleterious consequences to the cell (reviewed in [324]).

In addition to their functions preventing neurodegeneration, enzymes catalyzing U_{34} tRNA modifications have been described as regulators of certain types of cancer. However, they seem to have an opposite effect: whereas decrease of modified U_{34} is associated with the development of cognitive disorders, their overabundance is strongly associated with tumor cells survival and metastasis (Figure 46)[317,325].

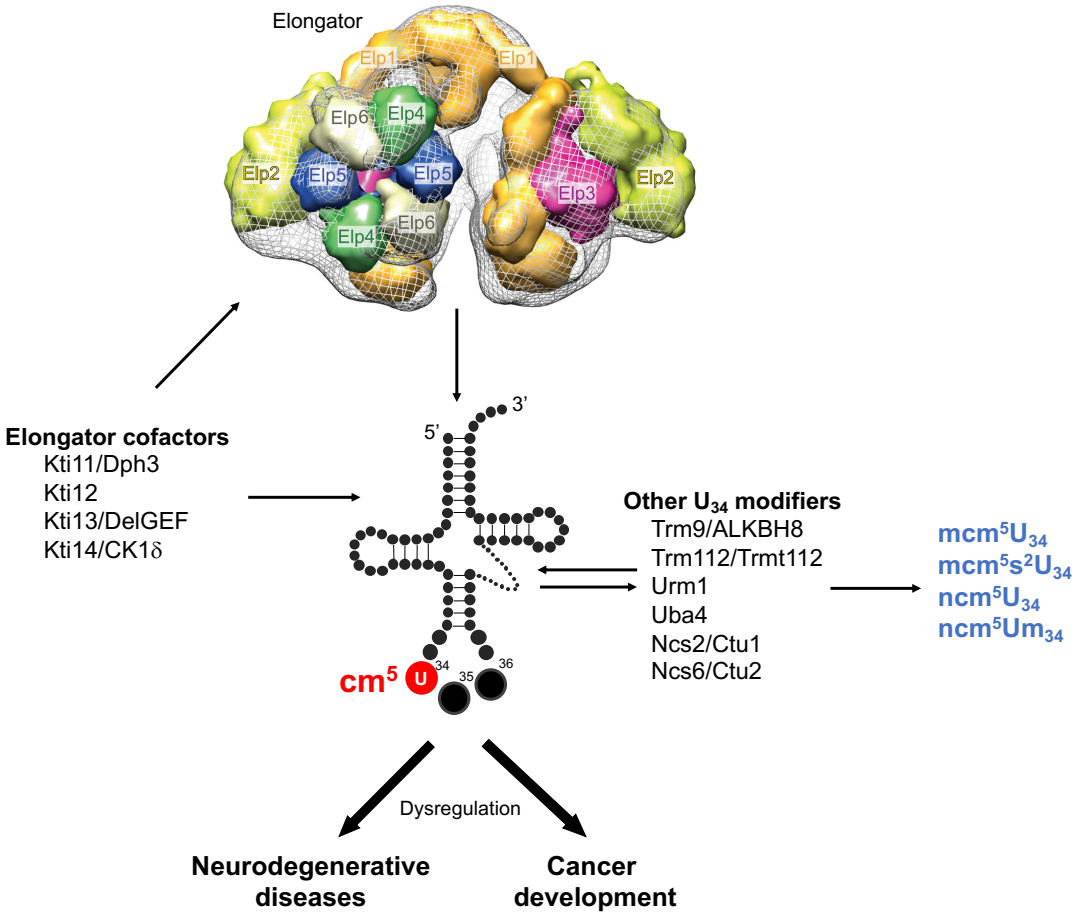


Figure 46. Scheme of Elongator-dependent modifications role in human diseases

Opposing roles, reduced or enhanced, of cm^5 modification in the development of neurodegenerative diseases and cancer respectively. Modified from [166].

1.3.5.1 Elongator in neurological disorders

Neurological disorders are becoming one of the biggest public health challenges. According to the World Health Organization (WHO) and the Global Burden of Disease (GBD), in 2015 neurological disorders represented 11.84% of the total deaths worldwide, and by 2030, the share is anticipated to increase to 12.22%[326].

In humans, Elongator germline mutations cause Hereditary Sensory and Autonomic Neuropathies III diseases, also known as Familial Dysautonomia (FD), an autosomal recessive disorder that affects 1/3600 individuals of the Ashkenazi Jewish population. FD is characterized by impaired development and progressive degeneration of the sensory and autonomic nerves (reviewed in [327]). Fibroblasts and brain tissue from FD patients contain reduced levels of the $mcm^5s^2U_{34}$ nucleoside due to a mutation in the Elp1 human homolog gene IKBKAP. In 99.5% of the cases, this homozygous mutation is a T to C substitution localized in the position 6 of the intron 20. It results in splicing out of exon 20 producing a frameshift that causes the formation of a premature stop codon. This aberrant splicing generates a dysfunctional truncated version of the protein IKAP (residues 1-914, instead 1332) resulting in inefficient translation caused by reduced levels of U_{34} modifications[328]. Interestingly, the expression of the truncated IKAP protein is tissue-

specific, as patients with FD were shown to present wild-type IKBKAP in lymphoblasts whereas the truncated mRNA appears primarily in brain tissue[307]. A study identified a small compound named RECTAS (rectifier of aberrant splicing) which promotes inclusion of exon 20 of IKBKAP by promoting exon recognition either through activating splicing enhancers or inhibiting splicing suppressors[308]. Mutations in IKBKAP have also been shown to increase risk of bronchial asthma in children[322].

Elongator has also been implicated in cerebellar ataxias, a group of severe neurodegenerative disorders characterized by the lack of voluntary coordination of muscle movements. Cerebellar ataxias are progressive and currently incurable (reviewed in [329]). A point mutation in Elp6 (L126Q) has been reported to cause Purkinje neuron degeneration and ataxia-like phenotype in mice. This Elp6 mutation destabilizes the assembly of the heterohexameric Elp456 ring, compromising tRNA modifications and promoting protein aggregation[214].

Elp3, the catalytic subunit of Elongator, was reported to modulate Amyotrophic Lateral Sclerosis (ALS), a fatal degenerative motor neuron disorder in adults, followed by paralysis and death. This disease is characterized by degeneration of motor neurons in the brainstem, spinal cord and motor cortex, resulting in muscle weakness, spasticity and atrophy (reviewed in [330]). The hereditary subtype of ALS is mostly caused by mutations in the Cu-Zn superoxide dismutase (SOD1) gene responsible for destroying free superoxide radicals in the cell. Mutations in SOD1 demonstrate neurotoxic properties via increasing aggregation and protein misfolding (reviewed in [330]). In a study of microsatellites markers, allelic variants of Elp3 were associated with ALS in three human populations[309]. Similarly, a mutagenesis screen for genes related to neural communication and survival in Drosophila identified two loss-of-function mutations in Elp3, involving the protein in neurodegeneration[309]. Zebrafish and mice models of ALS showed that expression of human Elp3 prevents the toxicity caused by SOD1 mutation. In counterpart, heterozygous depletion of Elp3 worsens ALS phenotype in mice associating low expression of the protein with an increasing disease risk, while homozygous deletion of Elp3 is embryonically lethal[310].

Autism spectrum disorder and intellectual disability (ID) are the two most common neurodevelopmental disorders. They are characterized by impairment in cognition and adaptative functioning. Two independent studies have associated rare Elp2 variants with ID[311,312]. Mouse models carrying homozygous variants of Elp2 (Elp2H206R or Elp2H206R/R464W) present very similar phenotypes including developmental delay, motor defects, hypotonia and vocalization impairment, suggesting that Elp2 mutations are the main genetic driver of the severe clinical phenotype[331]. Elp4 mutations have also been implicated with ID[313], autism[313] and Rolandic epilepsy[314].

1.3.5.2 Elongator in cancer

Cancer is a group of related diseases characterized by the accumulation of alterations in the cell genome leading to an uncontrolled cell division that promotes tumor formation[332]. Malignant cancer cells have the potential to travel through blood or lymph, invade distant tissues and generate a new tumor in a process named metastasis[333]. Cancer can originate from almost any tissue of the body and currently more than 100 different types have been identified depending on the origin tissue and cell characteristics[334]. In 2018, GLOBOCAN database estimated 18.1 million new cases and 9.6 million deaths caused by cancer worldwide, with lung, female breast, and colorectal cancer having the highest mortality[335].

Breast cancer is the most common invasive cancer and the first cause of death by cancer in women. In 2020, 2.3 million cases were diagnosed and the total of deaths 685 000 worldwide, representing 24.5% and 15.5% respectively of all female cancers[336]. Elp3 as well as the cytosolic thiouridylases subunits 1 and 2 (Ctu1 and Ctu2), the human homologs of the yeast Ncs2 and Ncs6, which are responsible for the tRNA U_{34} thiolation, are upregulated in human breast cancer and sustain metastasis by supporting cellular invasion through the translation of the histone chaperone DEK. DEK belongs to the chromatin topology modulators and it promotes the internal ribosome entry site (IRES)-dependent translation of the pro-invasive transcription factor LEF1[315].

Melanoma is a type of skin cancer which originates from the melanocytes. In 2020, it has an incidence of 3.4 and a mortality of 0.56 for each 100,000 habitants worldwide, increasing up to 18.9 and 1.5 respectively in western Europe[337]. Serine/threonine-protein kinase B-Raf encoded by the gene BRAF, is a protein kinase involved in signaling of the cell growth; it is considered as a proto-oncogene as its dysfunction is often related with cancer. The mutation substituting glutamic acid (E) for valine (V) in the codon 600 of BRAF kinase (BRAFV600E) is the most prevalent mutation in human melanoma with about 50% of melanoma cases carrying the substitution (reviewed in [338]). Elongator subunits Elp1 and Elp3, as well as Ctu1 and Ctu2, were found to be overexpressed in melanoma biopsies carrying BRAFV600E mutation[316]. Knockdowns of Elp3 and Cut1/Cut2 significantly reduced viability of human BRAFV600E melanoma cells, effect that is not observed in normal human melanocytes and other mutant BRAF cells (BRAFW)[317]. Moreover, depletion of Elp1, Elp3, Elp5 and Elp6 significantly decreased migration and tumorigenicity of melanoma-derived cells[318]. Additionally, silencing of Elp3 and Ctu2 in mouse melanoma cells expressing human BRAFV600E, increased sensitivity to vemurafenib, a BRAF inhibitor used as target treatment in melanoma patients. Elongator was also reported to promote glycolysis in melanoma cells trough a codon-dependent regulation of translation of the hypoxia-induced factor 1α (HIF1α) mRNA to maintain high protein levels. Development of resistance against BRAF therapy is associated to high levels of enzymes catalyzing U_{34} modification as well as HIF1α. Collectively, results demonstrate that Elongator promote survival and resistance to therapy of melanoma cells by regulating mRNA translation[317]. Additionally, depletion of Elp3 using CRISPR–Cas9 impaired the development of BRAFV600E melanoma in zebrafish[317].

Medulloblastoma is a type of brain cancer which develops in the cerebellum. Even though it can occur in children and adults, medulloblastoma is the most common malignant brain tumor in children, representing almost 20% of all pediatric brain tumors (reviewed in [339]). A study in 2019 identified Elp1 as an important marker of Sonic Hedgehog Medulloblastoma (MBSHH), the second most common subtype of medulloblastoma (reviewed in [340]). The authors found an association between pediatric MBSHH and germline loss-of-function (LOF) variants in IKBKAP[215]. Specifically, IKBKAP-associated MBSHH showed inactivation of the gene due to a somatic loss of chromosome arm 9q. The majority of patients carrying this genotype, also showed somatic alterations in the Protein Patched Homolog 1 (PTCH1) gene, a receptor for the Sonic Hedgehog protein (main regulator of embryonic morphogenesis, organogenesis and organization of the central nervous system) which acts as tumor suppressor[215]. These results suggest that IKBKAP-LOF variants together with the activation of the SSH pathways, predispose patients to develop MBSHH. Collectively, tumor of patients with IKBKAP-associated MBSHH are characterized by a destabilized Elongator complex and; therefore, loss of tRNA modifications resulting in the induction of unfolded protein response[215].

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer in adults; early metastasis represents the main cause of low survival rates and relapse among patients. Metastasis is a multistep process from which migration and invasion play a major role. These processes required diverse mechanisms including cytoskeleton dynamics and as well as expression of adhesion and proteolytic enzymes (reviewed in [341]). The matrix metalloproteinases (MMPs) are enzymes capable to degrade the extracellular matrix, characteristic that has vital importance in migration and proliferation processes[319]. Elongator proteins Elp3 and Elp4 were reported to be overexpressed in tumor tissue of patients with HCC. Moreover, overexpression of Elp3 or Elp4 increases the capability of cells to migrate and invade, effect that is abolished by silencing the genes[320]. Aberrant overexpression of Elp3 promotes phosphorylation of protein kinase B (AKT); which in turn enhances the expression of MMP-2 and MMP-9. Altogether, these results suggest that Elongator promotes migration and invasion of HCC cells by upregulating MMP-2 and MMP-9 through the PI3K/AKT signaling pathway[320]. A mice model also confirmed that overexpression of Elp3 enhances HCC tumor metastasis to lungs[320].

The methyltransferase ALKBH8, involved in the formation of the methyl group of $mcm^5s^2U_{34}$ is overexpressed in bladder cancer and its depletion showed to induce apoptosis due to the loss of expression of survivin, an anti-apoptotic protein[321].

1.4 PROJECT OUTLINE

My PhD thesis project is aimed at studying the contribution of the Kti12 cofactor to the function of the Elongator complex using yeast as a model system. Elongator is a protein complex present in eukaryotes whose main function is the post-transcriptional modification of U_{34} in tRNAs to form 5-carboxylmethiluridine (cm^5U) and its derivatives. It is a dodecamer of six subunits, Elp1–Elp6, each of which is present in two copies. In addition, to perform its enzymatic activity, Elongator requires four proteins of the killer insensitive toxin (Kti) protein family (Kti11- Kti14). Our laboratory previously demonstrated the roles of Kti11 and Kti13 as electron donors to the iron-sulfur (Fe-S) cluster in Elp3[285]. Also, Kti12 was proposed to regulate Elp1 phosphorylation together with Kti14[252]; however, the detailed Kti12 molecular function and its contribution to the Elongator-dependent modifications, remained ill-defined.

Based on the sequence similarities between Kti12 and its distant related protein *O*-phosphoseryl-$tRNA^{Sec}$ (PSTK)[289], an enzyme involved in the biosynthesis of the selenocysteine amino acid, we anticipated a structural and functional similarity between the two proteins. PSTK contains two domains: a kinase domain and a tRNA binding domain. Therefore, we made the hypothesis that Kti12 contains related domains and therefore similar activities: an N-terminal domain (NTD) with kinase signature and a C-terminal domain (CTD) with RNA binding capacity.

To assay the putative domains of Kti12 shares during my PhD, I focused on the following objectives:

(i) to assay Kti12 C-terminal domain interaction with RNA;
(ii) to assay the impact of Kti12 on yeast growth phenotype;
(iii) to assay Kti12 N-terminal domain kinase activity;
(iv) to assay interaction between Kti12 with possible protein partners.

First, I set out to verify our starting hypothesis that Kti12 contains two putative domains. For this purpose, I expressed and purified the recombinant *S. cerevisiae* Kti12 C-terminal domain protein. Using electrophoretic mobility shift assay (EMSA) and EMSA competition, I assayed the RNA binding capacity of this Kti12 domain.

Second, I tested the impact of Kti12 on yeast phenotypes constructing *S. cerevisiae* mutant strains lacking Kti12 full gene (ΔKti12), its N-terminal domain ($Kti12_{CTD}$) or its C-terminal domain ($Kti12_{NTD}$) among other mutations; and comparing them with the phenotypes of a yeast strain containing a dysfunctional Elongator (ΔElp3) or a functional one (wild-type).

Thirdly, I performed a kinase assay using Elp1-TAP tagged protein and their interacting partners from yeast extract. TAP IgG pull-downs were incubated with radiolabeled ATP to test whether Kti12 transfers phosphate groups to an Elongator subunit or to another KTI cofactor.

Finally, I used different techniques to test the interaction of Kti12 with other Elongator subunits or cofactors. Data pointed to Kti14 as possible partner of Kti12. Thus, I used the technique of co-immunoprecipitation to test whether Kti12 recruits Kti14, another Elongator

cofactor from the Kti family. Simultaneously, I used co-expression of recombinant Kti12 and Kti14 proteins, to test their interaction *in vitro.*

Elucidating the molecular details of Elongator activity and its modulation is important, due to the fact that the alterations in post-transcriptional events and translational control are extremely detrimental to the health of the organism. In particular, dysfunctions of enzymes involved in post-transcriptional modifications pathways (such as the Elongator complex) are linked to a number of various diseases, including cancer and neurological disorders. Cancer is a group of related diseases characterized by the accumulation of alterations in the cell genome leading to an uncontrolled cell division that promotes tumor formation[332]. Cancer can originate in almost any part of the body and currently more than 100 different types were identified depending on the origin tissue and cell characteristics[334]. Defects in enzymes catalyzing modifications at tRNA U34 were shown to promote metastasis of bladder and breast cancer cells, as well as improve the survival of melanoma cells (reviewed in [166,285]).

Similarly, neurological disorders are becoming one of the greatest threats to public health. Among the various cell types, neurons appear to be particularly sensitive to defects in tRNA modification, when slight reductions in modification levels cause dramatic cellular malfunction and stronger alterations become lethal[323].

Therefore, deeper and clearer understanding on the detailed mechanism in which the Elongator holoenzyme and its partners regulate the protein translation process, will open new avenues on possible therapeutic targets for diseases caused by a dysfunctional Elongator.

2. MATERIALS AND METHODS

2.1 *IN SILICO* ANALYSES

2.1.1 Basic Local Alignment Search Tool (BLAST)

Sequence similarities between Kti12 and PSTK were found using the protein-protein BLAST tool with the sequence of Kti12 of *S. cerevisiae* to find similarities to other proteins (https://blast.ncbi.nlm.nih.gov). FASTA protein sequence of *S. cerevisiae* Kti12 was "blasted" using the 'non-redundant protein sequence' (nr) database and the blastp (protein-protein BLAST) algorithm.

2.1.2 Multiple sequence alignments

Multiple sequence alignments between Kti12 and PSTK proteins from different organisms were made in ClustalX software using the complete alignment algorithm (http://www.clustal.org/clustal2/). Individual FASTA protein sequences from organisms, randomly chosen (in addition to yeast and human), were obtained from UniProt (https://www.uniprot.org). A total of 32 sequence were used for the alignment (21 of Kti12 and 11 of PSTK).

2.2 STRAINS AND MEDIA

2.2.1 Bacterial cultures

Bacterial cells were grown in liquid medium or on agar plates of:

- LB (Lysogeny Broth): 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl (Sigma L3022).
- Terrific Broth (Autoinduction medium): 12 g/l tryptone, 24 g/l yeast extract, 3.3 g/l, $(NH_4)_2SO_4$, 6.8 g/l, KH_2PO_4, 7.1 g/l Na_2HPO_4, 0.5 g/l glucose, 2.0 g/l -lactose, 0.15 g/l $MgSO_4$, 0.03 g/l (*Formedium* AIMTB0210).

Depending on the selection marker of the bacterial plasmid, medium was supplemented with adequate antibiotics in the following concentrations: ampicillin (100 µg/ml), kanamycin (50 µg/ml) and/or chloramphenicol (25 µg/ml).

2.2.2 Bacterial strains

Escherichia coli strain used for cloning and storing plasmids:

- MH1 (*araD39, lacX74, gal E, gal K, hsr, rpsL*)

Escherichia coli strains used for protein expression:

- BL21 CodonPlus-RIL (F- *ompT hsdS*$_B$($r_{B-}m_{B-}$) *dcm*+ Tetr *gal endA* Hte (DE3) [*argU ileY leuW* Cam^R])
- ROSETTA (F-*ompT hsdS*$_B$($r_{B-}m_{B-}$) *gal dcm* (DE3) pRARE (Cam^R))

Bacterial cultures were incubated at 37°C, unless otherwise stated, and shaken at 170 rpm for liquid cultures.

2.2.3 Yeast cultures

Yeast cells were grown in liquid medium or on agar plates (containing 2% bacto agar (*Becton Dickinson*) of:

- YPDA (yeast peptone dextrose adenine): 10 g/l yeast extract, 30 g/l bacto peptone, 20 g/l glucose, 40 mg/l adenine sulfate (*Formedium* CCM1010).
- CSM (complete supplement mixture): amino acids (full amino acids, -LEU, or -LEU-TRP) quantity defined by the manufacturer (*MP Biomedicals*), 6.7 g/l Yeast Nitrogen Base without amino acids (*Becton Dickinson*), 20 g/l glucose, Sorensen buffer 1x.
- Sorensen Buffer 1x: 10 mM Na_2HPO_4; 40 mM KH_2PO_4; pH 6.25

Yeast cultures were incubated at 30°C, unless otherwise stated, and shaken at 170 rpm for liquid cultures.

2.2.4 Yeast strains

Yeast strains were derived from:

- BMA64 *(MAT a, ade2-1, his3-11, 15 leu2-3, 112 trp1 delta ura3-1 can1-100)*
- SUP4 - UMY2893 *(MAT alpha, SUP4, leu2-3,112, trp1-1, can1-100, ura3-1, ade2-1, his3-11,15)*

Table M1. Yeast strains

ID	Genotype	Organism	Background	Reference
Phenotypic assays				
BSY2344	WT	*S. cerevisiae*	SUP4	Bystrom, UMY2893
BSY2339	ELP3::KanMX4	*S. cerevisiae*	SUP4	Céline Faux
BSY3360	Kti12::KanMX6	*S. cerevisiae*	SUP4	This project
BSY3362	Kti12 NTD::KanMX6	*S. cerevisiae*	SUP4	This project
BSY3401	Kti12 CTD::CRISPR	*S. cerevisiae*	SUP4	This project
BSY3404	Kti12 NTD::CRISPR	*S. cerevisiae*	SUP4	This project
BSY3387	Kti12::Kti12:K14A	*S. cerevisiae*	SUP4	This project
BSY3388	Kti12::Kti12:D85A	*S. cerevisiae*	SUP4	This project
BSY3512	Kti14:Kti14 FL-TAP	*S. cerevisiae*	SUP4	This project
BSY3523	Kti14:Kti14 (1-290)-TAP	*S. cerevisiae*	SUP4	This project
BSY3524	Kti14:Kti14 (1-394)-TAP	*S. cerevisiae*	SUP4	This project
Co-immunoprecipitations				
BSY3508	WT	*S. cerevisiae*	BMA64	Baudin-Baillieu *et al.*, 1997[342]

ID	Genotype	Organism	Background	Reference
pBS1479	TAP:TRP cassette			
pBS2438	TAP:TRP:t7 cassette			
BSY2367	ELP1::TAPt7::TRP	*S. cerevisiae*	BMA64	Céline Faux
BSY3523	Kti12::KanMX6	*S. cerevisiae*	BMA64	This project
BSY3524	ELP1::TAPt7::TRP, Kti12::KanMX6	*S. cerevisiae*	BMA64	This project
BSY3626	Kti12::Kti12 NTD	*S. cerevisiae*	BMA64	This project
BSY3630	ELP1::TAPt7::TRP, Kti12::Kti12 NTD	*S. cerevisiae*	BMA64	This project
BSY3642	Kti12::Kti12 CTD	*S. cerevisiae*	BMA64	This project
BSY3643	ELP1::TAPt7::TRP, Kti12::Kti12 CTD	*S. cerevisiae*	BMA64	This project
BSY3654	Kti14::Kti14:TAPt7:TRP	*S. cerevisiae*	BMA64	This project
BSY3656	Kti14::Kti14 (1-394):TAPt7:TRP	*S. cerevisiae*	BMA64	This project
BSY3658	Kti14::Kti14 (1-394):TAPt7:TRP	*S. cerevisiae*	BMA64	This project
Kinase assays				
BSY3389	Kti12::Kti12-TAP:TRP	*S. cerevisiae*	SUP4	This project
BSY3412	Kti12::Kti12 CTD-TAP:TRP	*S. cerevisiae*	SUP4	This project
BSY3413	Kti12::Kti12 NTD-TAP:TRP	*S. cerevisiae*	SUP4	This project
BSY3390	Kti12::KTI12_K14A:TAP-TRP	*S. cerevisiae*	SUP4	This project
BSY3392	Kti12::KTI12_D85A:TAP-TRP	*S. cerevisiae*	SUP4	This project
Eclipse assay				
BSY3378	WT (LEU2,TRP1 (Kl+K+))	*K. lactis*	AWJ137	K.Breuning. University of Halle, Germany
BSY3379	WT (ADE1, ADE2)	*K. lactis*	NK40	R. Schaffrat, University of Leicester, UK.
Yeast Two Hybrid				
BSY3146	WT (Mat alpha, leu2-3, 112, trp1-901, his3 delta 200, ade2-101, gal4 delta, gal80 delta, SPAL10::URA3, GAL1::lacZ, HIS3UAS GAL1::HIS3@LYS2, can1R, cyh2R)	*S. cerevisiae*	MAV203	Invitrogen (DB)

2.3 GENETIC MANIPULATION

2.3.1 Plasmids

Table M2. List of bacterial plasmids.

Organisms from which each protein is derived are abbreviated at the beginning of each gene name: protein organism can be found at the beginning of each gene: *Sc* (*Saccharomyces cerevisiae*), *Sp* (*Schizosaccharomyces pombe*), *Kl* (*Kluyveromyces lactis*) or *Ct* (*Chaetomium thermophilum*).

ID	Insert	Vector	Marker	Reference
Genetic manipulation				
pBS2379	KanMX6	—	Amp/G418	Mark Longtine
pBS1479	TAP:TRP	pBSK(-)	Amp/TRP1	Séraphin's Lab
pBS2438	TAP:TRP:t7	pBSK(-)	Amp/TRP1	Andrzej Dziembowski
CRISPR				
pBS5359	pCas	pCas	KanR	This project
pBS6135	*Sc*Kti12 guide1	pCas	KanR	This project
pBS6136	*Sc*Kti12 guide2	pCas	KanR	This project
pBS6136	*Sc*Kti12 guide3	pCas	KanR	This project
Protein expression				
pBS4603	His6-*Sc*Kti12 (1-314)	*E. coli* expression	KanR	Céline Faux
pBS5625	His6-*Sc*Kti12 NTD (1-182)	*E. coli* expression	KanR	Valerio Taverniti
pBS5626	His6-*Sc*Kti12 NTD (1-208)	*E. coli* expression	KanR	Valerio Taverniti
pBS5627	His6-*Sc*Kti12 CTD (183-314)	*E. coli* expression	KanR	Valerio Taverniti
pBS5628	His6-*Sc*Kti12 CTD (209-314)	*E. coli* expression	KanR	Valerio Taverniti
pBS5723	His6-*Sp*ScKti12	*E. coli* expression	KanR	This project
pBS5724	His6-*Kl*ScKti12	*E. coli* expression	KanR	This project
pBS5817	His6-*Ct*ScKti12CTD(255-378)	*E. coli* expression	KanR	This project
pBS5928	His6-TEV-*Sc*Kti12	*E. coli* expression	KanR	This project
pBS5929	GST-*Sc*Kti12	pGEX6P	AmpR	This project
pBS6005	His6-TEV-*Sc*Kti14 (1-394)	*E. coli* expression	AmpR	This project
pBS6006	*Sc*Kti14 (1-394) (no tag)	*E. coli* expression	AmpR	This project
pBS6194	GST-*Sc*Kti14	pGEX6P	AmpR	This project
pBS6195	GST-*Sc*Kti14 (1-394)	pGEX6P	AmpR	This project
pBS6196	GST-*Sc*Kti14 (1-290)	pGEX6P	AmpR	This project
pBS6257	*Sc*Kti12-His6	*E. coli* expression	AmpR	This project
pBS6258	*Sc*Kti12 NTD-His6	*E. coli* expression	AmpR	This project
Protein co-expression				
pBS5980	His6-TEV-*Sc*Kti12–*Sc*Kti14 FL	*E. coli* expression	KanR	This project
pBS5981	His6-TEV-*Sc*Kti14 (1-394)–*Sc*Kti12	*E. coli* expression	KanR	This project
pBS5982	His6-TEV-*Sc*Kti12–*Sc*Kti14 (1-394)	*E. coli* expression	KanR	This project
pBS5983	His6-TEV-*Sc*Kti12–*Sc*Kti14 (1-290)	*E. coli* expression	KanR	This project
pBS5984	His6-TEV-*Sc*Kti12 NTD (1-182)–*Sc*Kti14 (1-394)	*E. coli* expression	KanR	This project

ID	Insert	Vector	Marker	Reference
pBS5985	His6-TEV-*Sc*Kti12 CTD (183-314)–*Sc*Kti14 (1-394)	*E. coli* expression	KanR	This project
pBS6003	His6-TEV-*Sc*Kti12 CTD (183-313)–*Sc*Kti14 (1-394)	*E. coli* expression	KanR	This project
pBS6092	His6-TEV-*Sc*Kti14 (1-394)–K14A_*Sc*Kti1	*E. coli* expression	KanR	This project
pBS6093	His6-*Sc*Kti14 (1-394)–D85A_*Sc*Kti1	*E. coli* expression	KanR	This project
pBS6128	His6-*Sc*Kti14 K38R (1-394)–*Sc*Kti1	*E. coli* expression	KanR	This project
pBS6197	GST-*Sc*Kti12–*Sc*Kti14 FL	pGEX6P	AmpR	This project
pBS6198	GST-*Sc*Kti12–*Sc*Kti14 (1-394)	pGEX6P	AmpR	This project
pBS6199	GST-*Sc*Kti12–*Sc*Kti14(1-290)	pGEX6P	AmpR	This project
pBS6259	*Sc*Kti14 (1-394)–*Sc*Kti12-His6	*E. coli* expression	KanR	This project
pBS6260	*Sc*Kti14 (1-394)–*Sc*Kti12 NTD-His6	*E. coli* expression	KanR	This project
pBS6261	*Sc*Kti14 K38R (1-394)–*Sc*Kti12-His6	*E. coli* expression	KanR	This project
tRNA expression				
pBS6	sRNA U6 (1-57)	pUC19	AmpR	Bertrand Séraphin
pBS3508	dtRNA Glu1-TTC-1	pUC19	AmpR	This project
pBS5820	dtRNA Glu1-TTC-1 (Δ acceptor stem)	pUC19	AmpR	This project
pBS5821	dtRNA Glu1-TTC-1 (Δ D-loop)	pUC19	AmpR	This project
pBS5822	dtRNA Glu1-TTC-1 (Δ anticodon loop)	pUC19	AmpR	This project
pBS5823	dtRNA Glu1-TTC-1 (ΔTYC loop)	pUC19	AmpR	This project
pBS5824	dtRNA-Ser-TGA-1-1	pUC19	AmpR	This project
pBS5825	dtRNA-Leu-TAG-1-1	pUC19	AmpR	This project
Two Hybrid System				
pBS2609	pDONR221 (Gateway Technology)	pDONR221	KanR	Invitrogen (DB)
pBS4238	*Sc*Elp3	pDONR221	KanR	Céline Faux
pBS4239	*Sc*Elp1	pDONR221	KanR	Céline Faux
pBS4246	*Sc*Elp5	pDONR221	KanR	Céline Faux
pBS5106	*Sc*Kti11	pDONR221	KanR	Olga Kolaj-Robin
pBS5107	*Sc*Kti13	pDONR221	KanR	Olga Kolaj-Robin
pBS5726	*Sc*Kti12	pDONR221	KanR	This project
pBS5727	*Sc*Kti14	pDONR221	KanR	This project
pBS2354	pDEST22 (Gateway Technology)	pDEST22(AD)	Amp/TRP1	Invitrogen (DB)
pBS4178	*Sc*Elp2	pDEST22	Amp/TRP1	Céline Faux
pBS4180	*Sc*Elp4	pDEST22	Amp/TRP1	Céline Faux
pBS4182	*Sc*Elp5	pDEST22	Amp/TRP1	Céline Faux
pBS4183	*Sc*Elp6	pDEST22	Amp/TRP1	Céline Faux

ID	Insert	Vector	Marker	Reference
pBS4277	*Sc*Elp1	pDEST22	Amp/TRP1	Céline Faux
pBS4279	*Sc*Elp3	pDEST22	Amp/TRP1	Céline Faux
pBS5748	*Sc*Kti11	pDEST22	Amp/TRP1	This project
pBS5752	*Sc*Kti13	pDEST22	Amp/TRP1	This project
pBS5750	*Sc*Kti12	pDEST22	Amp/TRP1	This project
pBS5754	*Sc*Kti14	pDEST22	Amp/TRP1	This project
pBS2356	pDEST32 (Gateway Technology)	pDEST32(AD)	GenR/LEU2	Invitrogen (DB)
pBS4179	*Sc*Elp2	pDEST32	GenR/LEU2	Céline Faux
pBS4181	*Sc*Elp4	pDEST32	GenR/LEU2	Céline Faux
pBS4184	*Sc*Elp6	pDEST32	GenR/LEU2	Céline Faux
pBS4278	*Sc*Elp1	pDEST32	GenR/LEU2	Céline Faux
pBS4280	*Sc*Elp3	pDEST32	GenR/LEU2	Céline Faux
pBS4281	ScElp5	pDEST32	GenR/LEU2	Céline Faux
pBS5749	*Sc*Kti11	pDEST32	GenR/LEU2	This project
pBS5751	*Sc*Kti12	pDEST32	GenR/LEU2	This project
pBS5753	*Sc*Kti13	pDEST32	GenR/LEU2	This project
pBS5755	*Sc*Kti14	pDEST32	GenR/LEU2	This project
Co-immunoprecipitation				
pBS6069	*Sc*Kti14:HA endogenous prom+term	pRS305	AmpR/ LEU2	This project
pBS6070	*Sc*Kti12:HA endogenous prom+term	pRS305	AmpR/ LEU2	This project
pBS6148	*Sc*Kti14FL:HA endogenous prom+term	pRS426	AmpR/ URA3	This project
pBS6149	*Sc*Kti14FL:HA endogenous prom+term	pRS426	AmpR/ URA3	This project
pBS4919	*Sc*Elp4:His6-VSV3	pRS315	AmpR/LEU2	Céline Faux
pBS6173	*Sc*Kti12-HA3	pYX242	AmpR/LEU2	This project
pBS6174	*Sc*Kti14-HA3	pYX242	AmpR/LEU2	This project
pBS6175	HA3:*Sc*Kti14	pYX242	AmpR/LEU2	This project
pBS6176	*Sc*Elp4:HA3	pYX242	AmpR/LEU2	This project

2.3.2 Gene deletion introducing KanMX6 cassette

S. cerevisiae knock-out strains lacking Kti12, Kti12$_{NTD}$ or Kti12$_{CTD}$ were made by replacing the endogenous coding sequence of interest with a cassette marker (KanMX6) which provides yeast resistance to the G418 antibiotic. For this purpose, two oligos per deletion (forward and reverse) containing part of the sequence of the KanMX6 cassette (bold sequence presented in the Oligos 1 and two scheme below) and part of the sequence of the yeast genome flanking the coding sequence to be deleted of around 55 nucleotides long (N(55)) (Figure M1) were designed.

- Oligo 1 Forward: 5´ N(55)**CGGATCCCCGGGTTAATTAA**
- Oligo 2 Reverse: 5´ N(55)**GAATTCGAGCTCGTTTAAAC**

Oligos are used in a PCR containing the KanMX6 cassette containing-plasmid as template to amplify it, introducing the borders of the neighbor areas of the coding sequence to delete. PCR

products were verified in an agarose gel and PCR products were used for yeast transformation. Transformed cells are then incubated on YPDA plates containing G418 at 30ºC for two days to select those having integrated the selection marker. Afterward, the plate was replicated on a second selective plate to eliminate background colonies and the new plate incubated at 30ºC for two additional days. Then, 6-8 colonies of each transformation were subcloned in a fresh YPDA plate containing G418 and incubated as before. Then, one colony of each subclone was used to make a master plate and was also used to inoculate 3 ml of YPDA culture media. Liquid medium and plates were incubated at 30ºC with and without agitation respectively. From the master plate, new replica plates were made on CSM media lacking -TRP, -LEU, -HIS, -ARG, -URA, -LYS, -ADE or in YPDA or YPDA-G418 to test that the phenotypes of strains coincide with their genotype. From the liquid culture genomic DNA was isolated to check the correct insertion of the KanMX6 cassette by PCR using oligos located in the promoter and terminator sequence of the targeted gene (numbered 3 and a 4 respectively) to verify KanMX6 insertion site borders. Positive colonies were stored in glycerol stock containing 300µl of 100% autoclaved glycerol and 700 ml of yeast culture in YPDA. Stocks were stored at -80ºC until use.

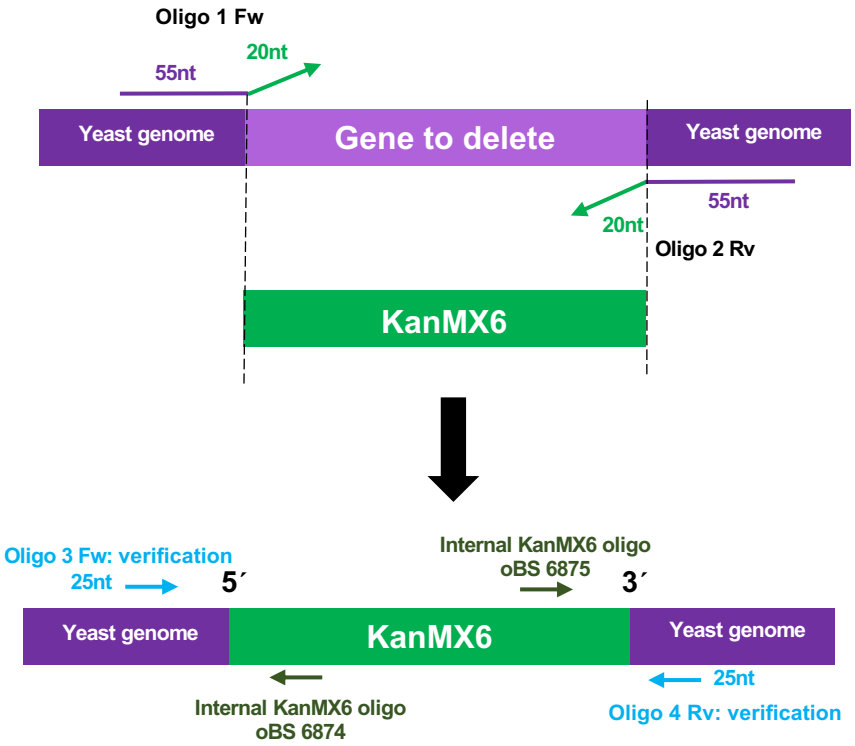


Figure M1. Diagram of gene deletion introducing KanMX6 cassette

2.3.3 Site-directed mutagenesis

To introduce features in some expression vectors (restriction sites, ribosome binding sites, His_6-tag or HA_3-tag), as well as point mutations ($Kti14_{K38R}$) in proteins, site directed mutagenesis was used. Whenever possible, a restriction site was introduced with the desired mutation to allow its quick monitoring. Complementary primers carrying the desired mutation(s) were designed to harbor 15-20 nucleotides on each side of the altered nucleotide(s). The target plasmid was

amplified, by using 40 nM of each primer, 200 ng plasmid, 0.2 mM dNTPs, PfuUltra II fusion HS DNA polymerase and PfuUltra II buffer in a 50 µl reaction. After an initial denaturing cycle at 95°C for 5 minutes, 16 cycles of amplification were performed at: 95°C 1 for minute, 58°C for 1 minute and 68°C for approximately 1.5 x the plasmid size (kb) in minutes, followed by a step of 10 minutes at 60°C. Then, 20 U of DpnI enzyme (*New England Biolabs*) was added to the reaction and incubated at 37°C for 1 hour, to remove the original plasmid. The product was used for MH1 bacteria transformation (section 2.4.9).

2.3.4 Gene mutation by CRISPR/Cas9

S. cerevisiae mutant strains Kti12$_{NTD}$, Kti12$_{K14A}$ or Kti12$_{D85A}$ were made using CRISPR/Cas9 technology. For this purpose, two repair oligos of 100 nucleotides containing the point mutation or flanking the region to delete were made. These oligos contain also a restriction site introduced by silent-mutagenesis (to assay clones by restriction digestion) and a mutation inactivating the specific PAM sequence (NGG) recognized by the *Streptococcus pyogenes* Cas9 (*sp*Cas9) to make a double strand break. Another oligo containing the guide RNA (gRNA) of 20 nucleotides long localized exactly next to PAM sequence was also designed and cloned into the plasmid containing pCas9-Tyr-HDV-LYP1 replacing the LYP1 guide by site directed mutagenesis. This approach allowed its expression with the tracRNA (Figure M2).

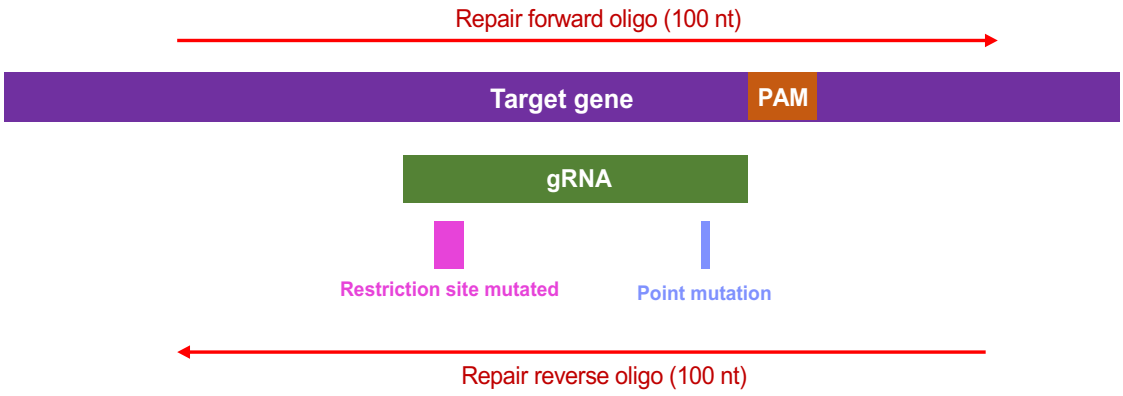


Figure M2. Organization of repair oligonucleotides used to introduce point mutation using CRISPR/Cas9

2.3.5 Generation of genomic C-terminus TAP-tagged constructs

To build *S. cerevisiae* mutant strains carrying the TAP-tag in the C-terminus side of the proteins of interest (Elp1-TAP, Kti12-TAP, Kti12$_{NTD}$-TAP or Kti12$_{CTD}$-TAP) two oligos containing 55 nucleotides of the flanking region of the C-terminus of the protein to tag and 17 nucleotides of the TAP:TRP cassette were designed. Oligos were used to amplified the TAP:TRP or TAP:TRP:T7 cassette (T7 for is detection with antibodies) by performing a PCR (Figure M3). PCR products were analyzed on a PCR and subjected to a PCI extraction and ethanol precipitation. Then, PCR products were used to transform the appropriate parental yeast strain. Transformed yeast cells were plated on CSM -TRP plates and incubated at 30 ºC for two days. Then, 6-8 colonies were subcloned in YPDA-G418 plates and incubated as before. After incubation time,

one colony of each subclone was used to make a master plate and was also used to inoculate 3 ml of YPDA culture media. Liquid medium and plates were incubated at 30ºC with and without agitation respectively. From the master plate, new replica plates were made on CSM media lacking -TRP, -LEU, -HIS, -ARG, -URA, -LYS, -ADE or in YPDA or YPDA-G418 to test that the phenotypes of strains coincide with their genotype. From the liquid culture genomic DNA was isolated to check the correct insertion of the TAP:TRP cassette by PCR using oligos located in the promoter and terminator sequence of the targeted gene (numbered 3 and a 4 respectively) to verify TAP insertion site borders. Positive colonies were used to perform Kushnirov rapid protein extraction (section 2.5.5) to test the expression of the TAP-tagged proteins on a western blot by using the peroxydase anti-peroxydase (PAP) antibody. Colonies expressing the TAP protein were stored in glycerol stock containing 300µl of 100% autoclaved glycerol and 700 ml of yeast culture in YPDA. Stocks were stored at -80ºC until use.

