



Structure, fonction et évolution de la famille universelle Sua5/YrdC impliquée dans la synthèse du nucléoside modifié t6A

Adeline Pichard

► To cite this version:

Adeline Pichard. Structure, fonction et évolution de la famille universelle Sua5/YrdC impliquée dans la synthèse du nucléoside modifié t6A. Biochimie, Biologie Moléculaire. Université Paris Saclay (COMUE), 2017. Français. [⟨NNT : 2017SACLS371⟩](#). [⟨tel-02724874⟩](#)

HAL Id: tel-02724874

<https://theses.hal.science/tel-02724874v1>

Submitted on 2 Jun 2020

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

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



HAL Authorization

NNT : 2017SACLS371

Thèse de doctorat

Structure, fonction et évolution de la famille universelle Sua5/YrdC impliquée dans la synthèse du nucléoside modifié t⁶A

Thèse de doctorat de l'Université Paris-Saclay
Préparée à l'Université Paris-Sud

École doctorale n°569 : Innovation thérapeutique :
du fondamental à l'appliqué (ITFA)
Spécialité de doctorat: Biochimie et Biologie structurale

Thèse présentée et soutenue à Orsay, le 16 Novembre 2017, par

Adeline Pichard

Composition du Jury :

Philippe Minard Professeur, Université Paris-Sud (– I2BC UMR9198)	Président
Béatrice Clouet-d'Orval DR, CNRS (– LMGM UMR5100)	Rapporteur
Sébastien Ferreira-Cerca Professeur Assistant, Universität Regensburg (–Institut für Biochemie, Genetik und Mikrobiologie)	Rapporteur
Nicolas Leulliot Professeur, Université Paris Descartes (–LCRB UMR8015)	Examineur
Ciaran Condon Directeur de recherche, Université Paris Diderot (–IBPC UMR8261)	Examineur
Tamara Basta Chargé de recherche, Université Paris-Sud (– I2BC UMR9198)	Directeur de thèse
Herman van Tilbeurgh Professeur, Université Paris-Sud (– I2BC UMR9198)	Co-Directeur de thèse

Remerciements

Je tiens à remercier tous les membres du jury pour avoir accepté d'évaluer mon travail ainsi que pour la discussion passionnante que nous avons eu lors de ma soutenance.

Je remercie l'ensemble des membres de l'équipe Biologie Cellulaire des Archées: Juliette Létoquart pour ton immense soutien durant ma première année de thèse, Danièle Gadelle pour tes conseils et ta patience, Jacques Oberto pour ton humour et tes arguments, Evelyne Marguet pour ta bienveillance, Catherine Badel, Paul Villain Matteo Cossu et Ryan Catchpole pour votre bonne humeur (ainsi que votre patience pour mes bavardages et mes bêtises...), Myriam Swist pour ton rire et tes larmes, Patrick Forterre pour tes conseils pleins de sagesse et tes histoires incroyables.

De même je souhaite remercier les membres de l'équipe Fonction et Architecture des Assemblages Macromoléculaires : Bruno Collinet pour tes conseils en enzymologie et ta gentillesse, Inès Gallay et Stéphane Plancqueel pour votre aide avec la cristallographie, Nouredine Lazar pour vos conseils concernant la purification de protéines.

Je souhaiterais remercier tout particulièrement Violette Da Cunha pour ton aide sur le travail réalisé sur l'étude de l'évolution de la famille Sua5/YrdC ainsi que Marie-Claire Daugeron et Dominique Liger pour votre aide avec la génétique de la Levure.

Je voudrais également remercier quelques amis qui ont été un moteur pour moi ces dernières années et qui ont traversé les mêmes épreuves que moi lors de leur thèse : Hanane Issa, Simon Veyron, Margarida Gomez et Dimitri Cherier.

Je remercie aussi toute l'équipe Modélisation et Ingénierie des Protéines qui m'a toujours accueillie comme l'une des leurs !

Je remercie également mes parents et ma sœur pour leur soutien infaillible durant toutes ces années.

Enfin je souhaiterais remercier chaleureusement mes directeurs de thèse, Herman van Tilbeurgh et Tamara Basta.

Merci Herman pour tous tes enseignements qui m'ont permis de me construire un esprit critique dont je suis fière et que je continuerai à affiner. Merci aussi pour ta bienveillance qui n'a jamais failli malgré mes doutes.

Un immense merci à Tamara pour un millier de raisons que je ne peux malheureusement pas toutes énumérer ici. En premier, je te suis reconnaissante d'avoir toujours cru en moi malgré des passages quelques peu difficiles lors de ma thèse. Je souhaiterais également te remercier d'avoir pris du temps pour m'écouter et répondre à mes questions et mes doutes, qui n'avaient pas toujours à voir avec la science. Je suis enchantée d'avoir fait ma thèse avec toi et je n'ai pas une seule fois regretté ce choix.

Bref, je souhaiterais remercier toutes les personnes qui ont contribué d'une manière ou d'une autre à faire de ma thèse ce qu'elle a été. Merci à tous ! Travailler et/ou vivre avec vous a été un vrai plaisir et m'a permis de grandir et de m'améliorer. Je vous serez toujours reconnaissante pour cela et je n'oublierai jamais ces dernières années et ce qu'elles m'ont appris. Grâce à vous je suis prête pour de nouvelles aventures !

Table des matières

INTRODUCTION	12
I- Aperçu général de la traduction.....	14
II- Les ARNt.....	24
a. Structure des ARNt	26
b. Maturation des ARNt.....	30
c. Aminoacylation des ARNt	32
III- Les nucléosides modifiés	36
a. Diversité et répartition des nucléosides modifiés.....	36
c. Modifications de la base 34	42
d. Modifications de la base 37	46
IV- Le nucléoside modifié universel t⁶A	50
a. Histoire et découverte	50
b. Propriétés physico-chimiques.....	52
c. Fonctions cellulaires	54
d. Essentialité chez les organismes	58
e. L'ARNt initiateur : l'exception à la règle.....	62
f. Variants et dérivés	64
g. Le dérivé cyclique ct ⁶ A.....	66

V- Les t⁶A synthétases	72
a. Identification des t ⁶ A synthétases	72
La famille universelle Sua5/YrdC	72
La famille universelle Kae1/YgjD/Qri7	78
Les partenaires de Kae1/YgjD	84
b. La nomenclature des t ⁶ A synthétases	88
c. Caractérisation de la voie de biosynthèse	90
d. Les homologues de Sua5/YrdC et Kae1/YgjD/Qri7	94
La O-carbamoylase TobZ	94
La S-carbamoylase HypF	96
La t ⁶ A synthétase des virus géants	98
Les paralogues YciO et Slr0006	100
VI- La famille Sua5/YrdC	104
a. Structures d'YrdC et Sua5	106
b. Les modes de liaison des substrats	110
c. Le L-thréonyl-carbamoyl-AMP	116
d. Le mécanisme enzymatique de formation du TC-AMP	120
e. D'autres fonctions pour Sua5/YrdC	122
VII- Modifications des ARNt et maladies humaines	124
Problématique	128
Objectifs	132
Choix de l'approche	132
Choix des modèles d'études	136

ARTICLE I	140
Structure function analysis of Sua5 protein reveals novel functional motifs required for the synthesis of the universal t⁶A tRNA modification	
Introduction.....	142
Materials and methods.....	150
Cloning procedures and mutagenesis.....	150
Recombinant gene expression and protein purification.....	150
tRNA substrate production and purification	152
Differential Scanning Calorimetry.....	154
Size-Exclusion-Chromatography coupled with Multi-Angle Laser Light Scattering	154
Crystallization, data acquisition and structure determination	154
Sequence Alignment and structure comparison	156
<i>In vitro</i> assay for the synthesis of t ⁶ A-modified tRNA.....	156
<i>In vitro</i> ATP hydrolysis assay	158
Functional complementation assay in <i>S. cerevisiae</i> cells.....	160
Colony PCR.....	160
Western blot.....	162
Results.....	164
<i>Crystallographic structure of Sua5 from P. abyssi reveals a novel linker conformation</i>	<i>164</i>
<i>The YrdC-like domain of Pa-Sua5 is highly unstable in absence of the SUA5 domain</i>	<i>174</i>
<i>Mutation of conserved residues of linker and SUA5 domain produce stable proteins</i>	<i>176</i>
<i>The conserved residues in the linker and SUA5 domain are functionally important</i>	<i>180</i>
<i>Effect of the presence of Pa-KEOPS on Pa-Sua5 in vitro ATPase activity.....</i>	<i>188</i>
<i>Crystallographic structure of Pa-Sua5^{H234A/Y235A}</i>	<i>190</i>
Discussion	192

ARTICLE II	202
Evolution of the universal Sua5/YrdC enzymes: when simplification is the ultimate sophistication	
Introduction.....	204
Materials and methods.....	212
Phylogenetic analysis.....	212
Cloning procedures and mutagenesis.....	212
Recombinant gene expression and protein purification.....	214
tRNA substrate production and purification	214
Differential Scanning Calorimetry.....	214
Structure prediction and analysis	216
<i>In vitro</i> activity assays	216
Results and discussion	218
<i>Phylogenetic analysis of Sua5/YrdC family</i>	<i>218</i>
<i>Sequence alignments reveal differences between Sua5 and YrdC catalytic domains</i>	<i>228</i>
<i>The signature residue is in the active site.....</i>	<i>234</i>
<i>An atypical Sua5 in Archaeoglobus profundus.....</i>	<i>238</i>
<i>An integrative evolution model for the Sua5/YrdC family.....</i>	<i>244</i>
Conclusion	252

ARTICLE III.....	254

A complete pathway for the universal t⁶A tRNA modification in pandoraviruses

Introduction

Materials and methods.....	264

Structure prediction	264

<i>In silico</i> analysis.....	264

<i>In vitro</i> t ⁶ A activity assay	264

Results.....	265

<i>Sequence analysis of the t⁶A synthetase of pandoraviruses</i>	266

<i>Structure prediction of Ps-t⁶A synthetase</i>	270

<i>Genomic context of the pandoraviruses t⁶A synthetase</i>	272

<i>In vitro</i> activity of <i>P. salinus</i> t ⁶ A synthetase	274

Discussion and perspectives	276

Conclusion et perspectives	282
Conclusion et perspectives du travail de thèse.....	284
Projets à venir	288
Annexes.....	294
Annexe 1 : La famille Kae1/YgjD/Qri7	296
a. Propriétés structurales	296
b. Le complexe DEZ chez les bactéries.....	298
Les partenaires d'YgjD	298
Structure quaternaire du complexe DEZ	300
c. Le complexe KEOPS chez les eucaryotes/archées.....	304
Les partenaires de Kae1	304
Structure quaternaire de KEOPS.....	310
D'autres fonctions pour KEOPS ?	314
d. L'orthologue Qri7 chez les mitochondries.....	318
e. Evolution de la famille Kae1/YgjD/Qri7	320
Annexe 2 : Chapitre : la synthèse de t⁶A chez les archées	322
Annexe 3 : Table des oligonucléotides	324
Annexe 4 : Propriétés biochimiques des protéines recombinantes	326
Liste des figures et de tables.....	328
Liste des abréviations	330
Références bibliographiques	334

INTRODUCTION

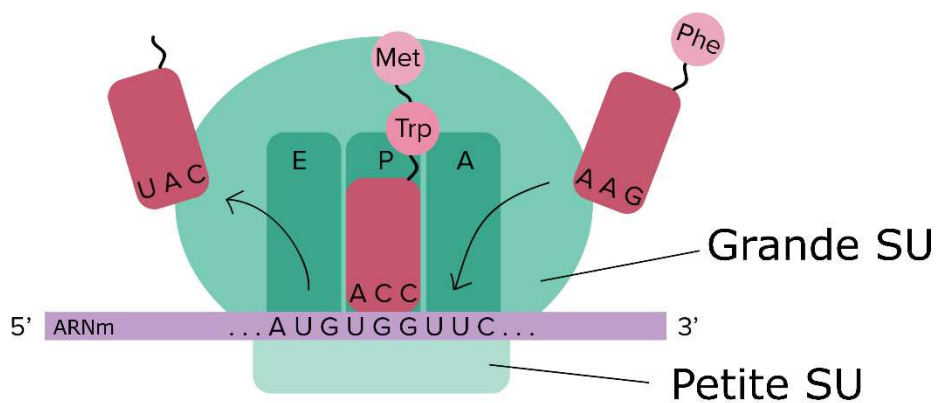


Figure 1. Organisation universelle des ribosomes lors de la traduction

Les grande et petite sous-unités sont en vert. L'ARNm est en violet. Les ARNt sont en rouge. Les sites de liaisons des ARNt, A, P et E, sont visibles au sein de la grande sous-unité. La description des rôles de chacun de ces éléments est explicitée dans le texte principal. SU = sous-unité. D'après <https://fr.khanacademy.org>.

I- Aperçu général de la traduction

La traduction est le processus cellulaire qui permet le passage entre l'information génétique, écrite sous forme de triplets de nucléosides appelés codons, et l'information protéique, écrite sous forme d'acides aminés (Crick et al., 1961; Nirenberg, 1965). Bien qu'il existe des variations spécifiques entre les archées, les bactéries et les eucaryotes, la traduction est un processus fondamental dont certains mécanismes moléculaires sont parmi les plus conservés au sein du vivant (Ganoza et al., 2002). La synthèse protéique fait intervenir de nombreux acteurs dont les principaux sont : (i) le ribosome, assemblage macromoléculaire complexe et dynamique qui joue le rôle de machinerie centrale où a lieu la synthèse des protéines, (ii) l'ARN messenger (ARNm), produit de la transcription des gènes qui porte l'information génétique à traduire, et (iii) les ARN de transfert (ARNt), qui jouent le rôle de décodeur entre langage génétique et langage protéique. A ces acteurs, il faut ajouter la cohorte de facteurs protéiques et d'enzymes de modification qui participent, directement ou indirectement, à la traduction. Cette dernière s'organise en quatre étapes : l'initiation, l'élongation, la terminaison et le recyclage.

Les ribosomes sont des particules ribonucléoprotéiques (RNP), c'est-à-dire qu'ils sont composés d'ARN et de protéines dites ribosomiques (protéines r). Ils possèdent deux sous-unités distinctes. La petite sous-unité permet le décodage de l'information génétique portée par l'ARNm. La grande sous-unité est quant à elle le siège de la synthèse des protéines puisqu'elle catalyse la formation des liaisons covalentes entre fonctions amine et carboxylique de deux acides aminés adjacents, appelées liaisons peptidiques. Les particules ribosomales sont nommées selon leur coefficient de sédimentation, de symbole S (pour Svedberg) (Baldwin, 1953). Le ribosome possède trois sites distincts de liaison aux ARNt: les sites *Aminoacyl* (A) et *Peptidyl* (P) (Watson, 1964) et le site *Exit* (E) (Noll, 1966) ([Figure 1](#)). Le site A accueille l'ARNt aminoacylé entrant dans le ribosome après reconnaissance spécifique de son codon. Le site P est occupé par l'ARNt lié à la chaîne polypeptidique en cours de synthèse. Enfin, le site E est un site de transit pour les ARNt déacylés avant d'être éjectés hors du ribosome.

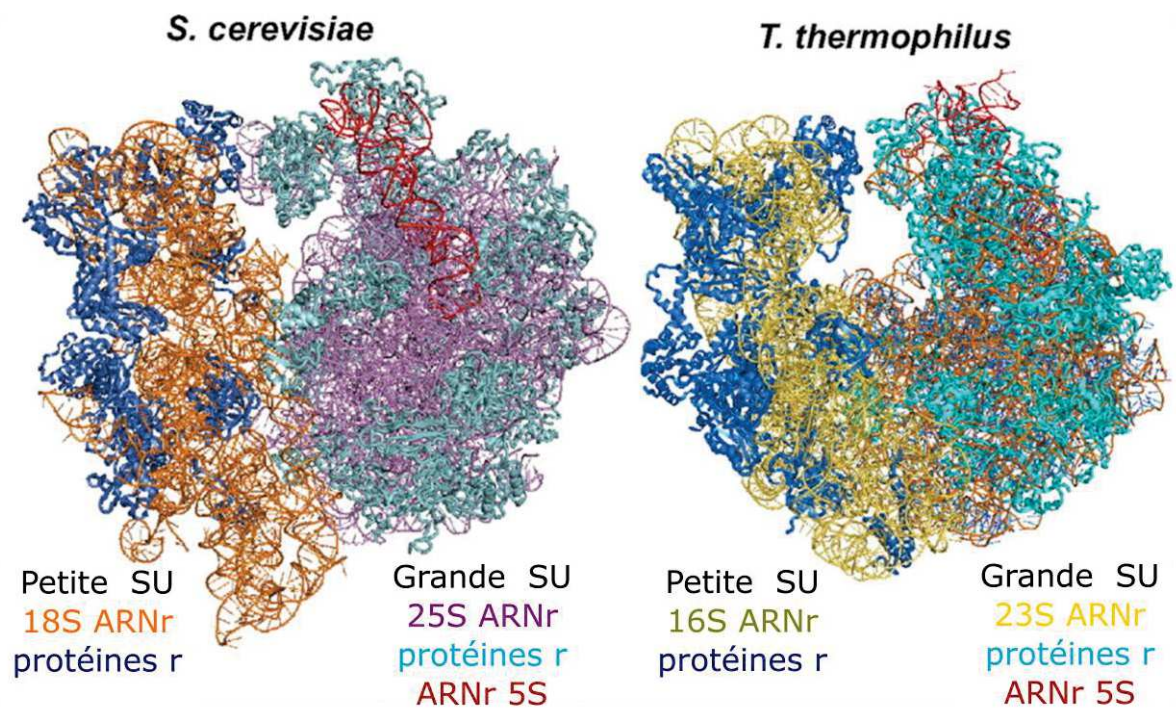


Figure 2. Les ribosomes Eucaryote et Bactérien

Structures des ribosomes de l'eucaryote *S. cerevisiae* (codes *protein databank* - PDB: 1S1H et 1S1I) et de la bactérie *Thermus thermophilus* (codes PDB: 2B64 et 2B66). Les ribosomes eucaryotes sont plus grands et possèdent généralement plus de protéines r que les ribosomes procaryotes. SU = sous-unité. D'après (Dinman, 2009)

Le cœur du ribosome est composé d'ARN ribosomiques (ARNr) et est très conservé au sein du vivant. Sa périphérie, qui comprend principalement les protéines r, l'est beaucoup moins (Fox, 2010; Petrov et al., 2015). Bien qu'il existe environ 32 protéines r universellement conservées, leur nature et leur nombre varient selon l'organisme étudié (Lecompte et al., 2002; Yutin et al., 2012). Chez *Escherichia coli* (*E. coli*), la petite sous-unité 30S est constituée de l'ARNr 16S et de 21 protéines r alors que la grande sous-unité 50S contient deux ARNr (5S et 23S) et 33 protéines r. Les deux sous-unités s'assemblent pour former un ribosome 70S. Chez les eucaryotes, les ribosomes sont plus volumineux et plus complexes et contiennent de manière générale un plus grand nombre de protéines r. Ils contiennent également des séquences d'ARNr supplémentaires appelées *expansion segments* en nombre et en taille plus importantes (Melnikov et al., 2012). Chez *Saccharomyces cerevisiae* (*S. cerevisiae*), la petite sous-unité 40S est constituée d'un ARNr 18S et de 33 protéines r. La grande sous-unité 60S se compose de 3 ARNr (5S, 5.8S, 25S) et de 49 protéines r (Figure 2). Petite et grande sous-unités forment un ribosome 80S après leur assemblage. Les ribosomes des archées ont une taille et une composition similaire à celle des bactéries : ils sont composés de deux sous-unités 30S et 50S qui s'associent pour former une particule 70S (Figure 2). Ils contiennent trois ARNr (16S, 23S et 5S) et entre 50 et 70 protéines r selon l'espèce. Cependant la séquence primaire des ARNr et des protéines r des archées est plus proche de celle des eucaryotes, donnant au système archée une complexité intermédiaire entre le système bactérien et le système eucaryote (Armache et al., 2013; Lecompte et al., 2002; Yutin et al., 2012).

L'étape d'initiation engage la synthèse protéique. Elle correspond à l'appariement du premier ARNt au niveau du premier codon du cadre de lecture de l'ARNm, appelé codon initiateur, au sein du site P du ribosome (Marintchev and Wagner, 2004). L'ARNt initiateur est spécifique et il porte une formyl-méthionine chez les bactéries et une méthionine chez les eucaryotes et les archées (RajBhandary, 1994). L'initiation est l'étape cinétiquement limitante de la traduction et donc la plus critique et la mieux contrôlée. Chez tous les organismes, cette étape implique les deux sous-unités ribosomales, l'ARNt initiateur et un certains nombres de facteurs d'initiation (*Initiation Factors* – IF) qui régulent le processus. Il s'agit de l'étape de la traduction qui présente le plus de divergences mécanistiques entre archées, bactéries et eucaryotes (Benelli and Londei, 2009).

Chez les eucaryotes, la petite sous-unité du ribosome 40S ne peut généralement pas lier directement de manière stable l'ARNm. A la place, elle se fixe au niveau de la coiffe 5' et commence un balayage (*scanning*) de l'ARNm en se déplaçant dans le sens 5' vers 3' (Kozak, 1989). Lorsqu'un codon initiateur est reconnu, la grande sous-unité est recrutée pour former le ribosome 80S. Ce processus implique près d'une douzaine d'*eucaryotic*-IF (*e*-IF), parmi lesquels seuls deux ont des homologues bactériens (Jackson et al., 2010). Chez les procaryotes, la traduction a lieu de manière concomitante avec la transcription des ARNm. Chez les bactéries, l'initiation est simple mais efficace. Elle fait intervenir 3 IF qui permettent à la petite sous-unité ribosomale 30S de lier directement l'ARNm, sans *scanning* (Gualerzi and Pon, 1990). Les archées ont un système d'initiation plus élaboré comprenant au moins 5 *archaeal*-IF (*a*-IF) dont 2 majeures qui ont des homologues eucaryotes (Londei, 2005).

Les ARNm possèdent une région non traduite en 5' (5'-*Untranslated Region* – 5'-UTR) qui contient des séquences ou des structures nécessaires pour la reconnaissance et la liaison du ribosome lors de l'initiation, appelées *Translation Initiation Region* (TIR). Chez les bactéries il s'agit typiquement d'une séquence riche en purines appelée séquence *Shine-Dalgarno* (SD) (Shine and Dalgarno, 1975). Certains ARNm bactériens ne possèdent cependant pas de séquence SD (Chang et al., 2006). De même les ARNm eucaryotes n'ont pas du tout de séquence SD mais possèdent à la place des modifications post-transcriptionnelles spécifiques, en particulier la coiffe en 5' mais également la queue poly-A en 3'. Une majorité des ARNm des archées ne possèdent pas non plus de séquence SD et certains n'ont pas de région 5'-UTR du tout (Brenneis et al., 2007). Ces derniers sont appelés ARNm *leaderless*. Bien qu'on ignore par quel(s) processus leur traduction est initiée, les ARNm *leaderless* sont répandus chez un certain nombre d'archées et de bactéries, où ils peuvent représenter jusqu'à 50 % des ARNm (Srivastava et al., 2016). Quelques cas d'ARNm *leaderless* ont également été reportés chez des eucaryotes (Jones et al., 1992; Li and Wang, 2004). Au bilan, les données accumulées ces dernières années semblent indiquer que le système SD ne soit en réalité pas le seul mécanisme utilisé pour l'initiation de la traduction chez les procaryotes (Nakagawa et al., 2017). De manière intéressante, certains ARNm *leaderless* peuvent être universellement traduit par des ribosomes *in vitro* provenant des trois domaines du vivant. Cette observation a permis de suggérer qu'il s'agissait peut-être de la forme ancestrale des ARNm (Grill et al., 2000).

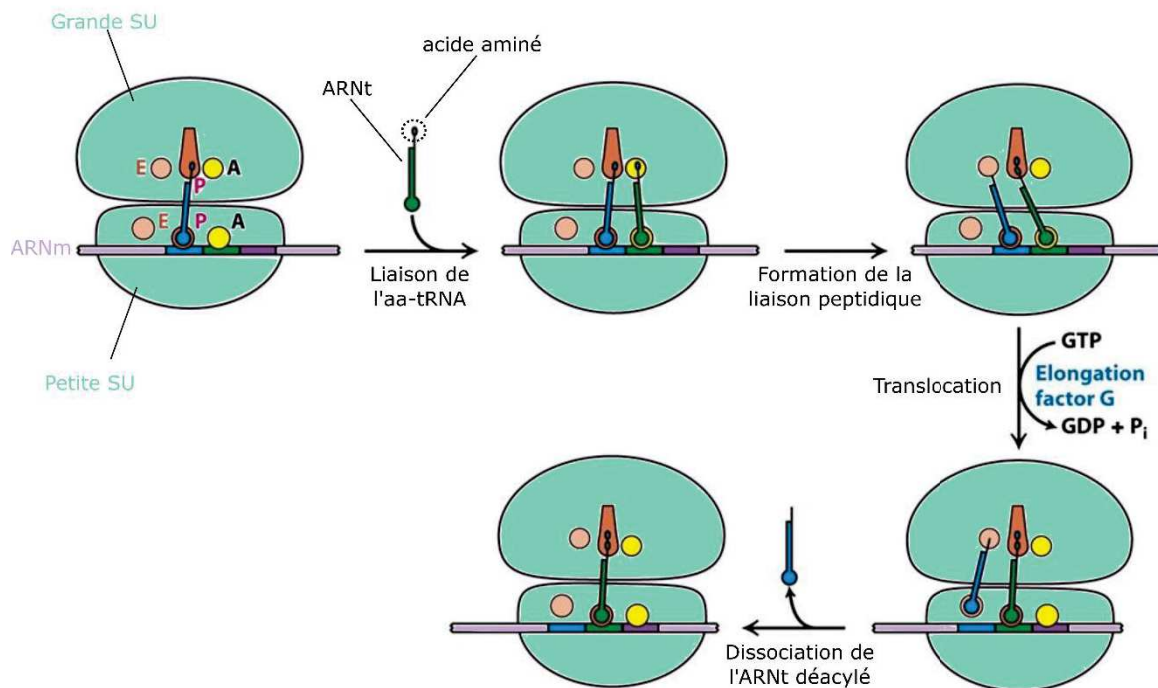


Figure 3. L'étape d'élongation lors de la synthèse protéique

Lors d'élongation, l'ARNT aminoacylé se lie au site A. La chaîne polypeptidique portée par l'ARNT précédent est transférée vers l'ARNT nouvellement entré et la liaison peptidique entre l'acide aminé qu'il porte et la chaîne en cours de synthèse est formée. La translocation assure le déplacement du ribosome sur l'ARNm-ARNT : l'ARNT en A se retrouve en P. Le site A est de nouveau libre pour que l'ARNT suivant puisse s'y lier. Ce cycle se répète jusqu'à l'identification d'un codon stop. Adapté de la figure 30-19 de Biochimie 6^{ème} édition (Berg et al., 2002).

L'élongation est l'étape la mieux conservée au niveau mécanistique ([Figure 3](#)). Elle permet le décodage de l'ARNm et dépend de facteurs d'élongation (*elongation factor* - EF). Une fois l'acide aminé initiateur mise en place, les acides aminés suivants sont successivement additionnés pour former une chaîne polypeptidique grâce aux ARNt dits élongateurs. Lorsqu'un nouvel ARNt élongateur se lie au site A du ribosome, le polypeptide en formation est transféré depuis l'ARNt au sein du site P vers l'acide aminé de l'ARNt élongateur entrant, grâce une réaction de transpeptidation (Dalglish, 1953). Cette réaction est catalysée par certaines bases de l'ARNr au sein d'une région de la grande sous-unité du ribosome (50S ou 40S) appelée *peptidyl transferase center* (PTC) (Staehelin et al., 1969). Le PTC est le cœur fonctionnel du ribosome et en est sa région la mieux conservée, et donc probablement l'une des plus anciennes. Le ribosome est donc un ribozyme, c'est-à-dire un ARN qui a une activité enzymatique (Fox, 2010), potentiellement la relique d'un ancien monde à ARN (*RNA World Hypothesis*). En effet l'existence des ribozymes, et tout particulièrement celle du PTC, appuie cette hypothèse, selon laquelle les premières formes de vie utilisaient l'ARN pour stocker l'information génétique et catalyser les réactions chimiques (Crick, 1968; Higgs and Lehman, 2015; Neveu et al., 2013; Orgel, 1968). Le recours à l'ADN et aux protéines, plus stables que l'ARN, aurait été sélectionné plus tard au cours de l'évolution.

Une fois la réaction de transpeptidation réalisée, le ribosome se déplace de 3 nucléosides, c'est-à-dire d'un codon, sur le complexe ARNm-ARNt, de telle sorte que le site A soit disponible pour accueillir un nouvel ARNt élongateur. Il s'agit de l'étape de translocation (Achenbach and Nierhaus, 2015). L'ARNt déacylé présent dans le site E est alors éjecté hors du ribosome. La précision de l'étape d'élongation est indispensable pour éviter les erreurs de lecture. Les taux d'erreurs varient d'un organisme à un autre : 10^{-4} pour *S. cerevisiae* (c'est-à-dire une erreur tous les 10 000 codons lus) (Kramer et al., 2010) et 10^{-3} à 10^{-4} pour *E. coli* (Kramer and Farabaugh, 2007). Pour arriver à une telle efficacité, les ARNt entrant au sein du ribosome sont soumis à une étape de vérification afin de s'assurer que leur conformation soit stable et que l'appariement codon/anticodon soit correcte (Cochella et al., 2007). Une erreur d'appariement peut aboutir à l'arrêt prématuré de la traduction ou à des décalages du cadre de lecture qui sont susceptibles d'avoir des effets délétères voire fatals pour la cellule.

La terminaison est la dernière étape de synthèse du polypeptide. Elle est enclenchée lorsqu'un codon stop se positionne au niveau du site A du ribosome. Contrairement aux étapes d'initiation et d'élongation, la reconnaissance des codons stop ne fait pas intervenir d'ARNt. Cette dernière dépend exclusivement de facteurs de terminaison (*release factor* – RF). Les RF discriminent les codons stop des codons sens avec une très grande spécificité, ce qui rend les erreurs de terminaison rares (Jørgensen et al., 1993). Bien que les codons stop soient les mêmes dans les trois domaines du vivant, les RF des eucaryotes/archées et des bactéries ne sont pas homologues et ont des structures différentes. Cependant le mécanisme derrière la terminaison de la traduction est similaire. Lorsque le codon stop est reconnu de manière spécifique, les RF entraînent l'hydrolyse de la liaison peptidyl-ARNt au niveau du site P afin de libérer le polypeptide (Nakamura et al., 1996).

Le recyclage est l'étape finale de la traduction qui consiste en la dissociation de la petite et de la grande sous-unité ribosomale afin qu'elles soient disponibles pour un nouveau cycle de traduction (Nürenberg and Tampé, 2013). Le recyclage des ribosomes a lieu après la synthèse réussite du polypeptide, à la fin de la terminaison. Il peut aussi avoir lieu dans les cas où la synthèse polypeptide est bloquée ou lorsque le ribosome s'est assemblé par erreur en absence d'ARNm. De manière intéressante, le système de recyclage eucaryote est capable de dissocier le ribosome 80S vide, c'est-à-dire sans ARNm. Il a donc été suggéré que l'assemblage-recyclage à vide puisse être utilisé comme contrôle qualité lors de la biogénèse des ribosomes (Pisareva et al., 2011). Les mécanismes de recyclage des ribosomes chez les bactéries et chez les archées/eucaryotes sont très éloignés et ont probablement une origine évolutive différente. Chez les bactéries, il dépend de facteurs de recyclage (*ribosome recycling factor* - RRF) (Zavialov et al., 2005). Chez les eucaryotes et les archées, il n'y a pas de RRF. Le recyclage est assuré par le facteur e-RF1 qui reste associé au ribosome après la terminaison et qui recrute l'ATPase ABCE1 afin d'assurer la dissociation (Pisareva et al., 2011).

II- Les ARNt

Après leur transcription, les ARN sont composés d'une succession de 4 types de nucléosides, adénosine (A), guanosine (G), cytidine (C) et uridine (U), reliés par des liaisons phosphodiester, pour former un polymère simple brin. Chacun de ces nucléosides est composé d'une base, d'un ribose et d'un phosphate. La présence d'un groupement hydroxyle -OH à la position 2' du ribose confère à l'ARN des propriétés chimiques et structurales bien différentes de celles de l'ADN (chez qui cette fonction est absente). La présence du groupement 2'-OH augmente la température de fusion (*melting temperature* – Tm) des ARN et permet à l'hélice d'ARN double brin d'adopter une géométrie de type A (plus stable que la géométrie de type B de l'ADN) (Lesnik and Freier, 1995). La réactivité de la fonction alcool en 2' la rend susceptible d'intervenir dans des liaisons avec des protéines partenaires spécifiques de l'ARN (Baidya and Uhlenbeck, 1995) mais aussi d'attaquer la liaison phosphodiester adjacente et d'entraîner l'auto-hydrolyse de l'ARN, surtout simple brin (Holley et al., 1965; Rich and RajBhandary, 1976). Ces caractéristiques expliquent la différence de reconnaissance et de stabilité entre ARN et ADN. Les brins d'ARN se replient souvent sur eux-mêmes, adoptant une structure secondaire stabilisée par la formation d'appariements internes, entre bases complémentaires (A avec U, G avec C et, parfois, G avec U), qui peut être très stable et compacte comme dans le cas des ARNt.

Les ARNt sont des petites molécules d'ARN faisant 76 à 98 nucléosides. Les ARNt interagissent avec l'ARNm et le ribosome lors de traduction afin de permettre la synthèse protéique. Ils assurent ainsi la correspondance entre la séquence nucléique et la séquence protéique, c'est-à-dire entre génotype et phénotype. Leur efficacité lors de la traduction dépend intégralement de leur structure et de leur stabilité.

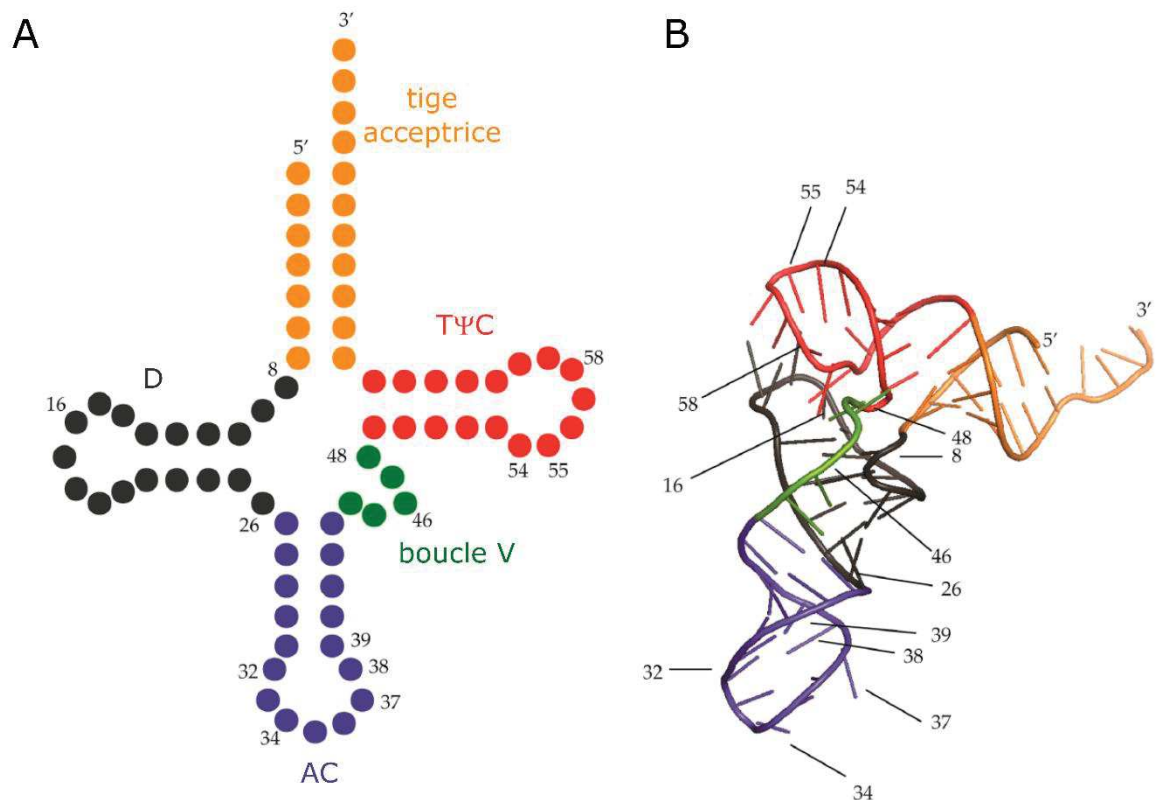


Figure 4. Structures secondaire et tertiaire des ARNt

(A) Structure en feuille de trèfle et **(B)** structure 3D canonique en forme de L (Code PDB : 1YFG) de l'ARNt initiateur de *S. cerevisiae* (Basavappa and Sigler, 1991). Les positions fréquemment modifiées chez les ARNt sont indiquées par leur numéro. La tige acceptrice est en orange, le bras D est en noir, le bras TΨC est en rouge, le bras AC est en bleu et la boucle variable (notée boucle V) est en vert. D'après (Väre et al., 2017) .

a. Structure des ARNt

Le brin d'ARNt mature se replie en une structure secondaire caractéristique en forme de feuille de trèfle (*cloverleaf*) ([Figure 4A](#)). Cette représentation permet d'identifier cinq régions (Powers and Peterkofsky, 1972a; Robertus et al., 1974): une hélice formée par les extrémités 5' et 3' appelée tige acceptrice (le pétiole du trèfle) et quatre structures tige/boucle composée chacune d'une courte hélice double brin qui se termine par une boucle simple brin (les feuilles du trèfle). Les trois principales sont respectivement appelées bras D, bras TΨC et bras anticodon (AC) (Powers and Peterkofsky, 1972a; Robertus et al., 1974). Les noms accordés aux différents bras reflètent leur composition caractéristique en nucléosides. Ainsi la boucle D possède plusieurs dihydrouridines (notée D) et la boucle TΨC porte systématiquement les nucléosides ribothymidine (notée T), pseudouridine (notée Ψ - Psi) et cytidine (notée C) en positions 54, 55 et 56 respectivement. Enfin, la boucle AC porte, en positions 34, 35 et 36, l'anticodon s'appariant au codon de l'ARNm. La taille de ces boucles varie peu d'un organisme à un autre. Chez certains ARNt, une quatrième boucle peut être présente entre les bras AC et TΨC. Cette boucle est dite variable car le nombre de nucléosides qui la compose n'est pas constant d'un ARNt à un autre.

La première structure cristallographique d'un ARNt, l'ARNt^{Phe(UUC)} de *S. cerevisiae*, a montré que les molécules d'ARNt forment une structure tridimensionnelle en forme de L (code PDB : 4TNA) (Kim et al., 1973) ([Figure 4B](#)). Les extrémités du L sont formées respectivement par la tige acceptrice et la boucle AC. Les autres régions forment le corps de l'ARNt. Cette structure canonique permet aux ARNt d'être reconnus spécifiquement durant la traduction. La stabilité de cette structure en L est assurée par l'interaction entre les boucle TΨC et D ainsi que par la conformation intrinsèque de la boucle TΨC (qui forment l'angle du L) (Quigley and Rich, 1976). La boucle AC et l'extrémité de la tige acceptrice sont distantes l'une de l'autre d'environ 75 Å (Ramakrishnan, 2002). Cette distance est cruciale car elle permet à l'ARNt de s'apparier à l'ARNm via le codon, à une extrémité, tout en délivrant l'acide aminé associé au ribosome, via l'autre extrémité lors de la traduction (Ramakrishnan, 2002).

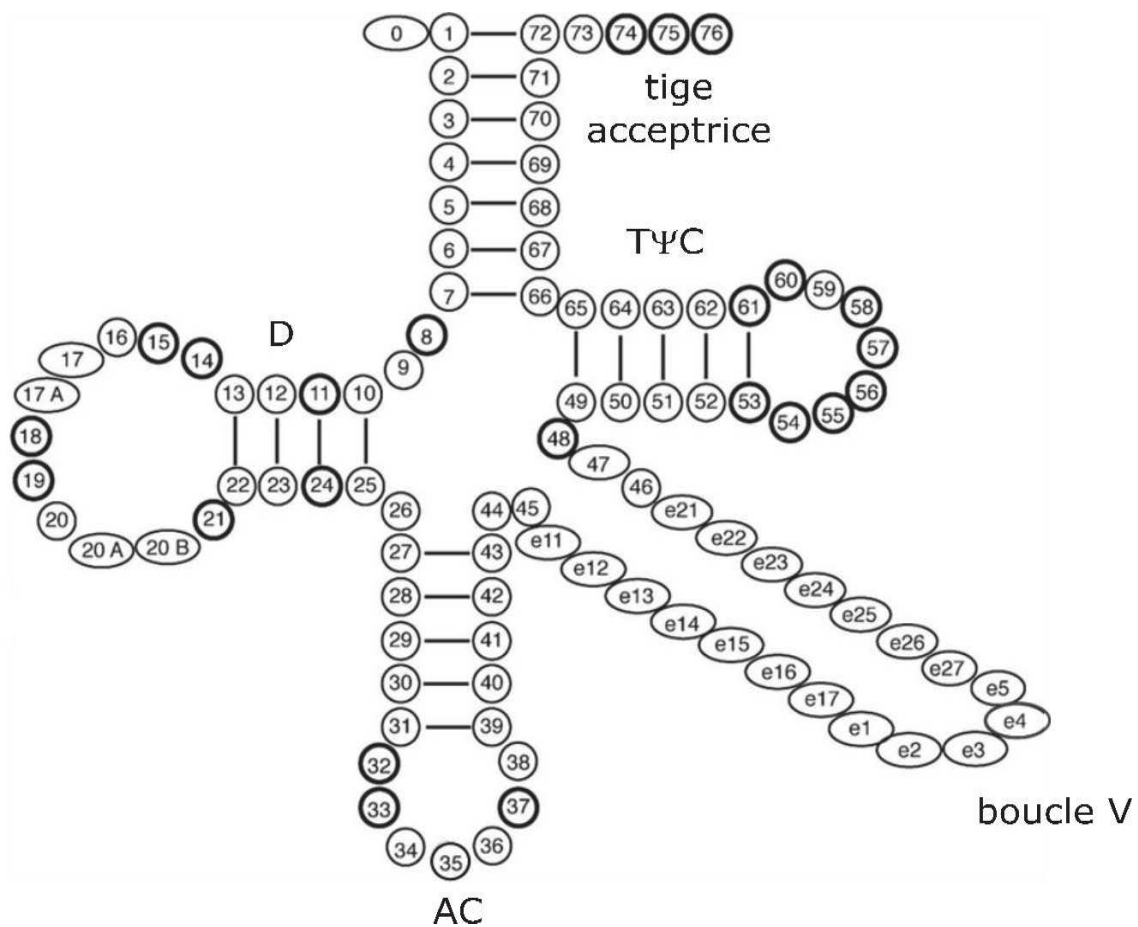


Figure 5. Système de numérotation des ARNt

Il y a 76 positions canoniques conservées chez tous les ARNt, numérotées de 1 à 76 (indiquées par des ronds dans la figure). Les positions auxquelles se trouvent des nucléosides invariants ou quasi-invariants sont indiquées par des contours plus épais. Chaque numéro est associé à une même position dans la structure tertiaire des ARNt. Les positions supplémentaires (indiquées par des ovals) correspondent à l'insertion de nucléosides dans la séquence de certains ARNt. Pour ces nucléosides, un système secondaire de numérotation avec des lettres est utilisé. Historiquement, ce système de numérotation est dérivé de la première structure d'ARNt résolue (Kim et al., 1973). D'après (Laslett and Canback, 2004; Sprinzl et al., 1989).