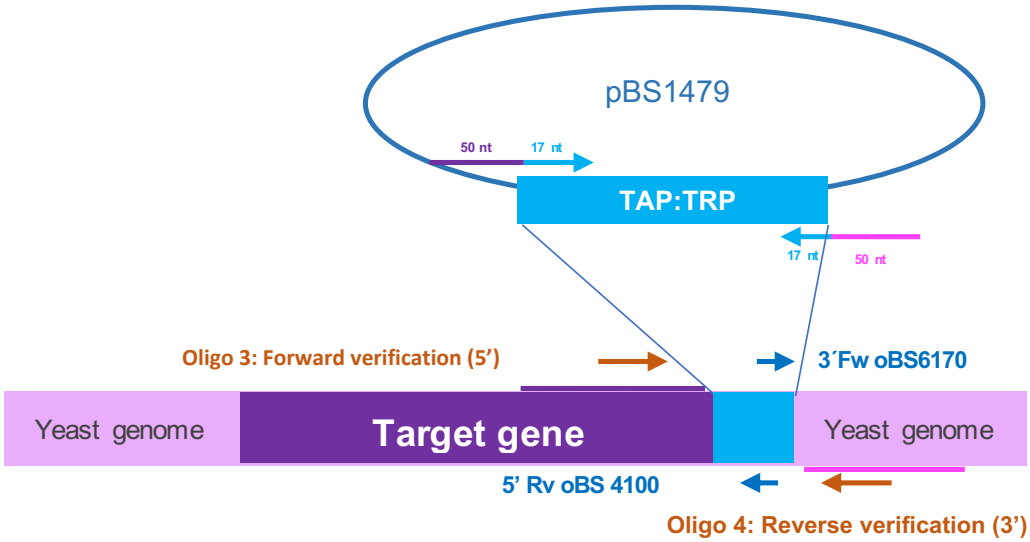


Figure M3. Diagram of C-terminus TAP tagging

2.4 NUCLEIC ACID METHODS

2.4.1 Determination of nucleic acid concentration

Nucleic acid concentration was measured using a NanoDrop, placing 1µl of DNA or RNA solution, recording absorbance at 260nm or 280nm, and deriving concentration. Ratio of absorbance 260/280 was also checked.

2.4.2 Plasmid DNA isolation from bacteria

Bacteria colonies containing the plasmid to be prepared were inoculated in 3 ml (miniprep) or 100 ml (midiprep) of LB medium containing the appropriate antibiotic and incubated overnight at 37ºC and 170 rpm. Saturated cultures were pelleted at 5000 rpm for 10 minutes. Plasmids were prepared by using the kits: NucleoSpin® Plasmid *EasyPure* (*Macherey Nagel* 740588.50) for

minipreps or NucleoBond® Xtra Midi (*Macherey Nagel* 740410.10) for midipreps according manufacturer's instructions. Plasmids were resuspended in 20-200 µl H_2O and store at -20ºC.

2.4.3 Genomic DNA isolation from yeast

1 ml of saturated yeast culture was pelleted at 14000 rpm for 5 minutes. Yeast pellet was resuspended in 200 µl of resuspension buffer (10 mM Tris-Cl pH 8.0, 100 mM NaCl, 1% SDS, 2% Triton, 1 mM EDTA pH 8.0). Then, 300 µl of glass beads and 300 µl of PCI were added to the yeast pellet and vortexed strongly for 3 minutes. After, 300 µl of H_2O were added and mixed energetically by hand. Samples were centrifuged at 1400 rpm for 5 minutes. Aqueous phase was recovered in a new tube and added with 1 ml of cold 100% ethanol and shake by hand. Samples were centrifuged at 14000 rpm for 30 min at 4ºC. Samples were dried and DNA was resuspended in 40 µl of H_2O.

2.4.4 PCI extraction and ethanol precipitation

Nucleic acids were purified by adding 1 volume of phenol:chloroform:isoamyl alcohol 25:24:1 mixture (PCI). Mixture was vortexed for 1 minute and centrifuged at 14000 rpm for 5 minutes. DNA/RNA in the upper phase was precipitated by addition of 1/10 volume of 3M NaAc pH 5.2, 2.5 volume of 100% ethanol (-20°C) and 1µl glycogen (20mg/ml) followed by 1 hour incubation at -80ºC or overnight incubation at -20°C. Mixture was centrifugated at 14000 rpm for 30 minutes at 4°C. DNA/RNA pellets were washed with 70% ethanol (-20°C) and centrifugated at 14000 rpm for 10 minutes at 4°C. Pellets were dry 5 min at room temperature or with a vacuum pump. After, nucleic acids were dissolved in 10µl of ultra-pure H_2O.

2.4.5 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was performed using either Phusion High-fidelity DNA polymerase and HF buffer (Finnzymes F-530) or PfuUltra II fusion HS DNA polymerase and the recommended buffer (Agilent 600670) following manufacturer's instructions. PCR products were analyzed on an agarose gel and then purified by PCI extraction and ethanol purification (section 2.4.4).

2.4.6 Restriction enzyme DNA digestion

DNA (genomic, plasmids or PCR products) was incubated with proper master mix (NEB CutSmart, 2.1 or 3.1) for 1 hour at 37ºC according to manufacturer's instructions *(New England Biolabs)*. After incubation, volume corresponding to 250 ng of DNA were loaded into 1% agarose gel for verification.

2.4.7 DNA agarose gel electrophoresis

Agarose gel electrophoresis was used to separate nucleic acids for analytical visualization. Gels were prepared containing 1% agarose, 0.5x Tris-borate-EDTA (TBE) buffer and 0.25 µg/ml of ethidium bromide (EtBr). DNA samples were mixed with 6x DNA Gel Loading Dye (*Thermo Scientific*™) and GeneRuler 1 kb Plus DNA Ladder (*Thermo Scientific*™) was used as molecular marker. Gels were run in 0.5x TBE buffer containing 0.25 µg/ml of EtBr at a constant voltage of 80-120 V. DNA bands were visualized in a UV Transilluminator at 254 nm.

2.4.8 In-gel ligation

Digested plasmids and PCR products were separated by using 1% Low-melting SeaPlaque® Agarose (*Lonza Rockland 50110*) in 1x Tris-acetate (TA) running buffer. Samples were run at 50 V for 3-5 hours at 4ºC. Then, gels were staining with 0.5% EtBr in TA buffer. DNA bands were detected by using a Safe Imager™ 2.0 Blue-Light Transilluminator (*Thermo Fisher Scientific*). Bands were cut with a scalpel and deposited in 1.5 ml Eppendorf tubes. Bands were melted at 68ºC for 10 minutes and then kept at 42ºC. Plasmid and inserts were mixed in 10µl volume for one insert (5 µl plasmid and 5 µl insert) or 15 µl (5 µl plasmid, 5 µl insert 1 and 5 µl insert 2) for two inserts and incubated with 1 volume of T4 DNA ligase mix (T4 DNA ligase 20 U/µl, 2x T4 DNA ligase buffer, *New England Biolabs* at 16ºC for at least 4 hours. Ligated reactions were melted at 68ºC for 10 minutes and then kept at 42ºC then 50 µl of 50 mM $CaCl_2$ was added and to half of the ligation reaction volume. The mixture was transferred on ice. Ligation products were then transformed into bacteria cells (section 2.4.9).

2.4.9 Bacterial transformation

E. coli cells were transformed by heat shock method. To 100 µl of frozen competent MH1 cells were mixed 100-250 ng of plasmid and cooled on ice for 30 minutes. Cells were then heat shocked at 42°C for 90 seconds and put back on ice. After the addition of 900 µl LB-medium cells were incubated at 37°C for 45 minutes with gentle shaking. Subsequently, cells were collected by 1 min centrifugation at 2000 rpm. The cell pellet was resuspended in 150 µl of LB-medium and plated on LB agar containing the selective antibiotics. Plates were incubated at 37ºC overnight.

2.4.10 Plasmid verification after ligation and mutagenesis

Transformed colonies were carefully selected and grow independently in 3 ml LB medium containing the appropriate antibiotic. Plasmids were purified by using miniprep (section 2.4.2). Plasmids were digested with appropriate restriction enzyme and analyzed in TBE 1% agarose gel containing 0.25% EtBr. 20 µl of plasmid containing 35-100 ng/ml from positive clones were sent for sequencing at Eurofins Scientific.

2.5 PROTEIN ANALYSES

2.5.1 Protein induction

Competent *E. coli* BL21-CodonPlus (or Rosetta DE3 for expression/purification trials) cells were transformed by heat shock method with 1 µg of plasmid (section 2.4.9). 100 ml of LB or Terrific Broth (autoinduction) media were then placed in 500 ml flask and inoculated with 5 colonies of the transformed bacteria. Cultures were added with the proper antibiotic (ampicillin 100 µg/ml or kanamycin 50 µg/ml) and chloramphenicol 25 µg/ml. For LB media cultures, cells were incubated at 37ºC, 170 rpm until they reached 0.5-0.6 OD_{600}, then IPTG was added to a final concentration of 1mM; then, cells were moved to another incubator at 18º C. For Terrific Broth (autoinduction) medium cultures, cells were incubated during 6 hours at 37 ºC, 170 rpm; then, moved to 18ºC. In both cases, cells were incubated overnight (~18h) for protein expression. Next day, cells were pelleted at 4ºC and 5000 rpm for 20 minutes. Cells were divided into two 50 ml pellets, then cells were washed with cold 1x PBS and centrifuged again as before. Pellets were flask frozen with liquid nitrogen and stored at -80ºC.

2.5.2 Purification of His_6-tagged proteins in batch

This protocol was used for small volume cultures for expression/purification trials. 50 ml pellets of bacterial culture were resuspended in 1.5 ml of buffer A (50mM Tris-Cl pH 7.5, 300 mM NaCl, 10 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x protease inhibitor cocktail (PIC) (*Roche*) and moved to a 5 ml plastic tube; aliquot 1 'total protein' was taken. Cells were incubated with 0.1 mg/ml lysozyme and 0.01 mg/ml DNase I during 10 minutes on ice. Cells were lysed by ultrasonication with a microtip Ultrasonic Liquid Processors (*Vibra-Cell™)* for 5 × 10 sec, pulse 2''on-2''off, 20% amplitude, keeping sample on ice between each sonication. Cells were moved to a 2 ml Eppendorf tube and centrifuged at 14000 rpm during 10 minutes at 4ºC. Supernatant was moved to a 1.5 Eppendorf tube; aliquot 2 'lysate' was taken. Purification was achieved by affinity chromatography using Ni-NTA beads (*Qiagen*). For each 50 ml of cell culture, 40 µl of Ni-NTA beads were used and washed by adding 1 ml of Buffer A supplemented with 1x PIC, spun 2,000 rpm for 2 minutes, beads were drained, and wash procedure was repeated two times more. Finally, beads were resuspended in 40 µl of Buffer A supplemented with 1x PIC. Equilibrated beads were added to each sample and rotated during 1h at 4ºC. Beads were spin-down at 2000 rpm for 2 min; aliquot 3 'flow through' was taken from supernatant. Beads were drained with a precision needle syringe and washed with 1 ml of Buffer B (50mM Tris-Cl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x PIC; cells were spin-down as before and aliquot 4 'first wash' was taken from supernatant. The washing step is repeated one more time. Bound proteins were eluted by adding 80 µl of Buffer C (50mM Tris-Cl pH 7.5, 300 mM NaCl, 200 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x PIC and incubating during 5 minutes with gentle shaking. Elution was recovered and move to a new Eppendorf tube; aliquot 5 'first elution' was taken. A second elution was made under the same conditions; aliquot 6 'second elution' was taken. All aliquots contained 30 µl of sample and

15 µl of 3x protein loading dye (PLD). Samples were quick frozen with liquid nitrogen and store at -80ºC.

2.5.3 Purification of His_6-tagged proteins in column

This protocol was used for bigger volume cultures for *in vitro* assays. Pellets of 1L bacteria culture were resuspended in 20 ml of Buffer A (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 10 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x protease inhibitor cocktail (PIC) (*Roche*) and moved to a 50 ml plastic tube; aliquot 1 'total protein' was taken. Cells were incubated with 0.1 mg/ml lysozyme and 0.01 mg/ml DNase I during 10 minutes on ice. Cells were lysed with a High-Pressure Cell Disruptors (*Constant Systems*®) at 1.5 kBar and 4ºC, passing cells 2 times. Lysated cells were moved to a new 50 ml tube and centrifuged 5000 rpm during 20 minutes at 4ºC; aliquot 2 'lysate' was taken. Purification was achieved by affinity chromatography using Ni-NTA beads (*Qiagen*) in a Poly-prep Chromatography column 0.8 x 4 cm. For each 1L of cell culture, 800 µl of Ni-NTA beads were used and washed by adding 5 ml of buffer A supplemented with 1x PIC, spun 2000 rpm for 2 minutes, beads were drained, and wash procedure was repeated two times more, beads were resuspended in 800 µl of buffer A supplemented with 1x PIC. Total protein extract was incubated with equilibrated beads during 1 hour at 4ºC. Then, column was emptied by gravity; aliquot 3 'flow through' was taken. Beads were washed with 7 ml of buffer B (50mM Tris-Cl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x PIC; aliquot 4 'first wash' was taken. Cells were washed again as before. Protein was eluted by adding 200 µl of Buffer C (50mM Tris-Cl pH 7.5, 300 mM NaCl, 200 mM imidazole, 1 mM β-ME and 0.1% NP40) supplemented with 1x PIC and incubating during 5 minutes with gentle shake. Elution was recovered and move to a new Eppendorf tube; aliquot 5 'first elution' was taken. A second elution was made under the same conditions; aliquot 6 'second elution' was taken. All aliquots were containing 30 µl of sample and 15 µl of 3x protein loading dye (PLD). Concentration of proteins elution were determined through Bradford Standard Curve with BSA. Samples were quick frozen with liquid nitrogen and store at -80ºC.

2.5.4 Purification of GST-tagged proteins in batch

Purification of GST-tagged proteins was performed as 'Purification of His_6-tagged proteins in batch' (section 2.5.2) but by using 40 µl of Glutathione Sepharose® 4B (*GE Healthcare Life Sciences*) beads per reaction instead of Ni-NTA beads. Samples were quick frozen with liquid nitrogen and store at -80ºC.

2.5.5 Kushnirov rapid protein extraction

This method was based on Kushnirov publication[343]. Yeast cells were grown in 3 ml YPDA media up to OD_{600} ~ 1. Then, 1 ml culture was pelleted at 14000 rpm for 1 minute and resuspended in 100 µl H20. 100 µl 0.2 M NaOH were added to the mix and incubated 5 minutes at room temperature. After, mixture was centrifugated again at 14000 rpm for 5 minutes. Pellet

was resuspended in 50 µl of 3x protein loading dye (PLD) and resolved by SDS-PAGE gel (section 2.5.6).

2.5.6 SDS-PAGE

Protein samples to be analyzed were previously stored in 3x Protein Loading Dye (Bromophenol blue 0.006%, 2-mercaptoethanol 15%, glycerol 30%, SDS 6%, Tris-HCl 0.187 M) at -80ºC. Samples were heated at 95ºC for 5 minutes and centrifuged at 14000 rpm for 5 minutes. Two main sizes of gels were used: small (8.5 x 6 cm) for western blot and protein purification and large (17 x 14 cm) for some protein purifications. Stacking phase gels contained 6% polyacrylamide (acrylamide/bis-acrylamide) in stacking buffer (0.5 M Tris-HCl, pH 6.8), whereas resolving phase gel contained 8-15% (depending on protein size) polyacrylamide in resolving buffer (0.5 M Tris-HCl, pH 6.8). Gels were run in 1x Tris-Glycine buffer. Stacking phase was run at 80 V, whereas resolving phase at 150 V (small gels) or 200 V (large gels). After, gels were stained either with Coomassie brilliant blue R-250 (Sigma Aldrich) as described below or stained with silver staining by using a SilverQuest kit (*Invitrogen LC6070*).

2.5.7 Coomassie staining

To visualize the protein bands, SDS-PAGE gels were first fixed for 1 hour or overnight with a solution containing 0.2% (w/v) Coomassie Brilliant Blue R-250 dissolved in ethanol and 45% (v/v) H_2O and 10% of acetic acid. Gels were distained overnight in a solution containing 20% ethanol and 10% acetic acid.

2.5.8 Silver staining

Silver staining was performed by using SilverQuest^{TM} Silver Staining Kit (Invitrogen DB, LC6070). Briefly, after gel electrophoresis gel was rinsed with ultra-pure H_2O. Protocol was carried out at room temperature with ultra-pure H_2O. Gel was fixed in 100 ml of fixative solution (40% ethanol, 10% acetic acid) for 20 minutes with gentle rotation. Solution was removed and gel was washed with 30% ethanol. After decanting, 100 ml of sensitizing solution (30 ml ethanol, 10 ml sensitizer, 60 ml H_2O) was added and incubated for 10 minutes. Liquid was decanted and gel was washed with 30% ethanol for 10 minutes and washed with H_2O for extra 10 minutes. Then gel was incubated in staining solution (1 ml stainer, 99 ml H_2O) for 15 minutes. Afterwards, the solution was removed, and the gel was rinsed with H_2O for 20-60 seconds. Then the gel was incubated with 100 ml of developing solution (10 ml developer, 1 drop developer enhancer, 90 ml H_2O) for 4 to 8 minutes until silver bands appear. Reaction was stopped by adding 10 ml of stopper solution and incubated for 10 min with gentle shaking. Gel was washed a final time with 100 ml of ultra-pure H_2O.

2.5.9 Western Blot

After SDS-PAGE, proteins were transferred from the polyacrylamide gel to a Protran nitrocellulose membrane (*Whatman 10401180*) in a wet/tank electroblotting system containing transfer buffer (25 mM Tris-Base, 20% ethanol and 190 mM glycine) at 80 V for 2 hours at 4°C. Then, membrane was rinsed in water and stained with Ponceau red to check the homogeneous protein transfer into the membrane. Membrane was rinsed again and incubated in blocking buffer (5% milk, 1x PBS, 0.3% Tween) during at least 1 hour. Membrane was washed 3 times with 1x PBS-Tween buffer and incubated with appropriated antibody for at least 1 hour (Table M3). Membrane was washed as previously described and incubated with second antibody (if required) for at least 1 hour. Membrane was washed as before and incubated for 1 minute with Immobilon® Crescendo Western HRP Substrate (Millipore) chemiluminescent reagent. Signals was detected by using Amersham Imager 600UV (*GE Healthcare Life Sciences*).

Table M3. List of antibodies for WB

Antibody/reagent	Against	Source	Ratio	Secondary Ab
PAP (Peroxydase anti-peroxydase)	Protein A of TAP-tag	Sigma	1:3000	None
HA.11 Clone 16B12 Monoclonal Antibody, Purified	HA-tag	Mouse, monoclonal Covance MMS-101P	1:2000	GAMPO (Goat anti-Mouse PerOxidase)
VSV	VSV-tag		1:1000	
AbBS8	Stm1	Rabbit, polyclonal	1:5000	GARPO (Goat anti-Rabbit PerOxidase)
GAMPO (Goat Anti-Mouse PerOxidase) IgG + IgM (H+L) Peroxidase-AffiniPure	mouse IgG + IgM secondary antibody	polyclonal Pierce 31460	1:5000	——
GARPO (Goat anti-Rabbit PerOxidase) IgG (H+L) Secondary Antibody, HRP conjugate	rabbit IgG + IgM secondary antibody	Jackson 115-035-068	1:5000	——

2.5.10 Mass spectrometry

Protein samples were run in a 12% acrylamide gel and stained with Coomassie blue. Gels were sent to the Proteomics Platform at the IGBMC where bands were cut and analyzed by nano LC-MS/MS.

2.5.11 *In vitro* translation

In vitro translation was done by using the NEBExpress® Cell-free *E. coli* Protein Synthesis System according to manufacturer's instructions. Before starting, DNA templates samples were subjected to PCI and ethanol precipitation, the DNA was digested and run in an agarose gel to check integrity and purity of the sample. All the components of the kit were thawed on ice; then, every reaction sample was mixed in a plastic tube containing: 12 µl NEBExpress S30 Synthesis

Extract, 25 µl Protein Synthesis Buffer, 1 µl T7 RNA Polymerase, 1 µl RNase Inhibitor Murine, 10µl H2O and 1µl 250 ng/µl of the DNA template. Reactions were incubated 37ºC with vigorous shaking for 4 hours. Translated proteins were analyzed in a SDS-PAGE gel (section 2.5.6).

2.6 RNA ANALYSES

2.6.1 Cold RNA *in vitro* transcription

Genes encoding *S. cerevisiae* snRNA U6, $tRNA^{Glu(UUC)}$, $tRNA^{Glu(CUC)}$, $tRNA^{Ser(UGA)}$, $tRNA^{Leu(UAG)}$, $tRNA^{Glu(UUC)}$ arm deletions (Δ acceptor stem, Δ D-arm, Δ anticodon arm or Δ TΨC-arm) or truncations (T-loop) were cloned into an expression plasmid. tRNA genes containing deletions (acceptor stem, anticodon arm, D-arm or TΨC-arm) were cloned in a TOPO vector. All tRNA genes were added with a T7 promoter in their 5' side plus a CCA and hammerhead ribozyme sequences in their 3' side. Plasmids encoding tRNA genes were expressed in *E. coli* MH1. Plasmid DNA was purified using NucleoSpin Easy Pure (Qiagen). Purified plasmid containing snRNA U6 was digested with XhoI in order to obtain a transcribed RNA with a length of 57nt. Plasmids containing tRNAs were digested with BbsI. All RNAs were *in vitro* transcribed with T7 RiboMAX Express Large Scale RNA Production System (Promega). Briefly, each reaction containing Transcription Optimized buffer, 10 µM DTT, 2.5 µM of ATP, GTP, CTP or UTP, 100 U RNasin, 40 U T7 RNA pol and 5 µg of digested plasmid was incubated 2.5 hours at 37ºC. Sample tubes were centrifuged at 14000 rpm for 10 minutes and supernatant subjected to a PCI extraction and ethanol precipitation. Transcripts were purified on an 8% polyacrylamide 7 M urea preparative gel. Gels were pre-run 20 min at 80 V and RNA samples were mixed with 2x RNA dye. RNA samples were heated at 95ºC for 3 minutes and loaded on the gel for an average migration of 2.5 hours at 20W. RNA bands were cut using UV shadowing and RNA samples were extracted from gel using the protocol described below (section 2.6.3). RNA was resuspended in water and stored at -20ºC.

2.6.2 Radiolabeled RNA *in vitro* transcription

Plasmid containing (snRNA U6, $tRNA^{Glu(UUC)}$, or $tRNA^{Glu(CUC)}$) were transcribed using the T7 RiboMAX Express Large Scale RNA Production System (Promega) as described in 'cold RNA *in vitro* transcription' (section 2.6.1), with the difference that to the transcription reaction was added CTP α^{32}P 20 mCi/ml 500-800 Ci/mmole. RNA transcripts were run on an 8% polyacrylamide 7 M urea gel. Gel was exposed for 15-25 minutes to a Kodak film for autoradiography and revealed in an autoradiography machine. Film was used as a template to detect and cut the radioactively labeled tRNA. RNA was extracted from gel as described below (section 2.6.3).

2.6.3 RNA extraction from preparative gel

300 µL of 1x PK buffer (100 mM Tris-Cl pH 7.5, 12.5 mM EDTA pH 8.0, 150 mM NaCl, 1% SDS) were added to each gel band containing RNA sample and incubated overnight. RNA was

eluted overnight in a Thermoblock at 24ºC and 450 rpm. Next day, supernatant was transfer to a new tube to be subjected to a PCI extraction and ethanol precipitation. RNA was precipitated with 7.5 M NH_4OAc and 2.5 volumes of 100% cold ethanol. RNA was resuspended in 50 µl of H_2O and 1 µl of it was used for Nanodrop measurement or was diluted in 500 µl of water to measure counts per minute (cpm) in a Cerenkov machine, for cold and radiolabeled RNA respectively.

2.6.4 Electrophoretic Mobility Shift Assay

Binding capacity of *S. cerevisiae* and *C. thermophilum* $Kti12_{CTD}$ to RNA was assayed using electrophoretic mobility shift assay (EMSA). Yeast RNAs (snRNA U6, $tRNA^{Glu(UUC)}$ or $tRNA^{Glu(CUC)}$) were radioactively transcribed *in vitro* in the presence of [α-^{32}P] CTP (section 2.6.2). $Kti12_{CTD}$ protein (residues 183-314) was purified and stored at -80ºC with 15% glycerol as described above (section 2.5.3). Different parameters and buffer conditions were tested to improve the binding and EMSA gel migration. The binding reactions for *S. cerevisiae* and *C. thermophilum* were performed during 30 min at 30ºC in 4 µl of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1000 cpm of radiolabeled RNA and increasing concentrations of $Kti12_{CTD}$ (0, 1, 3 ,10, 30, 100 µM). Samples were separated on a 6% polyacrylamide gel containing 0.5x TB buffer. After incubation, 1 µl of 5x loading dye (LD) was added to a final volume of 5 µl. Gel was run at 150 volts, 15 mA and 2 wats for 1 hour 20 minutes. Then gel was dried at 80ºC for one hour and exposed in a Phosphor Imager screen overnight. Signal was detected in a Typhoon FLA 7000 scanner at 100µm.

2.6.5 Electrophoretic Mobility Shift Assay competition

EMSA competition was used to assayed relative binding of Kti12 C-terminal domain to tRNA. Yeast $tRNA^{Glu(CUC)}$ was radioactively transcribed *in vitro* in the presence of [α-^{32}P] CTP (section 2.6.2). $Kti12_{CTD}$ protein (residues 183-314) was purified and stored at -80ºC with 15% glycerol as described above (section 2.5.3). Different parameters and buffer conditions were tested to improve the binding and EMSA gel migration. The binding reactions were performed during 30 min at 30ºC in 4 µl of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1000 cpm of radiolabeled RNA, 10 µM of $Kti12_{CTD}$ and increasing concentrations (0.2, 0.66, 2, 6.66, 20 µM) of cold-RNA *in vitro* transcribed (section 2.6.1). After incubation time, 1 µl of 5x loading dye (LD) was added to a final volume of 5 µl. Samples were separated on a 6% polyacrylamide gel containing 0.5x TB buffer. Gel was run at 150 volts, 15 mA and 2 wats for 1 hour 20 minutes. Then gel was dried at 80ºC for one hour and exposed in a Phosphor Imager screen overnight. Signal was detected in a Typhoon FLA 7000 scanner at 100µm.

2.6.6 Quantification of bands intensity

Intensity of bands obtained in EMSA assays as well as in co-immunoprecipitations were quantified using the ImageJ software. Briefly, squares containing the bands of interest were created with the 'rectangle' tool, to fix the rectangle size in the analyze menu was selected the

option: gels – select first lane. A copy of the rectangle was created and following the previous pathway the 'select next lane' option was selected. After adding all the rectangles to de desired bands. The option 'plot lanes' in the analyze-gels menu was selected to obtain the graphs of each band intensity. Every graph was 'closed by using the 'straight line' tool, and the area under the curve was obtained with the 'wand tracing' tool.

For EMSA, area of the bottom band represents free radiolabeled-$tRNA^{Glu(UUC)}$ and area of the top band represents complex of $Kti12_{CTD}$-radiolabeled-$tRNA^{Glu(UUC)}$; percentage of binding was calculated as follow:

- Percentage = (area top band/(area top band + area bottom band))*100.

For co-immunoprecipitations, correlation between Elp1-TAP and of Elp4-VSV or Kti14-HA_3 in which area of top band represents band intensity of Elp1-TAP, whereas area of bottom band represents band intensity of Elp4-VSV or Kti14-HA_3, was calculated as follow:

- $Ratio_{each\_sample}$ = area of top band / area of bottom band
- Percentage = ($ratio_{each\_sample}$ *100)/ $ratio_{Elp1\text{-}TAP}$

2.6.7 K_D determination

The dissociation constant (K_D) of $Kti12_{CTD}$ was calculated with percentages obtained after quantifications of bands obtained in EMSA (section 2.6.6) by using GraphPad Prism software. Briefly, an 'XY' data table was created with percentage values where 'X' was protein concentration (μM) and 'Y' percentage intensity of EMSA bands. Then data were analyzed with a nonlinear regression of a binding saturation equation and '*One site specific binding'* model using Least squares fit.

2.7 YEAST ANALYSES

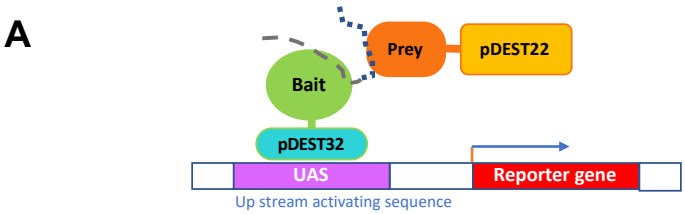
2.7.1 Yeast transformation

Precultures of desired yeast strains were diluted into 50 ml (enough for 8 transformations) of fresh YPDA media, calculating the volume needed to reach 1.0 OD_{600nm} next day with an average doubling time of 2h30. Cultures were incubated overnight at 37ºC and 170 rpm. Next day, cells were pelleted when reached 0.8-1.2 OD_{600nm} by centrifuging at 5000 rpm for 5 minutes at room temperature. Cells were then resuspended in 25 ml of sterile 10 mM Tris-Cl buffer pH 7.5 and immediately pelleted again as above. Yeast cells were resuspended in LiT (10 mM Tris-Cl pH 7.5, 100 mM LiOAc) buffer supplemented with 10 mM DTT and incubated at room temperature for 40 minutes with gentle shaking. After incubation, cells were pelleted as before and resuspended in 750 μl LiT buffer supplemented with 10 mM DTT. In 1.5 ml tubes, 100 μl of competent yeast were mixed with 1-2 μg of plasmid DNA, 5 μl of denatured salmon carrier DNA (10 mg/ml) and 50 μl of LiT buffer, followed by a 10 minutes incubation at room temperature. 300 μl of PEG solution (1:1 m/v PEG400 dissolved in LiT buffer) were added to each transformation and incubated for 10 minutes at room temperature. Then, transformations were incubated at 42°C for 15 minutes. Cells

were pelleted at 14000 rpm for 30 seconds, resuspended in 1 ml of YPDA medium and incubated at 30°C for 1 hour. Cells were pelleted at 14000 rpm for 60 seconds, resuspended in 200 µl 10 mM Tris pH 7.5 and plated 100µl on the appropriate media. Plates were incubated at 30ºC, unless other temperature was indicated.

2.7.2 Yeast two hybrid analysis

Two-hybrid yeast (Y2H) assay was performed with the ProQuestTM system by using the Gateway cloning strategy (Figure M4-A). A total of 24 constructs containing the coding sequences of *S. cerevisiae* Elp1, Elp2, Elp3, Elp4, Elp5, Elp6, Kti11, Kti12, Kti13, Kti14 proteins as well as a Pho92 and Not1 interaction as positive control (previously tested in the lab) were fused either to the pDEST32 vector containing the DNA-binding domain known as 'bait' or to the pDEST22 vector containing the activation-domain known as 'prey'. A total of 72 different reactions (36 in the 'bait→prey' and 36 in the 'prey→bait' direction) were tested (Figure M4-B). Plasmids carrying the fusion proteins were transformed into the *Saccharomyces cerevisiae* MAV203 strain. Transformed colonies were grown in Complete Supplemented Mixture (CSM) lacking tryptophan and leucine (-TRP, -LEU). Cells were incubated at 30ºC overnight to mid-log phase. 750 ml of culture were pelleted at 5000 rpm for 5 minutes and resuspended in 500 µl H_2O. The interaction of fusion proteins was monitored by luminescence using the luciferase beta-galactosidase dual reporter assay (*Promega*). Reactions were incubated at 30°C until color changed into bright yellow. Reaction was stopped by addition of 250 µl of 1M Na_2CO_3 and centrifuged for 5 minutes. OD_{420nm} was measured by using a Photospectrophotometer. Graphs were plotted with arbitrary units of activity (OD/luminescence).



B

	Background	Elongator subunits						KTI cofactors				Positive Control
	pDEST32	Elp1 pDEST32	Elp2 pDEST32	Elp3 pDEST32	Elp4 pDEST32	Elp5 pDEST32	Elp6 pDEST32	Kti11 pDEST32	Kti12 pDEST32	Kti13 pDEST32	Kti14 pDEST32	Pho92 pDEST32
pDEST22	X	X	X	X	X	X	X	X	X	X	X	X
Kti12 pDEST22	X	X	X	X	X	X	X	X	X	X	X	
Kti14 pDEST22	X	X	X	X	X	X	X	X	X	X	X	
A Not1 pDEST22	X											X
	pDEST22	Elp1 pDEST22	Elp2 pDEST22	Elp3 pDEST22	Elp4 pDEST22	Elp5 pDEST22	Elp6 pDEST22	Kti11 pDEST22	Kti12 pDEST22	Kti13 pDEST22	Kti14 pDEST22	Pho92 pDEST32
pDEST32	X	X	X	X	X	X	X	X	X	X	X	X
Kti12 pDEST232	X	X	X	X	X	X	X	X	X	X	X	
Kti14 pDEST32	X	X	X	X	X	X	X	X	X	X	X	
A Not1 pDEST32	X											X

Figure M4. Two-hybrid yeast assay ProQuestTM

(A) Scheme of the Gateway cloning strategy with the pDEST32 vector containing the bait (DNA-binding domain) and the pDEST22 vector containing the prey (activation domain). **(B)** Combination of the 26 different constructs made to test 72 different interactions between the Elongator subunits (Elp1-6), KTI cofactors (Kti11-14) and the positive (Pho92) and negative (Not1) controls.

2.7.3 Drop assay

Saccharomyces cerevisiae cells of tested strains (WT, ΔElp3, ΔKti12, $Kti12_{NTD}$, $Kti12_{CTD}$, $Kti12_{K14A}$ or $Kti12_{D85A}$) were grown in YPDA medium at 30ºC overnight. Next day, when cells reached mid-log phase (2-3 OD_{600nm}), cells were diluted in ultrapure autoclaved water to 0.1 OD_{600nm}. Then, three serial-dilutions of 10-fold (0.01, 0.001 and 0.0001) were performed diluting in water. 10µl of each dilution were plated in decreasing concentrations (from right to left) in the same row (Figure M5) in a plate containing the specific media (YPDA, CSM, CSM lacing adenine or CSM added with 120 mg/ml canavanine). Plates were incubated in the proper temperature (16, 25, 30 or 37ºC) during specific time.

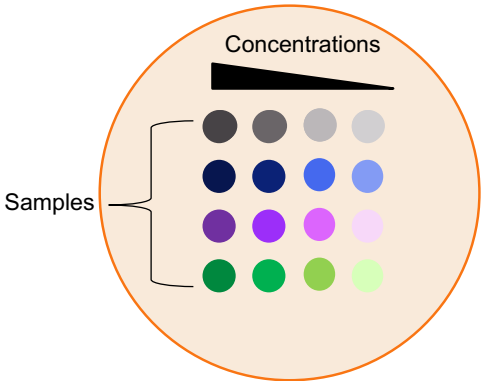


Figure M5. Disposition of cells for drop assay

Every line represents a different *S. cerevisiae* tested strain whereas every column represents a 10-fold serial dilution of yeast starting from the most concentrated one (left) at 0.1 OD_{600nm}.

2.7.4 Eclipse assay

Kluyveromyces lactis cells of the strains: AWJ137, a non-toxic (-) strain lacking the plasmid encoding zymocin toxin, and NK40, a toxic (+) strain containing the zymocin plasmid, were plated into YPDA plates and incubated at 30ºC for 2 days. In parallel, cells of target strains of *Saccharomyces cerevisiae* were grown in YPDA medium at 30ºC overnight. Next day, when cells reached mid-log phase (2-3 OD_{600nm}), *Saccharomyces cerevisiae* cells were diluted in ultrapure autoclaved water to 0.1 OD_{600nm}. Then, 10µl of each dilution were plated in a YPDA plate and let dry for 15 minutes. Once the drop was dried, a small amount of the proper *K. lactis* strain was added with a wood-stick to the top right part of each dried *S. cerevisiae* drop (Figure M6). Plates were incubated in at 30ºC for 2 days.

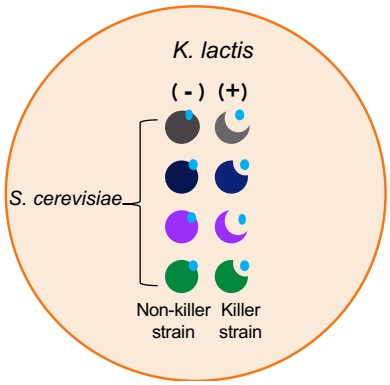


Figure M6. Disposition of cells for eclipse assay

Every line represents a different *S. cerevisiae* tested strain whereas the columns represent the different *K. lactis* strains added, minus sign (-) represents the column in which the non-killer *K. lactis* AWJ137 was added (top right blue circle). The plus sign (+) represents the column in which the killer *K. lactis* strain NK40 was added (top blue right circle).

2.7.5 Tandem Affinity Purification (TAP)

This protocol is based on the article published by Rigaut *et al.*[344]. Yeast cells were grown in 2 liters of complete supplement mixture lacking leucine (CSM -Leu) at 30ºC and 170 rpm until they reached 2 OD_{600nm}. Cells were pelleted at 5000 rpm for 20 minutes at 4°C, washed with 500 ml of cold water and spun as before. Cells were then resuspended in 20 ml H_2O, transferred to a 50 ml plastic tube and pelleted at 5000 rpm for 15 minutes at 4°C. Pellet was and frozen in liquid nitrogen and stored at -80ºC. All of the following steps were performed at 4°C unless otherwise indicated. Next day, cells were thawed and resuspended in 20 ml of buffer A (10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 0.5 mM DTT, 0.5 mM PMSF, 2 mM Benzamidine, 0.5 µg/ml (1 µM) Leupeptin, 1.4 µg/ml (2 µM) Pepstatin A, 2.4 µg/ml (4 µM) Chymostatin, 17 µg/ml (2.6 µM) Aprotinin 500 000KIE/ml). Cells were lysed with a High-Pressure Cell Disruptors (*Constant Systems*®) at 2 kBar and 4ºC, passing cells 2 times and moved to a new 50 ml tube. Cell lysate was supplemented with 2 ml of 2 M KCl to a final concentration of 0.2 M. Cell lysates were transferred to cold 45Ti tubes of an ultra-centrifuge and carefully balanced. Samples were centrifuged at 20.5 krpm for 30 minutes. Supernatant was transferred to a clean tube and centrifuged again for 1 hour 24 minutes. Supernatant was transferred to a clean 50 ml tube avoiding touching the upper lipidic part. Supernatant was then transferred into dialysis tube with a membrane MW cut-off 12-14 kDa, diameter 25 mm and incubated in buffer containing:10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol, 0.5 mM PMSF and 2 mM Benzamidine. Samples were incubated 3 hours in dialysis buffer, frozen in liquid nitrogen and stored at -80ºC. Next day, yeast extracts were thawed on ice. 200 µl IgG Sepharose (GE Healthcare) beads per sample (adding one extra aliquot to have enough material despites small pipetting errors) were washed by adding 5 ml of IPP150 buffer (10 mM Tris-HCl pH7.5, 150 mM NaCl, 0.1% NP-40) in a 15 ml tube and spun down, washes were repeated three times. After, yeast extract was put into a Poly-prep Chromatography column 0.8 x 4 cm and supplemented with 200 µl of washed beads. Columns were closed and rotated for 2 hours at 4ºC. Column were drained by gravity flow, then washed with 10 ml of IPP150 two times and an extra time with 10 ml

of TEV cleavage buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA). 10µl of 100 U TEV enzyme (Invitrogen 12575-015) were resuspended in 1 ml of TEV cleavage buffer and added to the column containing IgG beads. Beads were incubated in a rotating wheel for 2 hours at 16°C in the closed column. Eluate was collected, the dead volume was eluted by another 200 µl TEV cleavage buffer. To the eluate 3 volumes of calmodulin binding buffer and 3 volumes of 1 M $CaCl_2$ were added. It was then rotated with 200 µl calmodulin affinity resin (Agilent 214303), prewashed with 5 ml of calmodulin binding buffer, in a closed 10 ml column for 1 hour. The column was drained by gravity flow and washed with 30 ml of calmodulin binding buffer. Proteins were eluted by 5 consecutive additions of 200 µl calmodulin elution buffer to the resin. To verify if any protein was left bound to the IgG Sepharose or calmodulin affinity resin, additional elutions were performed using 200 µl 1% SDS. Protein loading dye (3x) was added to samples taken from cleared lysate, resuspended pellet, flow through, both washes and all elutions. Equal fractions of all stages were analyzed by western blot. The final elutions were concentrated by lyophilization and analyzed on a large SDS-PAGE.