La structure en *U-turn* est un motif caractéristique des ARN et correspond à un changement de direction abrupt du squelette de la chaîne nucléotidique causé par la présence d'une uridine. La boucle AC possède un motif *U-turn* canonique induit par la présence d'une uridine invariante en position 33 (U₃₃), en 5' de l'anticodon (Agris, 2008; Sundaram et al., 2000). Cette uridine, par des effets d'empilement (*stacking*) avec les bases 32 et 35, est essentielle pour la dynamique et la conformation du *U-turn* de la boucle AC (Ashraf et al., 1999). Cette conformation est cruciale pour plusieurs étapes de la traduction, telles que la reconnaissance codon-anticodon et la translocation (Phelps et al., 2002, 2004).

Le modèle type d'ARNt possède 76 nucléosides numérotés de 1 à 76. Les nucléosides supplémentaires, souvent présents au niveau de la boucle variable, ont un système de numérotation différent (Sprinzl et al., 1989). De cette façon, chaque numéro est associé à une même position précise dans la structure chez tous les ARNt (Figure 5).

b. Maturation des ARNt

Avant d'être fonctionnels, les ARN sont soumis à un processus de maturation spécifique et complexe, assuré par une multitude d'enzymes en aval de la transcription. Classiquement la maturation implique des clivages de la chaîne nucléotidique et l'addition de divers groupement chimiques (Betat et al., 2014). Cependant des variations existent selon le type d'ARN ciblé, l'organisme d'intérêt et son état physiologique. Les ARNt sont la classe d'ARN qui subissent le plus de transformations avant d'être mature (Phizicky and Hopper, 2010).

Les produits de transcription des gènes codant pour des ARNt possèdent en 5' et 3' des séquences nucléotidiques supplémentaires appelées respectivement séquence *leader* 5' et séquence *trailer* 3' (Betat et al., 2014). Lors de la maturation la séquence *leader* 5' est clivée par l'enzyme ribonucléase (RNase) P (Chamberlain et al., 1998) et la séquence *trailer* 3' par la RNase Z (Takaku et al., 2003). Chez les eucaryotes et les archées, certains ARNt non matures contiennent des introns qui doivent être excisés par une *splicing endonuclease* avant qu'une ARN ligase ATP-dépendante re-ligue les extrémités libérées lors de l'excision (Abelson et al., 1998; Tocchini-Valentini et al., 2011; Trotta et al., 1997). L'extrémité 3' des ARNt se termine systématiquement par le triplet CCA en position 74, 75 et 76 (Sprinzl and Cramer, 1979). Ce triplet joue plusieurs rôles critiques pour leur fonction dans la traduction. Il est le site de l'aminacylation (Sprinzl and Cramer, 1979). Il est également requis pour le positionnement correct de l'ARNt dans les sites A et P du ribosome et donc pour l'efficacité de la réaction de transpeptidation (Green and Noller, 1997). Enfin il est impliqué dans la terminaison de la traduction. En effet, cette séquence est importante pour la coordination des molécules d'eau qui hydrolysent la liaison ester liant la chaîne polypeptidique au dernier ARNt du site P (Simonović and Steitz, 2008). Le triplet CCA est codé par tous les gènes ARNt seulement chez certaines bactéries, comme *E. coli*, mais pas chez les autres procaryotes et il n'est jamais codé chez les eucaryotes. Dans le cas où le triplet n'est pas codé, ou est perdu suite à l'action d'une RNase, des ARNt nucléotidyltransférases sont chargées de le réintroduire (Zhu and Deutscher, 1987). L'addition du triplet CCA est cinétiquement plus rapide avec des ARNt matures correctement repliés qu'avec des ARNt défectueux. Les ARNt nucléotidyltransférases jouent donc un rôle dans le contrôle qualité des ARNt (Hou, 2010).

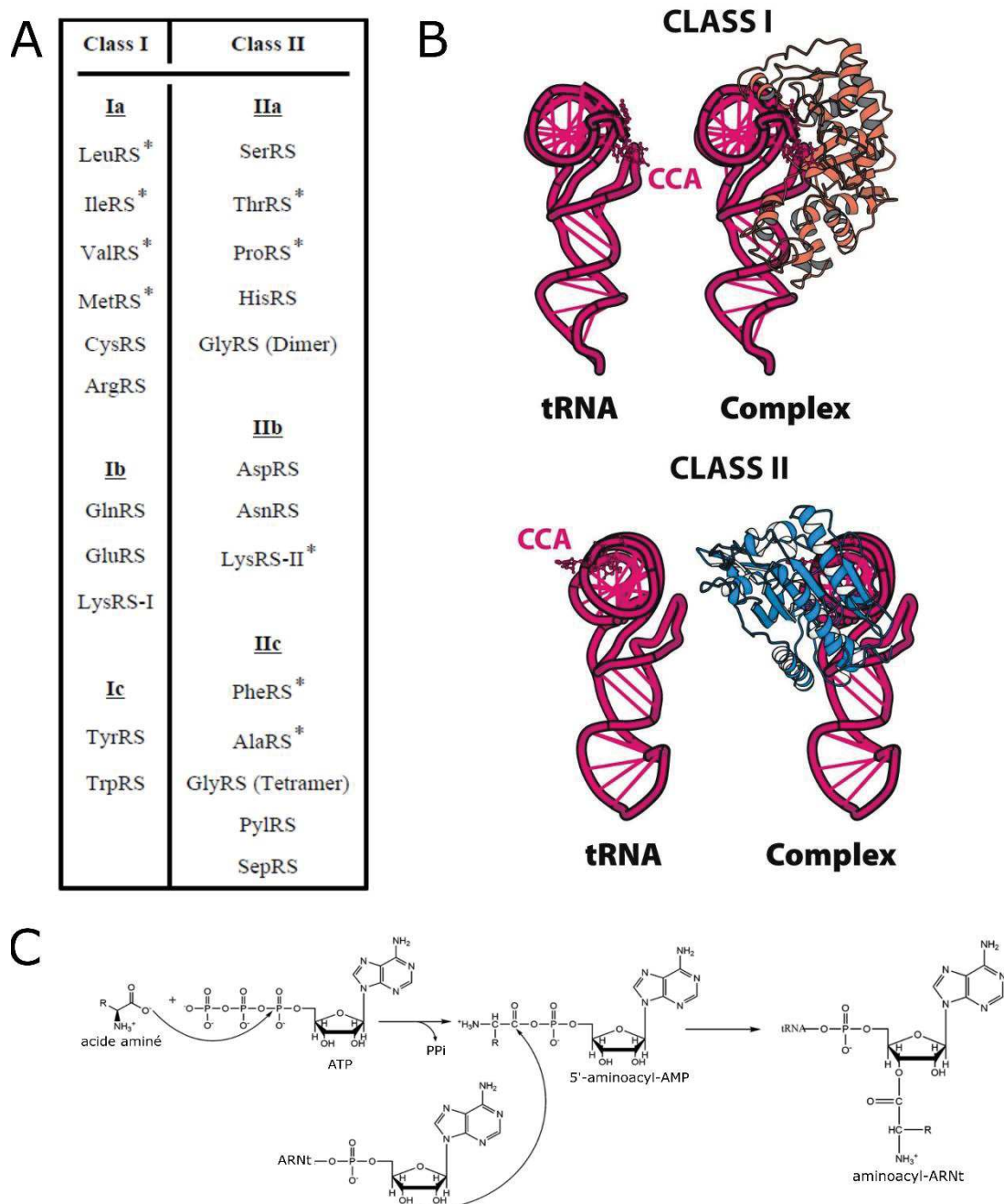


Figure 6. Les aminoacyl-ARNT synthétases

(A) Classification des aminoacyl ARNT synthétases. Les aaRS sont séparées en deux classes selon leurs propriétés chimiques ainsi que l'architecture et la séquence de leur domaine catalytique (Eriani et al., 1990). Chaque classe peut être divisée en sous-classes (a, b ou c) selon leurs caractéristiques mécanistiques et leur mode de liaison à l'ARNT ainsi que l'organisation de leurs domaines. L'astérisque * indique les aa-RS qui possèdent une activité d'édition en *cis*. D'après (Pang et al., 2014). **(B)** Exemples de complexes ARNT-aaRS, alignés par rapport à l'ARNT (en rose). Les enzymes de la classe I et II reconnaissent l'ARNT de manière significativement différente et n'ajoute pas l'acide aminé sur la même fonction hydroxyle de l'A₇₆. D'après (Berg et al., 2002) **(C)** Schéma réactionnel de l'aminoacylation. Malgré des différences mécanistiques, toutes les aaRS forment un intermédiaire aa-AMP avant de transférer l'acide aminé sur l'ARNT substrat.

c. Aminoacylation des ARNt

La nature de l'acide aminé porté par la tige acceptrice est directement liée à la séquence de l'anticodon porté par la boucle AC. Elle détermine le code génétique et est importante pour la fidélité de traduction. L'addition de l'acide aminé par estérification au niveau de l'A₇₆ de l'ARNt est appelée aminoacylation. Cette réaction est catalysée par les enzymes aminoacyl-ARNt synthétases (*aminoacyl-tRNA-synthetase* - aaRS). Chaque aaRS catalyse l'ajout spécifique d'un seul acide aminé sur le ou les ARNt substrat(s). Des ARNt ayant une séquence anticodon différente mais portant le même acide aminé sont appelés iso-accepteurs.

Les 24 familles d'aaRS sont classées en deux classes, I et II, selon leur similarité de séquence et de structure (Eriani et al., 1990; Pang et al., 2014) (Figure 6A). Le domaine catalytique de la classe I a un repliement de Rossmann (Irwin et al., 1976) alors que celui de la classe II a un repliement α/β atypique (Cusack et al., 1990). La classe I charge l'acide aminé sur la fonction 2'-OH alors que la classe II le charge sur la fonction 3'-OH (Sprinzl and Cramer, 1975). Enfin, les enzymes de la classe I reconnaissent la tige acceptrice de l'ARNt par le petit sillon (Rould et al., 1989) alors que ceux de la classe II le reconnaissent par le grand sillon (Ruff et al., 1991) (Figure 6B). Les domaines catalytiques de la classe I et celui de la classe II n'ont *a priori* aucune homologie (Cusack, 1995; O'Donoghue and Luthey-Schulten, 2003; Smith and Hartman, 2015).

Cependant, le mécanisme réactionnel global d'aminacylation des ARNt est le même pour toutes les aaRS (Figure 6C). L'aaRS active d'abord l'acide aminé en hydrolysant une molécule d'ATP pour former un aminoacyl-AMP (aa-AMP), puis catalyse le transfert de l'acide aminé depuis l'aa-AMP sur la fonction 2'-OH ou 3'-OH de l'adénosine terminale de l'ARNt substrat (Ibba and Soll, 2000). La reconnaissance spécifique de l'ARNt par l'aaRS est assurée en grande partie via la reconnaissance de la base en position 73, de la tige acceptrice et de l'anticodon (Giegé et al., 1998). Les aaRS sont des protéines multi-domaines qui possèdent, en plus du domaine catalytique d'autres domaines qui servent généralement à améliorer la spécificité et la fidélité de la réaction d'aminacylation (Pang et al., 2014).

L'addition du mauvais acide aminé lors de l'aminacylation n'a aucun impact sur la structure de l'ARNt et donc sur sa reconnaissance par le ribosome. Par conséquent, une telle erreur a une probabilité faible d'être détectée lors des étapes suivantes de la traduction. La spécificité des aaRS est donc critique pour la fidélité de la traduction (Ibba and Soll, 2000). certaines enzymes aaRS possèdent une activité d'édition en *cis* qui permet de détecter une erreur d'aminacylation et d'hydrolyser la liaison entre acide aminé et ARNt (Jakubowski and Goldman, 1992). Cette activité d'édition a lieu soit directement au sein du domaine catalytique responsable de l'aminacylation soit au sein d'un autre domaine distinct supplémentaire, appelé domaine d'édition (Perona and Gruic-Sovulj, 2013). Il existe également des protéines homologues au domaine d'édition des aaRS capables d'agir en *trans* en cas d'erreur d'aminacylation (Ahel et al., 2003). De plus, la présence de nucléosides modifiés au sein des ARNt, en stabilisant les ARNt dans leur structure tridimensionnelle fonctionnelle et/ou en favorisant une conformation locale particulière, contribue à l'efficacité de l'étape d'aminacylation (Jackman and Alfonzo, 2013; Naganuma et al., 2014).

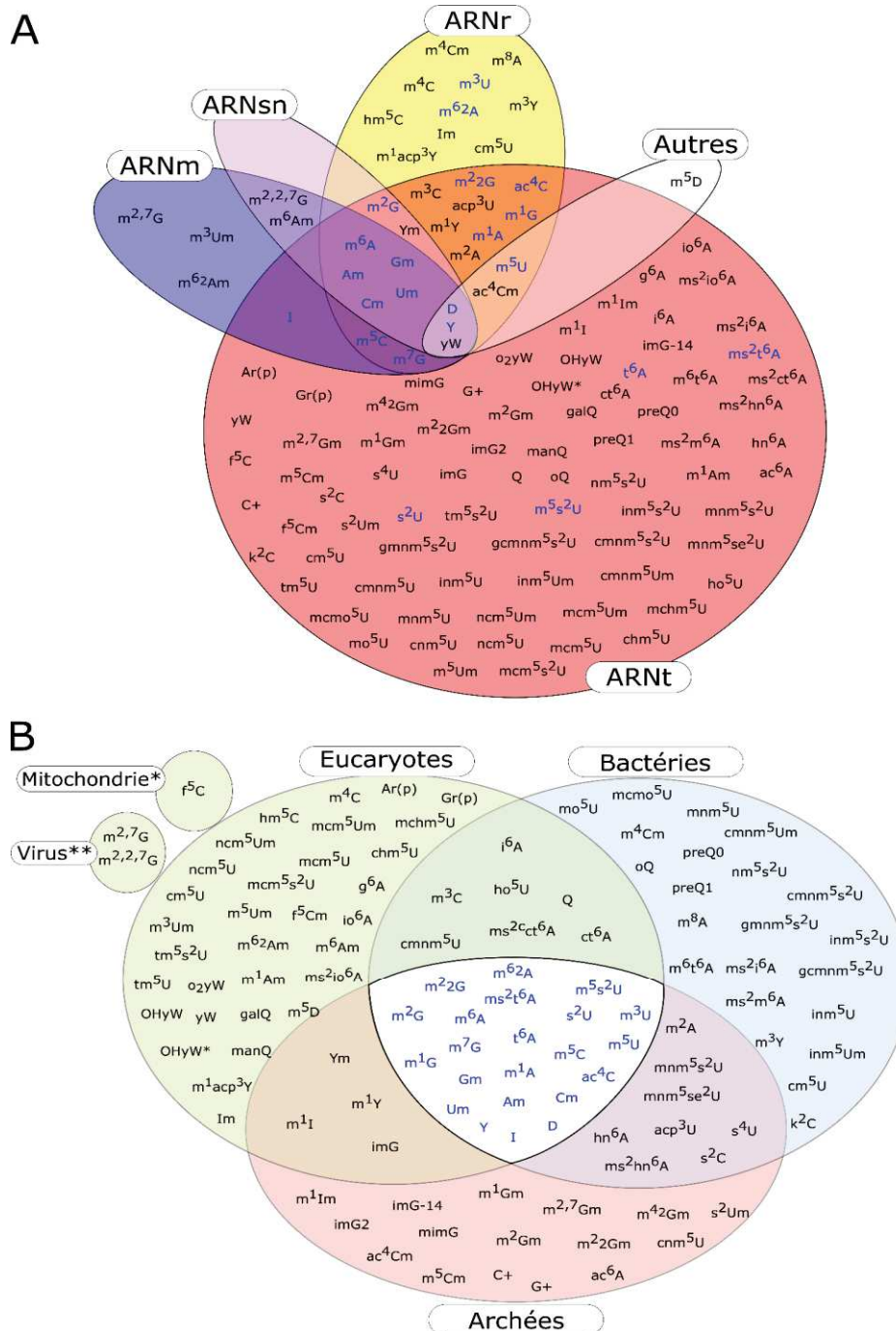


Figure 7. Distribution des nucléosides modifiés des ARN

(A) Par type d'ARN. **(B)** Par domaine du vivant. Chaque couleur de fond correspond à un type d'ARN ou un domaine du vivant. Les nucléosides retrouvés exclusivement dans les mitochondries ou les virus eucaryotes sont indiqués respectivement par * et **. Les nucléosides modifiés retrouvés dans les trois domaines du vivant sont en bleu. Les chiffres en exposant indiquent la position des groupements chimiques au niveau de la base azotée du nucléoside. Adapté des bases de données MODOMICS (Machnicka et al., 2013) et RNAMDB (Cantara et al., 2011). Les formules chimiques ainsi que la nomenclature des nucléosides modifiés peuvent être retrouvée sur MODOMICS et RNAMDB.

III- Les nucléosides modifiés

a. Diversité et répartition des nucléosides modifiés

A l'issu de la transcription, les ARN contiennent uniquement les bases canoniques A, G, C et U. Les nucléosides modifiés sont des dérivés post-transcriptionnels de ces quatre bases. Leur formation consiste dans la majorité des cas en une modification chimique plus ou moins complexe du nucléoside de départ. La présence de groupements polaires, chargés, hydrophobes ou stériquement encombrants modifie les propriétés physiques et chimiques des nucléosides classiques et leurs permet d'acquérir de nouvelles fonctionnalités.

Dès les années 50, des nucléosides différents au niveau chimique des quatre nucléosides canoniques furent identifiés. La pseudouridine Ψ , nommée le « 5^{ème} nucléoside », fut le premier identifié (Davis and Allen, 1957). Il est l'un des plus abondants et est retrouvé au niveau des ARNr, des ARNt et des ARN nucléaires. Actuellement, environ 115 nucléosides modifiés différents provenant d'ARN sont identifiés et répertoriés dans différentes banques de données (Cantara et al., 2011; Machnicka et al., 2013) ([Figure 7](#)). Une centaine de ces nucléosides est retrouvée dans les ARNt dont 76 de manière spécifique (c'est à dire qu'ils sont présents exclusivement dans les ARNt). Les progrès constants des méthodes de séquençage permettent de mettre à jour ces données régulièrement.

Avec en moyenne 17 % de positions modifiées, contre seulement 1-2 % chez les autres ARN (Jackman and Alfonzo, 2013), les ARNt sont les ARN les plus modifiés au sein du vivant. La prévalence des nucléosides modifiés au sein des ARNt varie selon les organismes bien qu'elle ait tendance à être supérieure chez les eucaryotes (Jackman and Alfonzo, 2013). Certains nucléosides modifiés sont spécifiques d'un organisme, d'un groupe d'organismes ou d'un domaine du vivant. Il existe cependant 22 nucléosides modifiés qui sont retrouvés de manière universelle chez tous les organismes dont 7 qui sont également présents au niveau des organelles (Grosjean and Westhof, 2016).

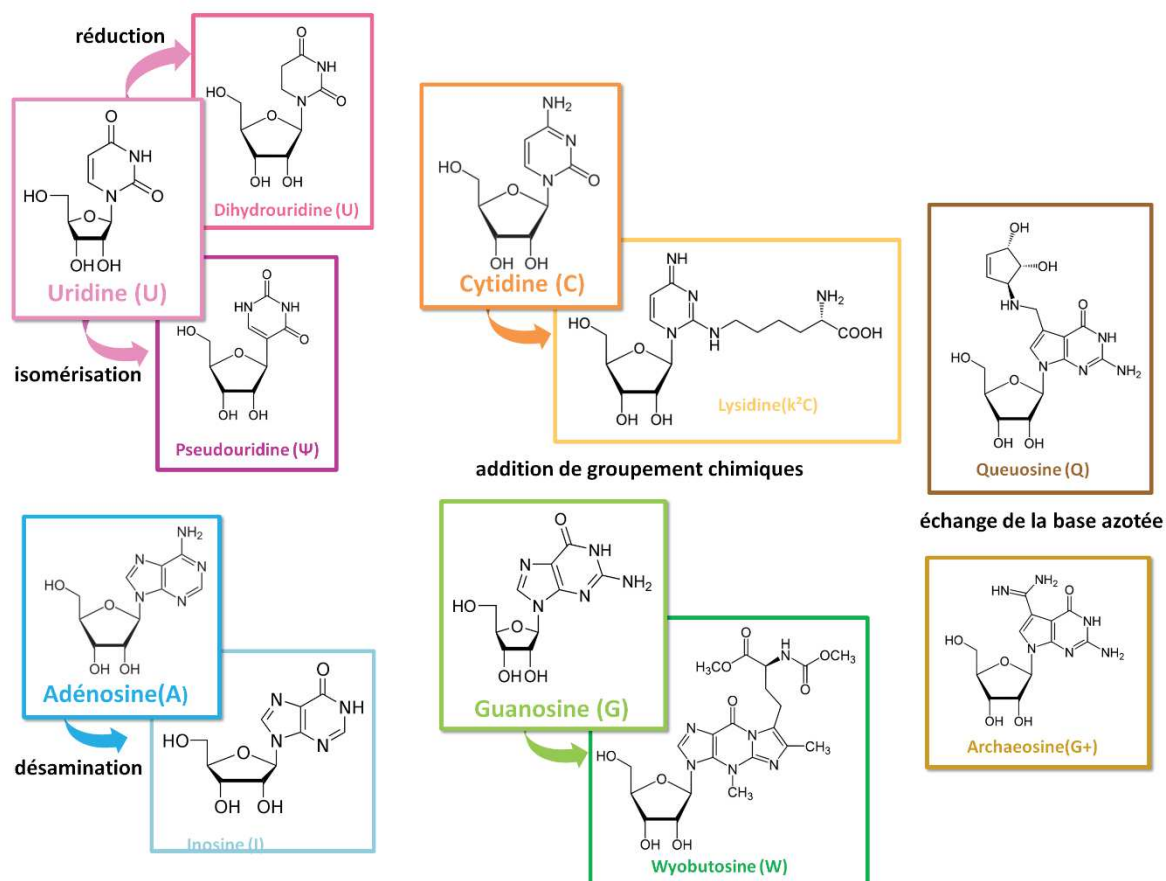


Figure 8. Exemples de modifications chimiques des nucléosides

Le nom et la structure des nucléosides initiaux, A (en bleu), U (en rose), C (en orange) et G (en vert), et finaux ainsi que la nature de la réaction chimique responsable de la modification sont indiqués. Extrait de la thèse de Ludovic Perrochia (2013).

b. Les nucléosides modifiés chez les ARNt

Dans la plupart des cas, les modifications dans le corps de l'ARNt affectent son repliement et sa stabilité, et donc indirectement sa fonction. C'est le cas de la m¹A₉ qui permet à elle seule la formation de la structure en feuille de trèfle de l'ARNt^{Lys(UUU)} mitochondrial humain (Helm et al., 1999). Les ARNt hypomodifiés sont moins stables (Motorin and Helm, 2010). C'est pourquoi l'un des premiers rôles des modifications des ARNt est d'éviter qu'ils entrent dans des voies de dégradation spécifiques (Phizicky and Hopper, 2010). Chez les organismes hyperthermophiles, il existe une corrélation entre la température de croissance et le nombre de méthylations au niveau des riboses des ARNt (Kowalak et al., 1994; McCloskey et al., 2000). Le groupement 2'-OH du ribose est très réactif à haute température. Sa méthylation permettrait de limiter cette réactivité et d'assurer une protection contre la dégradation de l'ARNt chez ces organismes (Motorin and Helm, 2010). Elle permet également d'augmenter leur température de fusion et donc de les stabiliser à haute température. A l'inverse, les ARNt provenant d'organismes psychrophiles (qui ne se divisent plus ou meurent à des températures supérieures à 15 °C) contiennent de nombreuses dihydrouridines D, connues pour augmenter la flexibilité des ARNt (Dalluge et al., 1997). La nature des nucléosides modifiés utilisés par les organismes est donc un paramètre déterminant de leur capacité à survivre dans des environnements particuliers.

Il existe une grande diversité de modifications chimiques au sein des ARNt ([Figure 8](#)). Il peut s'agir d'une isomérisation de la base azotée, comme dans le cas de la pseudouridine Ψ (un C-glycosyle isomère de l'uridine), de la perte d'un groupement chimique comme dans le cas de l'inosine I (une adénosine désaminée), ou d'une réaction de réduction comme chez la dihydrouridine D (où la double liaison du cycle aromatique est réduite). Dans de nombreux cas il s'agit de l'addition d'un ou plusieurs groupement(s) chimique(s) supplémentaire(s). Les modifications dites simples impliquent l'ajout d'un groupement de petite taille, tel qu'un méthyl (noté m) ou un groupement thiol (noté s), greffé sur la base ou le sucre du nucléoside (m¹G, m⁶A, Gm, Um, s²U, s²C...).

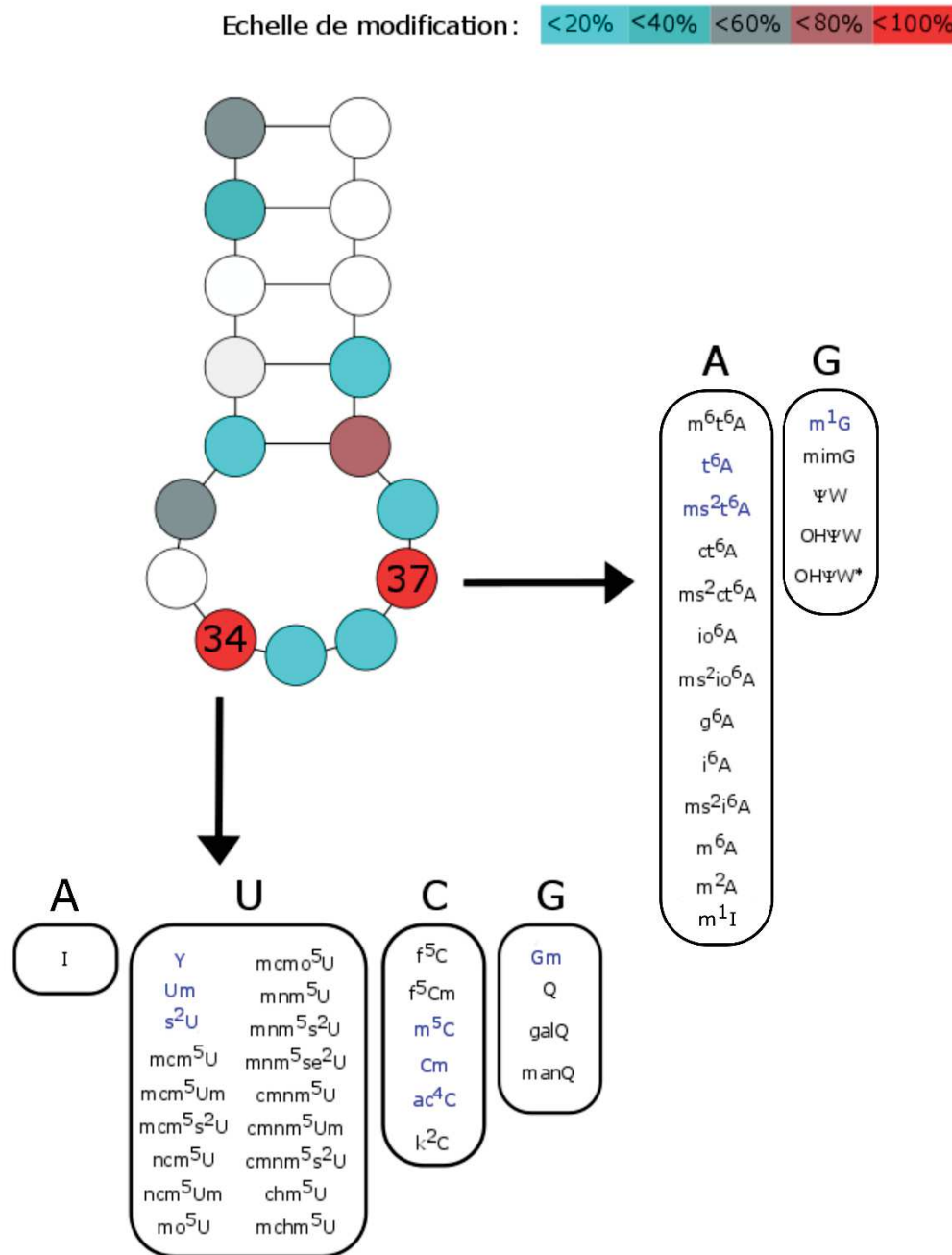


Figure 9. Les modifications des bases 34 et 37

Le bras AC des ARNt est représenté schématiquement. Chaque nucléoside est coloré en fonction du pourcentage d'ARNt chez lesquels cette position est modifiée. Les nucléosides modifiés trouvés aux positions 34 et 37 sont indiqués dans les cadres. La base initiale est indiquée au-dessus de chaque cadre. Les nucléosides universels sont indiqués en bleu. Une G en position 37 est toujours modifiée alors qu'une A₃₇ non modifiée est fréquente. A l'inverse, une A en position 34 est quasi-systématiquement modifiée, en Inosine. Adapté de la base de donnée MODOMICS (Machnicka et al., 2013).

Des modifications plus complexes existent également, telles que l'i⁶A (ajout d'un groupement isopentyle), la t⁶A et la g⁶A (ajout d'un acide aminé carbamoylé). La plupart de ces nucléosides portent plusieurs groupements différents, ajoutés successivement. Le nucléoside modifié final peut être très éloigné de celui d'origine comme dans le cas de la wybutosine (noté yW), un nucléoside retrouvé uniquement en position 37 des ARNt^{Phe} des eucaryotes et des archées (Urbonavičius et al., 2014). La queuosine et l'archaeosine sont des cas encore plus particuliers. La queuosine (noté Q) est retrouvée en position 34 des ARNt aspartate, asparagine, histidine, et tyrosine chez tous les organismes (Morris and Elliott, 2001). L'archaeosine (noté G+) est quant à elle un nucléoside spécifique des archées, retrouvé en position 15 des ARNt (Phillips et al., 2012). Ces deux modifications impliquent que la base azotée du nucléoside d'origine soit littéralement échangée par une autre molécule hétérocyclique synthétisée par une voie enzymatique complexe impliquant plusieurs enzymes. La queuosine peut être encore modifiée par l'ajout d'un mannose, pour former le manQ. Il existe également chez certaines bactéries des modifications qui impliquent l'ajout d'un groupement géranyle sur le nucléoside modifié 2-thiouridine (s²U) en position 34 des ARNt^{Glu(UUC)}, ARNt^{Lys(UUU)} et ARNt^{Glu(UUG)} (Dumelin et al., 2012). Il s'agit à ce jour du seul cas reporté d'oligoisoprénnylation d'un acide nucléique cellulaire. Un autre exemple marquant est l'ajout d'un groupement 2'-O-ribosyl-phosphate en position 64 des ARNt^{Mét} de *S. cerevisiae* qui sert de discriminant entre les ARNt^{Mét} initiateur et élongateurs (Aström and Byström, 1994)

Ces nucléosides sont qualifiés d'hypermodifiés en raison de leur complexité (Adamiak and Górnicki, 1985). Contrairement aux modifications simples qui sont retrouvées dans tous les types d'ARN, les nucléosides hypermodifiés sont caractéristiques des ARNt. Ils représentent 80 % des modifications ARNt-spécifique (Machnicka et al., 2013). Cependant ils ne sont pas répartis uniformément le long de la séquence. La plupart sont situés uniquement au niveau de la boucle AC et tout particulièrement aux positions 34 et 37 (Machnicka et al., 2014) ([Figure 9](#)). A l'inverse des modifications présentes dans le corps de l'ARNt, les modifications présentes en positions 34 et 37 influencent directement les interactions entre codon et anticodon. Elles constituent ainsi un paramètre crucial modulant la capacité de décodage de l'information génétique par les ARNt.

Hypothèse du <i>wobble</i> initiale		Hypothèse du <i>wobble</i> modifiée	
Anticodon N34	Codon N3	Anticodon N34	Codon N3
G	C, U	G	C, U
C	G	C	G
I	U, C, A	I	U, C, A
U	A, G	xm ⁵ U*	G
		s ² U**	A, G
		xo ⁵ U***	A, G, U, (C)

Table 1. Appariements possibles de la base 34 selon l'hypothèse du wobble initiale et sa version modifiée.

Les nucléosides trouvés en position 34 sont en jaune. Les nucléosides complémentaires trouvés en position 37 sont en bleu. Adapté de (Agris et al., 2007)

* = tous les dérivés de la 5-méthyluridine (m⁵U), à l'exception du 2-thio (s²m⁵U)

** = tous les dérivés de la 2-thiouridine (s²U)

*** = tous les dérivés de la 5-oxyuridine (o⁵U)

c. Modifications de la base 34

L'information génétique est codée sous forme de codon (Nirenberg, 1965; Söll et al., 1965). Trois codons sont utilisés comme signal stop de la traduction (Ambre AUG, Opale UGA et Ocre UAA). Les 61 restants se répartissent en plusieurs groupes de codons dits synonymes car ils codent pour un même acide aminé. On parle de dégénérescence du code génétique. Les codons synonymes divergent classiquement au niveau de la troisième base. Francis Crick estima que ces règles universelles avaient été établies aléatoirement très tôt dans l'évolution et n'avaient subi aucun changement depuis, comme gelées : c'est l'hypothèse du *frozen accident* (Crick, 1968). Pour expliquer cette dégénérescence au niveau mécanistique, il émit l'hypothèse que les deux premières bases du codon devaient s'apparier strictement avec les bases 35 et 36 de l'anticodon selon une interaction Watson-Crick ($A=U$, $G\equiv C$) alors que la troisième base du codon devait s'apparier de façon plus lâche avec la base 34, G pouvant alors s'apparier avec U : c'est l'hypothèse du *wobble* (vacillement en français) (Crick, 1966). Les données acquises depuis confirment que la position 34 (appelée *wobble base*) admet des appariements moins stricts que les autres bases de l'anticodon, permettant à un même ARNt de lire plusieurs codons dont la troisième base diverge. Cependant les propriétés physico-chimiques de la base 34 non modifiée n'étaient pas suffisantes pour expliquer toutes les modalités d'appariements codon-anticodon observées expérimentalement. Les travaux qui suivirent ont permis de comprendre que le phénomène de *wobble* repose en grande partie sur la présence de nucléosides modifiés en position 34 (Table 1).

En effet, la *wobble base* est l'une des positions les plus modifiées au sein des ARNt, en particulier lorsqu'il s'agit d'une uridine. Une trentaine de modifications différentes peuvent être trouvées à cette position, dont de nombreuses hypermodifications (Figure 6). Elles modifient ses propriétés chimiques, sans pour autant déstabiliser la conformation de la boucle AC, et influencent ainsi considérablement l'appariement avec le codon, en limitant ou au contraire en stimulant le phénomène de *wobble*. Dans ce second cas de figure, on parle alors d'*expanded wobble* (vacillement étendu). La caractérisation de l'impact des modifications de la base 34 est donc essentielle pour comprendre les modalités complexes qui permettent aux cellules de décoder correctement l'information génétique (Agris et al., 2007; Väre et al., 2017).

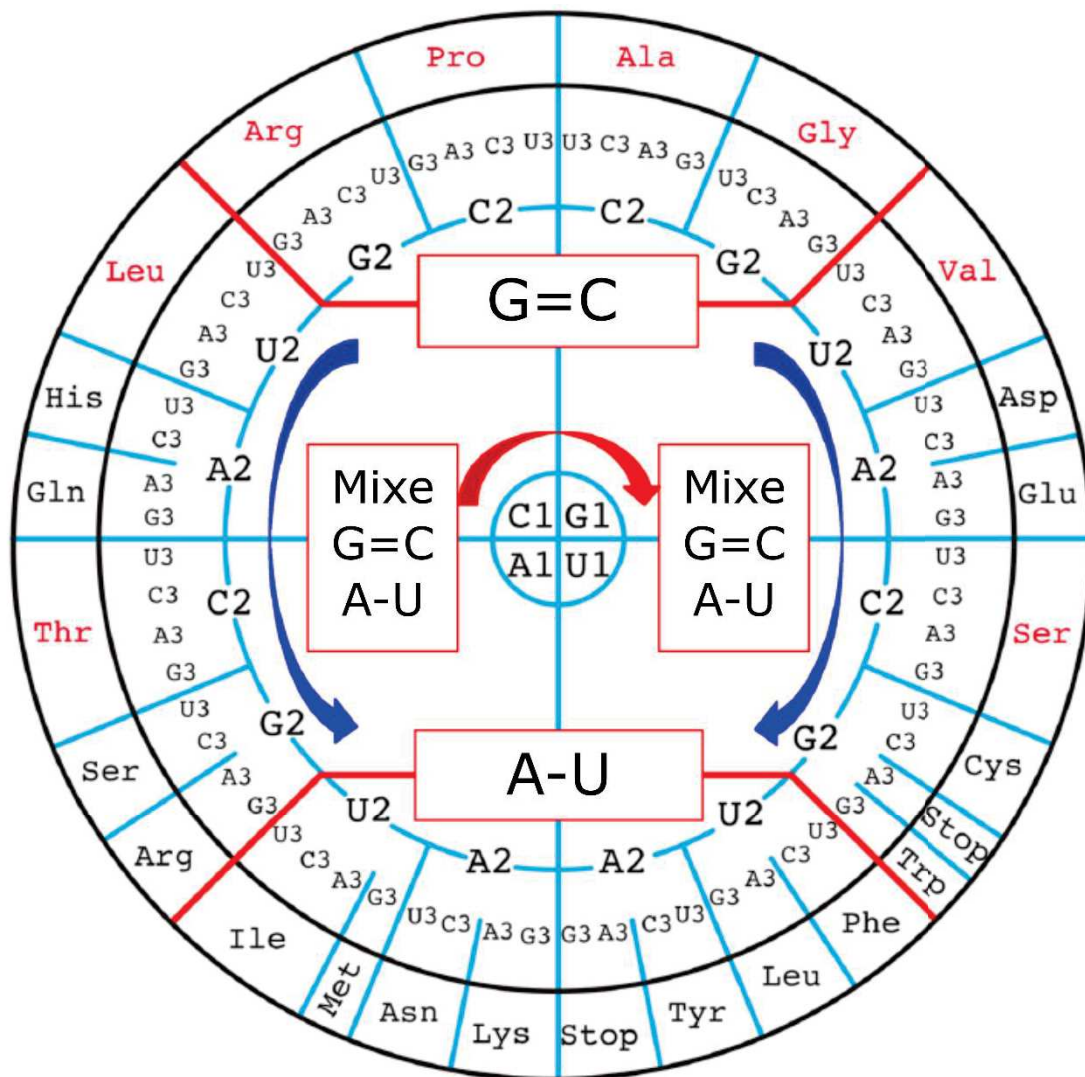


Figure 10. Représentation sphérique du code génétique.

Les codons contenant exclusivement des bases G=C aux deux premières positions sont au sommet et ceux contenant exclusivement des bases A-U à ces positions sont à la base. Les codons mixes sont entre les deux. La flèche rouge indique le sens de lecture pour la première base et les flèches bleues indiquent celui pour la seconde base. Les acides aminés correspondants sont à la périphérie. Ceux codés par les 4 codons d'une même *codon box* sont en rouge. D'après (Grosjean and Westhof, 2016), s'y référer pour d'autres représentations sphériques plus complexes intégrant, par exemple, la nature des nucléosides modifiés au sein du codon.

D'un point de vue évolutif, il a été proposé que la nécessité que la position 34 soit modifiée provienne d'une pression visant à prévenir les erreurs dans la reconnaissance des codons (Agris, 1991; Lim and Curran, 2001). La machinerie de traduction ancestrale, plus simple, n'était pas capable de discriminer correctement la dernière base des codons. Les codes génétiques primitifs auraient eu moins d'acides aminés, chacun d'entre eux étant alors codé par plus de codons. Réguler le *wobble* à la position 34 des ARNt, et donc contrôler la qualité de la reconnaissance des codons, aurait eu pour effet de permettre l'utilisation de nouveaux acides aminés au sein des protéines.

Le code génétique est conservé au sein du vivant et a probablement été hérité du dernier ancêtre commun universel (*Last Universal Common Ancestor* – LUCA), chez lequel ces règles étaient déjà appliquées avec succès. Cependant, certains organismes, notamment les Mycoplasmes, et les organelles ont réassignés certains codons et utilisent donc un code génétique légèrement différent (Bezerra et al., 2015; Grosjean et al., 2010; Ling et al., 2015; Watanabe and Yokobori, 2011). Le code génétique n'est donc pas irrémédiablement gelé.