2.7.6 TAP pull-down (First step of Tandem Affinity Purification)

This protocol was based on the article published by Rigaut *et al.*[344]. Five colonies of desired yeast were incubated in a 3 ml preculture of CSM (-Leu) medium and incubated overnight at 30ºC and 170 rpm. Next day, 200 ml CSM (-Leu) cultures were inoculated with preculture to reach the desired OD the next day based on estimated a doubling time of 2.5 hours. Cultures were incubated as previously described. When cells reached OD_{600nm} 1.0 they were pelleted at 5000 rpm for 20 minutes at 4°C, resuspended in 20 ml H_2O, transferred to a 50 ml falcon tube and centrifuged at 5000 rpm for 15 minutes at 4°C. The pellet was frozen in liquid nitrogen and stored at -80ºC. All of the following steps were performed at 4°C unless otherwise indicated. Pellets were thawed and resuspended in 0.6 ml of buffer A (10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 0.5 mM DTT, 0.5 mM PMSF, 2 mM Benzamidine, 0.5 µg/ml (1 µM) Leupeptin, 1.4 µg/ml (2 µM) Pepstatin A, 2.4 µg/ml (4 µM) Chymostatin, 17 µg/ml (2.6 µM) Aprotinin 500 000KIE/ml). Cells were transferred to a 15 ml glass Corex tubes and 600 µl of washed and sterile siliconized glass beads were added. Cells were lysed by 5 cycles of 30 seconds vortexing and keeping the samples on ice in between. Samples were centrifuged at 5000 rpm for 5 minutes. Supernatant was transferred to a 1.5 ml Eppendorf tubes and centrifuged twice at 14000 rpm for 20 minutes, transferring supernatant to a new Eppendorf tube after each centrifugation. Protein concentration of the extract was measured using Bradford assay and concentration of samples was adjusted to 15 mg/ml with buffer A. Adjusted extracts where aliquoted and frozen in liquid nitrogen and stored at -80ºC. When needed, yeast protein extracts were thawed on ice. 20µl of IgG beads Sepharose (GE Healthcare) per sample (adding one extra aliquot to have enough material despites small pipetting errors) were washed by adding 1 ml of buffer A, spinning down the tube and removing supernatant. The wash was repeated three times and beads where resuspended in 100 µl of buffer A per each original 20µl of beads. 300 µl of adjusted yeast extract were put in a new Eppendorf tube and mixed with 100 µl of washed IgG beads. Samples were incubated on a

rotating wheel for two hours at 4ºC. Western blot (WB) input aliquot was taken (20 µl of adjusted yeast protein extract plus 10 µl 3xPLD), frozen with liquid nitrogen and stored at -80ºC. After two hours incubation tubes were spun down and washed two times with IPP150 buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40) and supernatant was removed (for kinase assay see section 2.7.7, for co-immunoprecipitation continue with TEV cleavage). For co-immunoprecipitation, beads were washed a third time with TEV cleavage buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA) and added with 20 µl TEV cleavage buffer and 1 µl of 100U TEV enzyme (Invitrogen 12575-015). Samples were incubated two hours with gentle shaking at 16ºC. Supernatant was recovered in a new Eppendorf tube and supplemented with 3x PLD. WB pull-down aliquots were taken (20 µl of TEV cleavage sample plus 10 µl of 3x PLD). Samples were frozen in liquid nitrogen and stored at -80ºC. Samples were latter fractionated on an SDS-PAGE gel.

2.7.7 Kinase assay

For this assay, protein extract from *S. cerevisiae* cells was processed according to TAP pull-down (section 2.7.6). After washing two times with IPP150 buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40) the IgG were washed a third time with kinase buffer (10 mM Tris-HCL pH 7.5, 100 mM NaCl, 1.5 mM $MgCl_2$). Beads were resuspended with 24 µl of kinase buffer and WB pull-down aliquot was taken (4 µl of pull-down with 2 µl of 3x PLD). The following part was carried out in the special room to handle radioactivity. The remaining 20µl of pulldown were incubated with 100 ng of $tRNA^{Glu(UUC)}$ (only in samples to test phosphorylation of tRNA) and 300 µCi of [γ-^{32}P] ATP (3000 Ci/mmol) during 1 hour at 30ºC. Reaction was stopped by adding 10 µl of 3x PLD. Samples were stored at -20ºC. Next day, samples were heated at 95ºC for 5 minutes and loaded on a 10% acrylamide gel. For kinase samples 10 µl were loaded per well (30%) and for WB gels 3 µl (10 %) of input and 4.5 µl (15 %) of elution were loaded. Gels were run at 80 V stacking phase and 150 V resolving phase. Kinase gel was transfer to a Whatman paper and dried in a heat-vacuum drier for 45 minutes. Dried gel was exposed to a PhosphorImager screen overnight. Next day, signal in the screen was detected with a Typhoon 9500 FLA.

2.7.8 Co-immunoprecipitation

Plasmids containing Kti12-HA_3, HA_3-Kti14, Elp4-VSV were transformed into *S. cerevisiae* strains wild-type or Elp1-TAP according to the yeast transformation protocol (section 2.7.1). Protein yeast extract was obtained according TAP pull-down at (section 2.7.6). Pulldown samples were used for a WB (section 2.5.9).

3. RESULTS

3.1 Kti12 HAS SEQUENCE SIMILARITIES TO PSTK

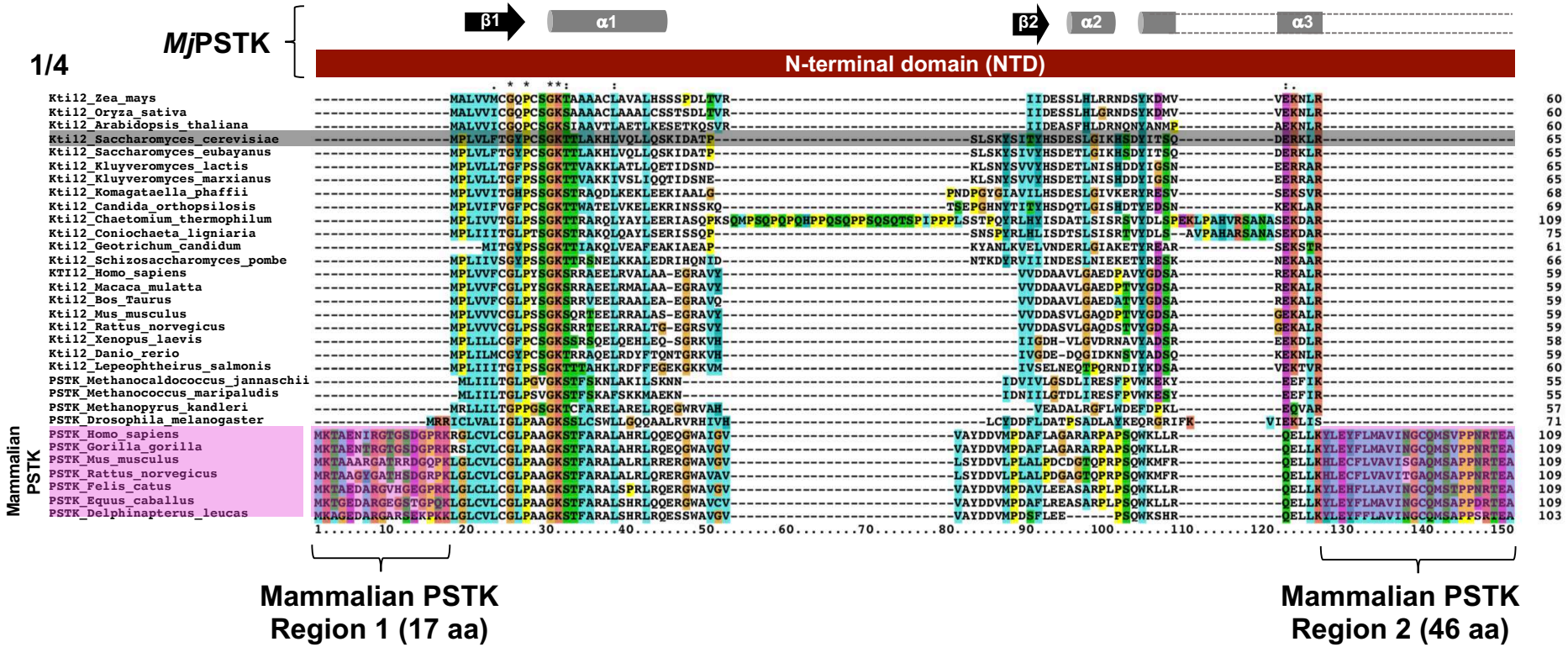
As was already mentioned, at the beginning of this project there was not much information about Kti12 and its similarity to PSTK. To investigate the potential function of Kti12 and its contribution to the Elongator-dependent tRNA modifications, I compared the *S. cerevisiae* Kti12 (*Sc*Kti12) protein sequence to various protein databases using the protein Basic Local Alignment Search Tool (BLAST). This identified numerous related proteins including Kti12 homologs. In particular, comparison of *Sc*Kti12 with proteins present in the Protein Data Bank (for which 3D structures are available) confirmed similarity to the PSTK protein from the thermophilic methanogenic archaea *Methanocaldococcus jannaschii.*

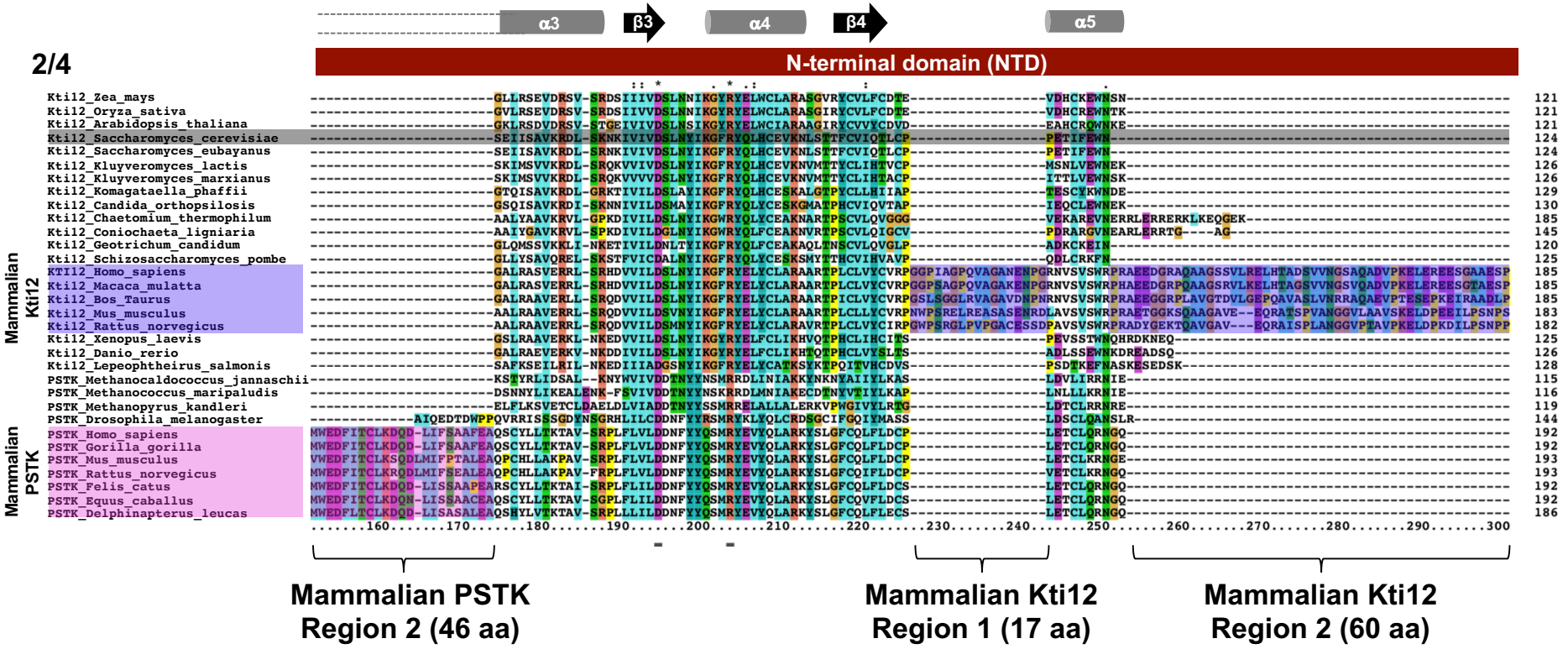
To gain a better understanding of the domain organization of the Kti12 protein, I performed a multiple sequence alignment comparing twenty-one Kti12 and eleven PSTK protein sequences from different prokaryotic and eukaryotic organisms using ClustalX. The alignment showed clear similarities between two proteins with very well-defined regions and highly conserved residues (Figure R1). A total of eight amino acids are fully conserved across all aligned sequences including putative catalytic residues in the kinase domain.

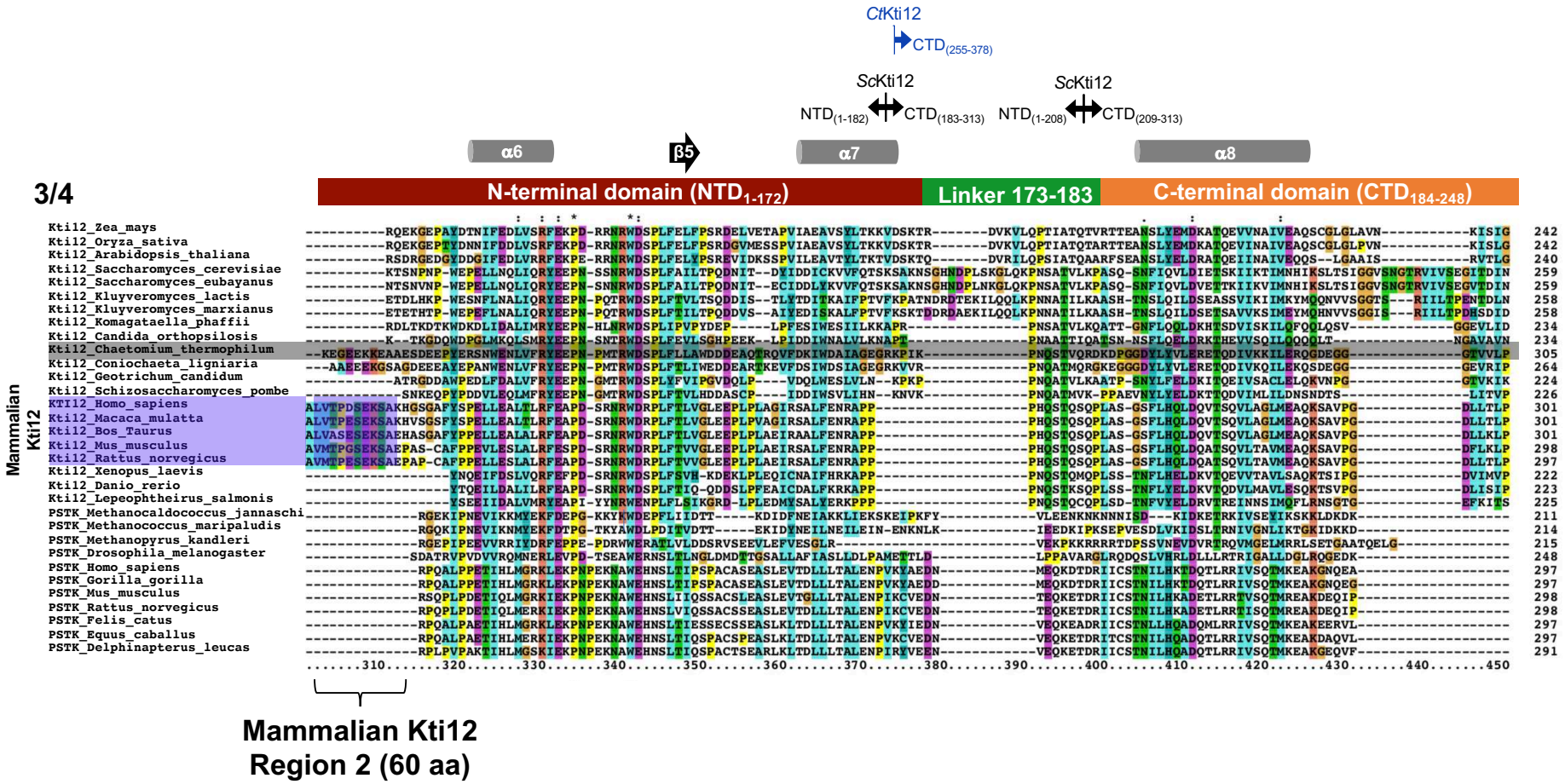
The sequence alignment also showed the presence or absence of specific motifs. Mammalian PSTK proteins contain two extra regions of amino acids compared to the rest of sequences. The first region, which is composed of 17 amino acids, is localized at the N-terminus the protein. The second region, containing 46 amino acids, is localized between the arginine 65 and the serine 66 residues (R65 and S66) of the *S. cerevisiae* Kti12 (*Sc*Kti12). In contrast, the archaeal PSTK do not contain these extended regions, being more similar to the yeast Kti12 (Figure R1, highlighted in pink).

Similarly, Kti12 sequences from mammals contain two extra amino acid regions compared to their *S. cerevisiae* homolog. The first one, composed of 17 amino acids, is localized between the proline 116 and the proline 117 (P116 and P117) residues of *Sc*Kti12. The second region containing 60 amino acids, is localized between asparagine 126 and lysine 127 (N126 and K127) of *Sc*Kti12 (Figure R1, highlighted in blue). Both extended regions are localized in the N-terminal kinase domain of the Kti12 and PSTK proteins according the *Methanocaldococcus jannaschii* PSTK sequence.

Using the *M. jannaschii* PSTK domains according to Chiba *et al.*[297] (Figure R1, N-terminal domain, linker and C-terminal domain represented in red, green and orange respectively) as 'template' we highlighted the alpha helices and beta sheets. We also used the *Mj*PSTK structure to delimit the borders of the *S. cerevisiae* Kti12 domains.


MjPSTK
β1
α1
β2
α2
α3
N-terminal domain (NTD)
Ktil2_Zea_mays
Ktil2_Oryza_sativa
Ktil2_Arabidopsis_thaliana
Ktil2_Saccharomyces_cerevisiae
Ktil2_Saccharomyces_eubayanus
Ktil2_Kluyveromyces_lactis
Ktil2_Kluyveromyces_marxianus
Ktil2_Komagataella_phaffii
Ktil2_Candida_orthopsilosis
Ktil2_Chaetomium_thermophilum
Ktil2_Coniochaeta_ligniaria
Ktil2_Geotrichum_candidum
Ktil2_Schizosaccharomyces_pombe
KTI12_Homo_sapiens
Ktil2_Macaca_mulatta
Ktil2_Bos_Taurus
Ktil2_Mus_musculus
Ktil2_Rattus_norvegicus
Ktil2_Xenopus_laevis
Ktil2_Danio_rerio
Ktil2_Lepeophtheirus_salmonis
PSTK_Methanocaldococcus_jannaschii
PSTK_Methanococcus_maripaludis
PSTK_Methanopyrus_kandleri
PSTK_Drosophila_melanogaster
Mammalian PSTK
PSTK_Homo_sapiens
PSTK_Gorilla_gorilla
PSTK_Mus_musculus
PSTK_Rattus_norvegicus
PSTK_Felis_catus
PSTK_Equus_caballus
PSTK_Delphinapterus_leucas
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150
Mammalian PSTK
Region 1 (17 aa)
Mammalian PSTK
Region 2 (46 aa)


α3
β3
α4
β4
α5
2/4
N-terminal domain (NTD)
Mammalian Kti12
Mammalian PSTK
Kti12_Zea_mays
Kti12_Oryza_sativa
Kti12_Arabidopsis_thaliana
Kti12_Saccharomyces_cerevisiae
Kti12_Saccharomyces_eubayanus
Kti12_Kluyveromyces_lactis
Kti12_Kluyveromyces_marxianus
Kti12_Komagataella_phaffii
Kti12_Candida_orthopsilosis
Kti12_Chaetomium_thermophilum
Kti12_Coniochaeta_ligniaria
Kti12_Geotrichum_candidum
Kti12_Schizosaccharomyces_pombe
KTI12_Homo_sapiens
Kti12_Macaca_mulatta
Kti12_Bos_Taurus
Kti12_Mus_musculus
Kti12_Rattus_norvegicus
Kti12_Xenopus_laevis
Kti12_Danio_rerio
Kti12_Lepeophtheirus_salmonis
PSTK_Methanocaldococcus_jannaschii
PSTK_Methanococcus_maripaludis
PSTK_Methanopyrus_kandleri
PSTK_Drosophila_melanogaster
PSTK_Homo_sapiens
PSTK_Gorilla_gorilla
PSTK_Mus_musculus
PSTK_Rattus_norvegicus
PSTK_Felis_catus
PSTK_Equus_caballus
PSTK_Delphinapterus_leucas
Mammalian PSTK Region 2 (46 aa)
Mammalian Kti12 Region 1 (17 aa)
Mammalian Kti12 Region 2 (60 aa)


CtKti12
CTD(255-378)
ScKti12
NTD(1-182)
CTD(183-313)
ScKti12
NTD(1-208)
CTD(209-313)
α6
β5
α7
α8
3/4
N-terminal domain (NTD1-172)
Linker 173-183
C-terminal domain (CTD184-248)
Mammalian Kti12
Kti12_Zea_mays
Kti12_Oryza_sativa
Kti12_Arabidopsis_thaliana
Kti12_Saccharomyces_cerevisiae
Kti12_Saccharomyces_eubayanus
Kti12_Kluyveromyces_lactis
Kti12_Kluyveromyces_marxianus
Kti12_Komagataella_phaffii
Kti12_Candida_orthopsilosis
Kti12_Chaetomium_thermophilum
Kti12_Coniochaeta_ligniaria
Kti12_Geotrichum_candidum
Kti12_Schizosaccharomyces_pombe
KTI12_Homo_sapiens
Kti12_Macaca_mulatta
Kti12_Bos_Taurus
Kti12_Mus_musculus
Kti12_Rattus_norvegicus
Kti12_Xenopus_laevis
Kti12_Danio_rerio
Kti12_Lepeophtheirus_salmonis
PSTK_Methanocaldococcus_jannaschi
PSTK_Methanococcus_maripaludis
PSTK_Methanopyrus_kandleri
PSTK_Drosophila_melanogaster
PSTK_Homo_sapiens
PSTK_Gorilla_gorilla
PSTK_Mus_musculus
PSTK_Rattus_norvegicus
PSTK_Felis_catus
PSTK_Equus_caballus
PSTK_Delphinapterus_leucas
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
Mammalian Kti12
Region 2 (60 aa)

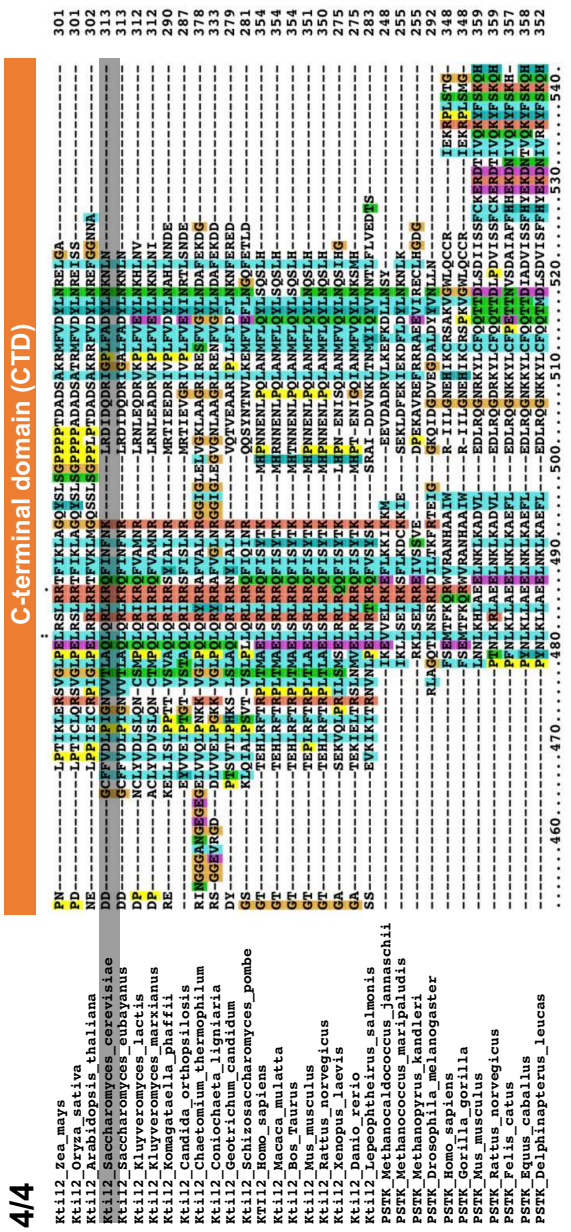


Figure R1. Multiple alignment of Kti12 and PSTK proteins

A multiple alignment of proteins from various organisms was produced with Clustal X (Materials and Methods section 2.1.2). Note that various alternatives exist for alignment of less conserved regions. The alignment reveals the presence of two extensions in the kinase domain of some proteins: the first extension is present in PSTK proteins from mammals (highlighted in pink) in comparison to the *S. cerevisiae* Kti12 (highlighted in grey). It is composed of two blocks: the first from the N-terminus contains 17 amino acids, whereas the second contains 46 residues. A second extension (highlighted in blue) concerns mammalian Kti12 proteins when compared to the *Sc*Kti12 (highlighted in black). Its first block contains 17 amino acids, whereas the second contains 60 residues. On the top of the alignment is presented a scheme of the protein organization with the N-terminal domain (red, residues 1-172), the linker (green, residues 173-183) and the C-terminal domain (blue, residues 184-248)[297] as well as beta sheets and alpha helixes based on the published data of *M. jannaschii* PSTK domain borders. The alpha helix 3 (α3) is indicated with dotted parts which represents inserted sequences in other organisms, and which may not be alpha helical. Part numbers of the sequence alignment are indicated in the top left.

Based on the sequence similarities of Kti12 and its distant related protein PSTK obtained by using multiple sequence alignments, we anticipated a structural and functional similarity between both proteins.

3.2 RNA BINDING PROPERTIES OF KTI12 C-TERMINAL DOMAIN

3.2.1 Protein purification

To elucidate the function of Kti12, the first step was to assess the activity of its two putative domains (Figure R2) by purifying recombinant *Saccharomyces cerevisiae* Kti12 protein expressed in *Escherichia coli*. I created constructs containing the Kti12 full-length protein containing residues 1-313 (Kti12) and truncated versions containing the N-terminal domain residues 1-182 ($Kti12_{NTD}$) or the C-terminal domain residues 183-313 ($Kti12_{CTD}$). Constructs were fused to a six-histidine tag (His_6) at their N-terminus.

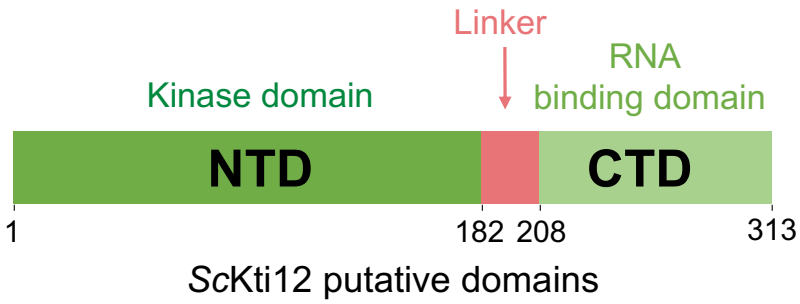


Figure R2. Putative domains of ScKti12
The full-length protein consisting of 314 amino acids is presented in green. The N-terminal domain is a putative kinase domain (dark green) and the C-terminal domain (light green) has putative RNA binding properties. They are separated by a short linker.

Plasmids containing protein constructs were expressed in *E. coli* cells and tagged proteins purified using nickel beads. Aliquots of different purifications steps were loaded on denaturing SDS-PAGE gels to evaluate the purity of the samples. (Materials and Methods section 2.5.2, Figure R3). $Kti12_{CTD}$ showed one major band on the SDS-PAGE gel and the yield was around 6 mg/ml. Denaturing polyacrylamide gel analyses of fractions recovered at different steps of the protein purification process showed the presence Kti12 and $Kti12_{NTD}$ in the total protein fraction but the proteins were lost in the lysate fraction during the centrifugation step, suggesting that these recombinant proteins were insoluble possibly as a result of protein aggregation. Thus, neither *S. cerevisiae* Kti12 nor $Kti12_{NTD}$ were obtained.

Given this difficulty, I then tried two different strategies to express and purify $Kti12_{NTD}$ and Kti12. On the one hand, for the N-terminal domain, I planned to test if increasing the length of the $Kti12_{NTD}$ might allow the protein to fold in a more stable way and facilitate its expression and purification. On the other hand, for the complete proteins, I wanted to test whether, due to their subtle differences in protein sequences, homologous proteins from other organisms might have different properties that may facilitate their expression and purification. In particular, due to their adaptation to high temperatures, proteins from thermophilic organisms are often more stable,

facilitating their production and study[345]. Thus, I made a new construct to express the N-terminal domain of Kti12, increasing its length from 1-182 ($Kti12_{NTD(1-182)}$) to 1-208 ($Kti12_{NTD(1-208)}$) amino acids. The corresponding C-terminal domain construct was also prepared ($Kti12_{NTD(209-313)}$). Similar to the previous constructs $Kti12_{NTD(1-182)}$ and $Kti12_{CTD(183-313)}$, the new ones $Kti12_{NTD(1-208)}$ and $Kti12_{NTD(209-313)}$ were selected according to a sequence alignment using *Mj*PSTK structure[297] to extrapolate the approximate domains of Kti12, based on the Kti12-PSTK sequence similarity.

Also, I prepared three Kti12 constructs by using DNA sequences of different organisms: the budding yeast *Kluyveromyces lactis,* the fission yeast *Schizosaccharomyces pombe* and the thermophilic fungus *Chaetomium thermophilum*. Plasmids constructs were built with a His_6-tag in the N-terminal end of the fusion proteins, transformed into *E. coli* and expressed and purified in batch (Materials and Methods section 2.5.2). The extended version of the N-terminal domain did not contribute to improve the protein purification. In contrast, the C-terminal domain showed a major band on the SDS-PAGE gel even after its shortening (Figure R3-A). Constructs using Kti12 homologs did not solve the solubility properties of the complete protein (Figure R3-B).

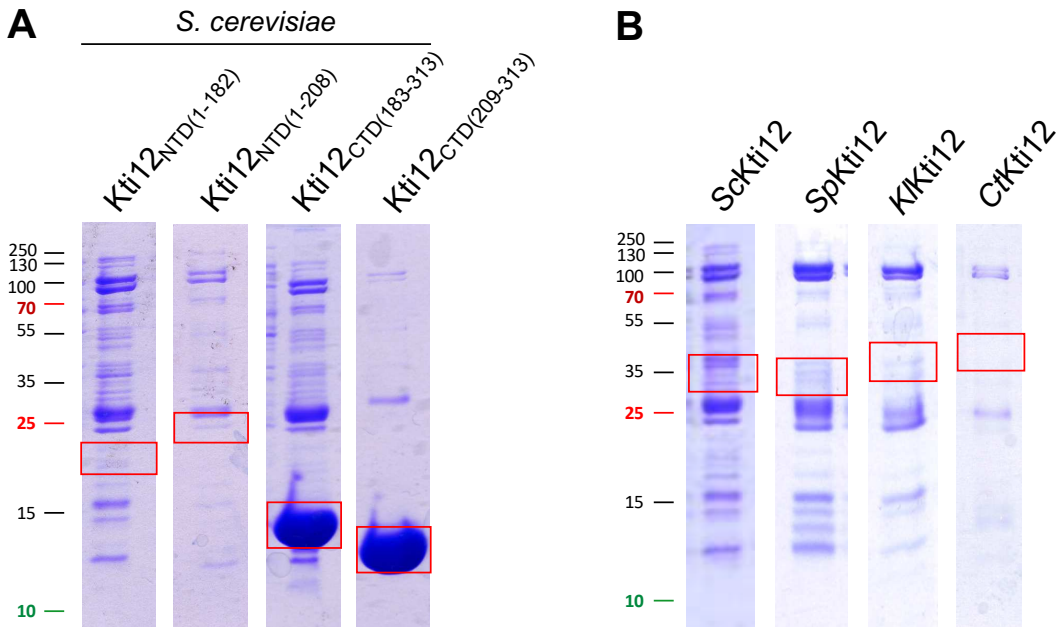


Figure R3. His_6-Kti12 purification

Protein profiles of the final elution fractions after purification are shown. 15% denaturing polyacrylamide gels were used. **(A)** Results for different constructs containing the NTD and CTD of *S. cerevisiae* Kti12. **(B)** Kti12 full-length homolog proteins form different organisms: *S. cerevisiae* (*Sc*Kti12) 36.15 kDa, *S. pombe* (*Sp*Kti12) 33.32 kDa, *K. lactis* (*Kl*Kti12) 39.79 kDa or *C. thermophilum* (*Ct*Kti12) 43.24 kDa. Red squares marked the size of the expected protein according the calculated molecular weight. Purifications shown were analyzed on different polyacrylamide gels. The complete gels of each purification can be found in the Appendix A1.

Given that these first optimization attempts still did not allow protein production, I tested different conditions of expression and purification to obtain recombinant Kti12 and $Kti12_{NTD}$ for *in vitro* analysis. Those are summarized in table R1.

Table R1. Tested conditions for recombinant protein expression of Kti12 and its domains

Variables	Tested conditions
Purification tag	His_6 or GST fused to the N-terminal side
E. coli competent strain for expression	BL21-CodonPlus, Rosetta (DE3)
Buffer components	DTT, β-met, NP40, $MgCl_2$
Protein induction	Autoinduction, IPTG (0.1-1 mM)
Growing media	Lysogeny broth (LB), Terrific Broth (Auto-Induction medium)
Growing temperature	16,18, 20, 25, 30, 37ºC
Incubation time after induction	16-20h
Co-expression to stabilize protein	Kti12-Kti14 and combination of their domains

Among the different constructs tested for the full-length protein, the one made by fusing the GST tag to the N-terminal side of Kti12 and purified with glutathione beads (Materials and Methods section 2.5.4) showed the presence in the elution of a clear band obtained with relatively high yield. Indeed, the profile of the purified elution fraction (Figure R4-A) revealed a band around 61 kDa which was identified as GST-Kti12 by mass spectrometry.

Unfortunately, cleavage of the GST tag led to protein degradation (Figure R4-B). Therefore, it was not possible to use it for further analyses.

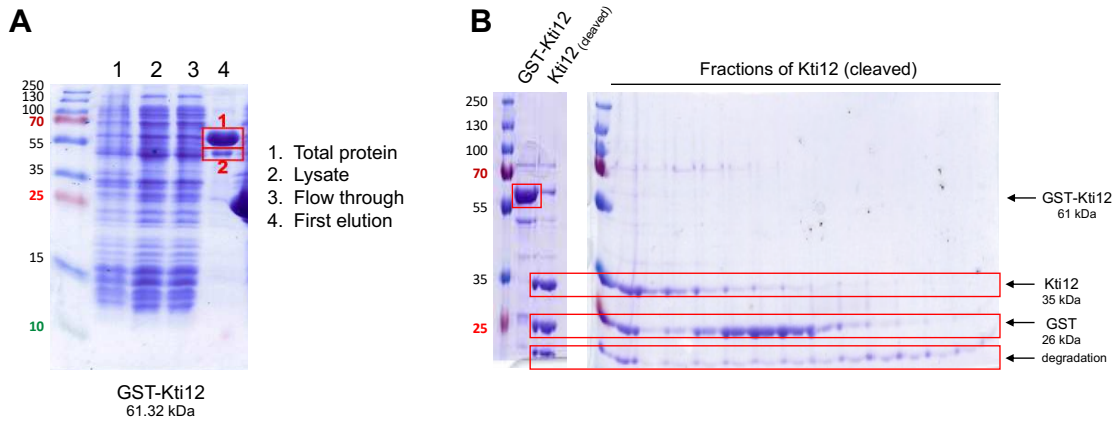


Figure R4. Purification of GST-Kti12

(A) SDS-PAGE gel analysis of f *S. cerevisiae* GST-Kti12 purified using glutathione beads. Each lane represents one step of the protein purification process as indicated on the right of the gel. 5µl of each aliquot and 5% of the first elution from a 50 ml culture, were loaded on a 15% polyacrylamide SDS-PAGE. Band marked with a red box were sent to mass spectrometry. Those were shown to contain GST-Kti12 with band 1 likely containing the full-length protein and band 2 a slightly degraded form **(B)** Ni-NTA purified GST-Kti12 (61.32 kDa) was digested with TEV protease to remove the tag and the resulting proteins fractionated by gel filtration. The different fractions were analyzed by SDS-PAGE. The left part shows the protein GST-Kti12 (61.32 kDa) before and after TEV cleavage, the right part the various gel filtration fractions. The resulting Kti12 should be 35.32 kDa. 12% denaturing polyacrylamide gel were used, and gels were stained with Coomassie blue.

Unfortunately, none of the other tested conditions allowed the purification of full length Kti12 nor $Kti12_{NTD}$. In contrast, $Kti12_{CTD}$ was well expressed and showed high purity on an SDS-PAGE gel.

To confirm the putative RNA binding properties of the C-terminal domain of Kti12, I purified a large amount of *Sc*Kti12$_{CTD}$ and *Ct*Kti12$_{CTD}$ (Materials and Methods section 2.5.3). Figure R5 shows an analysis on a polyacrylamide gel of the final profile of the purified proteins as well as fractions collected at different steps of the purification process. 10% glycerol was added to the purified proteins and aliquots were stored at -80ºC.

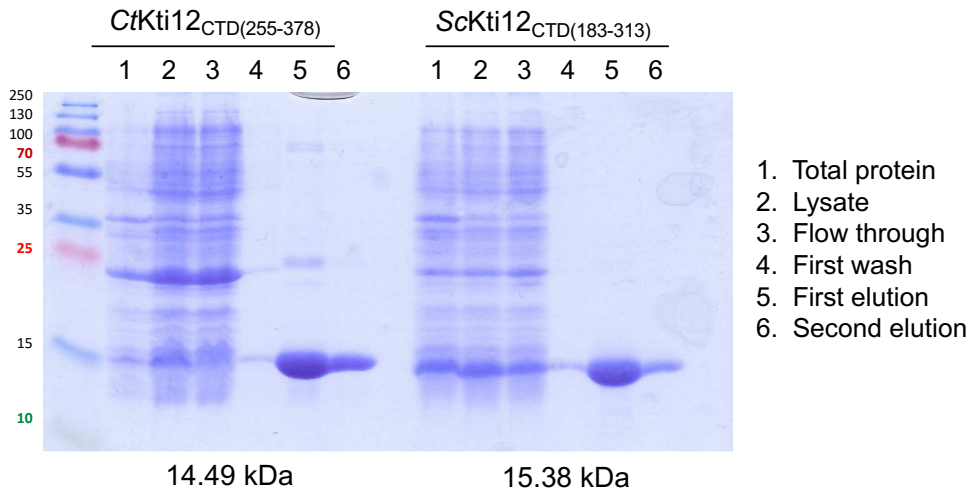


Figure R5. Purification of Kti12$_{CTD}$

Protein profiles of purified Kti12$_{CTD}$ proteins of *S. cerevisiae* (*Sc*Kti12$_{CTD}$) or *C. thermophilum* (*Ct*Kti12$_{CTD}$). Proteins were purified using Ni-NTA beads. Each lane represents a step of the protein purification as indicated on the right. Aliquots were loaded on a 15% polyacrylamide SDS-PAGE. After migration, the gel was stained with Coomassie blue. The concentration in the first elutions is around 5 mg/ml.

3.2.1.1 Purification of recombinant Kti12 and Kti12$_{NTD}$ proteins

Despite multiple unsuccessful attempts to produce the *S. cerevisiae* full-length Kti12 and Kti12$_{NTD}$ recombinant proteins, I continued trying to express and purify various constructs. Only near the end of my PhD, I succeeded to obtain soluble forms of Kti12 and Kti12$_{NTD}$ with a six-histidine (His$_6$) tag at the C-terminus instead of the N-terminus that exhibited a high purity on a SDS-PAGE. I used the His$_6$-Kti12$_{CTD}$ construct (containing the His$_6$-tag in the N-terminal side) as a control (Figure R6-A). Plasmids encoding protein constructs were transformed in *E. coli* cells and protein expression was induced with IPTG before purification in batch on Ni-NTA beads (Material and Methods section 2.5.2). The first elution fraction of the purification was loaded on a denaturing SDS-PAGE gel (Figure R6-B).

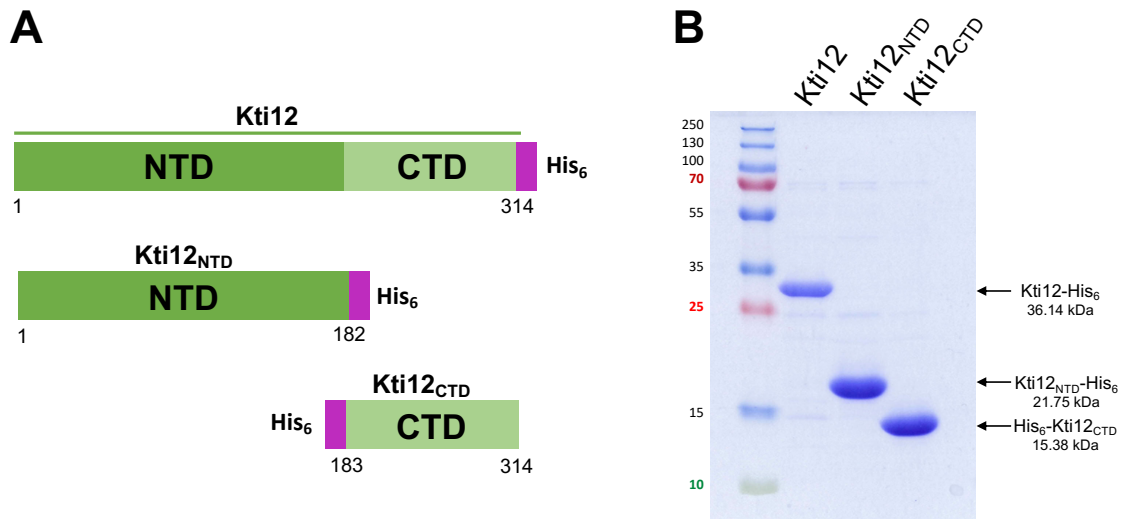
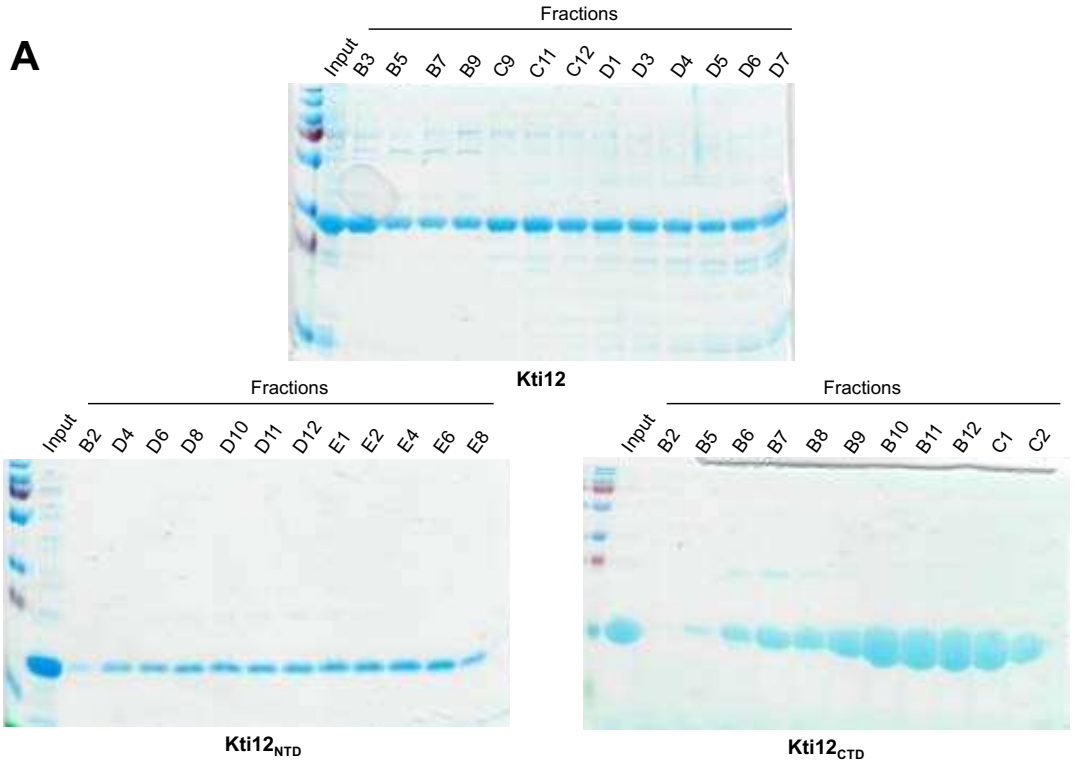


Figure R6. Purification of Kti12 and Kti12$_{NTD}$ by adding C-terminus His$_6$-tag
(A) Organization of expression constructs for *S. cerevisiae* Kti12-His$_6$, Kti12$_{NTD}$-His$_6$ or His$_6$-Kti12$_{CTD}$. **(B)** SDS-PAGE gel analysis of purified proteins after Ni-NTA affinity purification using a 15% denaturing polyacrylamide gel stained with Coomassie blue.