Il apparaît clairement que le code génétique ne repose pas uniquement sur les règles basiques d'appariement Watson-Crick entre codon et anticodon mais que d'autres forces ont participé et participent à sa mise au point et à son maintien. Traditionnellement, le code génétique est représenté sous forme d'une table composée de 16 *codon boxes* comprenant chacune 4 codons dont les deux premières bases sont les mêmes. Ces représentations prennent en compte uniquement la nature des nucléosides du codon. Récemment (Grosjean and Westhof, 2016) ont proposé des représentations sphériques du code génétique (Figure 10) intégrant différents éléments déterminants pour la reconnaissance correcte des ARNt par les ribosomes, notamment les niveaux d'énergie des différents duplex codon-anticodon, la structure de l'anticodon ou la présence de nucléosides modifiés en positions 34 et 37. Ce nouveau modèle a l'avantage de fournir une vision plus complète et plus réaliste, bien que plus complexe, que la table classique.

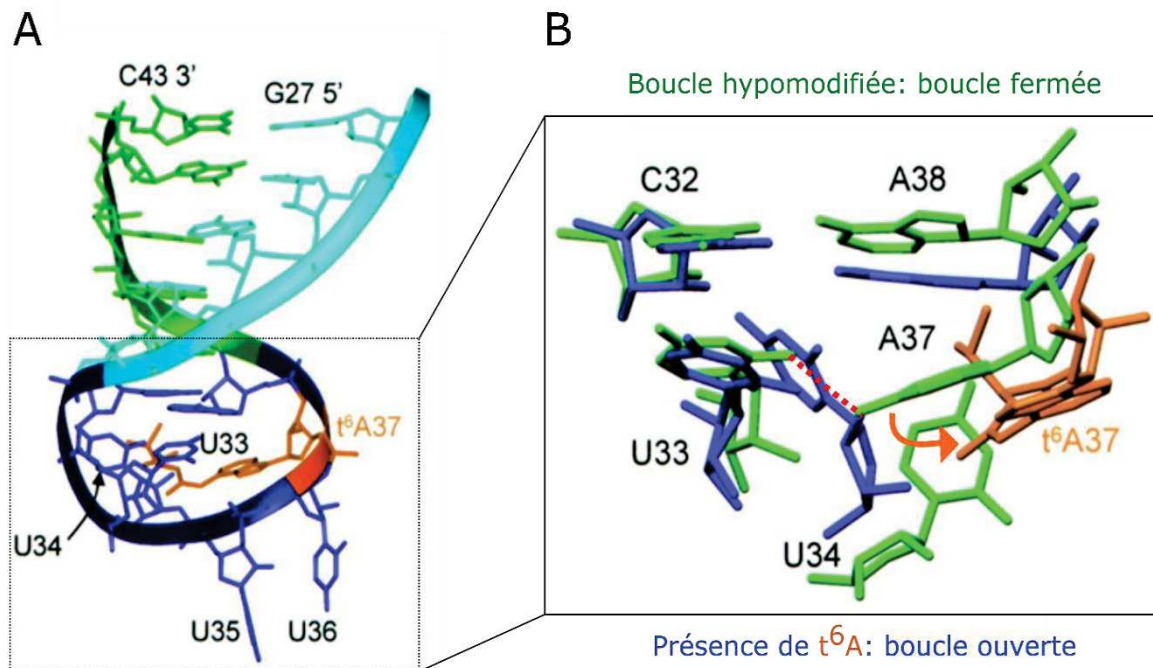


Figure 11. Rôle structural de t⁶A au niveau de la boucle anticodon

(A) Structure de la boucle AC de l'ARNt^{Lys(UUU)} modifiée. L'extrémité 5' est en verte et l'extrémité 3' en cyan. Le nucléoside modifié t⁶A en position 37, est en orange. **(B)** Superposition des bases C₃₂ à A₃₈ de l'ARNt^{Lys(UUU)} en absence (en vert) ou en présence (en bleu) de t⁶A. La présence d'une modification en position 37 empêche l'appariement intra boucle de A₃₇ avec U₃₃, symbolisé par des pointillés rouges, et assure ainsi la formation du *U-turn* canonique de cette boucle. Ceci permet aux bases de l'anticodon d'être orientés correctement pour lier l'anticodon. Adapté de (Stuart et al., 2000)

d. Modifications de la base 37

La base 37 est l'autre position stratégique au sein de la boucle AC. En 3' de l'anticodon, il s'agit presque toujours d'une purine (A₃₇ ou G₃₇). La purine 37 change de géométrie lors de la liaison à l'ARNt au sein du site A du ribosome ce qui n'est pas le cas du reste du domaine anticodon. Pour cette raison elle est appelée la *dangling base* (la base se balançant). En complexe avec l'ARNm au sein du ribosome, elle se positionne au-dessus du couple formé par la troisième base de l'anticodon, N₃₆, et la première base du codon, N₁ (Agris, 2008; Väre et al., 2017).

La boucle AC non modifiée a tendance à adopter une conformation fermée en raison de liaisons hydrogènes intra-boucle entre la base conservée U₃₃ et la base purique en 37 (Stuart et al., 2000). Les modifications en position 37 évitent ces appariements et permettent donc à la boucle d'adopter une conformation ouverte avec le motif *U-turn* canonique essentiel (Figure 11). L'ouverture de la boucle AC rend disponible et accessible les trois bases de l'anticodon pour l'appariement à l'ARNm. De plus, la structure du domaine anticodon étant stabilisée, le processus de recyclage des ARNt, qui implique de redonner une structure intègre à la boucle AC, est facilitée (Vendeix et al., 2008).

Lorsque les modifications sont de nature hydrophobe, elles renforcent l'effet de *stacking* de la base N₃₇ avec la base N₃₈ dans la boucle AC d'une part, et la base N₁ dans le codon sur l'ARNm d'autre part (Witts et al., 2013). Ceci a pour conséquence de renforcer l'appariement entre les bases N₃₆ de l'anticodon et N₁ du codon. Les propriétés physicochimiques des purines modifiées en position 37 participent donc au maintien du cadre de lecture. Elles jouent un rôle capital dans l'efficacité et la fidélité de la traduction. Tout comme la position 34, le nucléoside 37 de l'ARNt est la cible de nombreuses modifications aux effets diverses (Figure 9).

Les effets physico-chimiques stabilisateurs de l'A₃₇ et de ses modifications sont globalement comparables entre eux (Watts and Tinoco, 1978). La nature du nucléoside en position 37 est directement liée à la composition de l'anticodon adjacent, notamment de la base 36. Lorsque la première base du codon est U₁ ou A₁, respectivement A₃₆ ou U₃₆ sur l'anticodon, la position 37 des ARNt est une adénosine A₃₇. Une A₃₆ est généralement associée à la présence d'une N₆-isopentenyl-adénosine (i⁶A₃₇) chez les bactéries et les eucaryotes et à la présence d'une N₁-méthyl-guanosine (m¹G₃₇) chez les archées (Jühling et al., 2009). Une G₃₆ est également associée à une m¹G dans les trois domaines du vivant. Enfin, une U₃₆ est quant à elle souvent associée à une N₆-thréonylcarbamoyl-adénosine (t⁶A₃₇). La m¹G et la t⁶A sont les seuls nucléosides modifiés universels de la position 37 qui sont également présents chez les organelles (mitochondries et chloroplastes) (Grosjean and Westhof, 2016; Grosjean et al., 1995). En raison de leur position stratégique, et de leur caractère universel chacune de ces modifications a fait l'objet d'études afin de comprendre leur rôle au niveau de l'ARNt et leur mécanisme de biosynthèse.

La m¹G est présente chez presque tous les ARNt décodant les codons CNN (ou N = A, C, G ou U) mais aussi de manière fréquente sur ceux décodant les codons UNN (notamment chez les archées) où elle joue un rôle important en augmentant la fidélité de décodage au sein du ribosome (Lee et al., 2007). Les enzymes méthyltransférases responsables de sa synthèse sont appelées TrmD chez les bactéries et Trm5 chez les archées et les eucaryotes (Goto-Ito et al., 2008, 2017). Chez les eucaryotes, la protéine Trm5 est présente à la fois dans le cytoplasme et la mitochondrie, la version mitochondriale possédant un peptide d'adressage spécifique issu d'une initiation alternative (Lee et al., 2007). De manière intéressante, les deux familles de protéines TrmD et Trm5 ne sont pas homologues. Elles présentent des repliements, des modalités de reconnaissance de l'ARNt et également des mécanismes catalytiques différents (Christian and Hou, 2007; Goto-Ito et al., 2017). La raison pour laquelle deux familles ayant des origines évolutives différentes catalysent la synthèse d'un même nucléoside universel reste encore à comprendre et soulève des questions évolutives intéressantes telle que : LUCA possédait-il la m¹G ? D'un point de vue appliqué, les différences entre ces deux familles analogues permettent de développer des antibiotiques ciblant spécifiquement l'enzyme TrmD bactérienne qui n'affecterait pas l'enzyme humaine Trm5 (Lahoud et al., 2011).

IV- Le nucléoside modifié universel t⁶A

a. Histoire et découverte

La t⁶A fut découverte en 1969 dans des extraits d'ARNt totaux digérés issus de différents organismes. La t⁶A se trouve en position 37, en 3' de l'anticodon (Takemura et al., 1969) des ARNt reconnaissant exclusivement les codons ANN (Ishikura et al., 1969; Powers and Peterkofsky, 1972a, 1972b). C'est l'un des rares modifications impliquant un acide aminé ce qui lui a value d'attirer beaucoup d'attention. Elle a dans un premier temps été identifiée chez des bactéries comme *E. coli* (Chheda et al., 1969)(Chheda et al., 1969) puis également chez des eucaryotes comme *S. cerevisiae* et *Homo sapiens* (*H. sapiens*) (Chheda et al., 1969; Cunningham and Gray, 1974; Ishikura et al., 1969; Schweizer et al., 1969)(Chheda et al., 1969; Cunningham and Gray, 1974)(Chheda et al., 1969; Cunningham and Gray, 1974). Suite au séquençage du premier ARNt d'archée, l'ARNt^{Mét(CAU)} de *Thermoplasma acidophilum* (*T. acidophilum*), la présence de la t⁶A a été confirmée dans les trois domaines du vivant (Kilpatrick and Walker, 1981). Dans les années qui suivirent, les progrès techniques et la mise en place de banques de données régulièrement actualisées (Cantara et al., 2011; Machnicka et al., 2013; Sprinzl and Vassilenko, 2005) ont facilité l'identification et la classification des nucléosides modifiés. Il devient alors évident que la t⁶A est présente systématiquement chez tous les organismes, y compris chez des organismes aux génomes très réduits tel que les Mycoplasmes. Ce caractère universel indique l'importance de cette modification pour le fitness des organismes et suggère qu'elle était déjà présente chez LUCA.

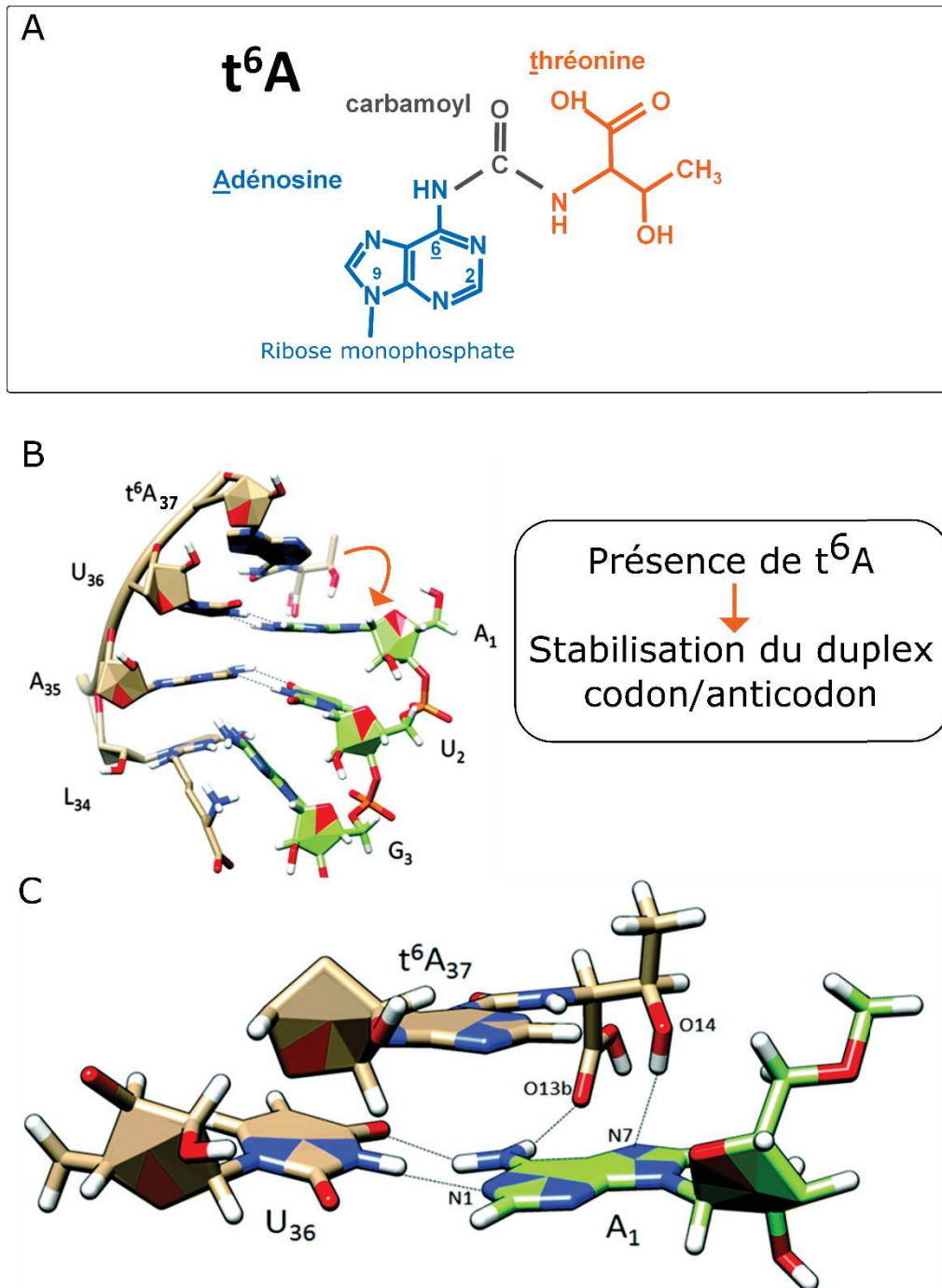


Figure 12. Structure et rôle de t⁶A au niveau du duplex codon-anticodon

(A) Structure de t⁶A. **(B)** Modélisation des bases 34 à 37 de l'ARNt^{Mét(CAU)} (en beige) et des bases de l'anticodon (AUG) (en vert). **(C)** Zoom sur les bases t⁶A, A₁ et U₃₆. La chaîne latérale de t⁶A forme des liaisons hydrogènes avec la base N₁ et favorise ainsi l'efficacité de reconnaissance des codons. Des effets de *stacking* existent également entre les cycles de la t⁶A et des bases N₁ et N₃₈ (non montrée) et permettent de renforcer d'avantages l'appariement codon-anticodon. Adapté de (Sonawane and Sambhare, 2015).

b. Propriétés physico-chimiques

La structure de la t⁶A est résolue en 1974 par cristallographie (Parthasarathy et al., 1974, 1977). Chimiquement, elle correspond à un groupement thréonyle qui est lié via sa fonction amine à un second groupement carbamoyle lui-même lié à l'atome d'azote en position 6 de l'A₃₇ (Schweizer et al., 1969) (Figure 12A). Lorsque l'A₃₇ est modifiée, elle n'est plus en mesure de former un appariement de type Watson-Crick avec une autre base. La présence de t⁶A empêche donc l'appariement intra-boucle entre A₃₇ et U₃₃ (Stuart et al., 2000; Yarian et al., 2002)(Agris et al., 2007)(Agris et al., 2007). Cela permet à la boucle AC d'adopter sa conformation ouverte et assure également la formation du motif *U-turn* canonique, favorable pour la liaison à l'ARNm (Figure 11). Par exemple, chez l'ARN^t_{Lys}(UUU), la présence simultanée de t⁶A et mnm⁵s²U₃₄ est requise pour former ce motif (Bénas et al., 2000; Sundaram et al., 2000).

Des liaisons hydrogènes intramoléculaires entre le groupement thréonine et le cycle de l'adénine de t⁶A donne une structure planaire allongée à ce nucléoside (Murphy et al., 2004). Cette structure favorise l'effet de *stacking* de t⁶A avec A₃₈ et la première base du codon A₁ (Figure 12) (Sonawane and Sambhare, 2015). La présence de t⁶A a pour conséquence de stabiliser l'appariement A₁=U₃₆, énergétiquement faible (Freier and Tinoco, 1975; Högenauer et al., 1972; Weissenbach and Grosjean, 1981). Les capacités de décodage de certains ARN^t sont grandement influencées par cette modification. Ainsi, l'ARN^t_{Lys}(UUU) a besoin de la présence simultanée de la t⁶A et de la mcm⁵U₃₄ afin de pouvoir décoder les codons AAG et AAA. Dans ce cas précis la t⁶A favorise le phénomène de *wobble* de la position 34 (Yarian et al., 2002). En conclusion, la présence de t⁶A limite le décalage du cadre de lecture, et joue donc un rôle critique dans l'efficacité de la traduction (Daugeron et al., 2011; El Yacoubi et al., 2011; Lin et al., 2010).

c. Fonctions cellulaires

Il faut faire la distinction entre les propriétés physico-chimiques élémentaires de la t⁶A, effets d'encombrement stérique et de *stacking* à l'échelle de la boucle AC, déterminés par des approches biophysiques *in vitro*, et les rôles du nucléoside modifié pour la fonction des ARNt dans la traduction *in vivo*. Contrairement aux premiers, qui sont relativement limités et bien connus, les seconds sont variables et restent encore largement incompris. La liste ci-dessous fournit quelques exemples des rôles de la t⁶A dans la traduction.

L'aminacylation des ARNt nécessite la reconnaissance spécifique de la boucle AC par les aaRS. Chez *E. coli*, une approche de *footprint*, a montré qu'elle se fait via les nucléosides G₃₄, A₃₅, U₃₆, t⁶A₃₇ et A₃₈ chez les ARNt^{Ile}(GAU) (Nureki et al., 1994). Une étude plus approfondie, utilisant des ARNt^{Ile} mutés, a confirmé que t⁶A est un déterminant majeur de l'aminacylation par les IleRS bactériennes, puisque son absence diminue significativement la constante de vitesse de cette réaction enzymatique. L'ARNt^{Met}(CAU) de *S. cerevisiae*, muté de telle sorte qu'il porte la séquence G₃₄ A₃₅ U₃₆ A₃₇ A₃₈ (similaire à celle des ARNt^{Ile} d'*E. coli*) au niveau de la boucle AC, est aminoacylé par IleRS d'*E. coli* lorsqu'il est synthétisé *in vivo* mais pas lorsqu'il est synthétisé *in vitro* (Nureki et al., 1994; Pallanck and Schulman, 1991). Cette différence est supposément due au fait que cet ARNt soit modifié par t⁶A *in vivo* mais pas *in vitro*. En revanche l'ARNt^{Thr}(CGU) d'*E. coli* est aminoacylé *in vitro* avec la même efficacité qu'il soit modifié ou pas par t⁶A (El Yacoubi et al., 2009). De même chez *S. cerevisiae*, l'aminacylation des ARNt^{Ile}(AAU/UAU) par les IleRS eucaryotes ne nécessite pas la présence de t⁶A (Senger et al., 1997). Son absence ne mène d'ailleurs pas à une accumulation ou à une dégradation majeure d'ARNt non aminoacylés dans la cellule (Lin et al., 2010).

En complexe avec le ribosome, la présence de t⁶A semble importante pour l'appariement de l'ARNt^{Ile}(GAU) et de l'ARNt^{Lys}(UUU) d'*E. coli* avec un ARNm (Miller et al., 1976; Yarian et al., 2002). Elle est cependant beaucoup moins importante pour l'appariement de l'ARNt^{Arg}(UCU) de *S. cerevisiae*, en présence ou en absence du ribosome (Weissenbach and Grosjean, 1981). Malgré son rôle stabilisateur sur l'interaction A₁=U₃₆, la présence de t⁶A n'est donc pas un prérequis absolu pour tous les appariements codon-anticodon, bien qu'elle le favorise.

En réalité, les modifications en 37, comme t⁶A, auraient une seconde fonction au niveau du ribosome. Elles réguleraient le temps d'occupation par les ARNt des sites A et P lors de l'étape d'élongation, afin qu'il soit constant d'un duplex codon-anticodon à un autre tout le long de la traduction. En effet, bien que les appariements codon-anticodon non modifiés aient des niveaux énergétiques différents, les constantes d'affinités des ARNt naturellement modifiés pour les ARNm au sein du ribosome sont similaires entre eux, indépendamment de la composition en G/C et A/U du codon (Fahlman et al., 2004; Grosjean et al., 1978; Houssier and Grosjean, 1985; Wohlgemuth et al., 2010, 2011). Le ribosome et les ARNt auraient co-évolué pour minimiser les différences énergétiques entre les différents couples codon-anticodon c'est-à-dire pour être ni trop serré ni trop relâché pendant toute la traduction. Cette constance permettrait une traduction optimale (Cedergren et al., 1981; Grosjean and Chantrenne, 1980; Grosjean et al., 1978).