Kti12-His$_6$, Kti12$_{NTD}$-His$_6$ or His$_6$-Kti12$_{CTD}$ proteins purified on Ni-NTA were further purified by gel filtration.

Gel filtration analysis revealed the presence of multiple peaks of high molecular weights indicative of the presence of protein aggregates for His$_6$-Kti12$_{CTD}$ (Figure R7-B). Gel filtration analysis of Kti12-His$_6$ and Kti12$_{NTD}$ -His$_6$ showed a constant signal along all fractions (Figure R7-A).

Analysis of the gel filtration fractions by SDS-page revealed a peak for His$_6$-Kti12$_{CTD}$, but Kti12-His$_6$ and Kti12$_{NTD}$-His$_6$ showed a constant signal across all fractions (Figure R7-A). Analyses of all chromatograms (Figure R7-B, His$_6$-Kti12$_{CTD}$ taken as example), revealed the presence of multiple peaks of high molecular weights indicative of the presence of protein aggregates.



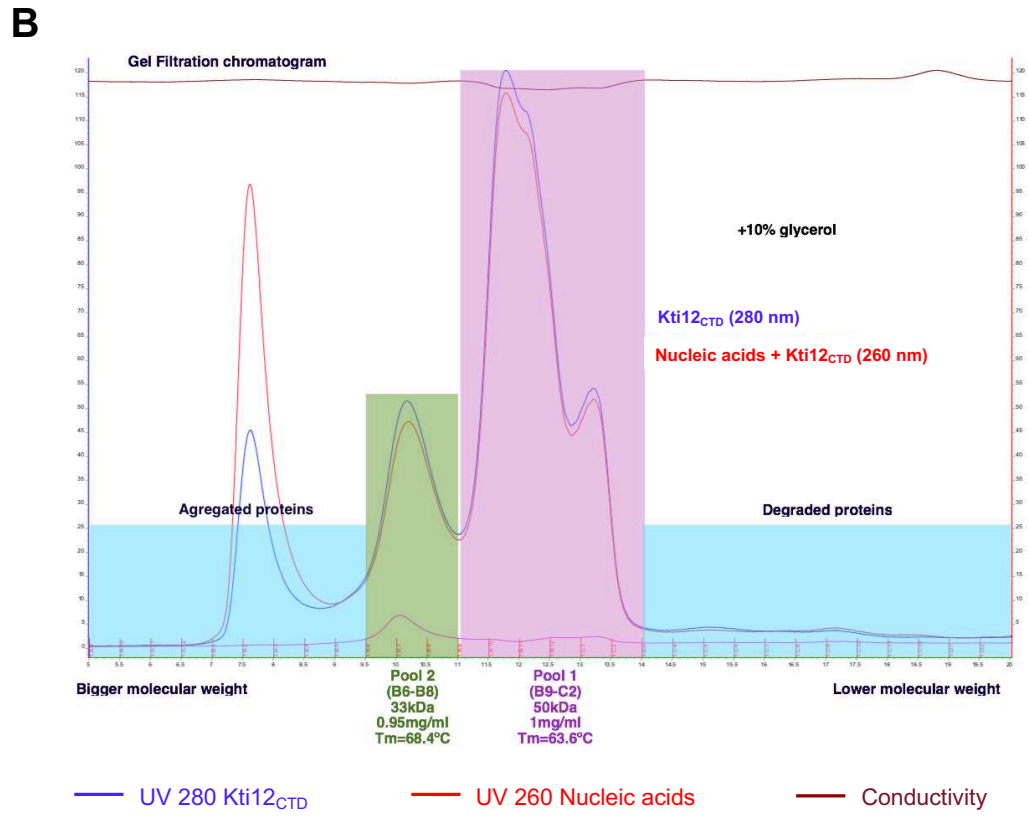


Figure R7. Purification of Kti12-His$_6$, Kti12$_{NTD}$-His$_6$, and His$_6$-Kti12$_{CTD}$ by gel filtration

(A) SDS-PAGE gels of Kti12, Kti12$_{NTD}$, and Kti12$_{CTD}$ after size exclusion chromatography (12%, 12% and 15% denaturing polyacrylamide gels were used respectively). Gels were stained with Coomassie blue. **(B)** Chromatogram of gel filtration of Kti12$_{CTD}$. Chromatograms of full-length Kti12 and the N-terminal domain are in the Appendix A2.

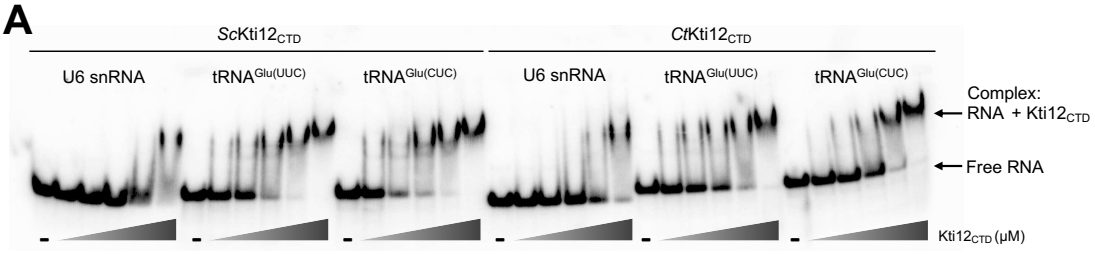
After optimization, purification of these recombinant proteins may open new possibilities to perform experiments to study the function of Kti12. In the next sections I will present the experiments performed with the Kti12 C-terminal domain.

3.2.2 The C-terminal domain of Kti12 binds RNA

To investigate the potential function of the C-terminal domain (CTD) of Kti12 (Kti12$_{CTD}$) in RNA-binding, I used an electrophoretic mobility shift assay (EMSA). As these experiments were done at the early stage of my PhD work, I used the Kti12$_{CTD}$ purified in batch on Ni-NTA (Material and Methods section 2.5.2). I incubated *S. cerevisiae* Kti12$_{CTD}$ (*Sc*Kti12$_{CTD}$) or *C. thermophilum* Kti12$_{CTD}$ (*Ct*Kti12$_{CTD}$) ranging from 0 to 100 µM concentration of protein, with ^{32}P-labelled *in vitro* transcribed U6 small nuclear RNA (U6 snRNA) or tRNA$^{Glu(UUC)}$, or tRNA$^{Glu(CUC)}$ and incubated for 30 minutes at 30 ºC (Materials and Methods section 2.6.4). Reaction mixtures were then loaded on a native electrophoresis gel (Figure R8-A). Increasing Kti12$_{CTD}$ concentrations led to increasing amounts of slow migrating bands or "shifted" RNA for all the tested RNAs, indicating the formation of RNA-protein complexes. Proteins from both organisms better interacted with tRNAs than with U6 snRNA: with tRNA a complex (shifted band) was detected at and above 1 µM of protein whereas for U6 snRNA this required at least 30 µM of protein. *Sc*Kti12$_{CTD}$ shifted almost all the input tRNA at 30 µM concentration, whereas 100 µM or more of *Ct*Kti12$_{CTD}$ was needed to produce the same effect. To gain a better understanding of the tRNA–protein binding properties, I determined the 'apparent' dissociation constant (K_D) of the different complexes by using values from band intensity quantification of the EMSA (Figure R8-B) (Materials and Methods section 2.6.7). These data showed that the *Sc*Kti12$_{CTD}$•tRNA$^{Glu(UUC)}$ complex has a slightly higher K_D (7.1 µM) compared to the one with tRNA$^{Glu(CUC)}$ (3.7 µM). In contrast, *Ct*Kti12$_{CTD}$ had a lower K_D for tRNA$^{Glu(UUC)}$ (8.1 µM) compared to tRNA$^{Glu(CUC)}$ (20.1 µM). In general, the 'apparent' K_D observed for *S. cerevisiae* protein are lower than for its *C. thermophilum* homolog. However, these values should be taken cautiously as more assays, possibly with more data points, should be performed to obtain precise values (thus the term 'apparent' K_D). This suggests that the *S. cerevisiae* protein may have a slightly higher affinity to tRNA than the *C. thermophilum* protein, but additional experiments would be necessary to ascertain this point. Also, the *C. thermophilum* protein may require higher temperature for its full activity. Most importantly, the Kti12 C-terminal domains of both S. cerevisiae and *C. thermophilum* can bind RNA and both proteins appear to bind tRNA better compared to U6 snRNA used as a control.

Due to the degeneracy of the genetic code, tRNAs displaying both UUC and CUC anticodons encode glutamate; however, only the tRNA with the UUC anticodon is modified by Elongator. In my EMSA assay, tRNA$^{Glu(UUC)}$ which is modified by Elongator *in vivo*, appears to bind Kti12$_{CTD}$ similarly to tRNA$^{Glu(CUC)}$ which is not modified by Elongator (Figure R8-A). This indicates that Kti12$_{CTD}$ does not display binding specificity for tRNAs that are modified by Elongator. This is consistent with the wide number of Elongator natural substrates and the fact that mutant tRNA suppressors, which have acquired a non-natural U at their wobble position (e.g., *SUP4*), get modified.

A



B

Determination of apparent K_D

	*Sc*Kti12$_{CTD}$			*Ct*Kti12$_{CTD}$		
	U6 snRNA	**tRNA$^{Glu(UUC)}$**	**tRNA$^{Glu(CUC)}$**	**U6 snRNA**	**tRNA$^{Glu(UUC)}$**	**tRNA$^{Glu(CUC)}$**
K_D (μM)	N/D	7.1	3.7	N/D	8.1	20.1
R^2	N/D	0.97	0.98	N/D	0.95	0.96

Figure R8. The Kti12 C-terminal domain binds RNA

(A) EMSA assay comparing the RNA-binding activities of the C-terminal domains of Kti12 from *S. cerevisiae* (*Sc*Kti12$_{CTD}$) and *C. thermophilum* (*Ct*Kti12$_{CTD}$). The indicator is migration shift of radiolabeled RNA (snRNA, tRNA$^{Glu(UUC)}$ or tRNA$^{Glu(CUC)}$) upon electrophoresis in a 6% polyacrylamide native gel. In the absence of RNA (sign '-'), free RNA migrates as a single band (bottom bands). Increasing protein concentrations (0, 1, 3, 10, 30, 100 μM), indicated by the black triangles, resulted in increasing amounts of RNA–protein complex (upper bands) (n=1). **(B)** Determination of apparent K_D of the different tested *S. cerevisiae* or *C. thermophilum* proteins. Values of dissociation constant (K_D) and coefficient of determination (R^2) are shown. N/D means not determined. Made with GraphPad Prism (Materials and Methods section 2.6.7).

RNA-binding properties of the C-terminal domain of *S. cerevisiae* Kti12 differing in length (*Sc*Kti12$_{CTD(183-313)}$ or *Sc*Kti12$_{CTD(209-313)}$) were also compared by EMSA. Shifts were detected in the presence of both proteins, however, at vastly different concentrations, i.e. 100 μM of *Sc*Kti12$_{CTD(209-313)}$ was required to see effect similar to 1 μM of *Sc*Kti12$_{CTD(183-313)}$ (data not shown). This may suggest that the shorter form is incomplete and/or poorly folded. Thus, for the following experiments investigating the RNA-binding properties of Kti12 C-terminal domain, I used the construct containing residues 183 to 313 of *S. cerevisiae* Kti12 (*Sc*Kti12$_{CTD(183-313)}$) and based on its 'apparent' K_D, I used a constant protein concentration of 10 μM.

3.2.3 tRNAs compete for binding to Kti12 C-terminal domain

To assess more accurately the relative binding of the C-terminal domain of Kti12 to tRNA, I performed an electrophoretic mobility shift (EMSA) competition assay. A constant concentration of *Saccharomyces cerevisiae* Kti12$_{CTD}$ (*Sc*Kti12$_{CTD}$) was incubated with a given amount of ^{32}P-labelled *in vitro* transcribed tRNA$^{Glu(UUC)}$ in the presence of increasing concentrations (0-20 μM) of unlabeled ('cold') *in vitro* transcribed U6 snRNA, tRNA$^{Glu(UUC)}$, tRNA$^{Ser(UGA)}$ or tRNA$^{Leu(UAG)}$ (Materials and Methods section 2.6.5). Complexes present in reaction mixtures were separated by native gel electrophoresis (Figure R9-A). All three competitor tRNAs successfully competed with radiolabeled-tRNA$^{Glu(UUC)}$ for binding Kti12$_{CTD}$ with almost complete displacement of

radiolabeled-tRNA at the highest concentration (Figure R9-C). Whereas the control U6 snRNA was able to displace only a fraction of radiolabeled-$tRNA^{Glu(UUC)}$ at 20 μM concentration.

Quantification of the band intensities from three different gels (Figure R9-B) (Materials and Methods section 2.6.6) showed only a modest degree of competition by U6 snRNA. For the tRNA competitors ($tRNA^{Glu(UUC)}$, $tRNA^{Ser(UGA)}$ and $tRNA^{Leu(UAG)}$) sigmoidal-like curves with negative slope indicative of a dose-response pattern are observed. From the different tRNA tested, $tRNA^{Leu(UAG)}$ was the best competitor followed by $tRNA^{Ser(UGA)}$ and then $tRNA^{Glu(UUC)}$. However, these differences should be interpreted with caution because they may be affected by minor errors in cold-tRNA quantification and by the proportion of folded/unfolded tRNA in each preparation which may itself depend upon intrinsic stability of the different tRNA species. The presence of multiple bands in the positive control as well as in some of the tRNA competitors, might be also explained by the proportion of folded/unfolded tRNA; however, this does not exclude the possibility of different protein conformations or protein aggregation.

Importantly, this reiterates the observation that the C-terminal domain of Kti12 does not discriminate tRNAs that are modified by Elongator ($tRNA^{Glu(UUC)}$ and $tRNA^{Ser(UGA)}$) from the ones that are not ($tRNA^{Leu(UAG)}$), confirming the results obtained in the first EMSA analysis with $tRNA^{Glu(UUC)}$ and $tRNA^{Glu(CUC)}$ (Figure R8).

Another fact that stood out was that the band corresponding to complex formation has higher intensity and is better defined in the presence of U6 snRNA competitor compared to the tRNA competitors. This effect was even visible at the lowest concentration of RNA competitor (0.2 μM) and in the positive control which lacked RNA competitor. A possible explanation for this is the competition effect in which in the presence of competitors ($tRNA^{Glu(UUC)}$, $tRNA^{Ser(UGA)}$ or $tRNA^{Leu(UAG)}$) complex formation is dynamic and the complex exists in a binding-unbinding state resulting diffused migration. However, this does not explain why, in the absence of competitor (positive control), the band is also fuzzy. Further analyses will be required to explain this phenomenon.

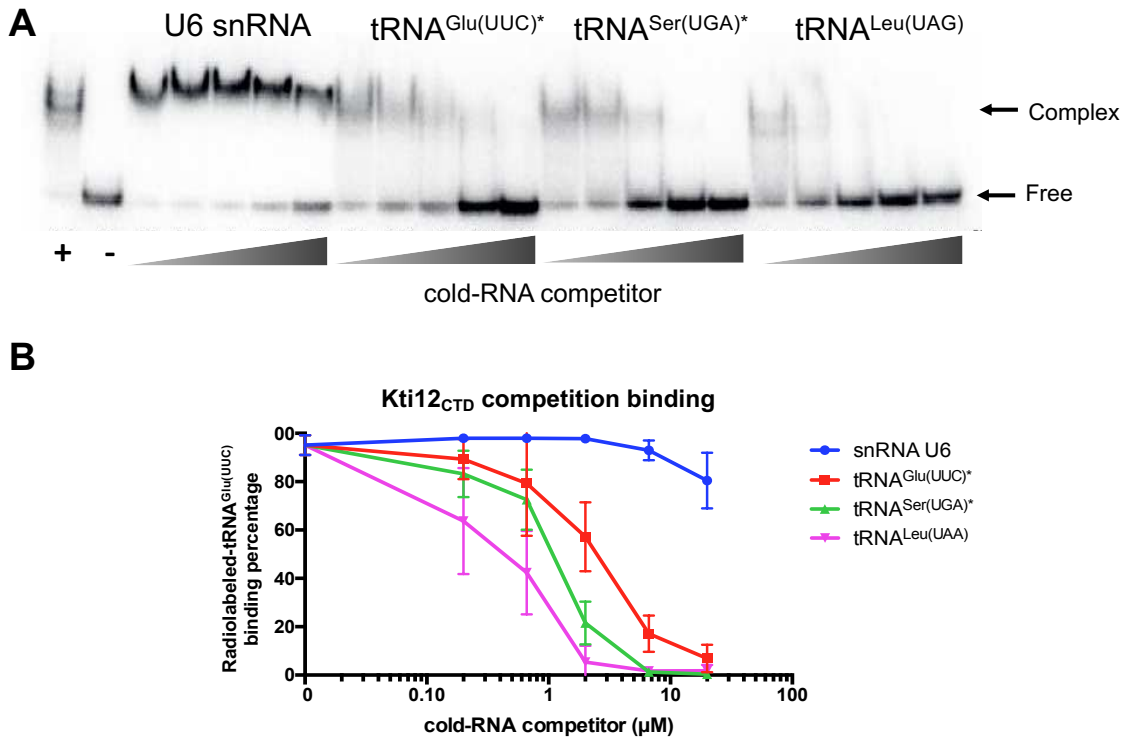


Figure R9. Kti12 C-terminal domain competition binding of tRNAs

(A) EMSA competition comparing the RNA binding activity of *S. cerevisiae* Kti12$_{CTD}$ to different RNAs (U6 snRNA, tRNA$^{Glu(UUC)}$, tRNA$^{Ser(UGA)}$ or tRNA$^{Leu(UAG)}$). In the absence of protein (sign '-'), free radiolabeled RNA migrates as a single band (bottom bands), whereas the presence of 10 µM of protein (sign '+') leads to formation of RNA–protein complex (upper bands); increasing concentrations of unlabeled 'cold-RNA' (0.2, 0.66, 2, 6.66, 20 µM) displace radiolabeled-RNA from the complex to the bottom bands. tRNAs modified by Elongator, are indicated with an asterisk. Electrophoresis was performed in 6% native polyacrylamide gel **(B)** Quantification of band intensity of the EMSA represented in a graph created by using ImageJ (n=3) (Materials and Methods section 2.6.6).

3.2.4 Lack of arms in tRNA affect Kti12 C-terminal binding affinity

To determine which parts of substrate tRNA were being recognized by the Kti12 C-terminal domain (Kti12$_{CTD}$), I tested different tRNA constructs derived from tRNA$^{Glu(UUC)}$ by deletion of specific arms in the EMSA competition assay. I incubated a range of concentrations, from 0-20 µM, of cold tRNAs constructs containing either the full-length of tRNA$^{Glu(UUC)}$ or truncated versions (Δ acceptor stem, Δ D-arm, Δ anticodon arm or Δ TΨC-arm) (Figure R10-A) with ^{32}P-labelled *in vitro* transcribed tRNA$^{Glu(UUC)}$ and a constant concentration of *Saccharomyces cerevisiae* Kti12$_{CTD}$. Complexes present in reaction mixtures were resolved by native gel electrophoresis (Figure R10-B). Similar to the previous EMSA competition assay, increasing concentrations of cold-tRNA competed with radiolabeled-tRNA$^{Glu(UUC)}$ for binding to Kti12$_{CTD}$. However, competition with the truncated tRNA species was less efficient compared with the full-length tRNA^{Glu}. In this experiment, the acceptor stem deletion (Δ acceptor stem) was the poorest competitor for Kti12$_{CTD}$ binding whereas deletion of the D-arm, anticodon arm or TΨC-arm, showed a similar intermediate competition (Figure R10-C). Other concentrations of competitors were also tested in two independent assays (data not shown) with similar result.

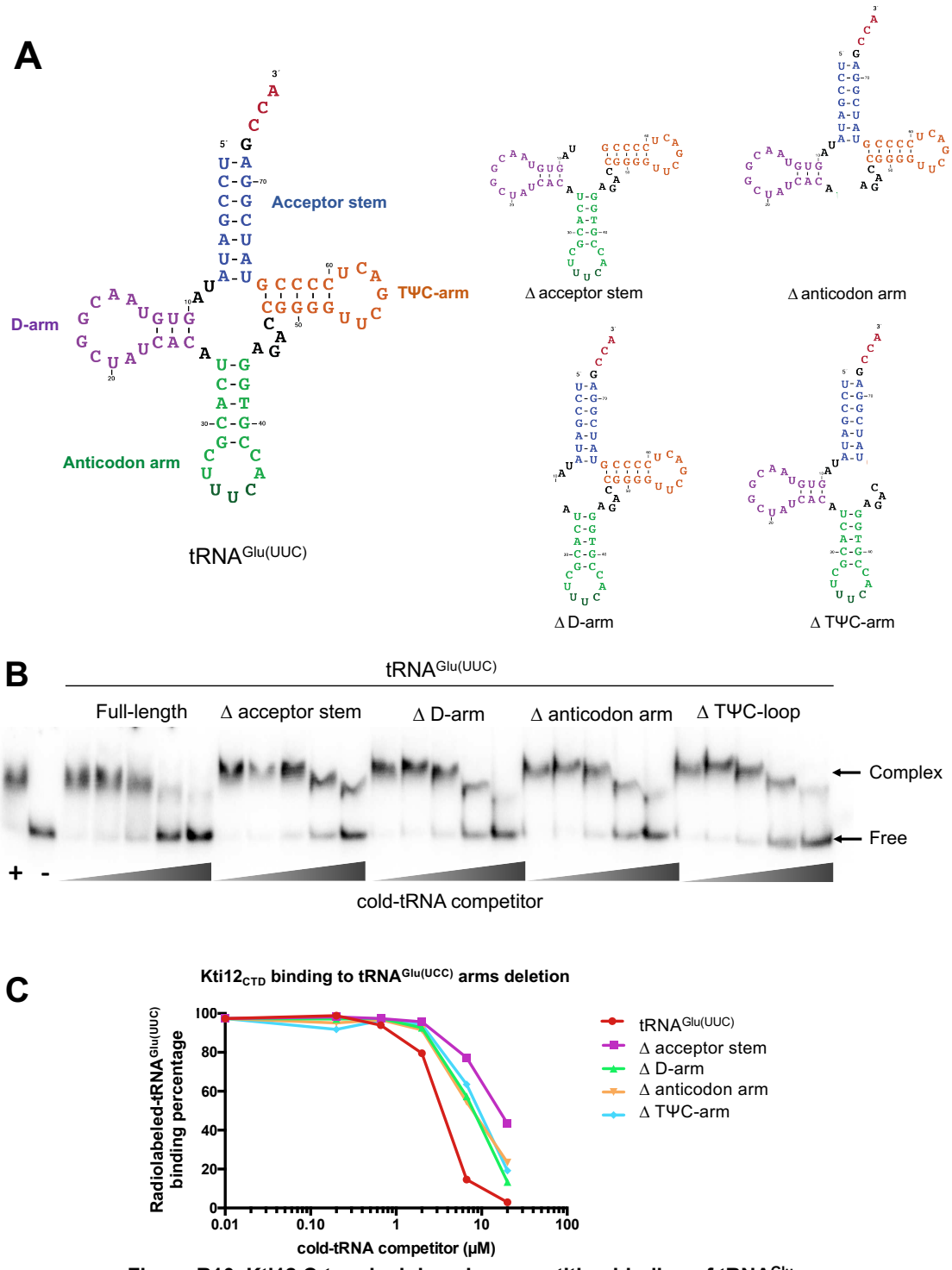


Figure R10. Kti12 C-terminal domain competition binding of tRNA^{Glu} arm deletions

(A) Scheme of five different tRNA^{Glu} arm-deletion constructs: full-length, Δacceptor stem, ΔD-arm, Δanticodon arm and ΔTΨC-arm. **(B)** EMSA competition assay gels comparing the binding of *S. cerevisiae* Kti12$_{CTD}$ to tRNA^{Glu} in the presence of deletion constructs. Complexes were detected by autoradiography after electrophoresis in a 6% polyacrylamide gel. In the absence of protein (sign '-'), free radiolabeled-RNA formed a single band (bottom bands), whereas the presence of 10 μM of protein (sign '+') led to RNA-protein complex formation (upper bands); increasing concentrations of unlabeled 'cold-RNA' (0.2, 0.66, 2, 6.66, 20 μM) displaced radiolabeled-RNA from the complex into the bottom bands. **(C)** Quantification of band intensity of the EMSA represented in a graph created using ImageJ (n=1) (Materials and Methods section 2.6.6).

At the highest tested concentrations of RNA competitor (6.66 and 20 μM) the band of the RNA-protein complex appeared at a lower level than in the lanes with lower competitor concentrations (0.2, 0.66 or 3 μM) forming a descending pattern. Several hypotheses may explain this effect:

(i) in presence of higher concentrations of RNA the conformation of the protein changes affecting its migration on the gel,

(ii) multimerization of proteins observed during purification (Figure R7) may result in multiple RNA molecules bound and give rise to heterogeneous population during competition,

(iii) protein degradation.

3.2.5 Kti12 C-terminal domain has specificity for RNA

To determine if the C-terminal domain of Kti12 ($Kti12_{CTD}$) binds DNA in addition to RNA or whether it has specificity for a particular nucleic acid, I performed an electrophoretic mobility shift assay (EMSA) competition using either the U6 snRNA, $tRNA^{Glu(UUC)}$ as well as a single strand DNA oligo with the sequence of the $tRNA^{Glu(UUC)}$ ($dtRNA^{Glu(TTC)}$). In parallel, since preliminary footprinting experiments suggested that $Kti12_{CTD}$ binds specifically the TΨC-arm of $tRNA^{Glu}$ (Appendix 3) we wanted to test the ability of the TΨC-arm of $tRNA^{Glu}$ alone to compete for binding against full tRNA. EMSA competitions were performed by incubating a range of concentrations (0-20 μM) of unlabeled: U6 snRNA, $tRNA^{Glu(UUC)}$, $dtRNA^{Glu(TTC)}$ or TΨC-arm of $tRNA^{Glu(UUC)}$; with ^{32}P-labelled *in vitro* transcribed $tRNA^{Glu(UUC)}$ and a constant concentration of *Saccharomyces cerevisiae* $Kti12_{CTD}$ (*Sc*$Kti12_{CTD}$). Complexes present in reaction mixtures were separated by native gel electrophoresis and the amount of RNA in complex was quantified (Figure R11-A). As previously shown in figure R9-A, only the highest concentrations of U6 snRNA were able to displace a small fraction (less than 20%) of the radiolabeled-$tRNA^{Glu(UUC)}$ from complex with $Kti12_{CTD}$. Competition of cold-$tRNA^{Glu(UUC)}$ showed the sigmoidal-like curve observed previously. The $dtRNA^{Glu(TTC)}$ competitor had a very similar competition profile as the U6 snRNA: only very large concentrations were able to displace a small fraction of the radiolabeled tRNA from the complex. Interestingly, increasing concentrations of TΨC-arm $tRNA^{Glu(UUC)}$ did not compete with the radiolabeled tRNA, showing that it is insufficient for binding.

In this gel (Figure R11-B), we see a very similar descending pattern of the migration of the RNA-protein complex band, as in figure R10-B, except for the case of TΨC-loop $tRNA^{Glu(UUC)}$, which shows an ascending pattern. This effect may be explained as excess of TΨC-loop competitor leads to an unspecific binding of more than one RNA molecule to the complex, which increases its size and produces a shift in the migration.

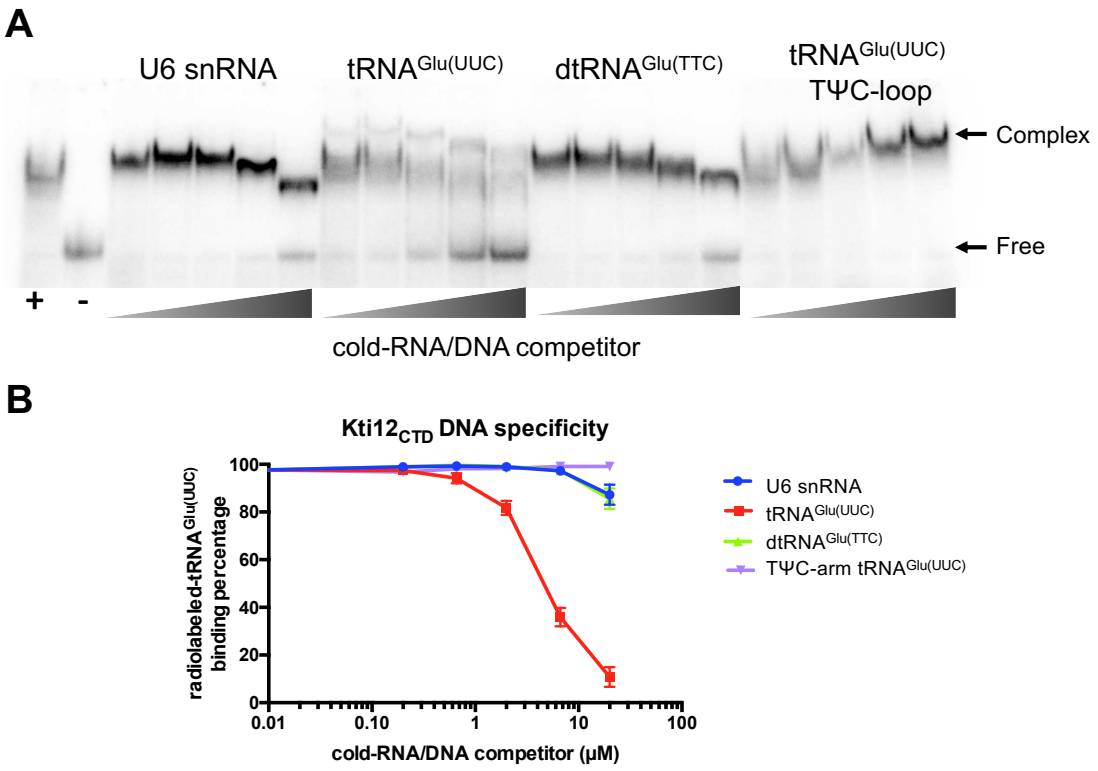


Figure R11. Kti12 C-terminal has specificity for tRNA

(A) EMSA competition comparing binding specificity of *S. cerevisiae* Kti12$_{CTD}$ to various nucleic acids: U6 snRNA, tRNA$^{Glu(UUC)}$, single stranded DNA (dtRNA$^{Glu(UUC)}$) or the TΨC-loop of tRNA$^{Glu(UUC)}$ (17 bp), using electrophoresis in an 8% native polyacrylamide gel. In the absence of protein (sign '-'), free radiolabeled RNA migrates as a single band (bottom bands), whereas the presence of 10 µM of protein (sign '+') leads to migration shift (upper bands); increasing concentrations of unlabeled 'cold-RNA/DNA' (0.2, 0.66, 2, 6.66, 20 µM) displaces radiolabeled-RNA from the upper to the bottom bands. **(B)** Quantification of band intensity of the EMSA represented in a graph created by using ImageJ (n=3) (Materials and Methods section 2.6.6).

3.3 PHENOTYPIC CHARACTERISTICS OF ΔKTI12 MUTANT STRAINS

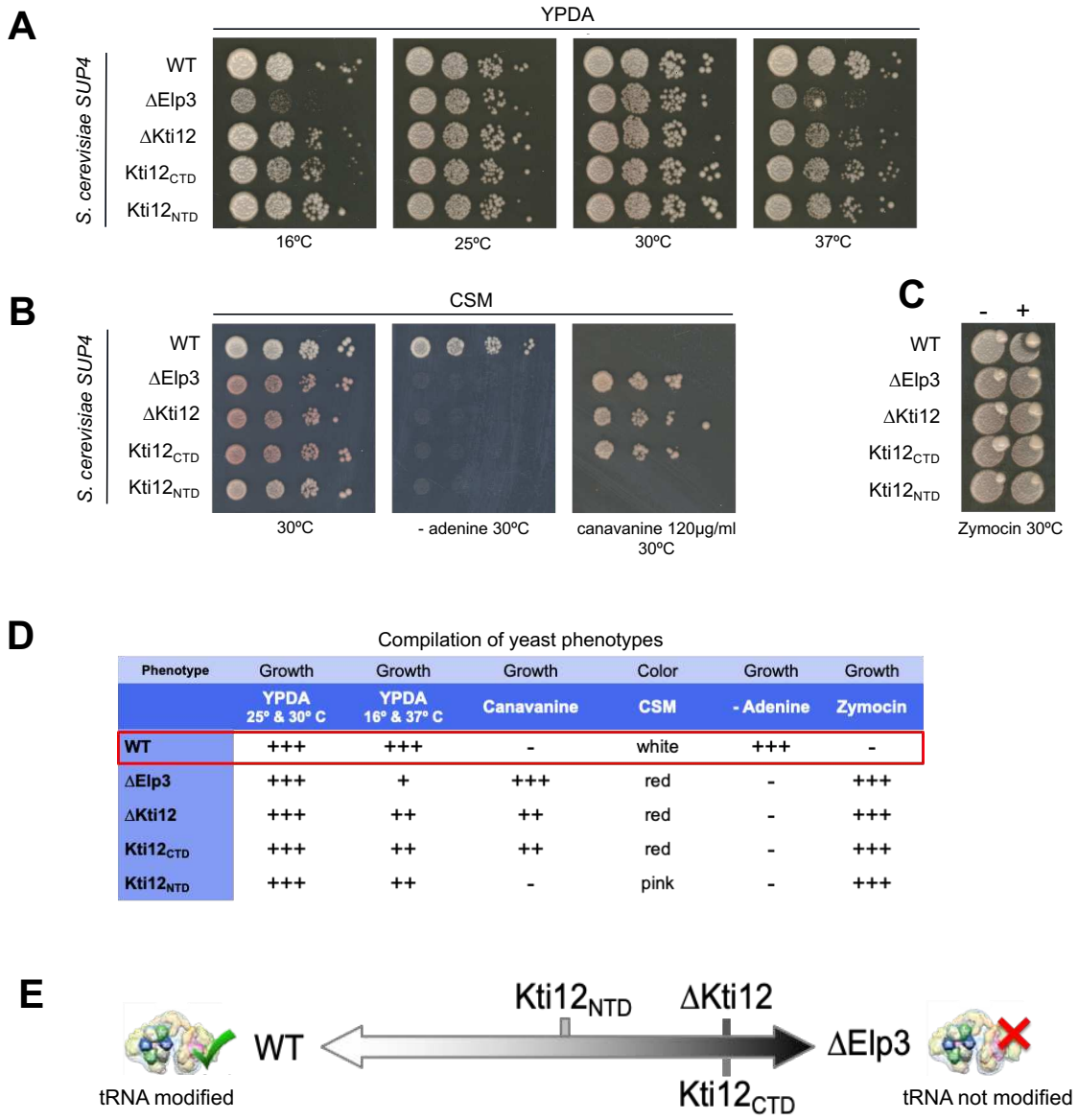
3.3.1 Deletion of Kti12 or Kti12$_{NTD}$ inactivates Elongator function *in vivo*

In parallel to analyzing the RNA binding properties of the C-terminal domain of Kti12, I wanted to investigate whether Kti12 or its two domains were functionally required for Elongator dependent tRNA modifications *in vivo*. For this purpose, I monitored Elongator activity using the following set-up: a yeast strain containing the SUP4 suppressor tRNA and mutant alleles of two genes, *ADE2* and *CAN1*, that are suppressed when SUP4 is active (Figure 19). The *SUP4* gene expresses a tyrosine tRNA with an anticodon complementary to the UAA stop codon. Tyrosine is thus often incorporated when ribosomes encounter this codon, but not always as SUP4 is in competition with translation termination factors. Efficiency of stop codon recognition is modulated by Elongator-dependent modifications of the tRNA. Thus, growth profile of this strain is directly dependent on Elongator function since cells with a defective Elongator will get a red coloration as subproduct of the adenine biosynthesis pathway and will not grow in media lacking adenine. In contrast, cells carrying a defective Elongator will be able to grow in media containing canavanine

as they are unable to incorporate this toxic amino acid (Introduction section 1.1.4.4.5.3 *SUP4* suppressor). Independently of the presence of SUP4, I could also assay the growth of this strain in the presence of the zymocin toxin that targets tRNAs that are modified by Elongator.

I created four different *S. cerevisiae* mutant strains: lacking the full-length of Kti12 (ΔKti12), containing only the N-terminal domain ($Kti12_{NTD}$) or the C-terminal domain ($Kti12_{CTD}$) as well as a strain lacking the catalytic subunit of Elongator (ΔElp3) in which Elongator-dependent modifications are defective. Mutant strains ΔElp3 and ΔKti12 were made by inserting a kanamycin resistance cassette KanMX6 (Materials and Methods section 2.3.2) whereas $Kti12_{NTD}$ and $Kti12_{CTD}$ mutant was made by using CRISPR/Cas9 (Materials and Methods section 2.3.4) to remove the N-terminal domain. During the course of the experiment, I observed that the technique used to make the CTD construct impacts the phenotype of the strain, these differences will be discussed in page 169 in the discussion section. Mutant strains together with a wild-type (WT) were grown under various conditions. To report Elongator-dependent effect, I tested YPDA (rich media) and CSM (complete supplemented media) with or without adenine, and CSM containing canavanine. I also assayed a range of temperatures (16-37ºC) as well as growth of the *S. cerevisiae* strains in the presence of two *K. lactis* strains from which one produces zymocin toxin and the other does not (Materials and Methods sections 2.7.3 and 2.7.4).

Under standard conditions (YPDA media and 25 or 30ºC) there was no visible phenotypic difference between the mutant and wild-type strains. However, under suboptimal growth conditions such as low or high temperature (16 or 37ºC) strains lacking Elp3 (ΔElp3) had a reduced grow rate whereas ΔKti12, and $Kti12_{CTD}$ showed a slightly impaired growth phenotype compared to wild-type strain (Figure R12-A). Different phenotypes between strains were more prominent in complete synthetic media at 30ºC. Under these conditions, the wild-type cells had white color whereas the ΔElp3, ΔKti12 and $Kti12_{CTD}$ strains showed a strong red coloration indicative of a poor ability to synthesize adenine and the $Kti12_{NTD}$ strain showed a light red color (Figure R12-B). In the absence of adenine, only the wild-type strain was able to proliferate, as a functional Elongator is needed to sufficiently suppress the premature stop codon in the *ADE2* gene leading to a normal cell proliferation. In an opposite manner, in the presence of canavanine, a toxic analog of the arginine amino acid, cells with defective Elongator (ΔElp3) or lacking Kti12 (ΔKti12) were able to survive (albeit to a lower extent in latter case), as suppression of the premature stop codon in the *CAN1* gene by functional Elongator-dependent modifications, leads to the uptake of the toxic canavanine and therefore, to a cellular death. Interestingly, like wild-type cells, those expressing only the N-terminal domain of Kti12 ($Kti12_{NTD}$) failed to grow on these plates in contrast to cells containing the C-terminal domain only ($Kti12_{CTD}$). This observation argues that the N-terminal domain in $Kti12_{NTD}$ retains some partial function of the complete protein.



Phenotype	Growth YPDA 25° & 30° C	Growth YPDA 16° & 37° C	Growth Canavanine	Color CSM	Growth - Adenine	Growth Zymocin
WT	+++	+++	-	white	+++	-
ΔElp3	+++	+	+++	red	-	+++
ΔKti12	+++	++	++	red	-	+++
Kti12$_{CTD}$	+++	++	++	red	-	+++
Kti12$_{NTD}$	+++	++	-	pink	-	+++

Figure R12. Phenotypic characteristics of ΔKti12 mutant strains

SUP4 S. cerevisiae strains (WT, ΔElp3, ΔKti12, Kti12$_{NTD}$ and Kti12$_{CTD}$) were tested under various conditions. ΔElp3 and ΔKti12 and strains were made by inserting kanamycin resistance cassette whereas Kti12$_{NTD}$ and Kti12$_{CTD}$ were made by using CRISPR/Cas9. **(A)** Drop assay in YPDA media to compare yeast growth under different temperatures (16, 25, 30 and 37ºC), cells were incubated for 6 days at 16ºC or 2 days for the other temperatures. **(B)** Drop assay to compare yeast growth and cell color in: CSM with adenine, CSM without adenine or CSM with canavanine (120 µg/ml). Cells were incubated at 30ºC for 2.5, 3 or 4.5 days, respectively. **(C)** Eclipse assay to compare zymocin resistance phenotype between five different *S. cerevisiae* strains in the presence of a *K. lactis* non-killer strain (AWK137) indicated with a (-) sign or a *K. lactis* killer strain (NK40) indicated with a (+) sign. **(D)** Compilation of characteristics obtained through the phenotypic assays. **(E)** Ranking of phenotypes representing the impact of Kti12 in the Elongator-dependent tRNA modifications, using wild-type and ΔElp3 strains as reference.

In the eclipse assay that monitors zymocin sensitivity, mutant *S. cerevisiae* strains as well as wild-type were grown in the presence of a non-toxic (-) or toxic (+) strain of the budding yeast *Kluyveromyces lactis.* Growth of the *S. cerevisiae* wild-type cells, containing a functional Elongator, is blocked by zymocin secreted by the toxic strain of *K. lactis*, with zymocin sensitive

cells forming a halo reminiscent of an 'eclipse' surrounding the *K. lactis* colony. In contrast, cells having a dysfunctional Elongator are not sensitive to zymocin and grow normally. Thus, in the presence of the non-toxic *K. lactis* strain, all five tested *S. cerevisiae* strains grew normally. However, in the presence of *K. lactis* killer strain, I observed that mutant strains ΔElp3, ΔKti12, $Kti12_{NTD}$ and $Kti12_{CTD}$ were resistant to zymocin (Figure R12-C).

Compiling the different phenotypes previously tested (Figure R12-D), I could clearly see that the different *S. cerevisiae* Kti12 mutant strains (ΔKti12, $Kti12_{NTD}$ and $Kti12_{CTD}$) present different phenotypes compared to the wild-type strain and were, at least for some assays, more similar to the ΔElp3 strain lacking a functional Elongator (Figure R12-C). Ranking phenotypes on the scale between the wild-type strain with functional Elongator-dependent modifications and the ΔElp3 strain with a dysfunctional Elongator, strains lacking Kti12 and the N-terminal domain (ΔKti12 and $Kti12_{CTD}$) will be closer to the dysfunctional-Elongator phenotype, whereas the strain containing the N-terminal domain ($Kti12_{NTD}$) will occupy an intermediate position (Figure R12-D). These data suggested that, even though both domains of Kti12 are important for the tRNA modification, the N-terminal domain has a stronger involvement in Elongator function under certain suboptimal conditions whereas the C-terminal domain might have an accessory stimulatory function.

3.3.2 The N-terminal domain of Kti12 is important for Elongator function but not its kinase catalytic residues

After demonstrating that the N-terminal domain of Kti12 had a stronger contribution to Elongator function than the C-terminal domain, I wanted to determine whether its putative kinase domain was active or not. Using the sequence alignment of Kti12 and PSTK homologues (Figure R1), we identified two highly conserved amino acids, which were previously identified as catalytic residues of PSTK[289] as well as in other ATPases. The lysine 17 (K17) is located in the P-loop motif (GxP(G/A)xGK(S/T)) and the aspartic acid 41 (D41) is located in the Walker B motif (DxxG) which is common to P-loop kinases[289]. According to the sequence alignment, these catalytic residues correspond to the lysine 14 (K14) and the aspartic acid 85 (D85) of *S. cerevisiae* Kti12 (Figure R13).

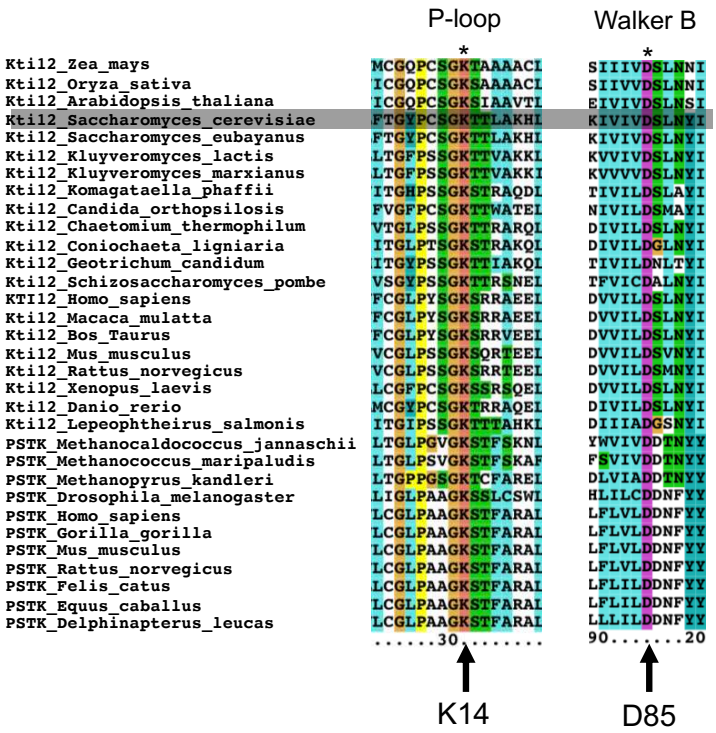


Figure R13. Conserved catalytic residues of Kti12

Sequence alignment of PSTK and Kti12 showed two conserved catalytic residues: lysine 14 (K14) and aspartic acid 85 (D85).

I created *S. cerevisiae SUP4* strains carrying mutations of those kinase catalytic subunits: the lysine 14 or the aspartic acid 85 were substituted by an alanine amino acid (K14A and D85A) by using the CRISPR-Cas9 method. Phenotypes of the strains were then tested through drop and eclipse assays.

Drop assay of strains grown in YPDA medium (Figure R14-A) showed that at 30ºC, the five tested mutant strains (ΔElp3, ΔKti12, Kti12$_{NTD,}$ Kti12$_{K14A}$ and Kti12$_{D85A}$) presented the same white color as the wild-type strain and ΔElp3 cells grew slower than the rest. At 25ºC, the ΔElp3, ΔKti12 and Kti12$_{NTD}$ strains present a slight red coloration compared to the wild-type strain, which is not present neither in Kti12$_{K14A}$ nor in Kti12$_{D85A}$. At 16 and 37ºC, the ΔElp3 strain has a marked reduction in cell growth compared to the wild-type; whereas the remaining mutant strains look very similar to the wild-type at 16ºC, and at 37ºC the ΔKti12 and Kti12$_{NTD}$ strains showed slightly reduced cell growth compared to wild-type and point mutant strains Kti12$_{K14A}$ and Kti12$_{D85A}$. Phenotypic differences between strains are more prominent in complete synthetic media (Figure R14-B). At 30ºC on CSM (with adenine) ΔElp3, ΔKti12 and Kti12$_{NTD}$ strains present a red pigmentation which contrasts with the white color of the WT, Kti12$_{K14A}$ and Kti12$_{D85A}$ strains. In the absence of adenine (-adenine), only cells of the WT, Kti12$_{K14A}$ and Kti12$_{D85A}$ strains are able to survive, meaning functional Elongator in these strains. In presence of the toxic amino acid analog canavanine, when only cells having a dysfunctional Elongator should be able to grow, only ΔElp3, ΔKti12 and Kti12$_{NTD}$ strains grew, but Kti12$_{NTD}$ cells presented a reduced growth compared to ΔElp3 or ΔKti12 cells. Importantly, Kti12$_{K14A}$ and Kti12$_{D85A}$ were unable to grow like the wild-type.

Eclipse assay showed that WT, Kti12$_{K14A}$ and Kti12$_{D85A}$ strains were sensitive to zymocin whereas ΔElp3, ΔKti12 and Kti12$_{NTD}$ strains were resistant to the toxin (Figure R14-C).

These data demonstrated that the *S. cerevisiae* strains carrying the K14A or D85A mutations in Kti12, maintain Elongator function indicative of proper tRNA cm^{5}U$_{34}$ derivatives modifications. This means that while the N-terminal domain of Kti12 is important for the Elongator-dependent modifications, its catalytic residues are not.

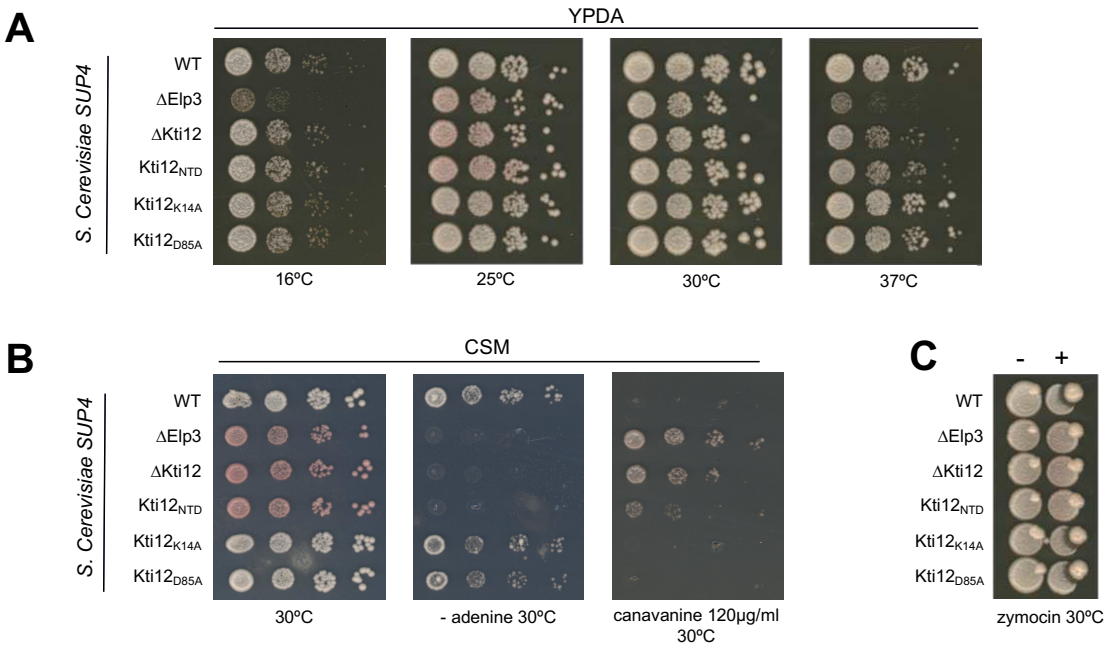


Figure R14. Phenotypic characteristics of Kti12 conserved catalytic residues

SUP4 S. cerevisiae strains (WT, ΔElp3, ΔKti12, Kti12$_{NTD}$, Kti12$_{K14A}$ and Kti12$_{D85A}$) were tested under various conditions. ΔElp3, ΔKti12, and Kti12$_{NTD}$ strains were made by inserting kanamycin resistance cassette whereas Kti12$_{K14A}$ and Kti12$_{D85A}$ were made by using CRISPR/Cas9. **(A)** Drop assay in YPDA media to compare yeast growth at different temperatures (16, 25, 30 and 37ºC), cells were incubated for 2.5 days or 5 days (at 16ºC). **(B)** Drop assay to compare yeast growth and coloration in: CSM with adenine, CSM without adenine or CSM with canavanine (120 µg/ml). Cells were incubated at 30 ºC for 3 days. **(C)** Eclipse assay to compare zymocin resistance phenotype between the six different *S. cerevisiae* strains in the presence of a *K. lactis* non-killer strain (AWK137) indicated with a (-) sign or a *K. lactis* killer strain (NK40) indicated with a (+) sign.

Interestingly, I noticed that the technique used to make the *S. cerevisiae* Kti12 N-terminal mutants impacted their phenotypes under some conditions. In YPDA media, there is no visible difference between the growth of a Kti12$_{NTD}$ mutant strain made by the insertion of the KanMX6 cassette (Kti12$_{NTD\text{-}KanMX6}$) and the Kti12$_{NTD}$ strain made by using CRISPR/Cas9 (Kti12$_{NTD\text{-}CRISPR}$) (Figure R15-A). However, in the CSM there is a clear phenotypic difference between these strains: in CSM supplemented with all amino acids Kti12$_{NTD\text{-}CRISPR}$ has a lighter red pigmentation compared to Kti12$_{NTD\text{-}KanMX6}$ and also in the presence of 60µg/ml of canavanine, Kti12$_{NTD\text{-}KanMX6}$ cells are able to grow showing a similar phenotype to the non-functional Elongator strains whereas Kti12$_{NTD\text{-}CRISPR}$ cells do not grow, similarly to the wild-type strain. These differences between Kti12$_{NTD\text{-}KanMX6}$ and Kti12$_{NTD\text{-}CRISPR}$ cells are abolished in the 120µg/ml CSM-canavanine

plates when canavanine concentration becomes too high to detect the residual N-terminal domain activity of the Kti12$_{NTD\text{-}CRISPR}$ strain (Figure R15-B). Both Kti12$_{NTD}$ strains were resistant to zymocin, similar to other cells with non-functional Elongator (Figure R15-C). These observations indicate that the Kti12$_{NTD\text{-}KanMX6}$ strain is less functional than the Kti12$_{NTD\text{-}CRISPR}$ strain. Such phenotypic differences between strains can be explained by their genomic organization (Figure R15-D):

(i) In the Kti12$_{NTD\text{-}KanMX6}$ strain, the KanMX6 marker inserted right after the truncated Kti12 alters the 3' processing of the mRNA, inducing production of aberrant mRNAs some of which are subjected to Nonsense Mediated Decay (NMD) because of the presence of aberrant 3' UTR. Indeed, the strategy to lower gene expression by introducing downstream cassettes, named DAmP for Decreased Abundance by mRNA Perturbation, has been extensively validated to generate weak alleles of essential genes[346].

(ii) In the Kti12$_{NTD\text{-}CRISPR}$ strain, there is no insertion and only the deletion of the C-terminal domain. Biosynthesis of the Kti12 mRNA is therefore unperturbed. This strain is thus 'cleaner' and should express 'normal' (not reduced) level of Kti12$_{NTD}$. Quantification of the amount of Kti12 produced in these different strains could help to solidify this but would require a dedicated antibody.

Altogether, the phenotypic data is consistent with reduction of Kti12 mRNA expression by the inserted KanMX6 marker synergizing with further reduction of the level of Kti12 protein due to mutations of critical residues of the kinase domain. These mutations probably affect protein stability either directly or by preventing the binding of stabilizing ligand such as ATP. In strains with both the KanMX6 marker and a kinase mutation, the Kti12 protein level probably decreases beyond a critical threshold abolishing Elongator function and displaying corresponding growth phenotypes. In contrast, the KanMX6 insertion in an otherwise wild type Kti12 background or mutation of conserved residues of the Kti12 kinase domain at an unaltered Kti12 loci maintain sufficient Kti12 activity for proper Elongator function. As the latter context only allows for rigorous genetic comparison of Kti12 catalytic site mutants with wild type counterparts, I can conclude that the corresponding residues are not necessary for Kti12 function *in vivo* even if I cannot formally exclude that they provide a positive stimulatory activity.

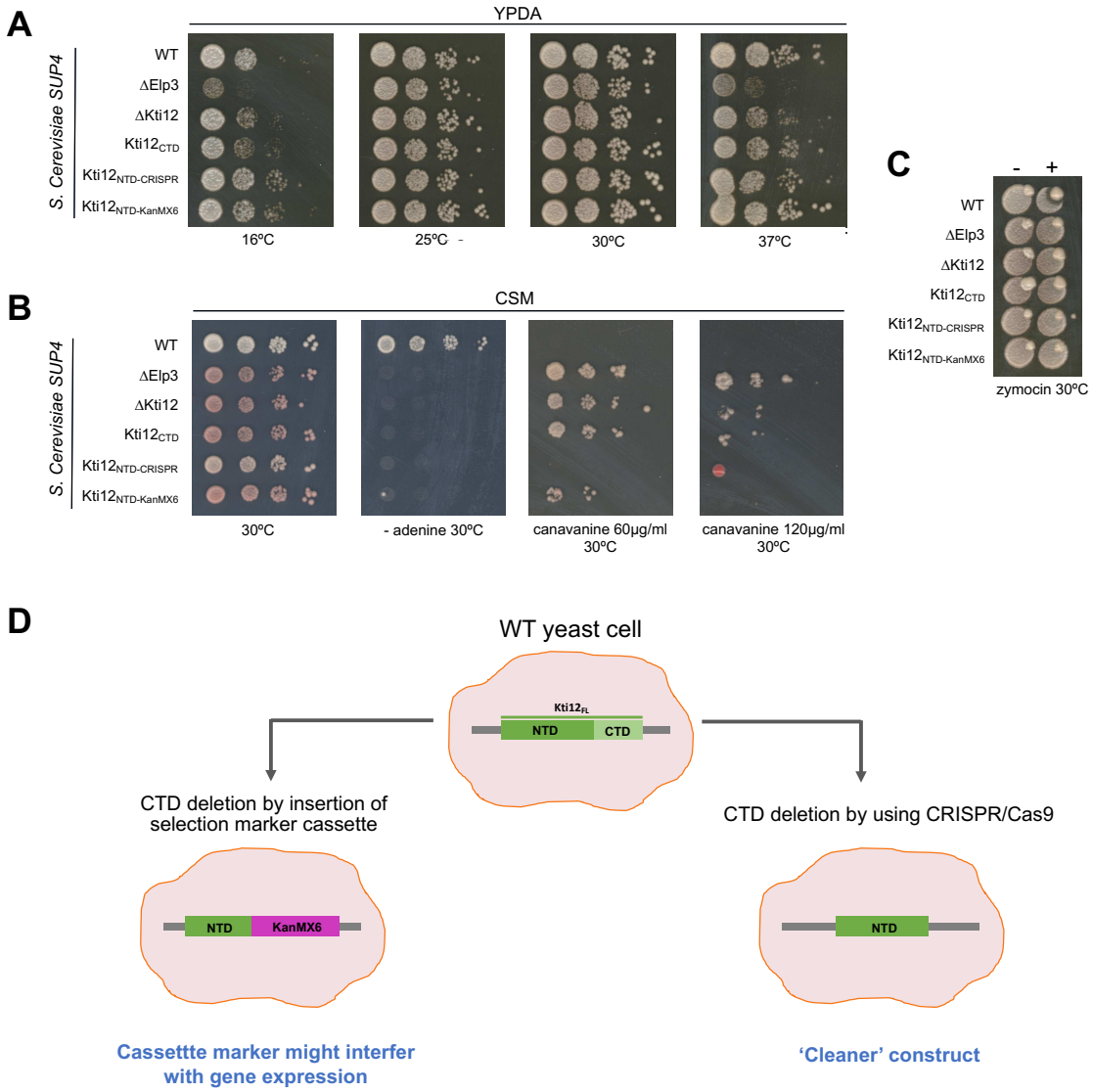


Figure R15. Phenotypic comparison of Kti12$_{NTD}$ made by CRISPR/Cas9 and KanMX6 cassette insertion

S. cerevisiae SUP4 strains (WT, ΔElp3, ΔKti12, Kti12$_{CTD}$, Kti12$_{NTD-CRISPR,}$ and Kti12$_{NTD-KanMX6}$) were tested under various conditions. **(A)** Drop assay in YPDA media to compare yeast growth under different temperatures (16, 25, 30 and 37ºC), cells were incubated for 4.5 days at 16ºC, 2 days at 25 and 30ºC or 2.5 days at 37ºC. **(B)** Drop assay to compare yeast growth and coloration in: CSM with adenine, CSM without adenine or CSM with canavanine (60 or 120 µg/ml). Cells were incubated at 30 ºC for 2.5, 3 or 4.5 days, respectively. **(C)** Eclipse assay to compare zymocin resistance phenotype between the six different *S. cerevisiae* strains in the presence of a *K. lactis* non-killer strain (AWK137) indicated with a (-) sign or a *K. lactis* killer strain (NK40) indicated with a (+) sign. **(D)** Scheme explaining the Kti12 loci organization difference between the two different techniques used to make the Kti12$_{NTD}$ mutant strain.

3.4 PROTEINS CO-PURIFYING WITH THE ELONGATOR COMPLEX

3.4.1 Tandem affinity purification

To determine the proteins which interact with Elongator complex, I performed a tandem affinity purification (TAP). I used a TAP-tag construct containing: a calmodulin binding peptide (CBP), a TEV cleavage site, and two IgG binding domains of protein A. The TAP-tag was fused to the C-termini of endogenous *S. cerevisiae* proteins Elp1, Kti12, Kti12$_{NTD}$ and Kti12$_{CTD}$ (Elp1-TAP, Kti12-TAP, Kti12$_{NTD(1-182)}$-TAP and Kti12$_{CTD(183-313)}$-TAP). TAP-tagged proteins and associated partners were purified by using IgG and calmodulin beads according to Rigaut *et al.*[344]. Purified TAP fractions were cleaned by a short migration on SDS-PAGE to remove small contaminants (salt, detergents, etc.). Briefly, purified TAP fractions were loaded on a 10% denaturing polyacrylamide gel and subjected to electrophoresis for a short time only allowing proteins to enter the gel for few millimeters before being stained. For each purification, a band containing the protein mixture (freed from other interfering molecules: salt, detergents, etc.) was cut and used for protein identification by mass spectrometry (Material and Methods sections 2.7.5 and 2.5.10). The data indicated that in the TAP purifications of Elp1, Kti12, Kti12$_{NTD}$ or Kti12$_{CTD}$. were identified 234, 511, 360 or 122 proteins respectively. In Elp1-TAP and Kti12-TAP, the six subunits of Elongator as well as Kti12 and Kti14 were detected in the purification. In contrast, in the Kti12$_{NTD}$-TAP purification, none of the Elongator subunits or other Kti cofactors were detected (except obviously Kti12). In the Kti12$_{CTD}$-TAP purification, only tagged Kti12 and Elp3 were detected by mass spectrometry. It is likely that in the absence of either the N- or C-terminal domains of Kti12, its interaction with Elongator is reduced explaining why Elp proteins are not found (Kti12$_{NTD}$-TAP) or barely detected (Kti12$_{CTD}$-TAP). Surprisingly, neither Kti11 nor Kti13 were present in the purified fractions (Figure R16-A). Other proteins present in purified fractions are often abundant factors such as ribosomal proteins and likely to be contaminants. A molecular function enrichment analysis (Figure R16-B) showed indeed that from the different proteins coprecipitating with Elp1-TAP and Kti12-TAP, the majority are ribosomal and ribosomal-associated proteins which have structural and RNA binding properties. Other molecular functions related to transcription and ATPase/ATP binding were also found and correspond also to abundant factors.

A

	Proteins	Elp1-TAP	Kti12-TAP	$Kti12_{NTD}$-TAP	$Kti12_{CTD}$-TAP
		Tandem Affinity Purification			
Elongator subunits	Elp1	✅	✅	X	X
	Elp2	✅	✅	X	X
	Elp3	✅	✅	X	✅
	Elp4	✅	✅	X	X
	Elp5	✅	✅	X	X
	Elp6	✅	✅	X	X
KTI cofactors	Kti11	X	X	X	X
	Kti12	✅	✅	✅	✅
	Kti13	X	X	X	X
	Kti14	✅	✅	X	X
	Total protein number	234	511	360	122

B

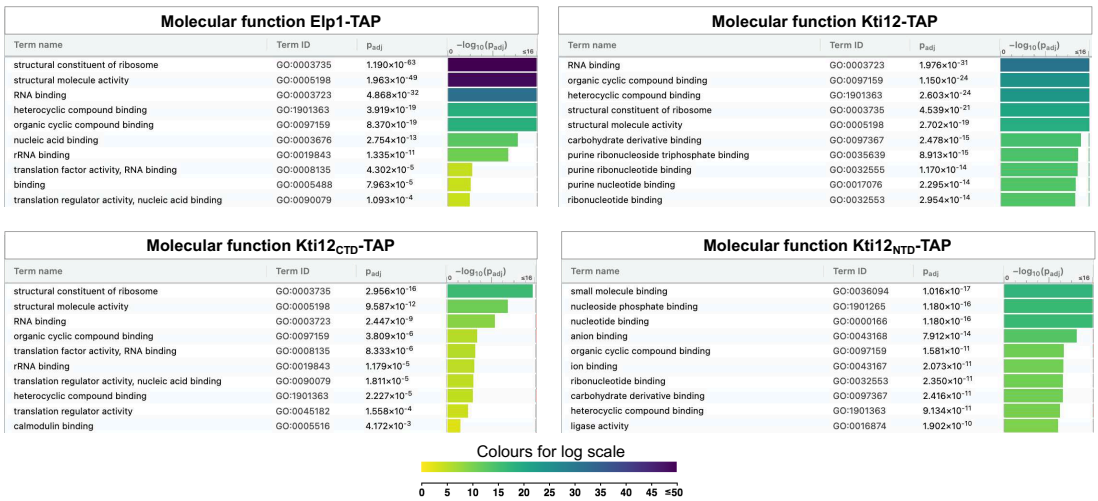


Figure R16. Mass spectrometry analysis of Elp1 and Kti12 TAP-tagged proteins

(A) Table summarizing the Elongator subunits and KTI cofactors copurifying with: Elp1-TAP, Kti12-TAP, $Kti12_{NTD}$-TAP or $Kti12_{CTD}$-TAP. **(B)** Functional enrichment analysis on molecular function of proteins copurifying in each TAP purification. In the columns are presented: name of the molecular function, gene ontology (GO) number, adjusted p-value (P_{adj}), logarithmic graph of P_{adj}. Only top 10 of molecular functions are presented for each sample.

3.5 KINASE ACTIVITY ASSOCIATED TO KTI12

3.5.1 Kinase assay development

To assay the properties of the kinase domain of Kti12, and know in particular whether it is endowed with kinase activity, I developed an *in vitro* kinase assay. For this, I used the strains carrying the TAP-tag fused to the C-terminus of endogenous *S. cerevisiae* proteins Elp1, Kti12, $Kti12_{NTD}$ and $Kti12_{CTD}$ (Elp1-TAP, Kti12-TAP, $Kti12_{NTD}$-TAP and $Kti12_{CTD}$-TAP). Total protein extracts of each strain were incubated with IgG beads to pull-down the tagged protein together with its interacting proteins (Figure R17-A). After washes, the beads and associated proteins were incubated with radiolabeled ATP [γ-32P]. After the reaction, proteins were released with SDS-containing buffer and products of each reaction were fractionated on a denaturing gel.

Phosphorylated proteins were identified by phosphorimaging (Figure R17-B) (Material and Methods section 2.7.7). If Kti12 has a kinase activity or if a kinase is associated with Elongator, we expected to find and detect phosphorylated protein(s) (and/or other molecules such as tRNA) among the molecules coprecipitating with the TAP fusion.

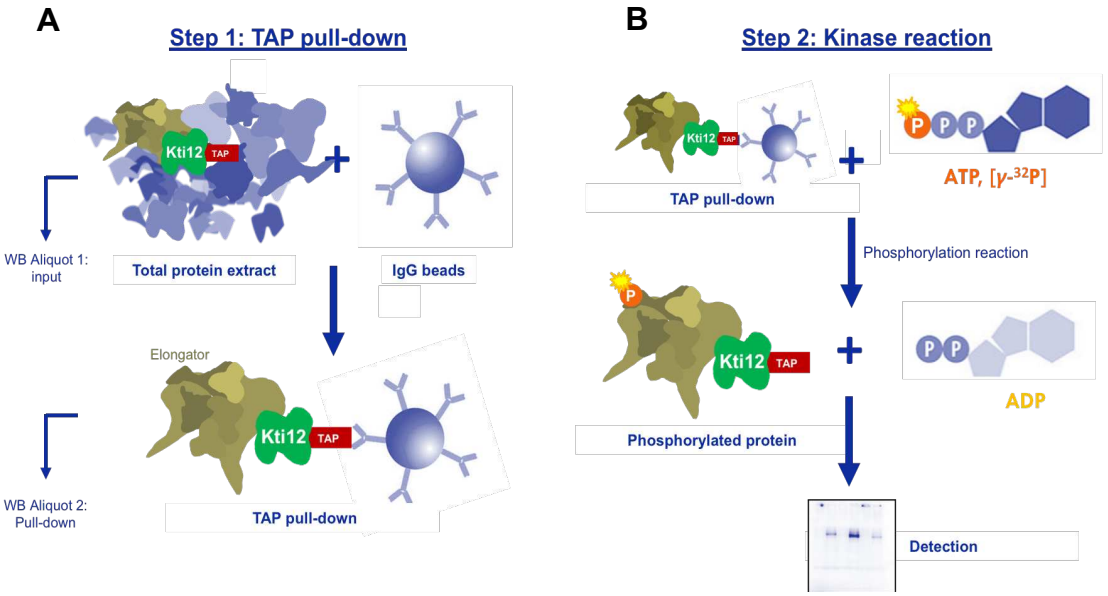


Figure R17. Scheme of the kinase assay
The kinase assay is divided into two steps: **(A)** TAP pull-down and **(B)** Kinase reaction.

3.5.2 A kinase phosphorylating a ~150 kDa protein is associated with Kti12 and Elongator

To validate the protocol described above, I first analyzed the presence of TAP tagged proteins in extracts and bead-associated fractions by western blot. This indicated that the target proteins bound to beads, thus validating the kinase assay strategy (Figure R17-B). The detection of the ^{32}P signal identified a phosphorylated protein of around 150 kDa present in the reactions with TAP-tagged strains (Elp1, Kti12 and Kti12$_{NTD}$) but not in the extracts obtained from untagged strains (BMA and SUP4). This signal is stronger in the Kti12-TAP and weaker in the Kti12$_{NTD}$-TAP strain (Figure R18-A). The 150 kDa phosphorylated band was absent in the reaction using the Kti12$_{CTD}$-TAP strain. Content of TAP-tagged proteins in the input and bead-associated fractions was controlled by western blot (Figure R18-B). Comparing the size of the phosphorylated protein (around 150 kDa) to list of Kti12-TAP interacting proteins obtained from mass spectrometry (Figure R16-A) and based on published evidence indicating that Elp1 is a phosphoprotein[252], we hypothesized that the phosphorylated protein was Elp1.

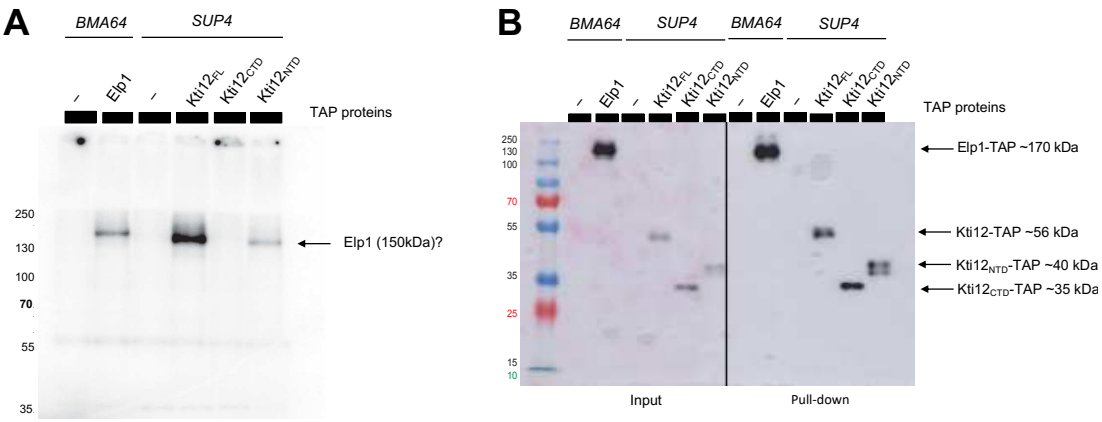


Figure R18. Kinase assay of Kti12

(A) Kinase reaction using extracts from *S. cerevisiae BMA* (WT and Elp1-TAP) or *SUP4* (WT, Kti12-TAP, Kti12$_{NTD}$-TAP or Kti12$_{CTD}$-TAP) strains. (-) signs mark the WT strain which does not carry the TAP-tag and thus demonstrate the background activity. A phosphorylated protein at around 150 kDa was detected after fractionation on 10% denaturing polyacrylamide gel. **(B)** Western blot TAP-tagged proteins using PAP reagent.

To test for the impact of tRNA on the phosphorylation, unlabeled tRNA was added to kinase reactions. Under the tested conditions, no tRNA phosphorylation was detected and no change in the reaction efficiency or products were noticed (data not shown).

To determine whether or not the ~150 kDa phosphorylated protein was Elp1, I performed a kinase assay by using extracts from Elp1-TAP or Kti12-TAP. To optimize/standardize the assay, two starting cellular concentrations were used: 1 OD_{600nm} (Elp1-A and Kti12-A) or 2 OD_{600} (Elp1-B and Kti12-B). Phosphorimaging of the polyacrylamide gel revealed that the phosphorylated product obtained from the Elp1-TAP pull-downs (at OD_{600nm} 1 or 2) had a slightly larger molecular weight than the phosphorylated product from the Kti12-TAP pull-down (both at OD_{600nm} 1 or 2) (Figure R19-A), corresponding to the addition of ~20 kDa of the TAP-tag to Elp1. Whereas the Kti12-TAP strain contains the endogenous version of Elp1 with a molecular weight of around 150 kDa. The western blot of different fractions used for the kinase assay (Figure R19-B) confirmed the presence of the expected TAP-tagged proteins (Elp1 and Kti12) in their respective strains. This strongly suggested that the phosphorylated protein was Elp1. This should be further confirmed by treating the reactions with TEV protease that should cleave the Elp1-TAP, restoring a molecular weight similar to the endogenous factor.

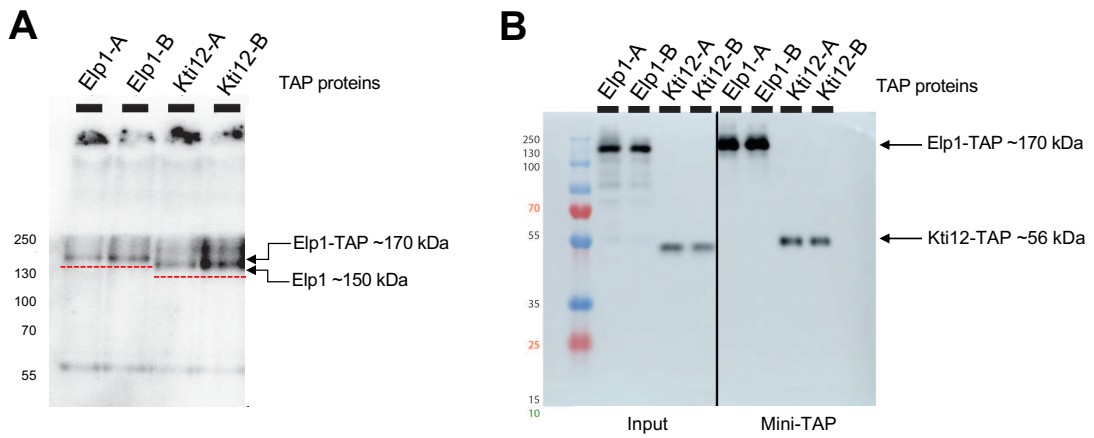


Figure R19. Identification of Elp1 as the 150 kDa phosphorylated protein
(A) Kinase reaction using extracts prepared from *S. cerevisiae* Elp1-TAP or Kti12-TAP, a phosphorylated protein at around 150 kDa was detected after fractionation of the products in 10% denaturing polyacrylamide gel. **(B)** Western blot TAP-tagged proteins using PAP reagent.

To get more information about the influence of the Kti12 N-terminal domain on the Elp1 phosphorylation, I performed a kinase assay by using Kti12 TAP-tagged strains carrying the previously described mutations of the catalytic residues of the Kti12 kinase domain (lysine 14 or aspartate 85 substituted by an alanine, K14A and D85A). Bead bound proteins from TAP-tagged strains (Kti12-TAP, Kti12$_{K14A}$-TAP or Kti12$_{D85A}$-TAP) as well as the wild type strain (-) were incubated with radiolabeled ATP [γ-^{32}P]. The phosphorimager image (Figure R20-A), revealed that phosphorylation of Elp1 was detected in all three strains. However, the observed band intensity was slightly weaker in reactions performed with the mutant strains (Kti12$_{K14A}$-TAP and Kti12$_{D85A}$-TAP) compared to the Kti12-TAP one. Western blot analysis of TAP-tagged proteins present in extracts and in the kinase assay (Figure R20-B) confirmed the presence of the three Kti12-TAP variants. The data indicated that mutations of the conserved kinase catalytic residues of Kti12 do not abolish Elp1 phosphorylation. The observed differences may indicate that the mutant Kti12 forms either have slightly impaired interaction with Elongator or support Elp1 phosphorylation less efficiently. For the Kti12$_{D85A}$ mutant, the lower signal may also result from lower protein amount in the reaction (as observed on the western blot, Figure R20-B). Further analyses would be required to assess whether the quantitative differences are significant. The qualitative difference is, however, informative and reinforces results obtained from the phenotypic assays (Figure R14-B) in which the same mutations did not affect Elongator-dependent function based on growth phenotype analyses. Altogether, these data suggest strongly that other kinase(s) associated with Kti12 and Elongator is(are) mediating Elp1 phosphorylation. The role of Kti12 in this process would be accessory and/or indirect. In agreement with (i) the presence of Kti14 in Elp1-TAP and Kti12-TAP purifications (see above), (ii) the fact that Kti14 is a kinase, and (iii) earlier suggestion that Kti14 phosphorylates Elp1[252], Kti14 was the prime candidate for mediating this reaction.

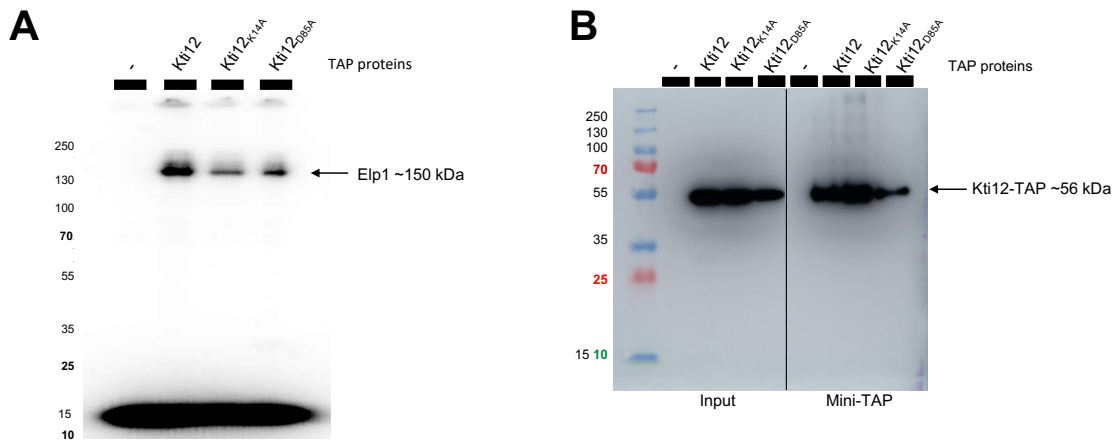


Figure R20. Kinase assay of Kti12 point mutants

(A) Kinase reaction with extracts of *S. cerevisiae* SUP4 wild-type or Kti12 TAP-tagged strains (Kti12-TAP, $Kti12_{K14A}$-TAP or $Kti12_{D85A}$-TAP). Phosphorimaging detects phosphorylated Elp1 after fractionation of the reaction products on 10% denaturing polyacrylamide gel. **(B)** Western blot TAP-tagged proteins using PAP reagent.

3.6 INTERACTIONS BETWEEN KTI12 AND KTI14

3.6.1 Two-Hybrid Assay

To test the interaction of Kti12 or Kti14 with Elongator subunits and other Kti cofactors, I used the ProQuest™ Two-Hybrid System. For this purpose, I cloned Elongator subunits or Kti cofactors into the pDEST32 vector (DNA-binding domain, 'bait') or into the pDEST22 vector (activation domain, 'prey') using the Gateway cloning strategy. The *S. cerevisiae* tester strain was transformed with a pair of vectors (pDEST32 and pDEST22) encoding either a Elongator subunit or a Kti cofactor. This allowed me to test 72 different protein interactions, 36 in each direction (prey→bait, bait→prey). Interaction between proteins was measured by luminescence detection (Material and Methods section 2.7.2). Results of the 36 'bait→prey' interactions plus controls (Figure R21) showed that, under the tested conditions, only a Kti14-Kti14 assay produced any signal; the corresponding luminescence activity was about 12000 units whereas for the positive control around 270000 units were detected. However, this wasn't very reproducible from experiment to experiment. Thus, whether Kti14 forms homodimers like PSTK remains to be confirmed. It was surprising that none of the other combinations of Elongators/Kti subunits gave a positive signal, as previous studies have shown Elongator subunits (Elp1) to coprecipitate with $Kti14^{288}$ and with $Kti12^{252}$ as well as co-purification of Kti12 and Kti14 by TAP[288]. Structural predictions using RoseTTAFold and AlphaFold also support an interaction between Elp2 and $Kti12^{347}$.

It is possible that the fusion to the transcription activation domain and/or DNA binding domain inactivates the proteins. Alternatively, they may not interact in the environment provided by the yeast nucleus. Altogether, negative results in the yeast two-hybrid assay are not informative and thus this assay failed to provide useful data. This prompted us to analyze protein association using co-immunoprecipitation assays and recombinant factors.

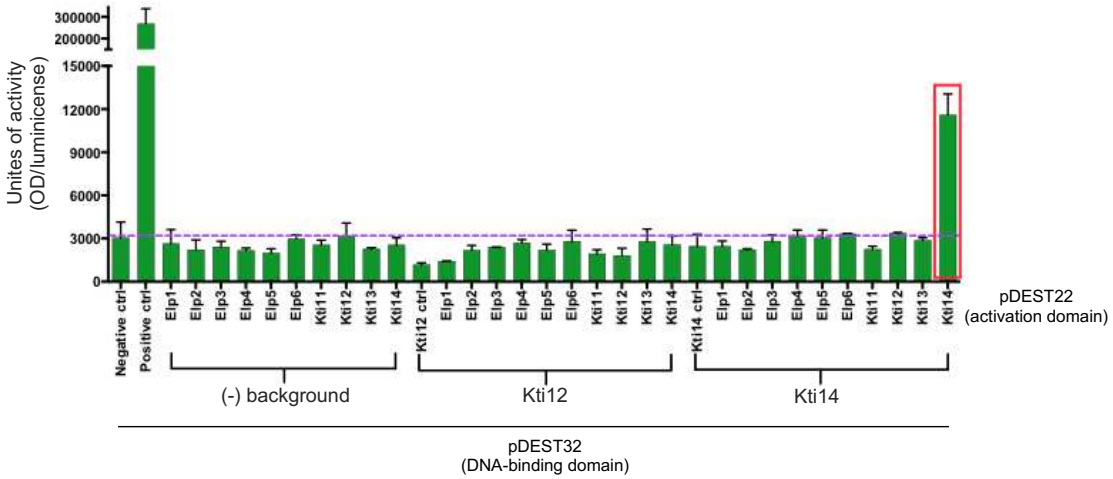


Figure R21. Interaction of Kti12 and Kti14 with Elongator subunits cofactors by using yeast two-hybrid

pDEST32 vector containing the bait (DNA-binding domain) and the pDEST22 vector containing the prey (activation domain). Green bars represent the units of activity of 36 'bait→prey' interactions between the Elongator subunits (Elp1-6), Kti cofactors (Kti11-14), positive control (Pho92 and Not1). (-) bars represent background activity in the absence of protein cloned into the pDEST32. Purple dot line represents the threshold of background, activities below it, were considered as noise; signals above it were considered significant. Red square highlights the interacting proteins. Units of activity were calculated arbitrary (n=2) (Material and Methods section 2.7.2).

3.6.2 Interaction of Kti12 and Kti14 by co-immunoprecipitation

3.6.2.1 Co-immunoprecipitation constructs selection

TAP purifications have shown that Kti12 and Kti14 can be co-purified[288] (Figure 16A). To facilitate analyses of mutants and comparison between experiments as well as to include internal controls, I developed a co-immunoprecipitation assay.