L'initiation est l'étape critique de la traduction. Chez la levure, l'absence de t⁶A provoque de nombreuses erreurs d'initiation, qui se font alors à des AUG non initiateurs, (Daugeron et al., 2011; El Yacoubi et al., 2011; Lin et al., 2010; Thiaville et al., 2016). Le défaut d'initiation s'explique par l'affinité faible entre l'AUG et l'ARNt^{Mét} initiateur hypomodifié. Le scénario proposé chez la levure est le suivant : lors du scanning de l'ARNm, le complexe de pré-initiation arrive au niveau du codon AUG initiateur authentique, cependant l'affinité de la paire codon-anticodon en absence de t⁶A est réduite et le complexe ne parvient pas s'arrêter. Il continue alors jusqu'à l'AUG suivant, non initiateur.

La translocation correspond au mouvement du complexe ribosomal sur l'ARNm vers le codon suivant lors de la traduction (Achenbach and Nierhaus, 2015). La qualité de cette étape dépend de façon critique de la conformation canonique en *U-turn* de la boucle AC des ARNt. Cette structure est stabilisée par la liaison hydrogène entre U₃₃ et N₃₅ lorsque ce dernier est une purine. Dans le cas où la position 35 est occupée par une pyrimidine des modifications en positions 34 et 37 sont nécessaires pour la formation de la structure en *U-turn* (Sundaram et al., 2000). Ainsi, la présence simultanée de mnm⁵U₃₄ et de t⁶A est requise pour l'étape de translocation pour l'ARNt^{Lys(UUU)} d'*E. coli* (Phelps et al., 2004).

d. Essentialité chez les organismes

La t⁶A confère un avantage sélectif aux organismes puisque sa voie de biosynthèse est retrouvée au sein de tous les génomes séquencés à ce jour (Thiaville et al., 2015a). Cependant ses fonctions *in vivo* varient selon la nature de l'ARNt substrat, l'état métabolique de la cellule et l'organisme étudié. Ces variations font que l'importance fonctionnelle de la t⁶A n'est pas la même chez tous les organismes.

Chez les eucaryotes, la t⁶A n'est pas essentielle *sensu stricto*. L'absence de t⁶A chez *S. cerevisiae* (Kisseleva-Romanova et al., 2006; Na et al., 1992) et *Schizosaccharomyces pombe* (*S. pombe*) (Kim et al., 2010; Spirek et al., 2010) n'est pas létale bien qu'elle engendre un phénotype qui affecte sévèrement le fitness de la cellule. Chez la levure, l'absence de t⁶A provoque de nombreuses erreurs d'initiation ainsi que des défauts d'arrêt lors de la traduction (Daugeron et al., 2011; El Yacoubi et al., 2011; Thiaville et al., 2016). Des graines mutantes de la voie de biosynthèse de t⁶A ont pu être obtenus chez *Arabidopsis thaliana* (*A. thaliana*) par la méthode des *T-insertions* (Alonso et al., 2003). Cependant l'approche systématique de cette étude ne permet pas de conclure clairement.

Plus récemment, des *knock-out* viables ont pu être obtenus par CRISPR/Cas9 chez *Danio rerio* (le zébrafish) et *Mus musculus* (*M. musculus*) (Braun et al., 2017). Chez ces eucaryotes pluricellulaires, ainsi que chez *Drosophila melanogaster* (*D. melanogaster*) (Lin et al., 2015), les effets délétères de l'absence de t⁶A se révèlent être associés au développement de l'organisme et sont tissu-spécifique. Ce second phénomène demande encore à être mieux compris mais deux hypothèses sont plausibles en se basant sur ce que l'on sait déjà de la t⁶A: les tissus les plus touchés sont ceux qui ont une activité de traduction importante et/ou ceux qui nécessitent l'expression tissu-spécifique de certains gènes possédant de nombreux codons ANN.

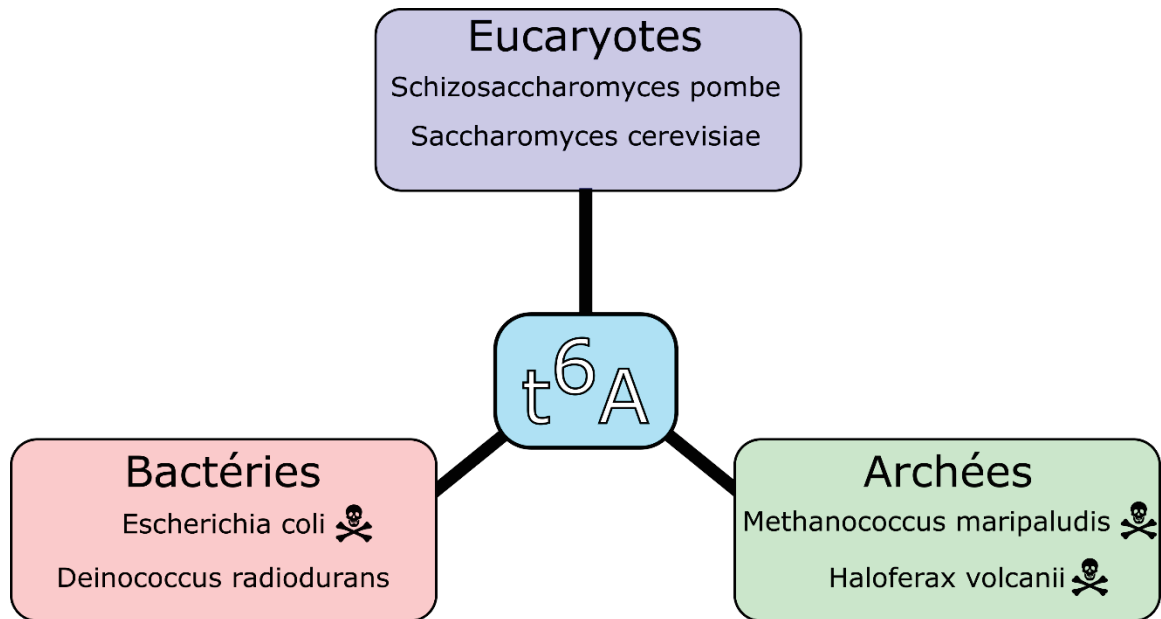


Figure 13. Essentialité de t^6A au sein des trois domaines du vivant

Malgré son universalité, l'absence de t^6A n'a pas les mêmes conséquences chez tous les organismes. Les modalités de l'essentialité de t^6A restent encore à déterminer.

Chez les procaryotes, la t⁶A est strictement essentielle pour la majorité des bactéries et des archées modèles comme *E. coli* (El Yacoubi et al., 2009) et *Haloferax volcanii* (*H. volcanii*) (Naor et al., 2012). Mais cela n'est pas le cas chez tous les procaryotes comme par exemple chez la bactérie *Deinococcus radiodurans* (*D. radiodurans*) (Onodera et al., 2013) (Figure 13). (Thiaville et al., 2015b) ont tenté d'expliquer ces différences de phénotype en se basant sur les données biochimiques connues de t⁶A. Trois déterminants de l'essentialité chez les procaryotes sont ainsi proposés, bien qu'ils n'excluent à aucun moment l'existence d'autres déterminants encore non-identifiés.

(i) L'aminacylation des ARNt^{Ile} par les IleRS bactériennes requiert la présence de t⁶A (Nureki et al., 1994) alors que ça n'est pas le cas pour les IleRS eucaryotes (Senger et al., 1997). Si cette différence est susceptible d'expliquer l'essentialité de t⁶A chez les bactéries, elle n'explique cependant pas celle des archées qui utilisent des IleRS homologues à celles des eucaryotes (Ibba and Soll, 2000).

(ii) Une seconde explication provient du fait que chez les procaryotes, mais pas chez les eucaryotes, l'ARNt^{Ile(CAU)} doit être modifié en position 34 en lysidine (k²C) par TilS chez les bactéries (Soma et al., 2003) ou en agmatidine (agm²C) par TiaS chez les archées (Ikeuchi et al., 2010; Mandal et al., 2010). TilS et TiaS sont essentielles (Blaby et al., 2010; Soma et al., 2003). La présence de t⁶A pourrait être requise pour la formation de k²C et agm²C. Cependant, cette hypothèse est ébranlée par le fait qu'*in vitro* l'ARNt^{Ile(CAU)} non-modifié est efficacement reconnu par TilS et TiaS (Ikeuchi et al., 2005, Ikeuchi et al., 2010, Osawa et al., 2011, Köhrer et al., 2008).

(iii) Chez *S. cerevisiae*, l'absence de t⁶A provoque des *read-through* de la terminaison lors de la traduction (Lin et al., 2010; Thiaville et al., 2016). La t⁶A pourrait donc avoir un rôle dans la terminaison, variable selon les organismes, et donc susceptibles d'être un paramètre déterminant de l'essentialité chez ces derniers (Dreyfus and Heurgué-Hamard, 2011).

	Organelles	Bactéries	Archées	Eucaryotes
Codon initiateur	AUG, GUG, UUG			AUG
ARNt ^{Mét} initiateur	formyl-méthionine		méthionine	
	pas de t ⁶ A			t ⁶ A
ARNt ^{Mét} élongateur	t ⁶ A			

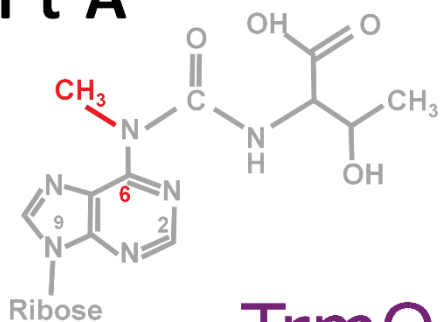
Table 2. Les ARNt initiateurs sont modifiés en t⁶A-ARNt uniquement par la machinerie eucaryote.

e. L'ARNt initiateur : l'exception à la règle

Les ARNt décodant des codons de type ANN portent la t⁶A ou un de ses dérivés. Ceci représente un quart du code génétique, soit 7 acides aminés différents : lysine, isoleucine, asparagine, thréonine, méthionine, sérine et arginine. Chez *E. coli*, par exemple, 9 ARNt portent la t⁶A et 2 ARNt portent son dérivé méthyle m⁶t⁶A sur les 47 espèces d'ARNt codés par cet organisme (Jühling et al., 2009). Il existe cependant une exception remarquable à cette règle. L'ARNt^{Mét(CAU)} initiateur des bactéries, des archées, des plastes et des mitochondries ne présente pas de modification t⁶A en position 37 contrairement à l'ARNt^{Mét} initiateur de *S. cerevisiae* et des eucaryotes en général (Table 2). La première hypothèse pour expliquer cette différence notable a été de proposer que la t⁶A aurait pour rôle de restreindre l'initiation au seul codon AUG chez les eucaryotes (Dube et al., 1968; Lescrinier et al., 2006). En effet l'ARNt initiateur des eucaryotes reconnaît spécifiquement le codon AUG *in vivo* (Baralle and Brownlee, 1978; Stewart et al., 1971) alors que l'ARNt^{Mét} initiateur des bactéries et des archées peut s'apparier, avec une affinité plus faible, au codon valine GUG et au codon leucine UUG (Files et al., 1975; Miller et al., 1976). Il avait été proposé que l'absence de t⁶A sur les ARNt^{Mét} initiateurs procaryotes puisse expliquer cet effet *wobble* atypique, au niveau de la première base du codon initiateur, qui permet la reconnaissance des anticodons non complémentaires GUG et UUG.

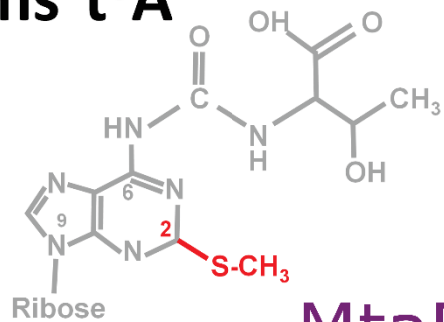
Différentes études donnent pourtant des résultats contradictoires avec cette hypothèse. L'ARNt^{Mét} initiateur modifié de *S. cerevisiae* est capable de reconnaître *in vitro* le codon AUG mais aussi le codon GUG, avec une plus faible affinité, en absence de ribosome (Freier and Tinoco, 1975; Högenauer et al., 1972). Une autre étude a montré que ce même ARNt^{Mét} initiateur, synthétisé *in vitro* cette fois-ci et donc non modifié, est en mesure de discriminer les codons non-AUG lorsqu'il est en complexe avec la sous-unité 40S (Kapp et al., 2006). La t⁶A n'est donc pas suffisante pour expliquer la différence de reconnaissance par l'ARNt^{Mét} initiateur entre procaryotes et eucaryotes. Les progrès réalisés ces dernières années semblent finalement indiquer que la restriction de l'initiation au codon AUG chez les eucaryotes repose en réalité sur plusieurs facteurs d'initiations spécifiques (Hinnebusch, 2017). On ignore donc encore pourquoi l'ARNt^{Mét} initiateur des procaryotes ne porte pas la t⁶A.

m⁶t⁶A



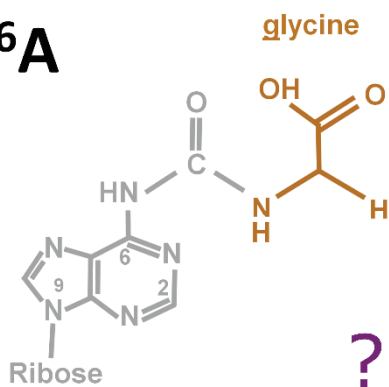
TrmO

ms²t⁶A



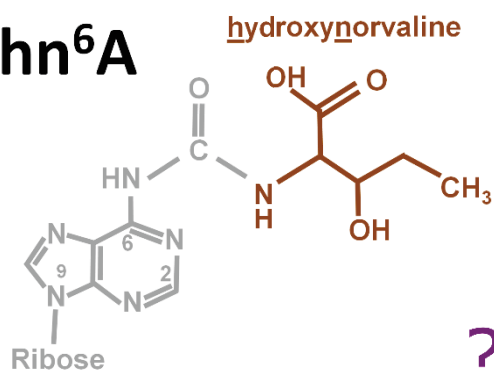
MtaB

g⁶A



?

hn⁶A



?

Figure 14. Exemples de variants et dérivés de la t⁶A

Les groupements chimiques supplémentaires des dérivés sont en rouge et les groupements différents des variants sont en marron. Les enzymes responsables de leur synthèse sont indiquées en violet.

f. Variants et dérivés

Il existe des variants naturels de la t⁶A, bien qu'ils soient beaucoup plus rares que cette dernière (Figure 14). La g⁶A contient une glycine à la place d'une thréonine et a été identifiée chez *S. cerevisiae* uniquement (Schweizer et al., 1970). La hn⁶A, et son dérivé méthyl-thiolé ms²hn⁶A, ont été retrouvés chez certaines bactéries et archées hyperthermophiles (Reddy et al., 1992). Ces modifications contiennent un groupement hydroxynorvaline, un analogue de la thréonine, synthétisé par certains micro-organismes (procaryotes et eucaryotes) ainsi que certaines plantes, qui a des propriétés antivirales et qui est toxique pour les cellules eucaryotes d'insectes et de mammifères (Louw et al., 2005; Yan et al., 2015). La g⁶A ou la hn⁶A n'ont cependant pas été détectées depuis, y compris dans les quelques 500 ARNt séquencés répertoriés par (Jühling et al., 2009). La répartition chez les organismes et la voie de biosynthèse de ces modifications demeurent donc complètement inconnues.

Il existe également des formes dérivées de la t⁶A produites à la suite d'une étape supplémentaire de modification. La N⁶-méthyl-t⁶A (m⁶t⁶A) est un dérivé méthylé de t⁶A retrouvé chez les bactéries, la mouche *D. melanogaster*, les plantes et les mammifères (Jühling et al., 2009; Kimura et al., 2014; Kimura-Harada et al., 1972). Chez *E. coli*, la présence de m⁶t⁶A dans deux ARNt^{Thr} est importante pour la traduction des codons ACY (où Y = C ou U) (Kimura et al., 2014). La 2-méthylthio-t⁶A (ms²t⁶A) est un dérivé thiométhylé retrouvé sur l'ARNt^{Lys} de *B. subtilis*, des *Trypanosoma*, des plantes et des mammifères (Ishikura et al., 1969; Sprinzl and Vassilenko, 2005). La ms²t⁶A est essentielle pour le bon décodage des codons par les ARNt^{Lys(UUU)}.

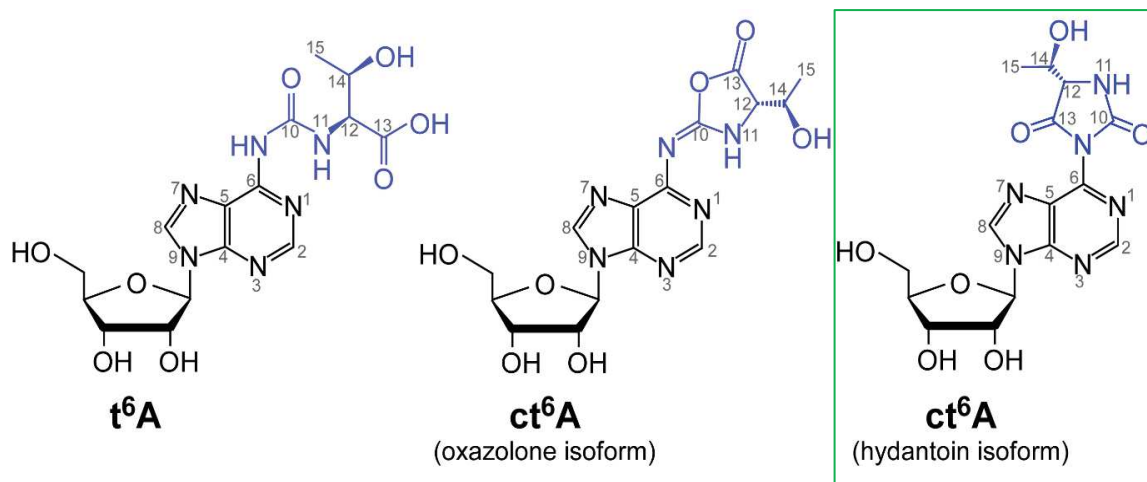


Figure 15. Structures proposées du dérivé cyclique de la t⁶A

D'abord identifié comme étant un isoforme oxazolone (au milieu), la structure cristallographique de ct⁶A révèle qu'il s'agit d'un isoforme hydantoine, dont les propriétés physico-chimiques sont différentes.

D'après (Matuszewski et al., 2017)

g. Le dérivé cyclique ct⁶A

Découverte

En 2013, une forme cyclique de la t⁶A, appelée ct⁶A pour *cyclic-t⁶A*, a été identifiée chez les bactéries *E. coli* et *B. subtilis*, les champignons *S. cerevisiae* et *S. pombe* ainsi que chez l'épinard *Spinacia oleracea* (Miyauchi et al., 2013). En revanche, elle n'est pas retrouvée chez les mammifères *M. musculus* et *H. sapiens*, chez les bactéries *Mycoplasma mobile* et *Acidimicrobium ferrooxidans* ni chez les archées *H. volcanii* et *T. acidophilum*. La ct⁶A est donc spécifique de certains groupes de bactéries et d'eucaryotes. L'identification d'une forme cyclique de t⁶A est une des découvertes majeures les plus récentes dans le domaine des modifications des ARNt. En utilisant des conditions douces pour l'extraction des ARNt totaux, la quantité de ct⁶A détectée devient majoritaire par rapport à celle de t⁶A chez *E. coli*. Les auteurs suggèrent donc que, chez certaines espèces, la modification authentique en position 37 est la ct⁶A, et non la t⁶A. La t⁶A détectée serait un artefact de l'hydrolyse de la ct⁶A, produit lors des extractions.

Structure

Initialement ct⁶A fut assignée par spectrométrie de masse et résonnance magnétique nucléaire (RMN) comme étant un ester actif cyclique possédant un cycle oxazolone (Matuszewski and Sochacka, 2014; Miyauchi et al., 2013). Cependant la structure cristallographique de ct⁶A isolé a mis en évidence la présence d'un cycle hydantoïne, un isoforme du cycle oxazolone non discriminable par les techniques citées ci-dessus (Figure 15) (Matuszewski et al., 2017). La découverte d'un cycle hydantoïne est significative. En effet l'arrangement de l'hydantoïne au sein de la ct⁶A lui confère une conformation tordue complètement différente de la conformation planaire de la t⁶A. Cette conformation planaire est usuellement associée au rôle stabilisateur de la t⁶A, par effet d'empilement stérique, sur l'appariement codon/anticodon. La structure de la ct⁶A ne permet pas de tels effets. Le rôle de la ct⁶A dans la traduction est donc susceptible d'être significativement différent de celui de t⁶A. Cependant, pour le moment, le seul rôle connu de ct⁶A est qu'elle agit de manière coopérative avec la modification ms²U₃₄ pour que les ARNt^{Lys} (UUU) puissent reconnaître les codons non substrats AGA and UAG par un effet de *expanded wobble* chez *E. coli*.