For this purpose, I made different constructs containing full-length proteins including: Kti12 and Kti14 fused to the HA_3 tag, and Elp4 fused to the VSV tag. Kti12 and Kti14 constructs were inserted into the overexpression plasmid pYX242 containing 2µ origin (high copy number) whereas Elp4-VSV was inserted in a low expression centromeric plasmid (Figure R22-A) (Material and Methods section 2.7.8). Overexpression plasmids were used for Kti12 and Kti14 because preliminary tests with low expression centromeric plasmid did not produce positive results in co-immunoprecipitation assays; in contrast, overexpressed Elp4-VSV generated some background (data not shown). *S. cerevisiae* wild-type (WT) and Elp1-TAP cells were transformed with the plasmids containing the tagged proteins, or the empty vector as a control. Co-immunoprecipitations were made by pulling down Elp1-TAP with IgG beads, resolving the beads-bound proteins in SDS-polyacrylamide gels, and detecting precipitated proteins by western blotting (Material and Methods sections 2.5.9).

Western blot using the PAP reagent (Figure R22-B) showed that Elp1-TAP protein (172 kDa) was detected in the input and elution fractions from Elp1-TAP strain and absent in the transformed WT strain. The Elp4-VSV protein (51 kDa) was detected in the input of the correspondingly

transformed WT and Elp1-TAP strains. Importantly, it was present only in the precipitated fraction from Elp1-TAP cells, but not in the wild-type one, demonstrating that Elp4 co-immunoprecipitated with Elp1-TAP. Similarly, HA_3-Kti14 (59 kDa) was detected in the input fraction of the correspondingly transformed WT and Elp1-TAP cells. As before, HA_3-Kti14 was also detected in the precipitated fraction of Elp1-TAP. HA_3-Kti14 was present in the elution fraction of WT cells; albeit the intensity of the band is much weaker than the one present the Elp1-TAP pull-down. I interpret this result as indicating that Kti14 co-precipitates with Elp1 but that unspecific binding of Kti14 to IgG beads results in a certain background level. Kti12-HA_3 was also detected in the input and elution fractions of the correspondingly transformed WT and Elp1-TAP cells; but despite the fact that the signal in the input fraction was good, very weak bands were detected in the precipitated fractions. The intensity of the bands of the elution fraction was very similar in the WT and Elp1-TAP cell pull-downs, indicating that these bands resulted from unspecific binding of Kti12 to the IgG beads. Thus, Kti12 interaction could not be tested in this assay. Stm1, an endogenous protein associated to the ribosome for which an efficient antibody is available in the lab, was used as a negative control in this assay. It was not precipitated with Elp1-TAP, confirming the specificity of the other positive results.

The fact that we did not detect co-precipitation of Kti12 in the Elp1-TAP pull-down under the tested conditions is not sufficient to conclude that these proteins do not coprecipitate/interact, since we already detected its interaction by mass spectrometry on the Kti12 and Elp1 TAP purifications (Figure R16-A). We already noticed that the position of the HA_3 tag on Kti14 (C- or N-terminus) impacts its co-precipitation efficiency as well as its unspecific binding to the IgG beads (data not shown); therefore, we cannot exclude that the same applies to Kti12 and that tagging it at its N-terminus might allow its co-precipitation with Elp1-TAP.

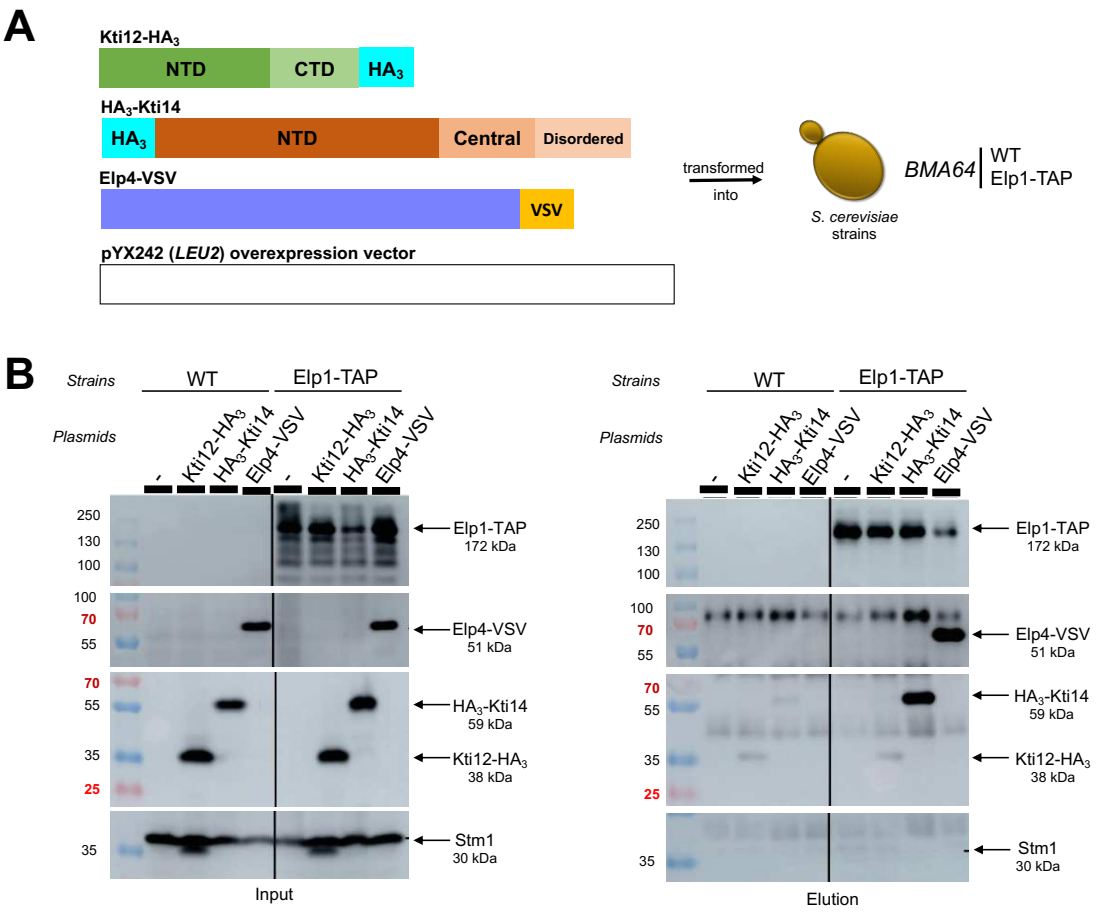


Figure R22. Selection of constructs for Co-IP

(A) Schemes of constructs used for Co-IP. *S. cerevisiae* BMA64 wild-type or expressing Elp1-TAP strains were transformed with either empty vector (pYX242) or plasmids encoding Kti12-HA_3, HA_3-Kti14 or Elp4-VSV. **(B)** Western blots analyses of Co-IP assays of Elp1-TAP by using IgG beads, showing input (left gel) or elution (right gel) fractions resolved on 8% denaturing polyacrylamide gels.

3.6.2.2 Kti12 recruits Kti14 through its N-terminal domain

Despite the fact that we could not detect co-precipitation of Kti12 in the Elp1-TAP pull-down (Figure R22-B), we wanted to decipher how the interaction between Kti12, Kti14 and Elp1 occurs. Thus, using a *S. cerevisiae* Elp1-TAP as parental strain, I created mutant strains lacking Kti12 (Elp1-TAP_ΔKti12) or containing only its N or C-terminal domains (Elp1-TAP_Kti12$_{NTD}$ or Elp1-TAP_Kti12$_{CTD}$) by using CRISPR/Cas9 mutagenesis method. Mutant strains as well as the wild-type were transformed with either HA_3-Kti14 or Elp4-VSV plasmids (Figure R23-A). I then performed co-immunoprecipitation and western blot analyses as above (Material and Methods sections 2.7.8 and 2.5.9).

On the one hand, western blot of yeast strains transformed with HA_3-Kti14 (Figure R23-B) showed the presence of Elp1-TAP (172 kDa) in the input and precipitated fractions of all strains except in the untagged WT when using PAP reagent for detection. HA_3-Kti14 protein (59 kDa) was detected in the input fractions from all of the correspondingly transformed yeast strains including WT; however, strong bands of HA_3-Kti14 were only detected in the precipitated fractions

from the Elp1-TAP and Elp1-TAP_Kti12$_{NTD}$ strains, whereas in Elp1-TAP_ΔKti12, Elp1-TAP_Kti12$_{CTD}$ as well as the untagged WT strain weaker bands were detected. The negative control Stm1 was detected in the input fraction of all the yeast strains and absent in the precipitated ones demonstrating that immunoprecipitations were specific. On the other hand, western blots from co-immunoprecipitations of yeast strains transformed with Elp4-VSV (Figure R23-C) showed that Elp4 was presented in the input fractions of all correspondingly transformed strains. The protein was also present in the precipitated fractions in all TAP tagged strains even though some variability in band intensity was observed. Stm1 was presented in all cells but only in the input fractions demonstrating that immunoprecipitations were specific.

To better evaluate the impact of the different Kti12 mutants on the co-precipitation of Kti14 and Elp4, I quantified intensities of the bands on the western blots and determined the ratios between HA$_3$-Kti14 or Elp4-VSV and the Elp1-TAP band intensities (Figure R23-DE, see Material and Methods: 2.6.6). Taking co-precipitation of Kti14 from Elp1-TAP strain as 100% (Figure R23-D), I could see a clear decrease of the Kti14 band intensity in the co-immunoprecipitations from Elp1-TAP_ΔKti12 (23.5%) and Elp1-TAP_Kti12$_{CTD}$ (16.6%) strains, whereas in Elp1-TAP_Kti12$_{NTD}$ (71.3%) case Kti14 was present in significant amount. In contrast, comparison of band intensities of Elp4 (Figure R23-E) co-immunoprecipitated from Elp1-TAP_ΔKti12 (113.7%), Elp1-TAP_Kti12$_{NTD}$ (88.6%) and Elp1-TAP_Kti12$_{CTD}$ (78.7%) strains with Elp1-TAP (100%) strain showed little fluctuation. Altogether, the Elp1-TAP/Elp4-VSV co-immunoprecipitation data indicated that Elongator was present and stable in all genetic backgrounds, i.e., independent of the presence or absence of Kti12 or specific domains thereof. This positive control allows me to exclude that my other observations are a mere consequence of Elongator disruption. In this context, the Elp1-TAP/HA$_3$-Kti14 co-immunoprecipitation data indicated that recruitment of Kti14 to Elongator complex occurs efficiently only in the presence of Kti12 or its N-terminal domain.

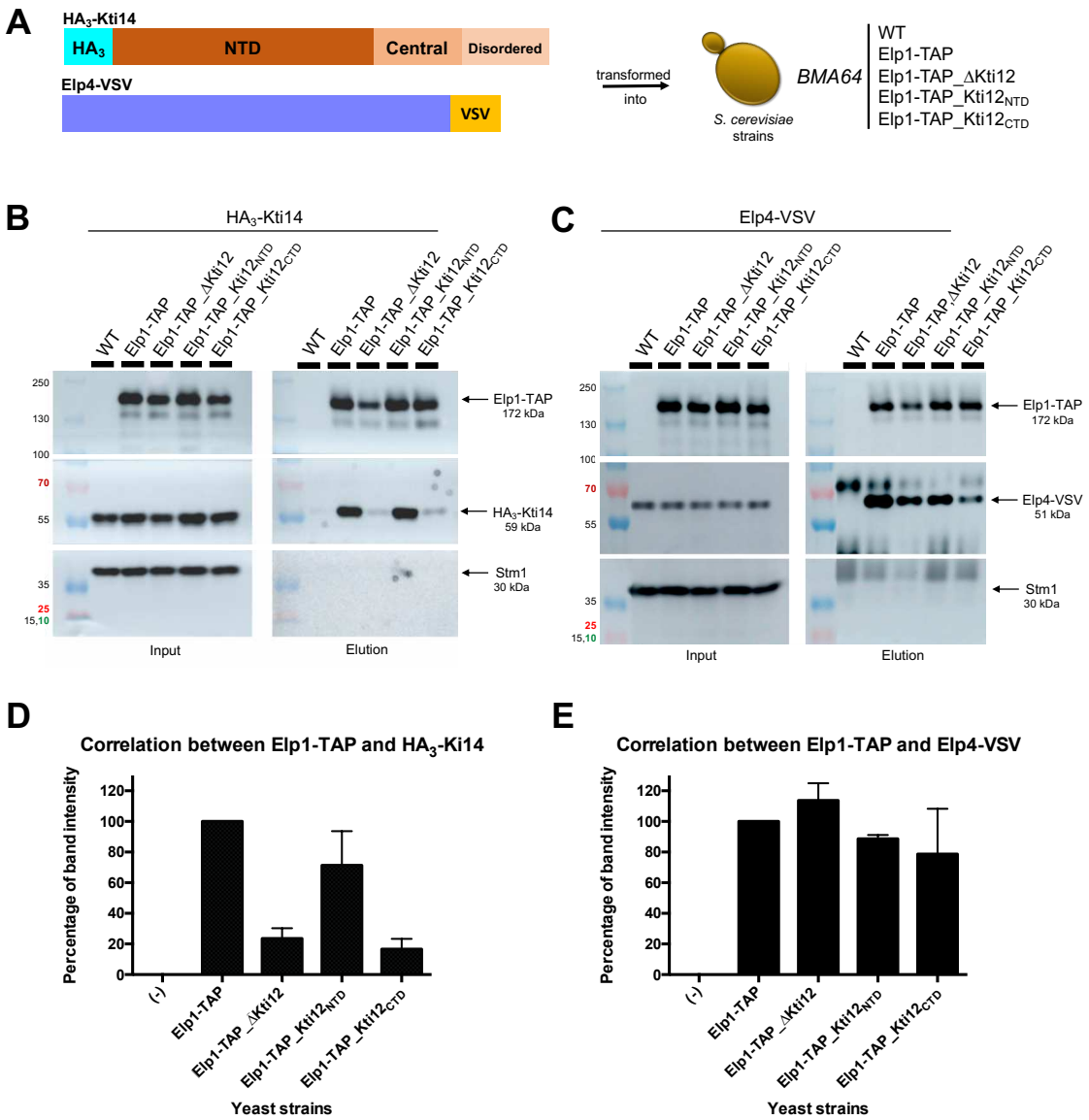


Figure R23. Co-IP demonstrating recruitment of Kti14 by the N-terminal domain of Kti12

(A) Scheme of plasmids and yeast strains used for Co-IP. The *S. cerevisiae* BMA64 WT, Elp1-TAP, Elp1-TAP_ΔKti12, Elp1-TAP_Kti12$_{NTD}$ or Elp1-TAP_Kti12$_{CTD}$ strains were transformed with plasmids encoding either HA$_3$-Kti14 or Elp4-VSV. Western blot of Co-IP pulling-down Elp1-TAP in the different strains transformed with HA$_3$-Kti14 **(B)** or Elp4-VSV **(C)** plasmids in 8% denaturing polyacrylamide gels. Quantification of band intensity. Percentage shows the ratio between the **(D)** HA$_3$-Kti14 and Elp1-TAP intensity, n=4 or **(E)** Elp4-VSV and Elp1-TAP intensity, n=3.

3.6.3 Co-expression of recombinant Kti12 and Kti14 proteins

3.6.3.1 Co-expression using GST tag

To obtain independent evidence confirming interaction of the N-terminal kinase domain of Kti12 with Kti14 and to demonstrate that this interaction is direct, I prepared different constructs to co-express recombinantly Kti12 and Kti14 or domains thereof. I constructed plasmids containing combinations of full or partial sequences of *S. cerevisiae* Kti12 and Kti14 organized as a mini-operon, with only one protein fused to a purification tag, to perform expression in *E. coli* before affinity purification (Figure R24-A). Despite the fact that my previous attempts to produce the recombinant Kti12 and Kti12$_{NTD}$ proteins were not successful, I hypothesized that their co-expression with the Kti14 putative partner could facilitate the protein expression and stabilization during purification. Because GST-Kti12 was better expressed than His$_6$-Kti12, I prioritized the former tag. In parallel, I tried to express various forms of GST-tagged Kti14 (Figure R24-A). *E. coli* cells were transformed with either the empty vector containing the GST tag (pGEX-6P) or with protein-coding constructs. Protein production was induced by IPTG before purification with glutathione beads (Material and Methods section 2.5.4). Analysis of the purified fraction by SDS-PAGE showed weak bands of around 61 kDa size in the constructs containing GST-Kti12 proteins. Similarly, a weak band of around 71 kDa size was detected in the construct containing GST-Kti14$_{1-394}$. However, no second band that could be attributed as Kti14$_{FL}$, Kti14$_{NTD}$ or Kti14$_{CTD}$ (57.34, 45.83 or 33.99 kDa, respectively) was detected in the constructs co-expressing Kti12-Kti14 (Figure R24-B). Protein expression from these constructs was also tested under different cell growth temperatures (16, 18, 25 or 37ºC); however, no additional bands could be detected neither with Coomassie blue nor with silver staining (data not shown).

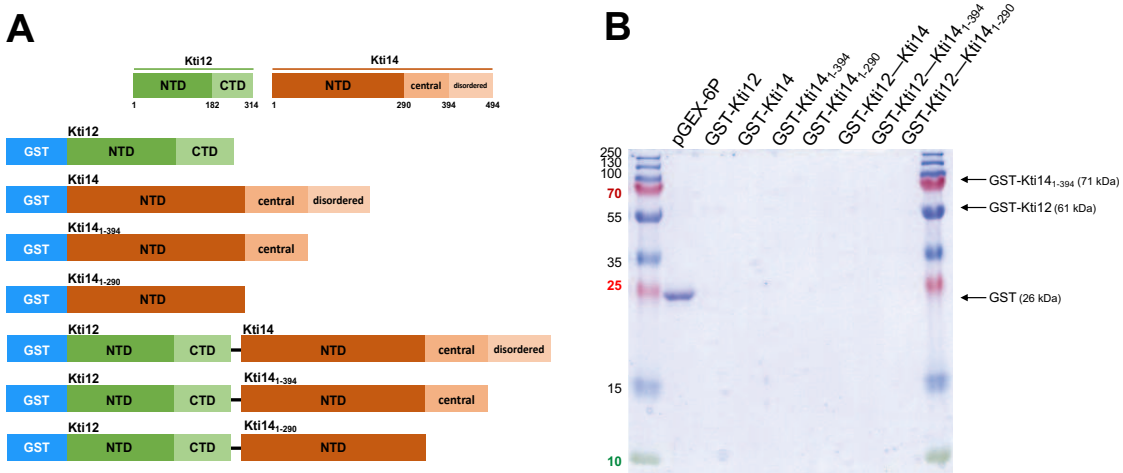


Figure R24. GST-tagged Kti12 and Kti14 co-expression

(A) Organization of co-expression GST-tagged (blue) constructs with Kti12 (green), Kti14 (orange) or their domains. **(B)** SDS-PAGE analysis of purified fractions. Expression was induced with 1mM of IPTG and analysis performed on 10% denaturing polyacrylamide gel.

3.6.3.2 Co-expression using N-terminal His_6 tag

In parallel, I created a second set of similar constructs to co-express Kti12-Kti14 with the first protein carrying this time a His_6 tag (Figure R25-A). Despite the presence of non-specific background, the profile of purified proteins showed two clearly detectable bands of approximately 47 kDa and 35 kDa size obtained from the construct containing $Kti14_{1\text{-}394}$–Kti12 (Figure R25-B). Protein identification by mass spectrometry demonstrated that these bands were indeed Kti14 and Kti12 respectively. Thus, I could observe that Kti12 interacts directly with $Kti14_{1\text{-}394}$. None of the other constructs show any relevant yield in the expression of both proteins.

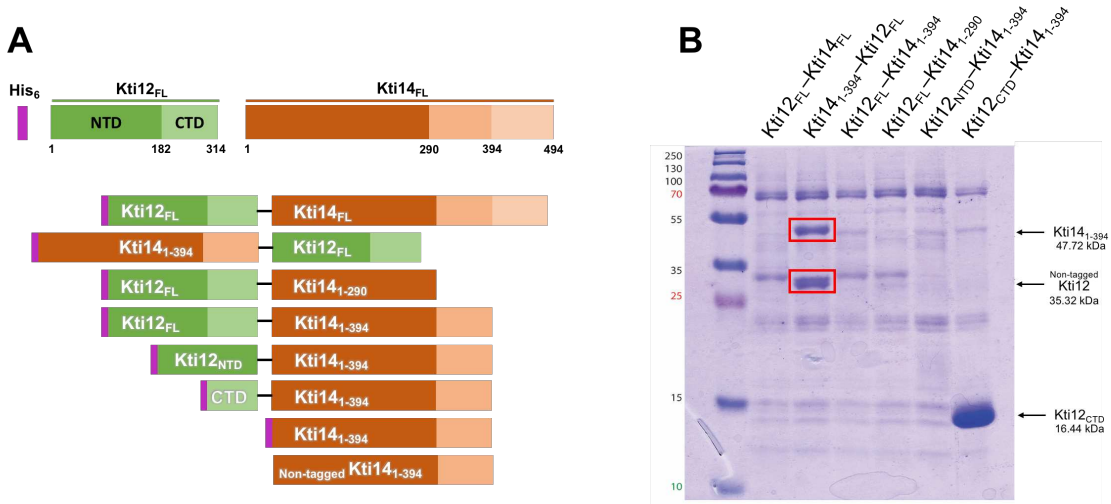


Figure R25. His_6-tagged Kti12 and Kti14 co-expression

(A) Organization of co-expression constructs with Kti12 (green), Kti14 (orange) and their different domains, and His_6 tag (purple). **(B)** Profiles of purified proteins. Expressions were induced with 1mM of IPTG and analyzed on a 15% polyacrylamide gel.

In order to improve the protein yield of $Kti14_{1\text{-}394}$–Kti12 for *in vitro* analyses, I created new constructs similar to the His_6-$Kti14_{1\text{-}394}$–Kti12 mini-operon described above. Each new construct contained a mix of different domains of Kti14 and Kti12 or carried a point mutation (Table R2).

Table R2. Constructs co-expressing Kti14-Kti12 fused to N-terminus His_6 tag

Construct	Mutations
His_6-$Kti14_{1\text{-}394}$–$Kti12_{NTD}$	—
His_6-$Kti14_{1\text{-}394}$–$Kti12_{CTD}$	—
His_6-$Kti14_{1\text{-}290}$–Kti12	—
His_6-$Kti14_{1\text{-}394}$–$Kti12_{K14A}$	Kti12 mutation K14A: inhibits ATP hydrolysis
His_6-$Kti14_{1\text{-}394}$–$Kti12_{D85A}$	Kti12 mutation D85A: inhibits ATP binding
His_6-$Kti14_{1\text{-}394\_K38R}$–Kti12	Kti14 mutation K38R: inhibits auto-phosphorylation

SDS-PAGE analyses of the protein profiles of the purified fractions failed to reveal constructs producing specific bands of interest above background, both by using Coomassie blue and silver staining (data not shown).

Independently, I also attempted to obtain $Kti14_{1\text{-}394}$ and $Kti14_{1\text{-}394}$–Kti12 by using an *in vitro* translation kit NEBExpress cell-free *E. coli* Protein Syntheses System (Material and Methods

section 2.5.11). However, apart from the positive control, no bands belonging to Kti12 or $Kti14_{1-394}$ were detected (data not shown).

3.6.3.3 Co-expression using C-terminal His_6 tag

Having observed that tagging Kti12 and $Kti12_{NTD}$ at their C-termini with a His_6 tag improves their expression and purification (see section 3.2.1.1), I created new mini-operon constructs to co-express Kti14-Kti12 proteins, containing either $Kti14_{1-394}$, or its mutant version K38R which inhibits auto-phosphorylation, and the different domains of Kti12 (Figure R26-A). These constructs were expressed in *E. coli*, proteins were purified on Ni-NTA as before. Silver staining of polyacrylamide gels of the purified factions revealed strong bands at the expected protein size of either Kti12-His_6 (36.14 kDa) or $Kti12_{NTD}$-His_6 (21.75 kDa) (Figure R26-B). An extra band (or doublet band) at approximately 45 kDa, was detected in all samples (Figure R26-C). Mass spectrometry analysis of the Coomassie-stained gel (Figure R23-D) revealed that the band appearing in the Kti12 purification (marked as 1 with red square) and present across the gel, did not contain Kti14 protein. This is consistent with the fact that no Kti14 should have been expressed in these cells. However, in the bands of the $Kti14_{1-394}$–Kti12, $Kti14_{1-394}$–$Kti12_{NTD}$ or $Kti14_{1-394\_K38R}$–Kti12 samples (marked as 2, 3 or 4 with a red square, respectively), the protein $Kti14_{1-394}$ was detected by mass spectrometry. Despite the fact that only low level of $Kti14_{1-394}$ is present, mass spectrometry confirmed its presence in all co-expression constructs encoding it. These results further confirm that Kti14 binds directly Kti12 and indicate further that the Kti12 N-terminal domain is sufficient to do so.

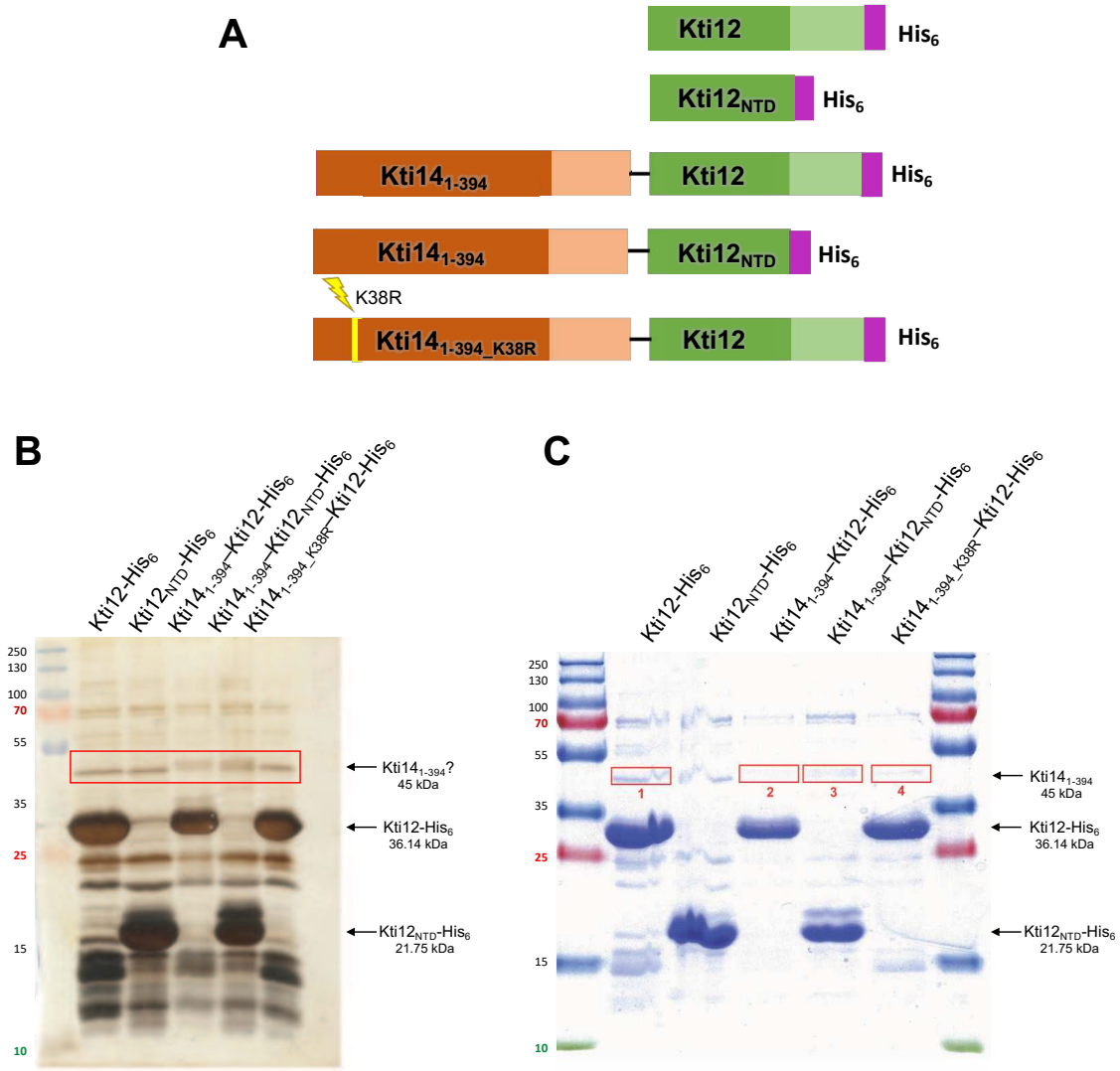


Figure R26. Purification using C-terminal His_6-tag of co-expressed Kti12-Kti14

(A) Organization of expression constructs with Kti12 (green), Kti14 (orange) and their different domains, and His_6 tag (purple). **(B)** Protein content of purified fractions after affinity chromatography on Ni-NTA beads analyzed on a 15% polyacrylamide SDS-PAGE. The gel was stained with silver. **(C)** As 'B' but the gel was stained with **C**oomassie blue. Red squares indicate proteins analyzed by mass spectrometry.

The successful attempt to obtain recombinantly expressed Kti12, $Kti12_{NTD}$ and $Kti12_{CTD}$ with a good yield opens a plethora of new possibilities to characterize the protein and its interaction with Kti14, Elongator as well as other cofactors in various *in vitro* assays.

4. DISCUSSION

Kti12 is a 35 kDa protein highly conserved among eukaryotes but with no homologue in archaea and bacteria. Kti12 has been identified as cofactor of the Elongator complex for post-transcriptional modification of tRNAs U_{34} to form 5-carboxylmethyluridine (cm^5U) moiety and its derivatives. Earlier studies have indicated a similarity of Kti12 with the archaeal protein *O*-phosphoseryl-$tRNA^{Sec}$ kinase (PSTK) which is involved in the biosynthesis of selenocysteine amino acid. Based on sequence similarities between these two distantly related proteins we expected a structural and functional similarity between both proteins. Our hypothesis, therefore, was that Kti12 has two putative domains: an N-terminal domain (NTD) with kinase signature and a C-terminal domain (CTD) with RNA binding capacity. In this thesis, I functionally characterized the two predicted domains of Kti12 and I also proposed a model describing how Kti12 modulates contribution of Kti14, another Elongator cofactor, to the Elongator function.

Our sequence alignment comparing thirty-two Kti12 and PSTK proteins, confirmed the presence of well conserved regions as well as highly conserved residues and motifs among the analyzed organisms, with the presence of extension regions that vary in length from kingdom to kingdom. Kti12-like protein sequences can be found annotated in different data bases including bacterial and archaeal ones; however, the evolutionary conservation of the Elongator holoenzyme to eukaryotic organisms as well as the phylogenetic divergence between Kti12 and PSTK suggests that the Kti12-like annotated proteins in bacteria and archaea are in fact PSTK-like sequences. However, this situation raises the question of why such a complex multi-subunit machinery needed to form 5-carboxylmethyluridine evolved in eukaryotes and moreover is evolutionarily conserved, when the same task can be achieved by a much simpler system in archaea and bacteria.

Our data revealed that the putative RNA-binding C-terminal domain of Kti12 has indeed tRNA binding properties. Different EMSA and competition EMSA indicated that $Kti12_{CTD}$ binds preferentially $tRNA^{Glu(UUC)}$ over U6 snRNA (Figure R8) or $dtRNA^{Glu(UUC)}$ (Figure R11) with only high excess of those nucleic acids (around 100 µM) capable to compete and displace partially the radiolabeled $tRNA^{Glu(UUC)}$. We also found that removal of any of the $tRNA^{Glu(UUC)}$ arms (delta acceptor stem, delta D-arm, delta anticodon arm and delta TΨC-arm) affected the formation of the tRNA-protein complex, requiring higher concentration of truncated constructs to match the competition capacity of the full-length tRNA (Figure R10). Interestingly, $Kti12_{CTD}$ did not show clear binding preference for tRNAs which *in vivo* are modified by Elongator ($tRNA^{Glu(UUC)}$, $tRNA^{Ser(UGA)}$) over the ones that are not modified ($tRNA^{Glu(CUC)}$, $tRNA^{Leu(UAG)}$) (Figures R8 and R9). These data suggest that Kti12 is not in charge of the tRNA discrimination to select the proper tRNA to be modified at wobble (U_{34}) position; and instead, I propose that it might just hold the tRNA for its proper accommodation on the Elp3 KAT catalytic site.

One of the biggest challenges we faced during this project was the inability to purify full-length Kti12 or its N-terminal domain (section 3.2.1). Since we got evidence of the presence of

the proteins in the 'total protein' fraction but not in the 'lysate' fraction at the following step of purification, we hypothesized that Kti12 and Kti12$_{NTD}$ were insoluble due to a possible protein aggregation. The structure of the PSTK (PDB ID 3ADC) revealed that the N-terminus of PSTK (MSELIILT) forms a β-sheet located in the buried internal part of the protein. In this foldi the 3 amino acids located at the very terminus (MSE) are exposed to solvent, allowing the tag to be located 'outside' of the protein core. The C-terminus of PSTK (DADRVLKEFKDLLNSY) forms an α-helix which is exposed to the surface of the protein. Due to the sequence similarities between Kti12 and PSTK, we assumed that tagging the N-terminal side of Kti12 would allow its proper purification as the N-terminal His$_6$-tag of PSTK was shown to express and purify without any problem[297]. However, this was not the case of Kti12 whose N-terminal tag (His$_6$ and GST) interfered with the ability to purify functional protein; resulting in the insolubility of the six-histidine tagged protein (Figure R3) or in its degradation after GST tag cleavage (Figure R4). A practical solution was found at the latter stage of this thesis by tagging Kti12 at its solvent exposed C-terminus. Despite the successful expression and purification of the Kti12 and Kti12$_{NTD}$ proteins using C-terminal His$_6$-tag, gel filtration chromatography demonstrated that Kti12 full length and its independently purified two domains had a tendency to form aggregates, since presence of the proteins was detected throughout the different fractions of the size exclusion chromatography (Figure R7).

Protein aggregation of Kti12$_{CTD}$ may explain the fuzzy bands detected in the EMSA competition experiments (Figures R9, R10 and R11), particularly in the positive control in which only Kti12$_{CTD}$ protein and radiolabeled tRNA$^{Glu(UUC)}$ were mixed. With this hypothesis, the presence of several fuzzy bands is the result of dimers, tetramers and other aggregated forms of Kti12$_{CTD}$, in which at least one protein binds a tRNA.

Similar to the Elongator subunits (including the catalytic subunit Elp3) and other Kti cofactors except Kti14, Kti12 is not an essential protein. Yeast cells lacking Kti12, Kti12$_{NTD}$, Kti12$_{CTD}$ or Elp3 are able to grow almost identically to wild-type cells under optimal growth conditions such as rich media YPDA at 25 or 30ºC (Figures R12-A, R14-A and R15-A). It is only under suboptimal conditions such as temperature (16 or 37ºC) or more demanding minimal media (CSM) that differences between strains become more evident, with the strongest effect in cells lacking Elp3 (Figures R12-B,C; R14-B,C and R15-B,C). These observations suggest that residual Elongator function may persist in the absence of Kti12 as its inactivation does not produce a phenotype as strong as complete inactivation of Elongator. Moreover, at the molecular level, the lack of cm^{5}U$_{34}$ derivatives at the wobble position of tRNAs as a consequence of an inactivated Elongator likely impacts the translation efficiency of the cells in all conditions. However, in optimal growth conditions of media and temperature, these perturbations on the general translation machinery do not result in reduced growth rate, since other processes within the cell remain optimal or are only modestly affected. Stress is a key factor that regulates protein synthesis, significant alterations to the proteome are required in order to allow cells to cope to more demanding conditions. *S. cerevisiae* for example, contains around 6000 open reading frames (ORFs), from which only a specific subset is expressed at a specific time (reviewed in [348]). Particularly in the

case of cellular stress caused by high suboptimal temperatures, only a few degrees above the optimal physiological conditions represent a challenge to the yeast cells because of the homeostasis perturbation caused by protein instability and conformational issues. In yeast, hundreds of genes are upregulated under stress temperatures[349], the majority of them are required to maintain proteostasis by balancing increased stress-dependent protein aggregation and degradation processes[349]. In this context, the high demand resulting from stress conditions, together with the perturbation in the translation efficiency and fidelity caused by a dysfunctional Elongator, boost the phenotypic differences between wild-type and mutant strains. Therefore, we can think that it is the cumulative effect of inefficient processes under suboptimal conditions, which finally differentiates the growth of yeast mutant strains compared to the wild-type (Figure D1).

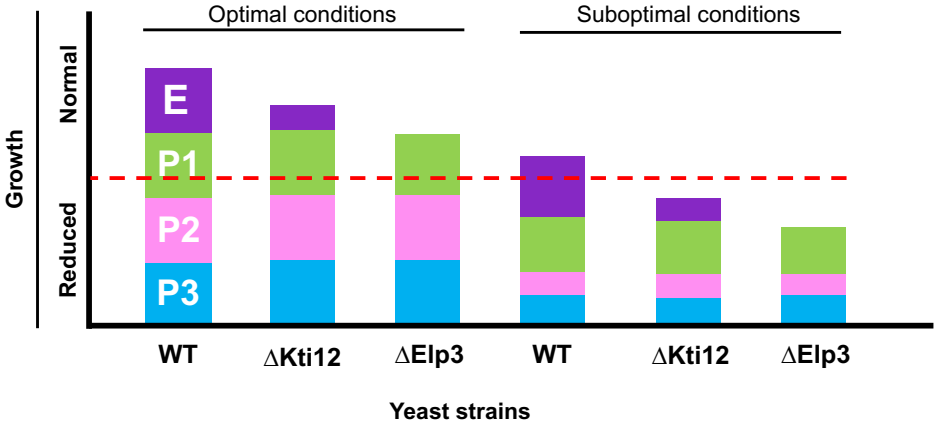


Figure D1. Scheme presenting interpretation of the cumulative impact different processes on growth efficiency

Contribution of Elongator dependent modification on cell growth is depicted with a purple rectangle labeled an 'E'. The cumulative contribution of different other cellular processes is symbolically represented by green, pink and blue rectangles labelled P1, P2 or P3. The dotted red line represents the threshold under which growth rate is reduced. Elongator function is reduced in a ΔKti12 strain and absent in a ΔElp3 strain. In suboptimal growth conditions the efficiency of some other processes (exemplified by P2) is reduced (smaller pink rectangle). Thus, even if the growth rates of the wild-type and mutant strains are comparable in optimal conditions, growth of the ΔKti12 and ΔElp3 strains is reduced compared to wild-type in suboptimal conditions.

Kti14 differs in some respect from the rest of Elongator subunits and Kti cofactors, due to the fact that it is an essential gene in yeast (section 1.3.4.3). Kti14 participates in many diverse pathways including but not limited to, ribosome biogenesis, DNA repair, autophagy, vesicle trafficking and microtubule assembly (reviewed in [300]). Thus, it is not possible to create a knockout yeast strain lacking Kti14 to demonstrate that the kinase activity detected in our kinase assays originates from this protein. Therefore, to show that Kti14 catalyzes directly the detected Elp1 phosphorylation, we propose a strategy to create a Kti14 mutant yeast strain in which the isoleucine 82 is replaced by a glycine (I82G) (Figure D2-A); this substitution is a 'gatekeeper' mutation that will structurally change the conformation of the Kti14 catalytic pocket rendering the mutant protein sensitive to bulky ATP-analogs (1NM-PPI or 3MB-PP). Using this strategy, we will be able to 'turn off' the kinase ability of Kti14 at will, *in vivo* or *in vitro* without interfering other pathways; and thus, identifying whether Kti14 is responsible for the detected Elp1 phosphorylation

(Figure D2-B). This approach has already been proven to be effective with many different kinases[252,350–352].

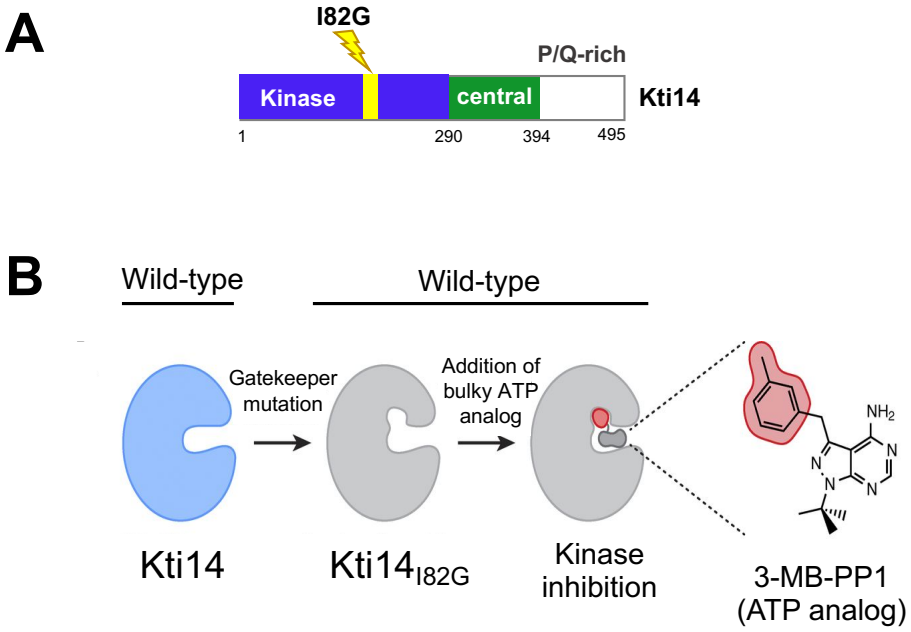


Figure D2. Kti14 gatekeeper mutation I82G

(A) Diagram of the Kti14 I82G mutation in its kinase domain. **(B)** Structural impact of I82G mutation on the catalytic pocket of Kti14.

Tandem affinity purification of Elp1, Kti12 or its domains identified a large number of proteins. Some are interacting with the Elongator complex or its cofactors (Figure R16-A), while others, in particular abundant proteins, are likely to be contaminants identified due to the high sensitivity of mass spectrometry. Except for the homodimer Kti11-Kti13, we could detect all the Elongator subunits and Kti cofactors in the elution fractions of Elp1-TAP and Kti12-TAP purifications. This observation suggests that Kti11-Kti13 interacts very transiently with Elongator, in agreement with their function of electron donor. Strikingly, none of the Elongator subunits or Kti cofactors (apart from Kti12 which was TAP-tagged) were detected in the Kti12$_{NTD}$-TAP elution fraction, and only Elp3 was detected in the Kti12$_{CTD}$-TAP elution fraction (Figure R16-A). The significance of the latter is unclear given that a single subunit of a large complex was identified. These findings are consistent with the results of the *in vivo* phenotypic assays in which deletion of the Kti12$_{NTD}$ or Kti12$_{CTD}$ resulted in phenotypes similar to the phenotype of cells lacking Elongator function completely exemplified by the ΔElp3 strain (Figures R12, R14 and R15): the absence of the Kti12$_{NTD}$ or Kti12$_{CTD}$ probably reduces the association of Kti12 to the Elongator complex.

The kinase assays developed in this project, allowed us to detect that Kti12 was not directly responsible for the phosphorylation of Elp1 since mutation of its kinase catalytic residues (K14A and D85A) did not abolish but limited the Elp1 phosphorylation (Figure R20). Instead, based on these results, Kti14 appears the most likely candidate to mediate such phosphorylation. The assay also supported our phenotypic results as the catalytic mutants K14A and D85A showed a wild-type growth phenotype and were still active (albeit at slightly reduced rate) in the kinase assay (Figure R14).

Data from the Elp1-TAP/HA_3-Kti14 co-immunoprecipitation (Figure R23) indicated that recruitment of Kti14 occurs in an efficient manner only in the presence of the N-terminal domain of Kti12. This result suggested the hypothesis that $Kti12_{NTD}$ recruits Kti14 to Elongator. Co-expression and purification of recombinant Kti12 and Kti14 proteins (Figures R25-B and R26-B) support a direct interaction between Kti12 and Kti14 and argue that this occurs through the N-terminal domain of Kti12. Signals were clear but faint and needed confirmation by mass spectrometry analysis since yields were limited. However, these observations together with the possibility to purify substantial amounts of full-length Kti12 and $Kti12_{NTD}$ with a C-terminal His_6 tag pave the way for optimization of the conditions to obtain stronger data.

As mentioned in the introduction, during the development of this project Krutyhołowa *et al.*[291] solved the structure of N-terminal domain of *Cheatomium thermophilum* Kti12 (*Ct*$Kti12_{NTD}$) In this publication, the authors also characterized Kti12 from *S. cerevisiae* and *C. thermophilum* biochemically and functionally. In this section I would like to go deeper in the similarities and differences of both works which are summarized in Figure D3.

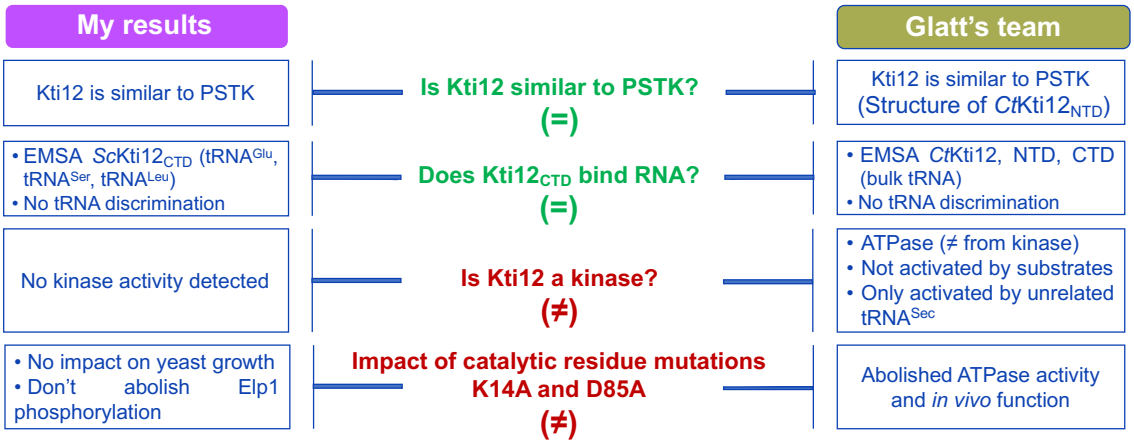


Figure D3. Comparison between my work and Glatt's team (Krutyhołowa *et al.*[291])

In the left column (purple) are summarized the results I obtained during my thesis work whereas in the right column (olive) are the results obtained by Glatt´s team. Topics are grouped by whether they agree (green) or disagree (red).

First, I would like to highlight the similarities between my results and the results obtained by Sebastian Glatt´s team. We both followed the assumption that Kti12 has similarities to PSTK; this assumption arose from two prior publications. In the first one, the authors observed an homology between Kti12 and PSTK which are evolutionary related proteins[289]. In the second one, the authors described sequence similarities and conservation of domains between the two proteins by using sequence alignments [290]. At the beginning of my work, limited information was available on Kti12 , whereas PSTK was a well-studied protein with a known crystal structure of an archaeal enzyme[297]. The starting research hypothesis was therefore to assume that Kti12 and PSTK share not only sequence but also functional similarities. So, we both extrapolated the PSTK domain organization into Kti12: a C-terminal domain that putatively binds RNA and an N-terminal domain

predicted as a kinase. Then, Glatt´s team solved the crystal structure of N-terminal domain of Kti12 from the fungus *Chaetomium thermophilum* (*Ct*Kti12*)*, the structure showed that Kti12 do not only resembles PSTK at a sequence level but also share structural similarities. The *Ct*Kti12 contains a globular ATPase domain formed by five parallel β-sheets and eight α-helices. Its N-terminal domain harbors a canonical P-loop motif which is located between the first alpha helix (α1) and the first beta sheet (β1). Similar to PSTK, *Ct*Kti12 display a short linker region between the NTD and CTD [291].

The structure confirmed structural similarities between Kti12 and PSTK. However, Glatt´s team could not solve the structure of the C-terminal domain of Kti12 and I faced a similar issue. I could produce large quantities of the protein. This was used and further purified by the structural biology platform that obtained some crystals that were not of sufficient quality to collect diffraction data. It is possible that the complexes between $Kti12_{CTD}$ and tRNA or even Kti12 and Kti14 will facilitate crystallization.

The similarity between the two projects was the discovery that the C-terminal domain of Kti12 binds RNA, preferentially tRNA and the protein does not discriminate between tRNAs which are modified or not by Elongator. I performed EMSA and competition EMSA using *Sc*Kti12 and *in vitro* transcribed glutamate ($tRNA^{Glu(UUC)}$) serine ($tRNA^{Ser(UGA)}$) or leucine ($tRNA^{Leu(UAG)}$) tRNAs of which glutamate and serine tRNAs are modified *in vivo* by Elongator but not the leucine tRNA. Glatt´s team also reached their conclusion by performing EMSA. They used three different versions of the *C. thermophilum* protein: Kti12 full-length, N-terminal or the C-terminal domain; and bulk yeast tRNA.

The next question that we have both asked: is Kti12 a kinase? Our answer differs from the conclusion of the Glatt team. I did not find any evidence pointing to Kti12 as a kinase protein, whereas Glatt´s team described Kti12 as an ATPase opening the possibility that it also has kinase activity. Indeed, I showed that Kti12 was not directly responsible for the Elp1 phosphorylation observed in the pull-downs of TAP-tagged Kti12 or $Kti12_{NTD}$ (but not of $Kti12_{CTD}$) (Figure R20) because the phosphorylation of Elp1 was not abolished by mutating the two putative kinase catalytic residues of Kti12 (Figure R20). Thus, I suggest that, instead, Kti12 plays a mediatory role by recruiting Kti14. In contrast, Glatt´s team presents Kti12 as an ATPase (which is different from a kinase) based on the structure of $Ct Kti12_{NTD}$ in which nucleotide binding pocket has the evolutionary conserved architecture typical of canonical P-loop ATPases (including kinases)[291]. The authors specifically tested the ATP hydrolysis activity (termed ATPase) of the N-terminal domain of Kti12, instead of kinase activity which is suggested by the similarities between Kti12 and PSTK. Interestingly, they found that the only tRNA able to activate the *Sc*Kti12 and *Ct*Kti12 ATPase activity in the presence of ATP and magnesium *in vitro* was *Homo sapiens* selenocysteine tRNA ($Hs tRNA^{Sec}$). None of the other tRNAs they tested were able to induce ATP hydrolysis. This result is puzzling because (i) neither *S. cerevisiae* nor *C. thermophilum* have $tRNA^{Sec}$, and (ii) known tRNAs requiring Kti12 for efficient modification by Elongator were not active. Despite the fact that the terms ATPase and kinase may be easily confused the reactions

catalyzed are different. On one hand, ATPases hydrolyze ATP using water to cleave the bond between the β and γ phosphates of ATP, producing phosphate ion (PO_4) and ADP (Figure D4-A). On the other hand, kinases transfer to a protein, tRNA or another substrate molecule (most often to a free hydroxyl group) the γ phosphate from ATP without the involvement of water (Figure D4-B). The fact that only a non-natural tRNA appeared to activate the hydrolytic ATPase reaction raises questions regarding the biological significance of this observation.

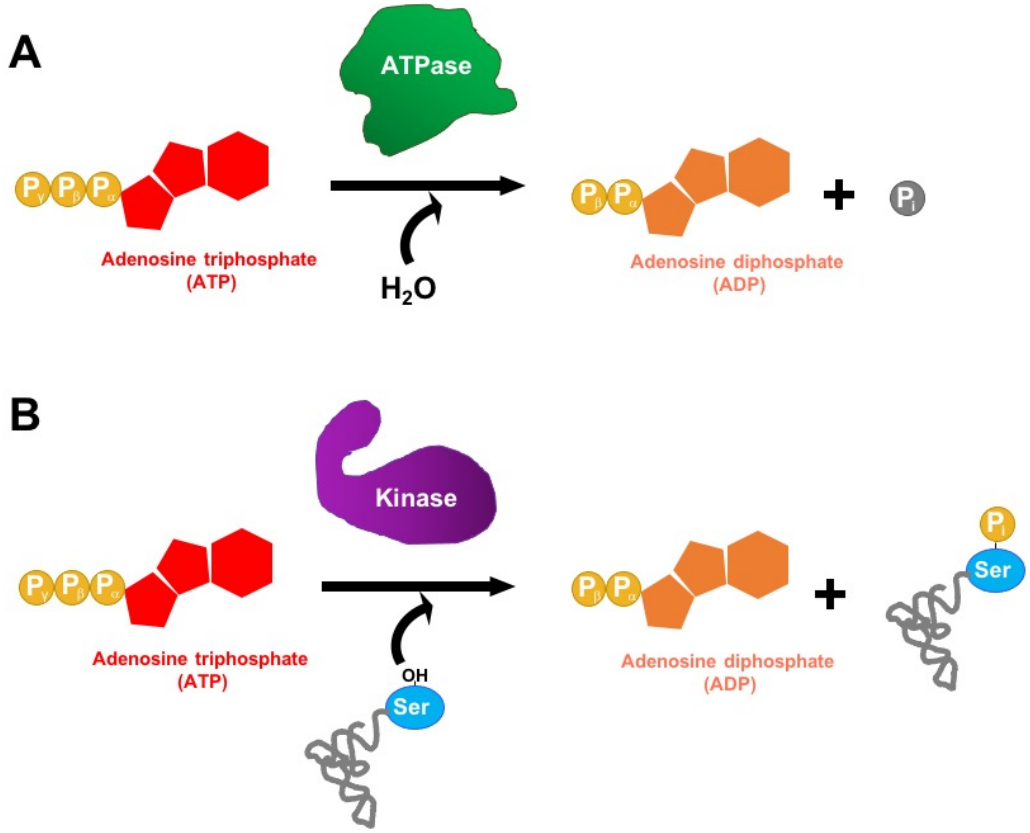


Figure D4. Difference in reaction between ATP hydrolysis (ATPase) and kinase

(A) Diagram of ATPase hydrolysis reaction. **(B)** Diagram of a kinase reaction exemplified by PSTK mediated phosphorylation of SectRNA-Ser.

Krutyhołowa *et al.* also characterized several mutants of catalytic residues of the *Sc*Kti12 kinase domain, including the K14A and D85A mutants. Interestingly, their phenotypic growth assay using *S. cerevisiae* mutants showed that the former mutation totally abolished Kti12 function, whereas the latter only decreased it. These results are in stark contrast with my findings where these same mutations did not have any effect on the yeast phenotype, the mutant strains being identical to wild-type (Figure R15).

As I discuss in the section 3.3.2, the strategy used to make a mutant strain impacts the phenotypic result. Introduction of a marker cassette downstream a gene is likely to lower its expression due to the production of aberrant mRNAs, some of which are subjected to NMD. The reduced expression of target gene following insertion of a downstream marker has been widely used in yeast to generate hypomorph alleles of essential genes. Therefore, constructing mutants using insertion of a downstream marker for selection may give incorrect results interpreted as loss

of activity[346].Schuldiner *et al.,* showed that a the 3' untranslated (3'UTR) region of a gene is disrupted by introducing an antibiotic resistance cassette (e.g., kanamycin), thus destabilizing its transcript and reducing the amount of mRNA by 2 to10 fold[353].

In comparison, the CRISPR/Cas9 technique allows to mutate the gene without interfering with its expression in an unwanted indirect manner since no nucleotides are changed except for the desired mutation. It is thus a "cleaner" strategy. Krutyhołowa *et al.* made their *Sc*Kti12 catalytic mutant strains by inserting the *Kl*TRP1 marker downstream of Kti12 whereas I made the same mutants by using CRISPR/Cas9. Given my observations of the difference of phenotypes between strains expressing the $Kti12_{NTD}$ constructed by the two methodologies, the difference between my phenotypic observations and their data is certainly explained by the different strategies.

Based on all the findings of this project, and published information about Kti12, Kti14 and the Elongator subunits, I propose a model for the Kti12 contribution to the Elongator complex (Figure D5). In this model $Kti12_{NTD}$ recruits Kti14 to phosphorylate Elp1 whereas $Kti12_{CTD}$ binds the tRNA to hold it in a proper position for the cm^5U_{34} modification at the Elp3 KAT catalytic site. According to Abdel-Fattah *et al.*[252] the sites phosphorylated by Kti14 might be the serines 1198 and 1202 of Elp1 which are localized close to its C-terminus and adjacent to the putative tRNA binding domain. This model leaves one open question: why does Kti12 maintain kinase catalytic residues in its NTD throughout the evolutionary process? One could imagine that those would have been lost if the only role of $Kti12_{NTD}$ is to recruit Kti14. Solving the structure of a $Kti12_{NTD}$-Kti14 heterodimer could possibly help to answer this question.

As mentioned before, elucidating the molecular details of enzymes catalyzing post-transcriptional modifications and their modulation is important, due to the fact that the alterations in post-transcriptional events and translational control are extremely detrimental to the human health. Deeper and clearer understanding on the detailed mechanism by which the Elongator holoenzyme and its partners regulate protein translation will open new search avenues for possible therapeutic targets against the diseases caused by a dysfunctional Elongator.

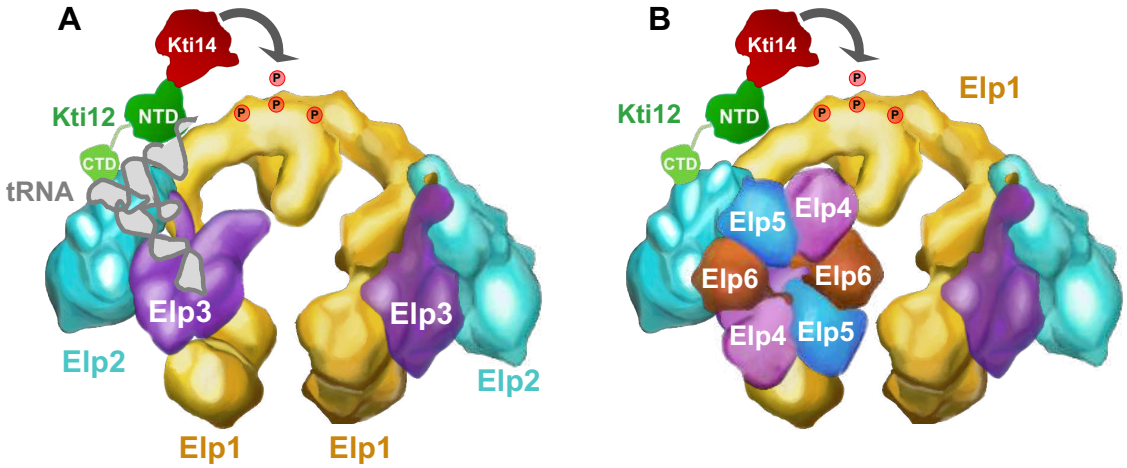


Figure D5. Proposed model for interaction of Kti12 and Kti14 with the Elongator complex

Two suggested functional states of the Elongator complex based on available information**. (A)** Model of Elp123 subcomplex bound to tRNA. The N-terminal domain of Kti12 (dark green) recruits Kti14 (red) to bring it closer to the C-terminal domain of Elp1 (yellow) to facilitate its phosphorylation. The C-terminal domain of Kti12 interacts with the TΨC-arm of tRNA. **(B)** In holoElongator the heterohexameric Elp456 ring (blue, pink and orange) occupies volume overlapping with the tRNA in (A) which suggests that the Elp456 participates in the release of tRNA. The order of binding of Kti12 and Kti14 remains unknown. The exact site of the Elongator complex contacted by Kti12 is not yet precisely defined.

5. REFERENCES

1. Crick, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–63 (1958).
2. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLOS Biol.* **15**, e2003243 (2017).
3. Karlson, P. The biological role of nucleic acids. *Trends Biochem. Sci.* **7**, 302–304 (1982).
4. Watson, J. D. Molecular Biology of the Gene James D. Watson. *Bioscience* **16**, 209–209 (1966).
5. Coffin, J. M. & Fan, H. The Discovery of Reverse Transcriptase. *Annu. Rev. Virol.* **3**, 29–51 (2016).
6. Crick, F. Central dogma of molecular biology. *Nature* **227**, 1970 (1970).
7. Anonymous. Central Dogma Reversed. *Nature* **226**, 1198–1199 (1970).
8. Wilczynski, S. P. Molecular Biology. in *Modern Surgical Pathology* 85–120 (Elsevier, 2009).
9. McCarthy, B. J. & Holland, J. J. Cultured Mammalian Cell Deoxyribonucleic Acid as a Template for in Vitro Protein Synthesis *. *Biochemistry* **5**, 1633–1637 (1966).
10. McCarthy, B. J., Holland, J. J. & Buck, C. A. Single-Stranded DNA as a Template for in Vitro Protein Synthesis. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 683–691 (1966).
11. McCarthy, B. J. & Holland, J. J. Denatured DNA as a direct template for in vitro protein synthesis. *Proc. Natl. Acad. Sci.* **54**, 880–886 (1965).
12. Pezo, V. *et al.* Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science (80-.).* **372**, (2021).
13. Maxwell. How do you draw a nucleotide and label its three basic parts? *Socratic* https://socratic.org/questions/how-do-you-draw-a-nucleotide-and-label-its-three-basic-parts#453384 (2017).
14. Guga, P., Maciaszek, A. D. & Radzikowska, E. Nucleotides and nucleic acids: mononucleotides. in 199–233 (2019).
15. Universe Review. DNA Structure. https://universe-review.ca/F11-monocell08.htm.
16. Nie, Y. & Stürzenbaum, S. R. Model Nematodes in Obesity Research. in *Animal Models for the Study of Human Disease* 267–280 (Elsevier, 2017).
17. Lewin, B. *Genes VIII*. *Chemistry & …* (Pearson Prentice Hall, 2004).
18. Meyers, R. a. *Encyclopedia of Molecular Cell Biology and Molecular Medicine*. *Molecular Cell* (2005).
19. Lodish, H., Berk, A., Zipursky, S. L. & Al., E. Molecular Cell Biology. in *W.H. Freeman* (2000).
20. Hubé, F. & Francastel, C. Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Front. Genet.* **9**, (2018).
21. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci.* **7**, 26 (2017).
22. Golden, B. L. Two distinct catalytic strategies in the HDV ribozyme cleavage reaction †. *Biochemistry* **50**, 9424–9433 (2011).
23. Hajnsdorf, E. & Kaberdin, V. R. RNA polyadenylation and its consequences in prokaryotes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, (2018).
24. Yeasmin, F., Yada, T. & Akimitsu, N. Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics. *Front. Genet.* **9**, 144 (2018).

25. Xing, J., Liu, H., Jiang, W. & Wang, L. LncRNA-Encoded Peptide: Functions and Predicting Methods. *Front. Oncol.* **10**, (2021).

26. Alberts, B. *et al.* How Cells Read the Genome: From DNA to Protein. in *Molecular biology of the cell* (eds. Anderson, M. & Granum, S.) (Garland Science, 2008).

27. Contributors, W. List of RNAs. *Wikipedia The Free Encyclopedia*.

28. Williams, R. The RNA Age: A Primer. *The Scientist* https://www.the-scientist.com/news-analysis/the-rna-age-a-primer-31520 (2017).

29. The Editors of Encyclopaedia Britannica. Ribosomal RNA. *Encyclopedia Britannica* (2011).

30. Long, E. O. & Dawid, I. B. Repeated Genes in Eukaryotes. *Annu. Rev. Biochem.* **49**, 727–764 (1980).

31. Hou, Y.-M. CCA addition to tRNA: Implications for tRNA quality control. *IUBMB Life* NA-NA (2010).

32. Szymanski, M., Deniziak, M. A. & Barciszewski, J. Aminoacyl-tRNA synthetases database. *Nucleic Acids Res.* **29**, 288–90 (2001).

33. Berg, M. D. *et al.* Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biol.* **16**, 1574–1585 (2019).

34. Parisien, M., Wang, X. & Pan, T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol.* **10**, 1853–1867 (2013).

35. Goodenbour, J. M. & Pan, T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res.* **34**, 6137–6146 (2006).

36. Database, G. tRNA. GtRNAdb Gene Symbol. http://gtrnadb.ucsc.edu/docs/naming/ (2019).

37. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).

38. Raina, M. & Ibba, M. tRNAs as regulators of biological processes. *Front. Genet.* **5**, 171 (2014).

39. Mustățea, G., Ungureanu, E. L. & Iorga, E. Protein acidic hydrolysis for amino acids analysis in food - progress over time: A short review. *Journal of Hygienic Engineering and Design* (2019).

40. Flatt, P. M. Peptide Bond Formation and Primary Protein Structure. in *Biochemistry - Defining Life at the Molecular Level* (Western Oregon University, 2019).

41. Alberts, B. *et al.* Proteins. in *Molecular biology of the cell* (eds. Anderson, M. & Granum, S.) (Garland Science).

42. Lewin, B. Proteins - GenesVIII. in *Chemistry & …* (eds. Carlson, G. & Challice, J.) (Pearson Prentice Hall, 2004).

43. Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry*. *Biochemistry* (2002).

44. Branden, C. I. & Tooze, J. *Introduction to Protein Structure*. *Introduction to Protein Structure* (2012).

45. McGraw-Hill Contributors. Protein Structure. *McGraw-Hill PreK-12* https://textflowprod-cf.prod.mheducation.com/prod/figures/007767670x/fig2_25.jpg.

46. Watson, J. D. *et al. Molecular Biology of the Gene*. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York* (2015).

47. Clancy, S. & Brown, W. Translation : DNA to mRNA to Protein. *Nat. Educ.* (2008).

48. Feklístov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* **68**, 357–376 (2014).

49. Busby, S. Promoter structure, promoter recognition, and transcription activation in

prokaryotes. *Cell* **79**, 743–746 (1994).

50. Macfarlane, W. M. Demystified ...: Transcription. *Mol. Pathol.* **53**, 1–7 (2000).

51. Amster-Choder, O. Transcriptional Regulation. in *Encyclopedia of Microbiology* 501–516 (Elsevier, 2009).

52. Seshasayee, A. S. N., Sivaraman, K. & Luscombe, N. M. An Overview of Prokaryotic Transcription Factors. in 7–23 (2011).

53. Venters, B. J. & Pugh, B. F. How eukaryotic genes are transcribed. *Crit. Rev. Biochem. Mol. Biol.* **44**, 117–141 (2009).

54. Sutherland, C. & Murakami, K. S. An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. *EcoSal Plus* **8**, ecosalplus.ESP-0004-2018 (2018).

55. Hocine, S., Singer, R. H. & Grunwald, D. RNA Processing and Export. *Cold Spring Harb. Perspect. Biol.* **2**, a000752–a000752 (2010).

56. Viktorovskaya, O. V. & Schneider, D. A. Functional divergence of eukaryotic RNA polymerases: unique properties of RNA polymerase I suit its cellular role. *Gene* **556**, 19–26 (2015).

57. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765 (2005).

58. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–43 (2015).

59. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* **1819**, 391–400 (2012).

60. Arimbasseri, A. G. & Maraia, R. J. Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. *Mol. Cell* **58**, 1124–1132 (2015).

61. Ghosh, A. & Lima, C. D. Enzymology of RNA cap synthesis. *Wiley Interdiscip. Rev. RNA* **1**, 152–172 (2010).

62. Baurén, G. & Wieslander, L. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183–92 (1994).

63. Vargas, D. Y. *et al.* Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054–65 (2011).

64. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–5 (2002).

65. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).

66. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* (2011).

67. Jurica, M. S. & Moore, M. J. Pre-mRNA Splicing. *Mol. Cell* **12**, 5–14 (2003).

68. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–21 (2014).

69. Patel, A. A. & Steitz, J. A. Splicing double: Insights from the second spliceosome. *Nature Reviews Molecular Cell Biology* (2003).

70. König, H., Matter, N., Bader, R., Thiele, W. & Müller, F. Splicing Segregation: The Minor Spliceosome Acts outside the Nucleus and Controls Cell Proliferation. *Cell* (2007).

71. Van Den Hoogenhof, M. M. G., Pinto, Y. M. & Creemers, E. E. RNA Splicing regulation and dysregulation in the heart. *Circulation Research* (2016).

72. van den Elzen, A. M. G. *et al.* Dissection of Dom34-Hbs1 reveals independent functions in two RNA quality control pathways. *Nat. Struct. Mol. Biol.* **17**, 1446–52 (2010).

73. Irimia, M. & Blencowe, B. J. Alternative splicing: Decoding an expansive regulatory layer. *Current Opinion in Cell Biology* (2012).

74. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. reports* **3**, 152–158 (2015).

75. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–6 (2008).

76. Hedberg, A. & Johansen, S. D. Nuclear group I introns in self-splicing and beyond. *Mob. DNA* **4**, 17 (2013).

77. Schmidt, C. A. & Matera, A. G. tRNA introns: Presence, processing, and purpose. *WIREs RNA* **11**, (2020).

78. Hajnsdorf, E. & Kaberdin, V. R. RNA polyadenylation and its consequences in prokaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* (2018).

79. Tudek, A., Lloret-Llinares, M. & Jensen, T. H. The multitasking polyA tail: nuclear RNA maturation, degradation and export. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, 20180169 (2018).

80. Dreyfus, M. & Régnier, P. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **111**, 611–3 (2002).

81. Neve, J., Patel, R., Wang, Z., Louey, A. & Furger, A. M. Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA Biol.* **14**, 865–890 (2017).

82. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* (2012).

83. Katahira, J. Nuclear export of messenger RNA. *Genes (Basel).* **6**, (2015).

84. Hopper, A. K. & Nostramo, R. T. tRNA Processing and Subcellular Trafficking Proteins Multitask in Pathways for Other RNAs. *Front. Genet.* **10**, (2019).

85. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nature Reviews Molecular Cell Biology* (2008).

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

87. Melnikov, S. *et al.* One core, two shells: Bacterial and eukaryotic ribosomes. *Nature Structural and Molecular Biology* (2012).

88. Kallberg, Y., Segerstolpe, Å., Lackmann, F., Persson, B. & Wieslander, L. Evolutionary Conservation of the Ribosomal Biogenesis Factor Rbm19/Mrd1: Implications for Function. *PLoS One* (2012).

89. Yusupova, G. & Yusupov, M. High-resolution structure of the eukaryotic 80S ribosome. *Annu. Rev. Biochem.* **83**, 467–86 (2014).

90. Greber, B. J. *et al.* The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283–6 (2014).

91. O'Brien, T. W. Properties of human mitochondrial ribosomes. *IUBMB Life* **55**, 505–13 (2003).

92. Boerema, A. P. *et al.* Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. *Nat. Plants* (2018).

93. Klinge, S. & Woolford, J. L. Ribosome assembly coming into focus. *Nature Reviews Molecular Cell Biology* (2019).

94. University of North Carolina Health Care. Researchers reveal hidden rules of genetics for how life on Earth began. http://astrobiology.com/2018/07/researchers-reveal-hidden-rules-of-genetics-for-how-life-on-earth-began.html (2018).

95. Hinnebusch, A. G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends in Biochemical Sciences* (2017).

96. Aitken, C. E. & Lorsch, J. R. A mechanistic overview of translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* **19**, 568–76 (2012).

97. Dever, T. E., Dinman, J. D. & Green, R. Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).

98. Agirrezabala, X. & Frank, J. Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Q. Rev. Biophys.* **42**, 159–200 (2009).

99. Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4**, a013706 (2012).

100. Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. Release factors differing in specificity for terminator codons. *Proc. Natl. Acad. Sci. U. S. A.* **61**, 768–74 (1968).

101. Nakamura, Y., Ito, K. & Isaksson, L. A. Emerging understanding of translation termination. *Cell* **87**, 147–50 (1996).

102. Adio, S. *et al.* Dynamics of ribosomes and release factors during translation termination in E. coli. *Elife* **7**, (2018).

103. Zhouravleva, G. *et al.* Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.* **14**, 4065–72 (1995).

104. Stansfield, I. *et al.* The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in Saccharomyces cerevisiae. *EMBO J.* **14**, 4365–73 (1995).

105. Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311–21 (2000).

106. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014–20 (1999).

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

108. Pisarev, A. V. *et al.* The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol. Cell* **37**, 196–210 (2010).

109. Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology* (2018).

110. Rodnina, M. V. Protein synthesis meets ABC ATPases: New roles for Rli1/ABCE1. *EMBO Reports* (2010).

111. Kakar, S. N., Zimmermann, F. & Wagner, R. P. Reversion behavior of isoleucine-valine mutants of yeast. *Mutat. Res. Mol. Mech. Mutagen.* **1**, 381–386 (1964).

112. Hawthorne, D. C. & Leupold, U. Suppressors in Yeast. in *Current Topics in Microbiology and Immunology* 1–47 (1974).

113. Kakar, S. N. Suppressor mutations for the isoleucine locus in Saccharomyces. *Genetics* **48**, 967–979 (1963).

114. Gaber, R. F. & Culbertson, M. R. The yeast frameshift suppressor gene SUF16-1 encodes an altered glycine tRNA containing the four-base anticodon 3'-CCCG-5'. *Gene* **19**, 163–172 (1982).

115. Sherman, F. An Introduction to the Genetics and Molecular Biology of the Yeast Saccharomyces cerevisiae. *Encycl. Mol. Biol. Mol. Medinicine* (1997).

116. Hendrick, J. L. *et al.* Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis. *Genetics* **157**, 1141 (2001).

117. Celis, J. E. & Piper, P. W. Nonsense suppressors in eukaryotes. *Trends Biochem. Sci.* **6**, 177–179 (1981).

118. Záhonová, K., Kostygov, A. Y., Ševčíková, T., Yurchenko, V. & Eliáš, M. An Unprecedented Non-canonical Nuclear Genetic Code with All Three Termination Codons Reassigned as Sense Codons. *Curr. Biol.* **26**, (2016).

119. Bröcker, M. J., Ho, J. M. L., Church, G. M., Söll, D. & Donoghue, P. O. Recoding the Genetic Code with Selenocysteine. *Angew. Chem. Int. Ed. Engl.* **53**, 319 (2014).

120. Stewart, J. W. & Sherman, F. Demonstration of UAG as a nonsense codon in bakers' yeast by amino-acid replacements in iso-1-cytochrome c. *J. Mol. Biol.* (1972).

121. Stewart, J. W., Sherman, F., Jackson, M., Thomas, F. L. X. & Shipman, N. Demonstration of the UAA ochre codon in bakers' yeast by amino-acid replacements in iso-1-cytochrome c. *J. Mol. Biol.* (1972).

122. Kim, D. & Johnson, J. Construction, expression, and function of a new yeast amber suppressor, tRNA(TrpA). *J. Biol. Chem.* (1988).

123. Porter, J. J., Heil, C. S. & Lueck, J. D. Therapeutic promise of engineered nonsense suppressor tRNAs . *WIREs RNA* (2021).

124. Temple, G. F., Dozy, A. M., Roy, K. L. & Wai Kan, Y. Construction of a functional human suppressor tRNA gene: An approach to gene therapy for β-thalassaemia. *Nature* (1982).

125. Bordeira-Carriço, R. *et al.* Rescue of wild-type E-cadherin expression from nonsense-mutated cancer cells by a suppressor-tRNA. *Eur. J. Hum. Genet.* (2014).

126. Gilmoee, R. A., Stewart, J. W. & Sherman, F. Amino acid replacements resulting from super-suppression of nonsense mutants of iso-1-cytochrome c from yeast. *J. Mol. Biol.* (1971).

127. Sussel, L., Vannier, D. & Shore, D. Epigenetic switching of transcriptional states: cis- and trans-acting factors affecting establishment of silencing at the HMR locus in Saccharomyces cerevisiae. *Mol. Cell. Biol.* (1993).

128. Yeast strains. https://www.phys.ksu.edu/gene/GENEFAQ.html.

129. Suizu, T. *et al.* L-canavanine resistance as a positive selectable marker in diploid yeast transformation through integral disruption of the can1 gene. *Agric. Biol. Chem.* (1989).

130. Rosenthal, G. A., Berge, M. A., Bleiler, J. A. & Rudd, T. P. Aberrant, canavanyl protein formation and the ability to tolerate or utilize L-canavanine. *Experientia* (1987).

131. Heck, A. M. & Wilusz, J. The Interplay between the RNA Decay and Translation Machinery in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **10**, a032839 (2018).

132. Luo, W., Johnson, A. W. & Bentley, D. L. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: Implications for a unified allosteric-torpedo model. *Genes Dev.* (2006).

133. Collart, M. A. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* **7**, 438–454 (2016).

134. Gorgoni, B. & Gray, N. K. The roles of cytoplasmic poly(A)-binding proteins in regulating gene expression: A developmental perspective. *Briefings in Functional Genomics and Proteomics* (2004).

135. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology* vol. 8 (2007).

136. Desrosiers, R., Friderici, K. & Rottman, F. Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci.* **71**, 3971–3975 (1974).

137. Furuichi, Y. *et al.* Methylated, blocked 5' termini in HeLa cell mRNA. *Proc. Natl. Acad. Sci. U. S. A.* (1975).

138. Wei, C. M. & Moss, B. Methylated nucleotides block 5'-terminus of vaccinia virus

messenger RNA. *Proc. Natl. Acad. Sci.* **72**, 318–322 (1975).

139. Chen, C. Y. A. & Shyu, A. Bin. Emerging Themes in Regulation of Global mRNA Turnover in cis. *Trends in Biochemical Sciences* vol. 42 (2017).

140. Karamyshev, A. L. & Karamysheva, Z. N. Lost in translation: Ribosome-associated mRNA and protein quality controls. *Frontiers in Genetics* vol. 9 (2018).

141. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell and Bioscience* vol. 7 (2017).

142. Kadaba, S. *et al.* Nuclear surveillance and degradation of hypomodified initiator tRNA Met in S. cerevisiae. *Genes Dev.* (2004).

143. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763–776 (2009).

144. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* (2005).

145. Vaňáčová, Š. *et al.* A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. *PLoS Biol.* **3**, e189 (2005).

146. Chernyakov, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5′-3′ exonucleases Rat1 and Xrn1. *Genes Dev.* (2008).

147. Whipple, J. M., Lane, E. A., Chernyakov, I., D'Silva, S. & Phizicky, E. M. The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. *Genes Dev.* (2011).

148. Alexandrov, A. *et al.* Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* (2006).

149. Cohn, W. E. & Volkin, E. Nucleoside-5′-phosphates from ribonucleic acid. *Nature* (1951).

150. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).

151. Franz, M., Hagenau, L., Jensen, L. R. & Kuss, A. W. Role of transfer RNA modification and aminoacylation in the etiology of congenital intellectual disability. *J. Transl. Genet. Genomics* (2020).

152. Pereira, M. *et al.* Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease. *International Journal of Molecular Sciences* (2018).

153. Lorenz, C., Lünse, C. E. & Mörl, M. Trna modifications: Impact on structure and thermal adaptation. *Biomolecules* (2017).

154. Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp. Mol. Med.* **52**, 400–408 (2020).

155. Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* (2017).

156. Nachtergaele, S. & He, C. The emerging biology of RNA post-transcriptional modifications. *RNA Biol.* **14**, 156–163 (2017).

157. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).

158. Saletore, Y. *et al.* The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13**, 175 (2012).

159. Grosjean, H. *DNA and RNA Modification Enzymes*. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (CRC Press, 2009).

160. Schubert, C. Technology Feature | Epitranscriptomics: RNA revisited. *Science (80-.).* **364**, 696.2-696 (2019).

161. Shi, H., Wei, J. & He, C. Where, When, and How: Context-Dependent Functions of RNA

Methylation Writers, Readers, and Erasers. *Mol. Cell* **74**, 640–650 (2019).

162. Esteve-Puig, R., Bueno-Costa, A. & Esteller, M. Writers, readers and erasers of RNA modifications in cancer. *Cancer Lett.* **474**, 127–137 (2020).

163. Jonkhout, N. *et al.* The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).

164. Arrondel, C. *et al.* Defects in t6A tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome. *Nat. Commun.* **10**, (2019).

165. Braun, D. A. *et al.* Mutations in KEOPS-complex genes cause nephritic syndrome with primary microcephaly. *Nat. Genet.* **49**, (2017).

166. Hawer, H. *et al.* Roles of Elongator Dependent tRNA Modification Pathways in Neurodegeneration and Cancer. *Genes (Basel).* **10**, 19 (2018).

167. Dawson, M. A. & Kouzarides, T. Cancer Epigenetics: From Mechanism to Therapy. *Cell* **150**, 12–27 (2012).

168. Adams, P. D. Epigenetics of cancer and aging: From mechanisms to therapies. *Exp. Gerontol.* **94**, 123 (2017).

169. Frye, M. & Blanco, S. Post-transcriptional modifications in development and stem cells. *Development* **143**, 3871–3881 (2016).

170. Blanco, S. *et al.* Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* **33**, 2020–39 (2014).

171. Chen, X. *et al.* 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat. Cell Biol.* **21**, 978–990 (2019).

172. Song, J. & Yi, C. Reading Chemical Modifications in the Transcriptome. *J. Mol. Biol.* **432**, 1824–1839 (2020).

173. Sergeeva, O. V., Bogdanov, A. A. & Sergiev, P. V. What do we know about ribosomal RNA methylation in Escherichia coli? *Biochimie* **117**, 110–118 (2015).

174. Stojković, V. & Fujimori, D. G. Radical SAM-Mediated Methylation of Ribosomal RNA. in *Methods in Enzymology* 355–376 (2015).

175. Holley, R. W., Everett, G. A., Madison, J. T. & Zamir, A. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J. Biol. Chem.* **240**, 2122–2128 (1965).

176. Machnicka, M. A., Olchowik, A., Grosjean, H. & Bujnicki, J. M. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* **11**, 1619–1629 (2014).

177. Krutyhołowa, R., Zakrzewski, K. & Glatt, S. Charging the code — tRNA modification complexes. *Curr. Opin. Struct. Biol.* **55**, 138–146 (2019).

178. Torres, A. G. *et al.* Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Res.* **43**, 5145–5157 (2015).

179. Motorin, Y. & Helm, M. tRNA Stabilization by Modified Nucleotides. *Biochemistry* **49**, 4934–4944 (2010).

180. Koh, C. S. & Sarin, L. P. Transfer RNA modification and infection – Implications for pathogenicity and host responses. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861**, 419–432 (2018).

181. Hou, Y. M., Gamper, H. & Yang, W. Post-transcriptional modifications to tRNA - A response to the genetic code degeneracy. *RNA* (2015).

182. Kopina, B. J. *et al.* Structure of a reaction intermediate mimic in t6A biosynthesis bound in the active site of the TsaBD heterodimer from Escherichia coli. *Nucleic Acids Res.* **49**, (2021).

183. Agris, P. F. *et al.* Celebrating wobble decoding: Half a century and still much is new. *RNA Biol.* **15**, 537–553 (2018).

184. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19**, 548–

555 (1966).

185. Agris, P. Wobble position modified nucleosides evolved to select transfer RNA codon recognition: A modified-wobble hypothesis. *Biochimie* **73**, 1345–1349 (1991).

186. Chen, C., Tuck, S. & Byström, A. S. Defects in tRNA Modification Associated with Neurological and Developmental Dysfunctions in Caenorhabditis elegans Elongator Mutants. *PLoS Genet.* **5**, e1000561 (2009).

187. Bjork, G. R., Huang, B., Persson, O. P. & Bystrom, A. S. A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. *RNA* **13**, 1245–1255 (2007).

188. Klassen, R. *et al.* Loss of Anticodon Wobble Uridine Modifications Affects tRNALys Function and Protein Levels in Saccharomyces cerevisiae. *PLoS One* **10**, e0119261 (2015).

189. Chen, Y.-T. *et al.* Loss of Mouse Ikbkap , a Subunit of Elongator, Leads to Transcriptional Deficits and Embryonic Lethality That Can Be Rescued by Human IKBKAP. *Mol. Cell. Biol.* **29**, 736–744 (2009).

190. Schaffrath, R. & Leidel, S. A. Wobble uridine modifications–a reason to live, a reason to die?! *RNA Biol.* **14**, 1209–1222 (2017).

191. Larsen, A. T., Fahrenbach, A. C., Sheng, J., Pian, J. & Szostak, J. W. Thermodynamic insights into 2-thiouridine-enhanced RNA hybridization. *Nucleic Acids Res.* **43**, 7675–7687 (2015).

192. Kurata, S. *et al.* Modified Uridines with C5-methylene Substituents at the First Position of the tRNA Anticodon Stabilize U·G Wobble Pairing during Decoding. *J. Biol. Chem.* **283**, 18801–18811 (2008).

193. Rozov, A. *et al.* Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, 10457 (2016).

194. Ranjan, N. & Rodnina, M. V. tRNA wobble modifications and protein homeostasis. *Translation* **4**, e1143076 (2016).

195. Lin, Z., Dong, M., Zhang, Y., Lee, E. A. & Lin, H. Cbr1 is a Dph3 reductase required for the tRNA wobble uridine modification. *Nat. Chem. Biol.* **12**, 995–997 (2016).

196. Kolaj-Robin, O. & Séraphin, B. Structures and Activities of the Elongator Complex and Its Cofactors. in *Enzymes* vol. 41 117–149 (2017).

197. Létoquart, J. *et al.* Insights into molecular plasticity in protein complexes from Trm9-Trm112 tRNA modifying enzyme crystal structure. *Nucleic Acids Res.* **43**, 10989–11002 (2015).

198. Jablonowski, D., Zink, S., Mehlgarten, C., Daum, G. & Schaffrath, R. tRNA Glu wobble uridine methylation by Trm9 identifies Elongator's key role for zymocin-induced cell death in yeast. *Mol. Microbiol.* **59**, 677–688 (2006).

199. Chen, C., Huang, B., Anderson, J. T. & Byström, A. S. Unexpected Accumulation of ncm5U and ncm5s2U in a trm9 Mutant Suggests an Additional Step in the Synthesis of mcm5U and mcm5s2U. *PLoS One* **6**, e20783 (2011).