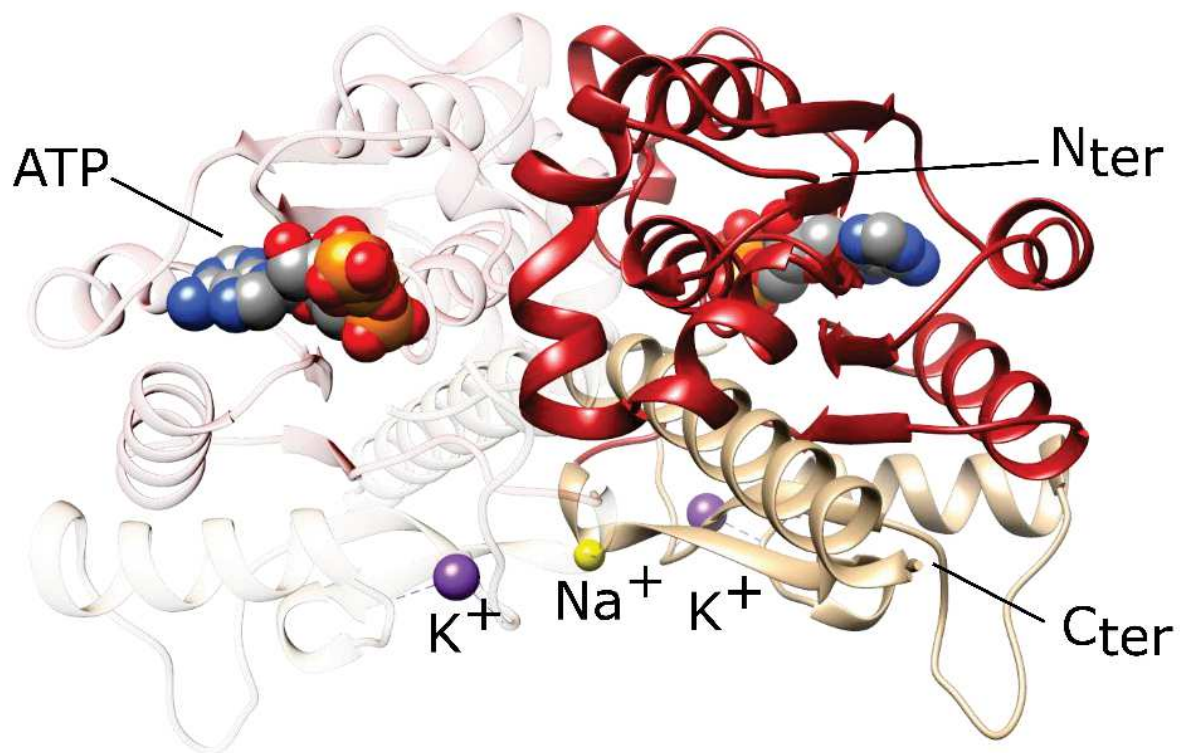


Figure 16. Structure cristallographique de TcdA d'*E. coli*

La structure de TcdA a été résolue avec une résolution de 1.77 Å (code PDB : 4D79) (López-Esteva et al., 2015). Elle forme des dimères en solution. Le domaine *E1-like* est en rouge. Le domaine de liaison à l'ARNt est en beige. Le monomère B est transparent. Deux ions K^+ et un ion Na^+ sont visibles (en violet et jaune respectivement). L'ion Na^+ est localisé au niveau de l'interface du dimère et est coordonné par des résidus des deux monomères. L'ATP est lié au sein d'une cavité qui est conservée.

L'enzyme de cyclisation TcdA

L'enzyme TcdA est responsable de la synthèse de ct^6A et a été identifiée en comparant les génomes des organismes qui présentent ce nucléoside modifié au sein de leur ARNt. L'enzyme TcdA (pour *threonylcarbamoyladenose dehydratase A* - anciennement nommée CsdL) d'*E. coli* est suffisante pour catalyser la réaction ATP-dépendante de déshydratation de t^6A en ct^6A *in vitro* (Miyauchi et al., 2013). La souche délétée pour *tcdA* a un phénotype de croissance normale seule mais ne survit pas dans une situation de compétition avec la souche sauvage. *S. cerevisiae* a deux homologues de TcdA, TCD1 et TCD2, qui ne sont pas non plus essentiels mais sont requis pour la respiration. TcdA possède également des homologues dans certains phylums bactériens (Protéobactéries, Verrucomicrobia et Firmicutes) ainsi que chez les plantes. TcdA fait partie de la famille des enzymes *E1-like* (ou *ubiquitin-activating-like*). Chez *E. coli*, il existe deux autres membres de la famille *E1-like*, MoeB et ThiF, qui sont impliqués respectivement dans les voies de biosynthèse des co-facteurs molybdène et thiamine. Ces protéines lient l'ATP et sont capable d'activer des protéines cibles, les *ubiquitin-like* MoaD et ThiS, grâce à une cystéine catalytique (Lake et al., 2001; Leimkühler et al., 2001; Xi et al., 2001). Cette réaction nécessite l'hydrolyse d'un ATP et est très proche de celui des enzymes eucaryotes *ubiquitin-activating E1*, qui activent également l'ubiquitine via l'une de leurs cystéines.

Cette proximité avec les enzymes *E1-like* a été confirmée avec les structures cristallographiques de TcdA d'*E. coli*, en complexe avec l'AMP ou l'ATP (Figure 16) (Kim et al., 2015; López-Esteva et al., 2015). Le domaine N-terminal de TcdA est similaire au domaine liant l'ATP de MoeB et ThiF. Cela suggère que TcdA utilise un mécanisme proche de celui de ces enzymes, impliquant une cystéine en C-terminal, lors de la déshydratation de t^6A en ct^6A . L'analyse de la structure suggère que Cys²³⁴ est le résidu catalytique potentiel de TcdA (López-Esteva et al., 2015). La ressemblance structurale entre TcdA et MoeB/ThiF s'arrête là. Le domaine C-terminal de TcdA adopte quant à lui une conformation inédite. L'étude de sa structure couplée à une approche biochimique de mutagenèse, suggère qu'il est impliqué dans la reconnaissance et la liaison de l'ARNt. Comme la plupart des enzymes *E1-like*, TcdA forme un homodimère. Cette dimérisation dépend de la liaison forte de deux cations, Na⁺ et K⁺ chez TcdA et Zn⁺ chez la plupart des enzymes *E1-like*. TcdA s'associe à l'ARNt^{Lys(UUU)} *in vitro* avec une stœchiométrie de 2:2 : deux molécules d'ARNt sont liées par dimère, indépendamment l'une de l'autre.

TcdA (ou CsdL) est connue pour interagir avec CsdE, une protéine acceptrice de soufre, elle-même capable d'interagir avec CsdA, l'une des trois cystéine-désulfurases d'*E. coli* (López-Esteva et al., 2015; Trotter et al., 2009). Les cystéines-désulfurases sont la source primaire de soufre dans la cellule et participent à la biogénèse des clusters Fe/S de certaines enzymes ou à la thiolation de certains cofacteurs ou ARNt (Jüdes et al., 2016; Mueller, 2006). TcdA est capable de recevoir un soufre de CsdA *in vitro*, soit directement soit via CsdE (López-Esteva et al., 2015; Trotter et al., 2009). Bien que TcdA puisse synthétiser seule *in vitro* du ct⁶A, l'abondance de ct⁶A produit *in vivo* chute significativement dans les souches $\Delta csdE$ (2%) et $\Delta csdA$ (12%). La synthèse optimale de ct⁶A dépend donc de la présence d'un système de transfert du soufre CSD (pour *csdAEL gene*) fonctionnel (Miyachi et al., 2013). Les modalités du fonctionnement moléculaire du système CSD sont encore mal comprises. Il ne semble cependant n'avoir aucun rôle, en plus de celui de synthèse de ct⁶A, dans la formation des clusters Fe-S ou dans la thiolation des ARNt.

Récemment, le dérivé 2-methylthio cyclic t⁶A (ms²ct⁶A) a été identifié dans des ARNt issus de plantes, de *B. subtilis* et de l'eucaryote *Trypanosoma brucei* (Kang et al., 2017). L'ajout supplémentaire du groupement 2-méthylthiol est catalysé par l'enzyme MtaB sur t⁶A ou ct⁶A de manière indépendante puisque les souches $\Delta tcdA$ et $\Delta mtaB$ produisent respectivement ms²ct⁶A et ct⁶A.

V- Les t⁶A synthétases

a. Identification des t⁶A synthétases

La famille universelle Sua5/YrdC

La protéine Sua5 fut identifiée en 1991 chez *S. cerevisiae* comme un gène suppresseur d'un défaut d'initiation de la traduction, engendré suite à une mutation du gène *cyc1* qui code pour la protéine Cyc1 (iso-1-cytochrome c) (Hampsey et al., 1991). L'allèle muté *cyc1-1019* se caractérise par l'apparition d'un AUG aberrant en amont du cadre de lecture fonctionnel. Dans ce contexte génétique, la production de Cyc1 chute à 2 % de la quantité produite par une souche sauvage. Huit gènes suppresseurs de cet effet ont été identifiés et appelés *SUA* (pour *suppressor of upstream AUG*) 1 à 8. L'inactivation individuelle de ces gènes restaure partiellement la quantité de Cyc1 produite. En particulier, lorsque le gène *SUA5* est inactivé, le niveau de Cyc1 produit passe de 2 à 60 %. Pour en apprendre plus sur le gène *SUA5*, de fonction inconnue à cette époque, ce dernier a été isolé et séquencé, en 1992 (Na et al., 1992). Cette même étude tenta de décortiquer son rôle suppresseur *in vivo*. En déterminant que les sites d'initiation des transcrits de *cyc1-1019* (le sauvage et l'aberrant) ne sont pas modifiés par l'effet suppresseur de *SUA5*, les auteurs conclurent que la fonction de *SUA5* n'était pas au niveau transcriptionnel. Un alignement de séquence avec les protéines homologues hypothétiques de *H. influenzae*, *E. coli*, *B. subtilis* et *Mycobacterium leprae* révéla des résidus conservés entre ces protéines et suggéra une fonction conservée entre tous ces organismes.

L'inactivation de *SUA5* chez la *S. cerevisiae* provoque des effets pléiotropes. En particulier la croissance est sévèrement ralentie (temps de division de 10 h au lieu de 90 min) et la levure est incapable de croître avec du glycérol ou du lactate comme seule source de carbone ce qui indique un défaut de la chaîne respiratoire mitochondriale (Na et al., 1992). Un phénotype aussi marqué chez la levure suggéra que ce gène soit impliqué dans des processus cellulaires fondamentaux et renforça l'intérêt de son étude.

L'orthologue humain de Sua5, YrdC_H, fut identifié en 2003 suite à la recherche de partenaires de la protéine RBBP10 (*retinoblastoma binding protein 10*, un partenaire du suppresseur de tumeur Rb1) par double hybride (Chen et al., 2003). Cette interaction n'a cependant jamais été validée par une autre approche. YrdC_H est exprimé de façon ubiquitaire dans tous les tissus humains. Son taux d'expression est particulièrement élevé dans les tissus ayant une activité traductionnelle élevée tels que le foie, le cerveau et les tumeurs. En recherchant chez la souris des gènes activés en conditions d'ischémie (c'est-à-dire de carence d'oxygène), l'homologue murin d'YrdC a été identifié et renommé IRIP (*ischemia/reperfusion inducible protein*). Comme chez l'Homme, cette protéine est produite de façon ubiquitaire (Jiang et al., 2005).

L'orthologue YrdC d'*E. coli* a été identifié en 2004 comme étant un gène suppresseur d'une mutation de *prfA* (Kaczanowska and Rydén-Aulin, 2004). *prfA* code l'un des facteurs de terminaison de la traduction. Son allèle muté, *prfA1*, confère un phénotype sensible à la température de croissance. Une mutation ponctuelle en 5' du gène *yrdC*, entraînant la perte de son expression génique, annule le phénotype thermosensible de *prfA1*. L'absence d'YrdC entraîne une diminution du nombre de ribosomes actifs ce qui, directement ou indirectement, est responsable de la suppression du phénotype sensible de *prfA1*. Dans le but de mieux comprendre cet effet, une étude génétique ciblée d'YrdC chez *E. coli* a été réalisée par les mêmes auteurs l'année suivante (Kaczanowska and Rydén-Aulin, 2005). Aucun mutant délétère $\Delta yrdC$ ne peut être obtenu chez des souches sauvages d'*E. coli* n'ayant pas la mutation *prfA1*. Le gène *yrdC* est donc essentiel dans les conditions habituelles de cultures en laboratoire. La quantité d'ARN 17S, un précurseur immature de l'ARN 16S, augmente chez *E. coli* lorsqu'*yrdC* et *prfA* sont tous les deux mutés. Cette observation a poussé les auteurs à suggérer qu'*yrdC* ait un rôle dans la maturation des ARNr bien qu'aucune interaction avec des ARNr n'ait pu être mise en évidence. Le nom RimN (*ri*bosome *ma*turation *N*) fut alors proposé pour remplacer celui d'YrdC.

Au cours des années qui suivent, l'accumulation de nouveaux génomes séquencés permet de montrer que les enzymes Sua5/YrdC forment une famille universelle, désormais classée dans le COG0009 (*Cluster of Orthologous Genes*). Sa fonction reste cependant inconnue jusqu'en 2009, l'année où l'équipe de Valérie de Crécy-Lagard établit un lien entre Sua5/YrdC et t⁶A (El Yacoubi et al., 2009) en suivant le raisonnement suivant: (i) la t⁶A étant universelle, les enzymes de sa synthèse le sont également, (ii) la fonction de ces enzymes n'est pas encore connue bien qu'elles semblent jouer un rôle fondamental pour la cellule, (iii) ces enzymes sont susceptibles d'avoir une affinité pour l'ARNt. Enfin, la nature chimique de la t⁶A (une thréonyl-carbamoylation de la base A₃₇ des ARNt) suggérait fortement que sa synthèse implique une réaction de carbamoylation. C'est pourquoi la présence d'un domaine orthologue à YrdC dans des enzymes connues pour avoir une activité carbamoyl-transférase, HypF et TobZ (ces enzymes seront détaillées plus tard), a fait de la famille Sua5/YrdC la candidate idéale.

Son rôle dans la synthèse de t⁶A a été confirmé par des approches génétiques et biochimiques (El Yacoubi et al., 2009). La délétion de *SUA5* chez *S. cerevisiae* abolit la synthèse de t⁶A. Chez les eucaryotes, la protéine Sua5/YrdC est présente à la fois dans le cytoplasme et la mitochondrie, la version mitochondriale possédant un peptide d'adressage spécifique issu d'une initiation alternative de la traduction (Thiaville et al., 2014). L'attribution du rôle d'YrdC d'*E. coli* dans la synthèse de nucléoside modifié t⁶A a permis de comprendre la sévérité des phénotypes observés lors de sa mutation. En particulier son rôle hypothétique dans la maturation des ARNr, évoqué ci-dessus, est fortement remis en question (Kaczanowska and Rydén-Aulin, 2005). En effet, les défauts de maturation des ARNt sont connus pour mener à l'accumulation de précurseurs ribosomiques immatures (Slagter-Jäger et al., 2007). L'expression hétérologue des orthologues YrdC d'*E. coli*, de *B. subtilis* et de l'archée *M. maripaludis* chez la souche Δ sua5 de *S. cerevisiae* rétablit le niveau de t⁶A initial et restaure ainsi le phénotype sauvage. Contrairement à la levure, la délétion d'YrdC chez *E. coli* est létale. Cette délétion peut cependant elle-aussi être compensée par l'expression hétérologue des orthologues de *S. cerevisiae*, *B. subtilis* et *M. maripaludis*. La fonction de la famille Sua5/YrdC est donc conservée entre les trois domaines du vivant. En revanche aucune synthèse de t⁶A n'est observée *in vitro* avec la protéine purifiée YrdC d'*E. coli* ce qui suggère qu'un ou plusieurs autre(s) partenaire(s) non identifié(s) soi(en)t requis dans la synthèse de t⁶A.



Figure 17. Phylogénie de Kae1/YgjD/Qri7 et YeaZ

Les branches de l'arbre sont décrites dans le texte principal. D'après (Hecker et al., 2007)

La famille universelle Kae1/YgjD/Qri7

En 1987, le séquençage d'un fragment du génome d'*E. coli* identifia un nouveau cadre de lecture ouverte, OrfX, appelé par la suite gène *ygjD* (Nesin et al., 1987). Dans les années suivantes, il apparaît que ce gène appartient à une famille qui est présente chez tous les génomes nouvellement séquencés, à l'exception de quelques espèces parasites au génome extrêmement réduit comme *Carsonella rudeii* et *Sulcia mulleri* (Arigoni et al., 1998; Freiberg et al., 2001; Sloan and Moran, 2012). Les eucaryotes présentent deux orthologues (appelés respectivement Kae1 et Qri7 chez *S. cerevisiae*) alors que les archées n'en présentent qu'un seul (Kae1). Les bactéries présentent deux homologues distincts (YgjD et YeaZ). La phylogénie de cette famille a permis de mieux comprendre cette répartition et les liens de parenté entre ces différents homologues (Figure 17) (Hecker et al., 2007).

L'arbre phylogénétique se divise en 3 grands ensembles. Une première branche regroupe les orthologues de Kae1 des eucaryotes et des archées. Une seconde branche regroupe les orthologues bactériens d'YgjD et les orthologues eucaryotes de Qri7. La proximité de Qri7 avec YgjD d'Alpha-Protéobactéries et le fait qu'elle soit systématiquement importée dans la mitochondrie chez *S. cerevisiae* et *Caenorhabditis elegans* (Haussuehl et al., 2009; Oberto et al., 2009) indiquent que Qri7 est une version mitochondriale qui a probablement divergé depuis l'orthologue bactérien ancestral, acquis suite à l'évènement d'endosymbiose. Une troisième branche à part regroupe tous les orthologues bactériens du gène *yeaZ* d'*E. coli*. Ces gènes sont des paralogues d'YgjD spécifiques aux bactéries.

En raison de leur forte similarité de séquence, les orthologues Kae1, YgjD et Qri7 sont regroupés au sein du COG0533. Tout comme Sua5/YrdC, Kae1/YgjD/Qri7 fut placée en 2004 dans le Top 10 des protéines conservées à étudier en priorité car leur fonction était encore inconnue (Galperin and Koonin, 2004). Ces deux familles d'enzymes étant essentielles, elles furent classées au sein du « génome minimal » théorique c'est-à-dire de l'ensemble minimal de gènes requis pour une vie cellulaire indépendante (Koonin, 2003). De même, elles font partie à juste titre du groupe de protéines probablement présentes chez LUCA (Ranea et al., 2006).

Initialement annotée comme métallo-protéase, aucune activité de ce type n'a jamais été détectée chez aucun orthologue de la famille Kae1/YgjD/Qri7 (Hecker et al., 2007; Mao et al., 2008). Chez *S. cerevisiae*, la délétion de *KAE1* affecte fortement la croissance cellulaire, et cette protéine, associée au sein d'un complexe multiprotéique, a un rôle dans l'homéostasie des télomères d'une part (Downey et al., 2006), et dans la régulation de la transcription de gènes essentiels d'autre part (Kisseleva-Romanova et al., 2006). Chez les bactéries *E. coli* et *B. subtilis*, le gène *ygiD* est essentiel (Arigoni et al., 1998). Une déplétion progressive d'YgjD chez *E. coli* conduit à des phénotypes pléiotropes : une répartition inhabituelle de l'ADN, une perte du nucléoïde (Oberto et al., 2009), ainsi qu'à un changement de la morphologie cellulaire (Handford et al., 2009). L'orthologue d'YgjD chez *D. radiodurans* n'est pas essentiel et a été associé à la réparation des *cross-links* d'ADN (Onodera et al., 2013). Enfin, la délétion de la version mitochondriale Qri7 chez *A. thaliana* conduit à des graines anormales et un arrêt très précoce du développement (Haussuehl et al., 2009). Chez *S. cerevisiae* et *C. elegans*, l'absence de Qri7 a pour conséquence une fragmentation du génome mitochondrial ainsi qu'une morphologie aberrante de cet organelle, phénotype similaire à celui rencontré chez *E. coli* (Oberto et al., 2009).

Répartition universelle, mutants à phénotypes pléiotropes et fonction indéterminée : ces caractéristiques rappellent celles qui menèrent à l'identification de Sua5/YrdC comme étant une enzyme de la voie de biosynthèse de t⁶A. La présence d'un domaine structural orthologue à YgiD, fusionné au domaine YrdC, chez les carbamoyl-transférases HypF et TobZ fut l'ultime indice suggérant que cette famille est impliquée dans cette voie métabolique. Cette hypothèse fut prouvée par des approches génétiques menées par plusieurs équipes (Daugeron et al., 2011; El Yacoubi et al., 2011; Srinivasan et al., 2011). Chez les eucaryotes, un mutant $\Delta kae1$ de *S. cerevisiae* présente une quantité faible mais détectable de t⁶A. En revanche, le double mutant $\Delta kae1$ et $\Delta qri7$ ne présente plus aucune trace de t⁶A (Srinivasan et al., 2011). La délétion du signal d'adressage mitochondrial de Qri7 chez le mutant $\Delta kae1$ rétablit la présence de t⁶A au niveau des ARNt cytoplasmique (El Yacoubi et al., 2011). L'orthologue mitochondrial Qri7 semble donc intervenir également dans la synthèse de t⁶A et ce de manière indépendante de Kae1.