200. Glatt, S. *et al.* The Elongator subcomplex Elp456 is a hexameric RecA-like ATPase. *Nat. Struct. Mol. Biol.* **19**, 314–320 (2012).

201. Otero, G. *et al.* Elongator, a Multisubunit Component of a Novel RNA Polymerase II Holoenzyme for Transcriptional Elongation. *Mol. Cell* **3**, 109–118 (1999).

202. Wittschieben, B. O. Overlapping roles for the histone acetyltransferase activities of SAGA and Elongator in vivo. *EMBO J.* **19**, 3060–3068 (2000).

203. Wittschieben, B. Ø. *et al.* A Novel Histone Acetyltransferase Is an Integral Subunit of Elongating RNA Polymerase II Holoenzyme. *Mol. Cell* **4**, 123–128 (1999).

204. Winkler, G. S., Kristjuhan, A., Erdjument-Bromage, H., Tempst, P. & Svejstrup, J. Q. Elongator is a histone H3 and H4 acetyltransferase important for normal histone

acetylation levels in vivo. *Proc. Natl. Acad. Sci.* **99**, 3517–3522 (2002).

205. Kouskouti, A. & Talianidis, I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *EMBO J.* **24**, 347–357 (2005).

206. Svejstrup, J. Q. Elongator complex: how many roles does it play? *Curr. Opin. Cell Biol.* **19**, 331–336 (2007).

207. Kim, J.-H., Lane, W. S. & Reinberg, D. Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proc. Natl. Acad. Sci.* **99**, 1241–1246 (2002).

208. Winkler, G. S. *et al.* RNA Polymerase II Elongator Holoenzyme Is Composed of Two Discrete Subcomplexes. *J. Biol. Chem.* **276**, 32743–32749 (2001).

209. Rahl, P. B., Chen, C. Z. & Collins, R. N. Elp1p, the Yeast Homolog of the FD Disease Syndrome Protein, Negatively Regulates Exocytosis Independently of Transcriptional Elongation. *Mol. Cell* **17**, 841–853 (2005).

210. Creppe, C. *et al.* Elongator Controls the Migration and Differentiation of Cortical Neurons through Acetylation of α-Tubulin. *Cell* **136**, 551–564 (2009).

211. Li, Q. *et al.* The Elongator Complex Interacts with PCNA and Modulates Transcriptional Silencing and Sensitivity to DNA Damage Agents. *PLoS Genet.* **5**, e1000684 (2009).

212. Versées, W., De Groeve, S. & Van Lijsebettens, M. Elongator, a conserved multitasking complex? *Mol. Microbiol.* **76**, 1065–1069 (2010).

213. Walker, J. *et al.* Role of Elongator Subunit Elp3 in Drosophila melanogaster Larval Development and Immunity. *Genetics* **187**, 1067–1075 (2011).

214. Kojic, M. *et al.* Elongator mutation in mice induces neurodegeneration and ataxia-like behavior. *Nat. Commun.* **9**, 3195 (2018).

215. Waszak, S. M. *et al.* Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature* **580**, 396–401 (2020).

216. Chen, C., Huang, B., Eliasson, M., Rydén, P. & Byström, A. S. Elongator Complex Influences Telomeric Gene Silencing and DNA Damage Response by Its Role in Wobble Uridine tRNA Modification. *PLoS Genet.* **7**, e1002258 (2011).

217. Esberg, A., Huang, B., Johansson, M. J. O. & Byström, A. S. Elevated Levels of Two tRNA Species Bypass the Requirement for Elongator Complex in Transcription and Exocytosis. *Mol. Cell* **24**, 139–148 (2006).

218. Karlsborn, T. *et al.* Elongator, a conserved complex required for wobble uridine modifications in Eukaryotes. *RNA Biol.* **11**, 1519–1528 (2014).

219. HUANG, B. An early step in wobble uridine tRNA modification requires the Elongator complex. *RNA* **11**, 424–436 (2005).

220. Selvadurai, K., Wang, P., Seimetz, J. & Huang, R. H. Archaeal Elp3 catalyzes tRNA wobble uridine modification at C5 via a radical mechanism. *Nat. Chem. Biol.* **10**, 810–812 (2014).

221. Lin, T.-Y. *et al.* The Elongator subunit Elp3 is a non-canonical tRNA acetyltransferase. *Nat. Commun.* **10**, 625 (2019).

222. Glatt, S. *et al.* Structural basis for tRNA modification by Elp3 from Dehalococcoides mccartyi. *Nat. Struct. Mol. Biol.* **23**, 794–802 (2016).

223. Benítez-Páez, A., Villarroya, M. & Armengod, M. E. The Escherichia coli RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy. *RNA* (2012).

224. Moukadiri, I. *et al.* Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Res.* (2009).

225. Zinshteyn, B. & Gilbert, W. V. Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling. *PLoS Genet.* **9**, e1003675 (2013).

226. Ranjan, N. & Rodnina, M. V. Thio-Modification of tRNA at the Wobble Position as Regulator of the Kinetics of Decoding and Translocation on the Ribosome. *J. Am. Chem. Soc.* **139**, 5857–5864 (2017).

227. Rezgui, V. A. N. *et al.* tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proc. Natl. Acad. Sci.* **110**, 12289–12294 (2013).

228. Bauer, F. *et al.* Translational Control of Cell Division by Elongator. *Cell Rep.* **1**, 424–433 (2012).

229. Schäck, M. A. *et al.* Eukaryotic life without tQCUG: the role of Elongator-dependent tRNA modifications in Dictyostelium discoideum. *Nucleic Acids Res.* **48**, 7899–7913 (2020).

230. Padgett, L. R. *et al.* Elp3 and RlmN: A tale of two mitochondrial tail-anchored radical SAM enzymes in Toxoplasma gondii. *PLoS One* **13**, e0189688 (2018).

231. Kaneko, T. Wobble modification differences and subcellular localization of tRNAs in Leishmania tarentolae: implication for tRNA sorting mechanism. *EMBO J.* **22**, 657–667 (2003).

232. Berná, G., Robles, P. & Micol, J. L. A Mutational Analysis of Leaf Morphogenesis in Arabidopsis thaliana. *Genetics* **152**, 729–742 (1999).

233. Nelissen, H. *et al.* The elongata mutants identify a functional Elongator complex in plants with a role in cell proliferation during organ growth. *Proc. Natl. Acad. Sci.* **102**, 7754–7759 (2005).

234. An, C., Wang, C. & Mou, Z. The Arabidopsis Elongator complex is required for nonhost resistance against the bacterial pathogens Xanthomonas citri subsp. citri and Pseudomonas syringae pv. phaseolicola NPS 3121. *New Phytol.* **214**, 1245–1259 (2017).

235. DeFraia, C. T., Zhang, X. & Mou, Z. Elongator subunit 2 is an accelerator of immune responses in Arabidopsis thaliana. *Plant J.* **64**, 511–523 (2010).

236. Wang, C. *et al.* Arabidopsis Elongator subunit 2 positively contributes to resistance to the necrotrophic fungal pathogens Botrytis cinerea and Alternaria brassicicola. *Plant J.* **83**, 1019–1033 (2015).

237. Silva, K. J. P. *et al.* The Arabidopsis ELP3/ELO3 and ELP4/ELO1 genes enhance disease resistance in Fragaria vesca L. *BMC Plant Biol.* **17**, 230 (2017).

238. Irving, P. *et al.* New insights into Drosophila larval haemocyte functions through genome-wide analysis. *Cell. Microbiol.* **7**, 335–350 (2005).

239. Bolukbasi, E. *et al.* Drosophila poly suggests a novel role for the Elongator complex in insulin receptor–target of rapamycin signalling. *Open Biol.* **2**, 110031 (2012).

240. Yu, D., Tan, Y., Chakraborty, M., Tomchik, S. & Davis, R. L. Elongator complex is required for long-term olfactory memory formation in Drosophila. *Learn. Mem.* **25**, 183–196 (2018).

241. Solinger, J. A. *et al.* The Caenorhabditis elegans Elongator complex regulates neuronal alpha-tubulin acetylation. *PLoS Genet.* **6**, e1000820 (2010).

242. Kawamuraa, K. & Maruyama, I. N. The Elongator Complex is Required to Maintain Locomotor Healthspan in 1 Caenorhabditis elegans. *bioRxiv* (2018).

243. Yoo, H., Son, D., Jang, Y.-J. & Hong, K. Indispensable role for mouse ELP3 in embryonic stem cell maintenance and early development. *Biochem. Biophys. Res. Commun.* **478**, 631–636 (2016).

244. Morini, E. *et al.* Developmental regulation of neuronal gene expression by Elongator complex protein 1 dosage. *bioRxiv* (2021).

245. Yajima, H., Tokunaga, M., Nakayama-Murayama, A. & Hishinuma, F. Characterization of IKI1 and IKI3 Genes Conferring pGKL Killer Sensitivity on Saccharomyces cerevisiae. *Biosci. Biotechnol. Biochem.* **61**, 704–709 (1997).

246. Xu, C. & Min, J. Structure and function of WD40 domain proteins. *Protein Cell* **2**, 202–214 (2011).

247. Lamb, J. R., Tugendreich, S. & Hieter, P. Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20**, 257–259 (1995).

248. D'Andrea, L. TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655–662 (2003).

249. Di Santo, R., Bandau, S. & Stark, M. J. R. A conserved and essential basic region mediates tRNA binding to the Elp1 subunit of the Saccharomyces cerevisiae Elongator complex. *Mol. Microbiol.* **92**, 1227–42 (2014).

250. Lange, A., McLane, L. M., Mills, R. E., Devine, S. E. & Corbett, A. H. Expanding the Definition of the Classical Bipartite Nuclear Localization Signal. *Traffic* **11**, 311–323 (2010).

251. Xu, H. *et al.* Dimerization of elongator protein 1 is essential for Elongator complex assembly. *Proc. Natl. Acad. Sci.* **112**, 10697–10702 (2015).

252. Abdel-Fattah, W. *et al.* Phosphorylation of Elp1 by Hrr25 Is Required for Elongator-Dependent tRNA Modification in Yeast. *PLoS Genet.* **11**, e1004931 (2015).

253. Waszak, S. M. *et al.* Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature* **580**, 396–401 (2020).

254. Fellows, J., Erdjument-Bromage, H., Tempst, P. & Svejstrup, J. Q. The Elp2 Subunit of Elongator and Elongating RNA Polymerase II Holoenzyme Is a WD40 Repeat Protein. *J. Biol. Chem.* **275**, 12896–12899 (2000).

255. Dong, C. *et al.* The Elp2 Subunit Is Essential for Elongator Complex Assembly and Functional Regulation. *Structure* **23**, 1078–1086 (2015).

256. Dauden, M. I. *et al.* Architecture of the yeast Elongator complex. *EMBO Rep.* **18**, 264–279 (2017).

257. Krogan, N. J. & Greenblatt, J. F. Characterization of a Six-Subunit Holo-Elongator Complex Required for the Regulated Expression of a Group of Genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **21**, 8203–8212 (2001).

258. Glatt, S. & Müller, C. W. Structural insights into Elongator function. *Curr. Opin. Struct. Biol.* **23**, 235–242 (2013).

259. Collum, R. G., Brutsaert, S., Lee, G. & Schindler, C. A Stat3-interacting protein (StIP1) regulates cytokine signal transduction. *Proc. Natl. Acad. Sci.* **97**, 10120–10125 (2000).

260. Mencalha, A. L. *et al.* Inhibition of STAT3-interacting protein 1 (STATIP1) promotes STAT3 transcriptional up-regulation and imatinib mesylate resistance in the chronic myeloid leukemia. *BMC Cancer* **14**, 866 (2014).

261. Abbassi, N.-H., Biela, A., Glatt, S. & Lin, T.-Y. How Elongator Acetylates tRNA Bases. *Int. J. Mol. Sci.* **21**, 8209 (2020).

262. Arragain, S. *et al.* Nonredox thiolation in tRNA occurring via sulfur activation by a [4Fe-4S] cluster. *Proc. Natl. Acad. Sci. U. S. A.* **114**, (2017).

263. Trievel, R. C. *et al.* Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc. Natl. Acad. Sci.* **96**, 8931–8936 (1999).

264. Tanner, K. G., Langer, M. R., Kim, Y. & Denu, J. M. Kinetic Mechanism of the Histone Acetyltransferase GCN5 from Yeast. *J. Biol. Chem.* **275**, 22048–22055 (2000).

265. Dauden, M. I. *et al.* Molecular basis of tRNA recognition by the Elongator complex. *Sci. Adv.* **5**, 2326 (2019).

266. Lin, Z. *et al.* Crystal Structure of Elongator Subcomplex Elp4–6. *J. Biol. Chem.* **287**, 21501–21508 (2012).

267. Gasior, S. L. *et al.* Assembly of RecA-like recombinases: Distinct roles for mediator proteins in mitosis and meiosis. *Proc. Natl. Acad. Sci.* **98**, 8411–8418 (2001).

268. Setiaputra, D. T. *et al.* Molecular architecture of the yeast Elongator complex reveals an unexpected asymmetric subunit arrangement. *EMBO Rep.* **18**, 280–291 (2017).

269. Dauden, M. I., Jaciuk, M., Müller, C. W. & Glatt, S. Structural asymmetry in the eukaryotic

Elongator complex. *FEBS Lett.* **592**, 502–515 (2018).

270. Strak, M. J. R., Boyd, A., Mileham, A. J. & Ramonos, M. A. The plasmid-encoded killer system ofKluyveromyces lactis: A review. *Yeast* **6**, 1–29 (1990).

271. Jablonowski, D. & Schaffrath, R. Zymocin, a composite chitinase and tRNase killer toxin from yeast. *Biochem. Soc. Trans.* **35**, 1533–1537 (2007).

272. LU, J. The Kluyveromyces lactis -toxin targets tRNA anticodons. *RNA* **11**, 1648–1654 (2005).

273. Huang, B., Lu, J. & Bystrom, A. S. A genome-wide screen identifies genes required for formation of the wobble nucleoside 5-methoxycarbonylmethyl-2-thiouridine in Saccharomyces cerevisiae. *RNA* **14**, 2183–2194 (2008).

274. Kawamoto, S. *et al.* Isolation and characterization of mutants of Saccharomyces cerevisiae Resistant to Killer Toxin of Kluyveromyces lactis. *J. Ferment. Bioeng.* **70**, 222–227 (1990).

275. Kishida, M. *et al.* Isolation and Genetic Characterization of pGKL Killer-insensitive Mutants (iki) from Saccharomyces cerevisiae. *Biosci. Biotechnol. Biochem.* **60**, 798–801 (1996).

276. Butler, A. R. *et al.* Two Saccharomyces cerevisiae genes which control sensitivity to G1 arrest induced by Kluyveromyces lactis toxin. *Mol. Cell. Biol.* **14**, 6306–6316 (1994).

277. Fichtner, L. & Schaffrath, R. KTI11 and KTI13, Saccharomyces cerevisiae genes controlling sensitivity to G1 arrest induced by Kluyveromyces lactis zymocin. *Mol. Microbiol.* **44**, 865–875 (2002).

278. Schaffrath, R. & Meinhardt, F. Kluyveromyces lactis zymocin and other plasmid-encoded yeast killer toxins. in 133–155 (2004).

279. Fichtner, L. *et al.* Elongator's toxin-target (TOT) function is nuclear localization sequence dependent and suppressed by post-translational modification. *Mol. Microbiol.* **49**, 1297–1307 (2003).

280. Br, C., Zabel, R., Liu, S., Stark, M. J. R. & Schaffrath, R. A versatile partner of eukaryotic protein complexes that is involved in multiple biological processes: Kti11Dph3. *Mol. Microbiol.* (2008).

281. Su, X., Lin, Z. & Lin, H. The biosynthesis and biological function of diphthamide. *Crit. Rev. Biochem. Mol. Biol.* **48**, 515–521 (2013).

282. Schaffrath, R., Abdel-Fattah, W., Klassen, R. & Stark, M. J. R. The diphthamide modification pathway from Saccharomyces cerevisiae--revisited. *Mol. Microbiol.* **94**, 1213–26 (2014).

283. Glatt, S. *et al.* Structure of the Kti11/Kti13 Heterodimer and Its Double Role in Modifications of tRNA and Eukaryotic Elongation Factor 2. *Structure* **23**, 149–160 (2015).

284. Zabel, R., Bär, C., Mehlgarten, C. & Schaffrath, R. Yeast α-tubulin suppressor Ats1/Kti13 relates to the Elongator complex and interacts with Elongator partner protein Kti11. *Mol. Microbiol.* **69**, 175–187 (2008).

285. Kolaj-Robin, O., McEwen, A. G., Cavarelli, J. & Séraphin, B. Structure of the Elongator cofactor complex Kti11/Kti13 provides insight into the role of Kti13 in Elongator-dependent tRNA modification. *FEBS J.* **282**, 819–833 (2015).

286. Sun, J. *et al.* Solution Structure of Kti11p from Saccharomyces cerevisiae Reveals a Novel Zinc-Binding Module † , ‡. *Biochemistry* **44**, 8801–8809 (2005).

287. Dong, M. *et al.* Dph3 Is an Electron Donor for Dph1-Dph2 in the First Step of Eukaryotic Diphthamide Biosynthesis. *J. Am. Chem. Soc.* **136**, 1754–1757 (2014).

288. Mehlgarten, C., Jablonowski, D., Breunig, K. D., Stark, M. J. R. & Schaffrath, R. Elongator function depends on antagonistic regulation by casein kinase Hrr25 and protein phosphatase Sit4. *Mol. Microbiol.* **73**, 869–881 (2009).

289. Sherrer, R. L., O'Donoghue, P., Sö Ll, D. & Söll, D. Characterization and evolutionary history of an archaeal kinase involved in selenocysteinyl-tRNA formation. *Nucleic Acids*

Res. **36**, 1247–1259 (2008).

290. Mehlgarten, C. *et al.* Use of a Yeast tRNase Killer Toxin to Diagnose Kti12 Motifs Required for tRNA Modification by Elongator. *Toxins (Basel).* **9**, 272 (2017).

291. Krutyhołowa, R. *et al.* Kti12, a PSTK-like tRNA dependent ATPase essential for tRNA modification by Elongator. *Nucleic Acids Res.* **47**, 4814–4830 (2019).

292. Fichtner, L. *et al.* Molecular analysis of KTI12/TOT4 , a Saccharomyces cerevisiae gene required for Kluyveromyces lactis zymocin action. *Mol. Microbiol.* **43**, 783–791 (2002).

293. Krutyhołowa, R., Reinhardt-Tews, A., Chramiec-Głąbik, A., Breunig, K. D. & Glatt, S. Fungal Kti12 proteins display unusual linker regions and unique ATPase p-loops. *Curr. Genet.* **66**, 823–833 (2020).

294. Smejda, M. *et al.* Same but different — Molecular comparison of human KTI12 and PSTK. *Biochim. Biophys. Acta - Mol. Cell Res.* **1868**, 118945 (2021).

295. Schmidt, R. L. & Simonović, M. Synthesis and decoding of selenocysteine and human health. *Croat. Med. J.* **53**, 535–550 (2012).

296. Serrão, V. H. B. *et al.* The unique tRNASec and its role in selenocysteine biosynthesis. *Amino Acids* **50**, 1145–1167 (2018).

297. Chiba, S., Itoh, Y., Sekine, S. & Yokoyama, S. Structural Basis for the Major Role of O-Phosphoseryl-tRNA Kinase in the UGA-Specific Encoding of Selenocysteine. *Mol. Cell* **39**, 410–420 (2010).

298. Araiso, Y. *et al.* Structure of a tRNA-dependent kinase essential for selenocysteine decoding. *Proc. Natl. Acad. Sci.* **106**, 16215–16220 (2009).

299. Vancura, A., Sessler, A., Leichus, B. & Kuret, J. A prenylation motif is required for plasma membrane localization and biochemical function of casein kinase I in budding yeast. *J. Biol. Chem.* (1994).

300. Ye, Q., Ur, S. N., Su, T. Y. & Corbett, K. D. Structure of the Saccharomyces cerevisiae Hrr25:Mam1 monopolin subcomplex reveals a novel kinase regulator. *EMBO J.* **35**, 2139–2151 (2016).

301. Corbett, K. D. & Harrison, S. C. Molecular Architecture of the Yeast Monopolin Complex. *Cell Rep.* **1**, 583–589 (2012).

302. Fukuchi, S., Hosoda, K., Homma, K., Gojobori, T. & Nishikawa, K. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct. Biol.* (2011).

303. Longenecker, K. L., Roach, P. J. & Hurley, T. D. Three-dimensional structure of mammalian casein kinase I: Molecular basis for phosphate recognition. *J. Mol. Biol.* (1996).

304. Truitt, M. L. & Ruggero, D. New frontiers in translational control of the cancer genome. *Nature Reviews Cancer* (2016).

305. DeOcesano-Pereira, C. *et al.* Post-Transcriptional Control of RNA Expression in Cancer. in *Gene Expression and Regulation in Mammalian Cells - Transcription From General Aspects* (2018).

306. Monies, D., Vågbø, C. B., Al-Owain, M., Alhomaidi, S. & Alkuraya, F. S. Recessive Truncating Mutations in ALKBH8 Cause Intellectual Disability and Severe Impairment of Wobble Uridine Modification. *Am. J. Hum. Genet.* **104**, (2019).

307. Slaugenhaupt, S. A. *et al.* Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* (2001).

308. Yoshida, M. *et al.* Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc. Natl. Acad. Sci.* **112**, 2764–2769 (2015).

309. Simpson, C. L. *et al.* Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. *Hum. Mol. Genet.* **18**, 472–481 (2009).

310. Bento-Abreu, A. *et al.* Elongator subunit 3 (ELP3) modifies ALS through tRNA modification. *Hum. Mol. Genet.* **27**, 1276–1289 (2018).

311. N, A., SP, O., MT, B., J, M. & M, H. Whole Exome Sequencing Reveals A Mutation in ELP2 Gene in Iranian Family Suffering from Autosomal Recessive Mental Retardation. *J. Mol. Genet. Med.* **12**, (2018).

312. Cohen, J. S. *et al.* ELP2 is a novel gene implicated in neurodevelopmental disabilities. *Am. J. Med. Genet. Part A* **167**, 1391–1395 (2015).

313. Addis, L. *et al.* Microdeletions of ELP4 Are Associated with Language Impairment, Autism Spectrum Disorder, and Mental Retardation. *Hum. Mutat.* **36**, 842–850 (2015).

314. Strug, L. J. *et al.* Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4). *Eur. J. Hum. Genet.* **17**, 1171–1181 (2009).

315. de Albuquerque Oliveira, A. C., Kappes, F., Martins, D. B. G. & de Lima Filho, J. L. The unique DEK oncoprotein in women's health: A potential novel biomarker. *Biomedicine and Pharmacotherapy* (2018).

316. Patton, E. E. *et al.* BRAF Mutations Are Sufficient to Promote Nevi Formation and Cooperate with p53 in the Genesis of Melanoma. *Curr. Biol.* **15**, 249–254 (2005).

317. Rapino, F. *et al.* Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* (2018).

318. Close, P. *et al.* DERP6 (ELP5) and C3ORF75 (ELP6) Regulate Tumorigenicity and Migration of Melanoma Cells as Subunits of Elongator. *J. Biol. Chem.* **287**, 32535–32545 (2012).

319. Jabłońska-Trypuć, A., Matejczyk, M. & Rosochacki, S. Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs. *J. Enzyme Inhib. Med. Chem.* **31**, 177–183 (2016).

320. Xu, Y. *et al.* Elongator promotes the migration and invasion of hepatocellular carcinoma cell by the phosphorylation of AKT. *Int. J. Biol. Sci.* **14**, 518–530 (2018).

321. Shimada, K. *et al.* A Novel Human AlkB Homologue, ALKBH8, Contributes to Human Bladder Cancer Progression. *Cancer Res.* **69**, 3157–3164 (2009).

322. Takeoka, S. *et al.* Amino-acid substitutions in the IKAP gene product significantly increase risk for bronchial asthma in children. *J. Hum. Genet.* **46**, 57–63 (2001).

323. Bednářová, A. *et al.* Lost in Translation: Defects in Transfer RNA Modifications and Neurological Disorders. *Front. Mol. Neurosci.* **10**, (2017).

324. Vaquer-Alicea, J. & Diamond, M. I. Propagation of Protein Aggregation in Neurodegenerative Diseases. *Annu. Rev. Biochem.* **88**, 785–810 (2019).

325. Delaunay, S. *et al.* Elp3 links tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis in breast cancer. *J. Exp. Med.* (2016).

326. World Health Organization. *Neurological disorders: public health challenges*. *WHO Library Cataloguing-in-Publication Data* (2017).

327. Dietrich, P. & Dragatsis, I. Familial dysautonomia: Mechanisms and models. *Genetics and Molecular Biology* (2016).

328. Karlsborn, T., Tükenmez, H., Chen, C. & Byström, A. S. Familial dysautonomia (FD) patients have reduced levels of the modified wobble nucleoside mcm5s2U in tRNA. *Biochem. Biophys. Res. Commun.* **454**, 441–445 (2014).

329. Manto, M., Gandini, J., Feil, K. & Strupp, M. Cerebellar ataxias: an update. *Curr. Opin. Neurol.* **33**, 150–160 (2020).

330. Redler, R. L. & Dokholyan, N. V. The complex molecular biology of Amyotrophic Lateral Sclerosis (ALS). in *Progress in Molecular Biology and Translational Science* (2012).

331. Kojic, M. *et al.* Elp2 mutations perturb the epitranscriptome and lead to a complex neurodevelopmental phenotype. *Nat. Commun.* **12**, 2678 (2021).

332. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* (2011).

333. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* (2000).

334. National Institutes of Health. NIH Curriculum Supplement Series [Internet]. *Biological Sciences Curriculum Study* (2007).

335. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* (2018).

336. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).

337. Globocan. Melanoma of skin cancer. (2020).

338. Alqathama, A. BRAF in malignant melanoma progression and metastasis: potentials and challenges. *Am. J. Cancer Res.* **10**, 1103–1114 (2020).

339. Rossi, A., Caracciolo, V., Russo, G., Reiss, K. & Giordano, A. Medulloblastoma: From molecular pathology to therapy. *Clinical Cancer Research* (2008).

340. Garcia-Lopez, J., Kumar, R., Smith, K. S. & Northcott, P. A. Deconstructing Sonic Hedgehog Medulloblastoma: Molecular Subtypes, Drivers, and Beyond. *Trends Genet.* **37**, 235–250 (2021).

341. Tahtamouni, L., Ahram, M., Koblinski, J. & Rolfo, C. Molecular Regulation of Cancer Cell Migration, Invasion, and Metastasis. *Analytical cellular pathology (Amsterdam)* (2019).

342. Baudin-Baillieu, A., Guillemet, E., Cullin, C. & Lacroute, F. Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast* **13**, 353–356 (1997).

343. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857–860 (2000).

344. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032 (1999).

345. Razvi, A. & Scholtz, J. M. Lessons in stability from thermophilic proteins.

346. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. (2008).

347. Humphreys, I. R. *et al.* Computed structures of core eukaryotic protein complexes. *Science (80-.).* (2021).

348. Crawford, R. A. & Pavitt, G. D. Translational regulation in response to stress in Saccharomyces cerevisiae. *Yeast* **36**, 5–21 (2019).

349. Mühlhofer, M. *et al.* The Heat Shock Response in Yeast Maintains Protein Homeostasis by Chaperoning and Replenishing Proteins. *Cell Rep.* **29**, 4593-4607.e8 (2019).

350. Ghalei, H. *et al.* Hrr25/CK1δ-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J. Cell Biol.* **208**, (2015).

351. Zhang, B. *et al.* The activity-dependent regulation of protein kinase stability by the localization to P-bodies. *Genetics* **203**, (2016).

352. Nemec, C. M. *et al.* Noncanonical CTD kinases regulate RNA polymerase II in a gene-class-specific manner. *Nat. Chem. Biol.* **15**, 123–131 (2019).

353. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).

6. APPENDIX

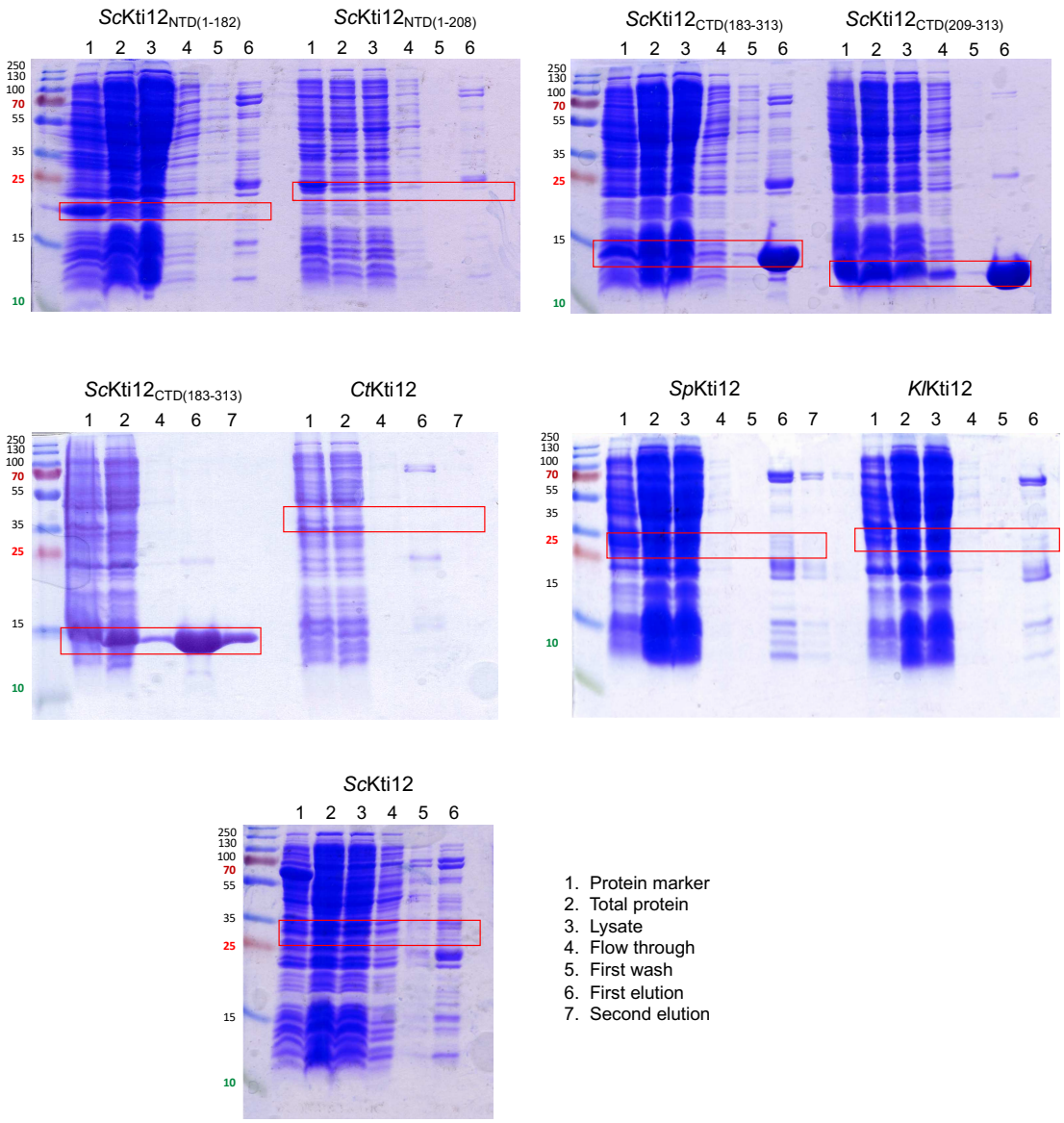


Figure A1. His_6-Kti12 purification

Acrylamide gels of protein purifications of constructs containing six-histidine tag (His_6) of *S. cerevisiae* (*Sc*Kti12), *S. pombe* (*Sp*Kti12), *K. lactis* (*Kl*Kti12), *C. thermophilum* (*Ct*Kti12).

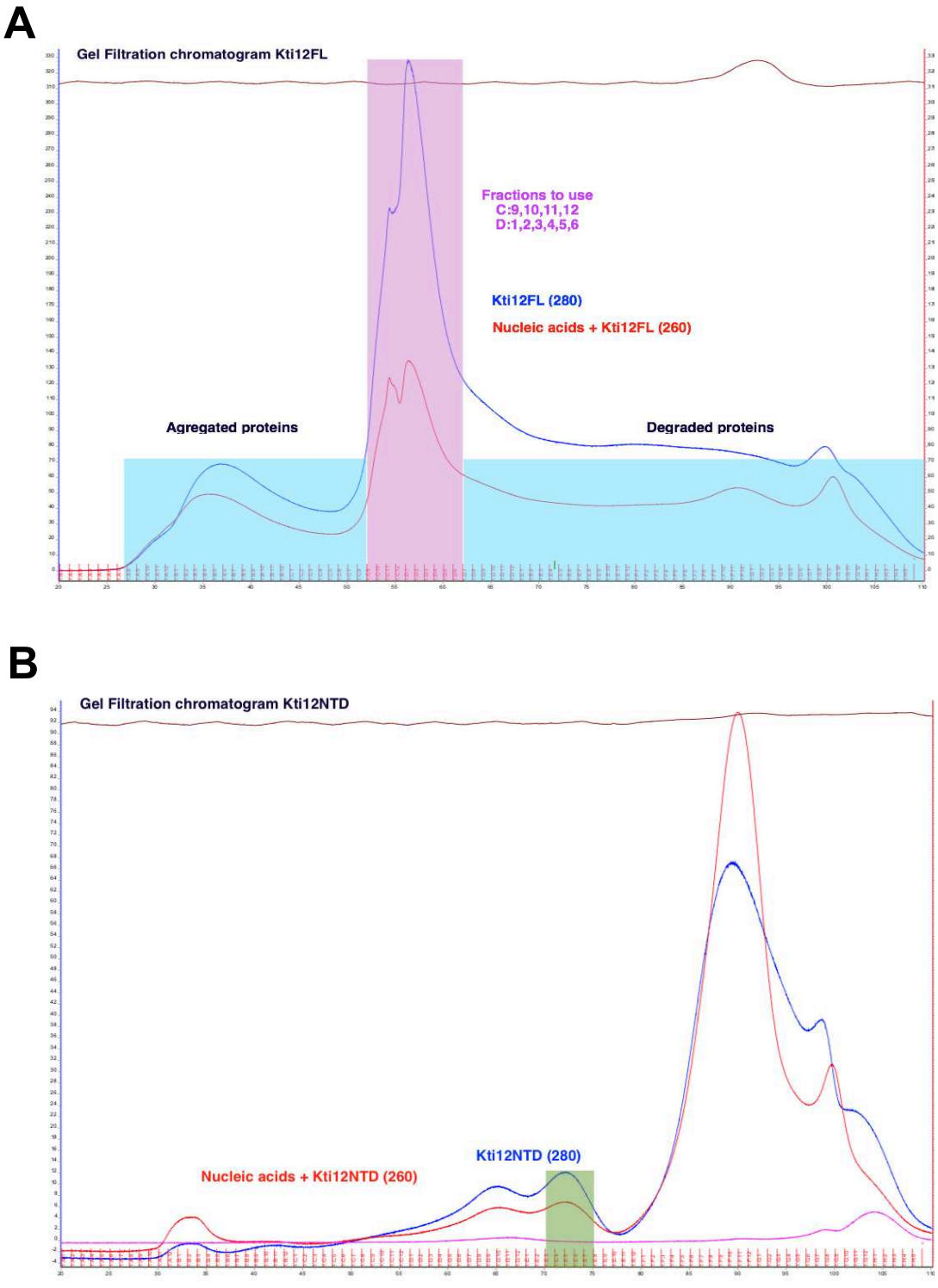


Figure A2. Chromatogram of gel filtration

Chromatogram of gel filtration after Nickel beads purification of **(A)** Kti12 full-length ($Kti12_{FL}$) and **(B)** Kti12 N-terminal domain ($Kti12_{NTD}$).

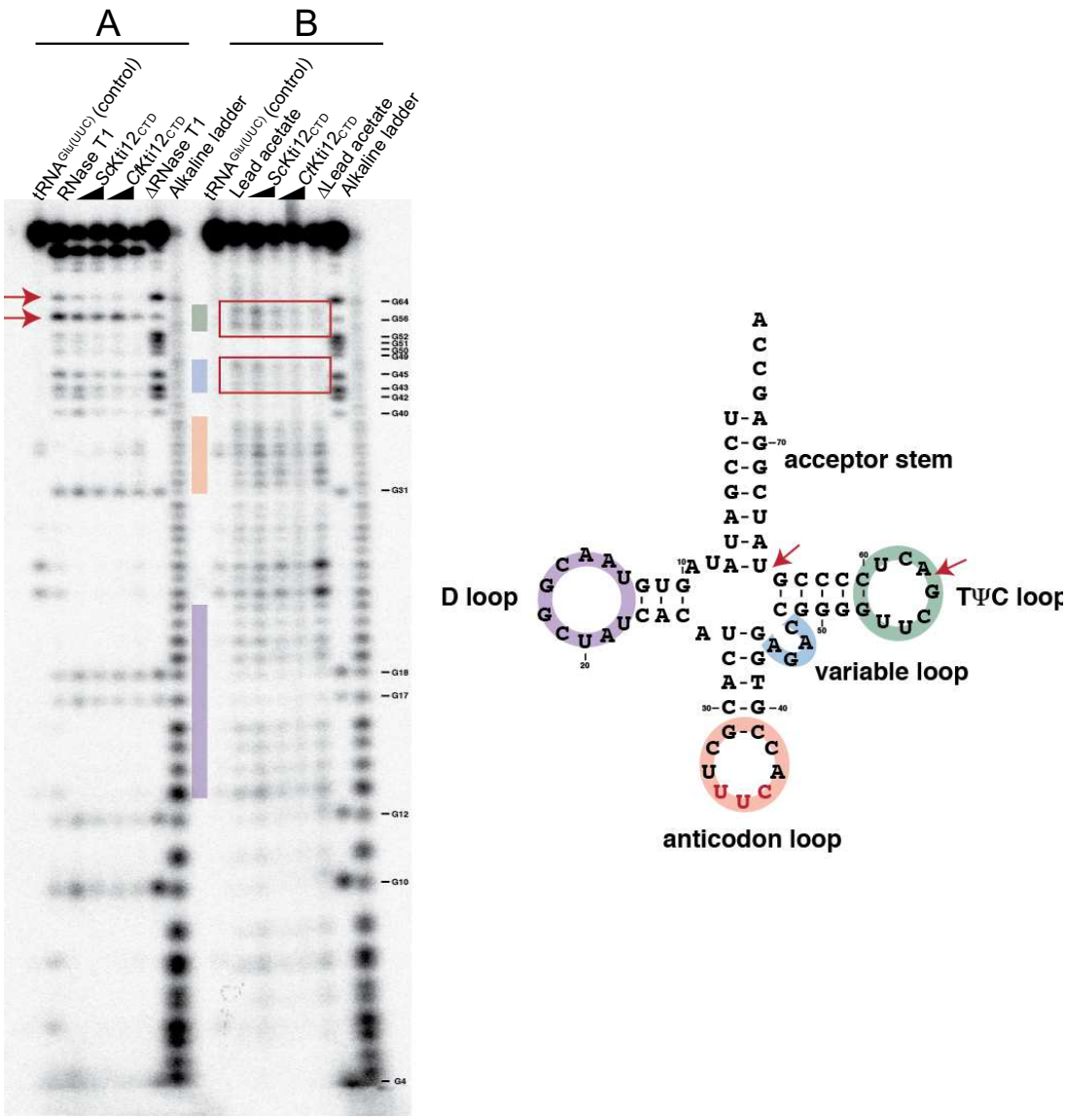


Figure A3. Autoradiogram of footprint experiments

$tRNA^{Glu(UUC)}$ was subjected to RNAaseT1 (panel **A**) or lead acetate (panel **B**) treatment. Addition of *S. cerevisiae* and *C. thermophilum* $Kti12_{CTD}$ appears to protect from modification the region of the TΨC-loop (highlighted in green and marked with red arrows).

Mónica PEÑA LUNA

Functional study of the Elongator complex from activity to diseases

Characterization of Kti12, a cofactor with a kinase domain

Résumé

Les altérations de modifications post-transcriptionnelles de l'ARN sont impliquées dans plusieurs troubles neurologiques et cancer. Ainsi, des mutations du complexe Elongator, dont le rôle principal est de modifier les U à la position « wobble » des ARNt (U_{34}), sont en effet particulièrement importantes car elles modulent la qualité de la traduction. Elongator est un complexe protéique de 6 sous-unités dont l'activité nécessite les cofacteurs Kti11-14. Outre l'absence de modifications d'ARNt, les défauts d'Elongator et de ses cofacteurs entraînent une mauvaise traduction et une agrégation des protéines. Du fait de la forte conservation de ces facteurs chez les eucaryotes, la levure *S. cerevisiae* s'est révélée être un système pertinent pour élucider leurs fonctions. Les rôles de Kti12 et Kti14 restant mal définis, mon projet vise à déchiffrer leurs activités moléculaires en utilisant la levure comme modèle. Mes résultats indiquent que Kti12 est composé de 2 domaines : un domaine N-terminal (NTD) présentant des similitudes avec les kinases et un domaine C-terminal (CTD) liant l'ARN. Basé sur mes données, nous proposons un nouveau modèle impliquant le recrutement de Kti14 par le NTD de Kti12 pour phosphoryler Elp1, stimulant ainsi Elongator. Mes résultats établissent une base solide pour déchiffrer rapidement les fonctions de Kti12 et Kti14 et le lien avec du complexe Elongator.

Résumé en anglais

Alterations of post-transcriptional modifications of RNA play a major role in several neurological disorders and cancer. A subset of such alterations results from mutations affecting the Elongator complex, whose main function is to modify uridine at the wobble position of tRNAs. Such tRNA modifications are particularly important because they modulate translation accuracy. Moreover, disruptions of some of Elongator subunits in humans have been associated with various neurological disorders and cancer types. Elongator is a 6-subunit protein complex whose activity requires the Kti11-14 cofactors. Besides reduced tRNA modifications, defects in Elongator and Kti-cofactors activities lead to poor translation and protein aggregation. As the precise roles of Kti12 and Kti14 remain ill-defined, my project aims at deciphering their molecular activities using yeast as a model. Given the high conservation of these factors in eukaryotes, the yeast *Saccharomyces cerevisiae* has proven to be a relevant system to elucidate their functions and their contribution to growth control. My current results indicate that Kti12 comprises 2 domains: an N-terminal domain (NTD) with sequence similarities to kinases and a C-terminal domain (CTD) with RNA binding properties. Based on our data, we propose a new model where Kti14 is recruited by the $Kti12_{NTD}$ in order to phosphorylate Elp1, stimulating thereby Elongator function. My current results provide a solid basis to define the exact function Kti12 and Kti14 and their link with the Elongator complex.