Toujours chez *S. cerevisiae*, la version humaine de Kae1 peut compléter partiellement le phénotype de retard de croissance observé chez la souche $\Delta kae1$, suggérant une conservation de la fonction de Kae1 entre ces deux eucaryotes (Kisseleva-Romanova et al., 2006). Chez les bactéries, la déplétion d'YgjD chez *E. coli* et *Staphylococcus aureus* provoque une réduction significative de la quantité de t⁶A indiquant que les orthologues bactériens sont également impliqués dans cette voie de biosynthèse (Lei et al., 2012; Srinivasan et al., 2011). Cependant, les enzymes YrdC et YgjD purifiés d'*E. coli* ne suffisent pas à produire de la t⁶A *in vitro*, ce qui suggère que d'autres enzymes sont encore requises pour cette synthèse (El Yacoubi et al., 2011).

La répartition phylogénétique de Kae1/YgjD/Qri7 montre clairement une distinction entre une version bactérie/mitochondrie et une version archée/eucaryote. Les résultats des expériences de transcomplémentation acquises au cours des années sont venus creuser d'avantage cette distinction. Chez la souche $\Delta ygjD$ d'*E. coli*, l'expression de KAE1 de *S. cerevisiae* ou de l'archée mésophile *Methanococcus maripaludis* (*M. maripaludis*) ne permet pas de restaurer un phénotype normal (El Yacoubi et al., 2011; Oberto et al., 2009). Les protéines YgjD et Kae1 semblent donc avoir des différences fonctionnelles significatives, dépendante d'un environnement protéique très précis et de la présence de certains partenaires qui s'avèrent être domaine-spécifique.

Les partenaires de Kae1/YgjD

Afin d'identifier ces partenaires, on s'intéressa en priorité aux interactants connus de Kae1/YgjD/Qri7. Chez les bactéries, des approches de double-hybride ont montré qu'YgjD est capable d'interagir avec son paralogue YeaZ mais aussi avec une autre protéine appelée YjeE (Handford et al., 2009; Msadek, 2009). Ainsi YgjD forme avec ces deux protéines accessoires un complexe nommé DEZ (Deutsch et al., 2012; Zhang et al., 2015a). Les protéines accessoires YeaZ et YjeE sont conservées et essentielles chez les bactéries (Arigoni et al., 1998; Teplyakov et al., 2002). Elles ne possèdent pas d'orthologue chez les eucaryotes. En 2012, la biosynthèse de t⁶A fut reconstituée *in vitro* pour la première fois avec les enzymes YrdC, YgjD, YeaZ et YjeE purifiées d'*E. coli*, confirmant le rôle du complexe DEZ dans cette voie de biosynthèse (Deutsch et al., 2012).

Chez *S. cerevisiae*, Kae1 a été identifié par double-hybride comme étant un partenaire potentiel d'une petite sérine-thréonine kinase, appelée Bud32 (Lopreiato et al., 2004). D'autre part, la protéine Kae1 est fusionnée à Bud32 chez certaines archées comme chez *Methanocaldococcus jannaschii* (*M. jannaschii*) suggérant un lien physique et fonctionnel conservé (Hecker et al., 2008). La délétion de Bud32 chez la levure provoque une altération de la croissance, une survie réduite en phase stationnaire, une incapacité des homozygotes diploïdes à entrer en sporulation, aucune division méiotique visible, des altérations de la structure de la paroi, ainsi qu'un bourgeonnement aléatoire. Ce phénotype sévère n'est pas sans rappeler celui de la souche Δ kae1. L'orthologue Bud32 de l'archée *H. volcanii* est essentiel pour la cellule (Naor et al., 2012). Comme Kae1, la protéine Cgi121 fut également identifié comme étant un interactant connu de Bud32 (Miyoshi et al., 2003). Cgi121 est présente chez les eucaryotes et chez les archées, mais pas chez les bactéries, ce qui correspond à la distribution de Bud32 dans le monde vivant, suggérant un lien entre Kae1, Bud32 et Cgi121.

En 2006, deux études parallèles chez *S. cerevisiae* confirmèrent l'interaction entre Bud32, Kae1 et Cgi121 et complétèrent la liste des interactants. La première étude montra que ces trois protéines forment dans la cellule un complexe stable et isolable, contenant également une petite protéine spécifique des champignons appelée Gon7 (Downey et al., 2006). Ce complexe fut appelé KEOPS pour *K*inase (correspondant à Bud32), *E*ndopeptidase (correspondant à l'activité initiale supposée de Kae1) and *O*ther *P*roteins of small *S*ize (désignant les petites protéines Cgi121 et Gon7).

La seconde étude identifia une sous-unité supplémentaire de KEOPS, appelée Pcc1 (Kisseleva-Romanova et al., 2006). Cette dernière, tout comme Bud32 et Cgi121, est conservée chez les eucaryotes et les archées, mais absente chez les bactéries. L'association entre les protéines Kae1, Bud32, Cgi121 et Pcc1 fut par la suite observée également chez les archées (Mao et al., 2008; Perrochia et al., 2013a) et chez l'humain (Costessi et al., 2012). Le rôle du complexe KEOPS dans la synthèse de t⁶A fut confirmé par des approches génétiques et grâce à la reconstitution *in vitro* de la voie de biosynthèse (Daugeron et al., 2011; Perrochia et al., 2013a; Srinivasan et al., 2011). Récemment, un cinquième élément similaire à Gon, nommé C14ORF142 (ou C14) a été identifié chez les métazoaires dont *H. sapiens*.

La version mitochondriale Qri7 de *S. cerevisiae* est quant à elle capable de synthétiser la t⁶A seule en présence de Sua5 et ne requiert donc aucun partenaire pour être active (Wan et al., 2013).

	BACTERIES		EUCARYOTES		ARCHEES	Nomenclature utilisée dans ce manuscrit
	<i>Escherichia coli</i>	<i>Bacillus subtilis</i>	<i>Saccharomyces cerevisiae</i>	<i>Homo sapiens</i>	<i>Pyrococcus abyssi</i>	
Universelle	YrdC/RimN (TsaC)	YwlC (TsaC2)	Sua5 (Tcs2)	YrdC ^H (Tcs1)	Sua5 (Tcs2)	Sua5/YrdC
	YgjD (TsaD)	YdiE (TsaD)	Kae1 (Tcs3)	OSGEP (Tcs3)	Kae1 (Tcs3)	Kae1/YgjD/Qri7
			[Qri7]mt (Tcs4)	[OSGEPL1]mt (Tcs4)		
Eucaryotes/Archées spécifique			Bud32 (Tcs5)	PRPK (Tcs5)	Bud32 (Tcs5)	Bud32
			Pcc1 (Tcs6)	LAGE3 (Tcs6)	Pcc1 (Tcs6)	Pcc1
			Cgi121 (Tcs7)	TPRKB (Tcs7)	Cgi121 (Tcs7)	Cgi121
			Gon7 (Tcs8)			Gon7
				C14ORF142		C14
Bacteries specifique	YeaZ (TsaB)	YdiC (TsaB)				YeaZ
	YjeE (TsaE)	YdiB (TsaE)				YjeE

complexe
KEOPS (TCTC)

complexe DEZ
(TsaBDE)

Table 3. Résumé des nomenclatures existantes pour les t⁶A synthétases

Pour chaque protéine et chaque espèce représentée dans cette table, le nom original est écrit en premier. Toutes les anciennes nomenclatures trouvables dans la bibliographie ne sont pas indiquées. La nouvelle nomenclature, d'après (Thiaville et al., 2015a) est indiquée en dessous entre parenthèses. Les noms écrits en gras dans la colonne de droite correspondent à ceux utilisés pour la suite de ce manuscrit.

b. La nomenclature des t⁶A synthétases

Leur fonction exacte étant restée inconnue pendant plusieurs décennies, les enzymes de synthèse de t⁶A ont reçu des noms différents, selon l'organisme auquel elles appartiennent, reflétant souvent les caractéristiques de leur découverte. Par exemple, Bud32 est appelé ainsi car sa délétion chez *S. cerevisiae* a un impact négatif sur le bourgeonnement (*budding*) de la cellule. Cependant, l'accumulation des appellations rend la nomenclature des t⁶A synthétases complexe et parfois déroutante (Table 3). Suite à l'identification de leur rôle universel dans la synthèse de t⁶A (Deutsch et al., 2012; Thiaville et al., 2015a) ont proposé une nomenclature alternative en vue de la simplifier. Les enzymes bactériennes sont renommées Tsa pour *threonylcarbamoyl-6-adenosine* B à E, par rapport à l'enzyme TsaA, responsable de la méthylation de t⁶A en m⁶t⁶A chez *E. coli*, et identifiée 10 ans avant les autres enzymes (Qian et al., 1998). Chez les eucaryotes/archées, les noms Tcs 1 à 8, pour *threonyl-carbamoyl synthesis*, sont proposés. Par soucis de clarté et pour rester constante, je garderai cependant l'ancienne nomenclature dans ce manuscrit.

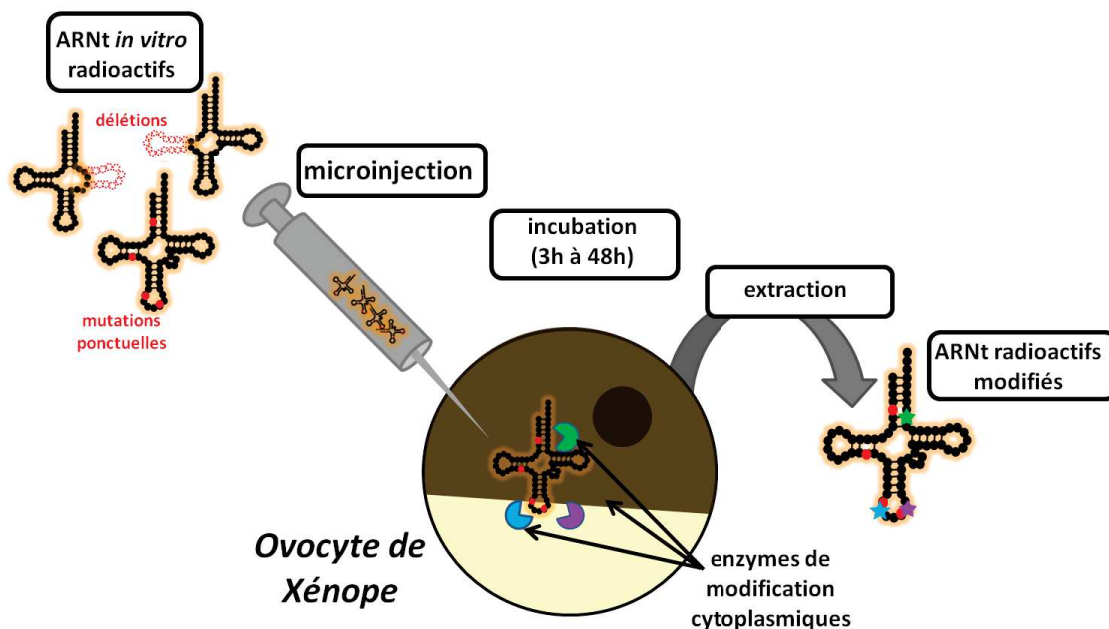


Figure 18. Etude de la voie de biosynthèse de t⁶A *in cellulo*

La plupart des nucléosides modifiées ont été identifiés dans les années 1970-1980. Cependant, la majorité des enzymes responsables de ces modifications sont restées inconnues pendant des décennies, rendant leur étude compliquée. Pour contourner cet obstacle, (Grosjean et al., 1987) ont développé une méthode basée sur la microinjection d'ARNt transcrits *in vitro* et radiomarqués, pouvant porter des mutations ponctuelles ou des délétions, dans le cytoplasme d'ovocytes de Xénope non fécondés. Après incubation, les ARNt sont extraits, digérés et analysés afin de quantifier les nucléosides modifiés radioactifs, traduisant l'action d'enzymes de modifications sur les substrats injectés. Extrait de la thèse de Ludovic Perrochia (2013).

c. Caractérisation de la voie de biosynthèse

Les travaux qui suivirent la découverte de la t⁶A en 1969 ont permis d'identifier rapidement les substrats précurseurs nécessaires à sa synthèse (Powers 1971, Chheda 1972, Elkins and Keller 1974, Korner and Soll 1974). Ces études parvinrent à montrer que de la L-thréonine, du carbonate, de l'ATP et du magnésium Mg²⁺ sont requis pour la formation de t⁶A (Chheda et al., 1972). Cependant, les enzymes responsables de cette synthèse étaient encore inconnues à cette époque et ne pouvaient pas être isolées. De plus, le rendement de la réaction *in vitro* fait à partir d'extraits cellulaires restait faible et variable.

Pour contrer cet obstacle, des approches originales ont été mises au point. En utilisant un système *in vitro* reconstitué, avec des extraits cellulaires solubles cytoplasmiques d'ovocytes de Xénope et des ARNt extraits de cellules humaines, il a été montré que les ARNt ne pouvaient être modifiés en t⁶A-ARNt qu'après la maturation de leurs extrémités 5' et 3' (Harada et al., 1984; Koski and Clarkson, 1982). Quelques années plus tard une nouvelle approche *in cellulo* utilisant des ovocytes de Xénope cette fois-ci comme « tube à essai cellulaire » fut développée (Figure 18) (Grosjean et al., 1987).

Des substrats ARN radio-marqués sont fournis à l'ovocyte par microinjection. Les produits radiomarqués sont ensuite récupérés par simple extraction avant d'être digérés et leur contenu analysé. Cette approche a l'avantage de pouvoir suivre la cinétique des réactions de synthèse de t⁶A et d'identifier les modalités de cette synthèse *in vivo* sans connaître les enzymes qui en sont responsables. L'injection d'ARNt de *S. cerevisiae* mutés à différentes positions, afin d'altérer la séquence ou la structure tridimensionnelle de la molécule, a permis de mettre en évidence que la composition de la boucle AC et la structure tertiaire de l'ARNt sont les deux paramètres critiques pour la synthèse de t⁶A (Grosjean et al., 1987; Morin et al., 1998).

U₃₆ et A₃₇ sont strictement requis pour que la réaction ait lieu et la présence d'une adénosine en 38 augmente la vitesse. Cependant, la composition de la tige AC, la taille de la boucle AC et le type de nucléosides en position 34 et 35 n'influencent pas la réaction. Ceci explique le caractère universel de la t⁶A malgré des séquences d'ARNt variables d'un organisme à un autre. Des perturbations significatives de la structure générale de l'ARNt, comme la délétion des bras D et TΨC, abolissent totalement sa synthèse de t⁶A. Les t⁶A synthétases appartiennent donc à la classe II des enzymes de modifications car elles requièrent une conformation tridimensionnelle intacte.

L'ARNt^{Mét(CAU)} initiateur de *S. cerevisiae*, naturellement modifié chez cet organisme, est modifié en t⁶A très rapidement dans les ovocytes de Xénope (Grosjean et al., 1987; Morin et al., 1998). Un ARNt^{Val(UAC)} du même organisme, qui ne porte pas naturellement t⁶A, dont la boucle AC est mutée pour être identique à celle de l'ARNt^{Mét} initiateur devient lui aussi un excellent substrat de la réaction. Chez les bactéries, l'ARNt^{Mét} initiateur n'est pas modifié bien qu'il possède tous les éléments de séquence et de structure requis. Fait notable, ce même ARNt, lorsqu'il est injecté dans l'ovocyte de Xénope, est modifié en t⁶A-ARNt (Grosjean et al., 1987). Cet ARNt est donc substrat pour la machinerie eucaryote alors qu'il ne l'est pas pour la machinerie bactérienne.

La spécificité de reconnaissance semble donc dépendre des systèmes enzymatiques de l'organisme plus que des particularités intrinsèques des ARNt substrats. L'identification des t⁶A synthétases, près de quarante ans après celle du nucléoside modifié lui-même, donna de nouvelles opportunités pour la compréhension de cette voie de biosynthèse. Une fois encore, l'étude des enzymes HypF et TobZ, réalisant des réactions de carbamoylation similaires à celle observée pour la synthèse de t⁶A, a permis de progresser dans la compréhension des mécanismes de cette synthèse.

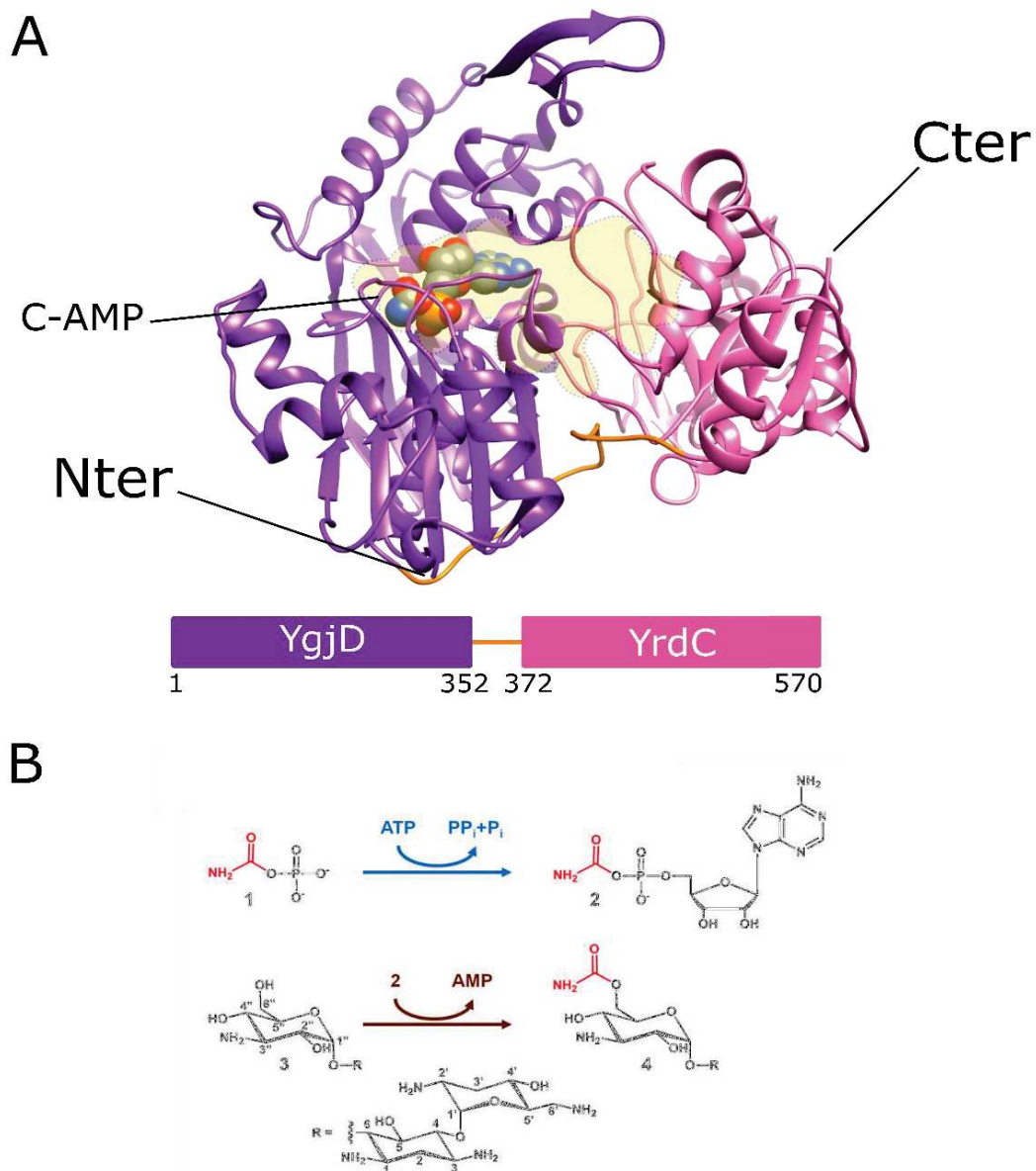


Figure 19. La O-carbamoylase TobZ

(A) Structure de TobZ de *S. tenebrarius* (codes PDB : 3VEZ, 3VEX). Le domaine YrdC-like est en rose et le domaine YgjD-like en violet. Une molécule de carbamoyl-AMP (C-AMP) est liée au sein du domaine YgjD. Le tunnel formé par les domaines YrdC et YgjD est indiqué en jaune. **(B)** Réaction catalysée par TobZ. Un carbamoyl-AMP (2) est formé par le domaine YrdC à partir de carbamoyl-phosphate (1) et ATP. Cet intermédiaire est ensuite transmis au domaine YgjD qui se charge de transférer le groupement carbamoyl (en rouge) sur la tobramycine (3) afin de former de la nébramycine (4). D'après (Parthier et al., 2012)

d. Les homologues de Sua5/YrdC et Kae1/YgjD/Qri7

La O-carbamoylase TobZ

Les protéines de la famille CmcH/TobZ/NodU sont composées d'un domaine N-terminal homologue à YgjD et d'un domaine C-terminal homologue à YrdC. A ce jour, elles sont retrouvées uniquement chez une dizaine d'espèces de bactéries (principalement des rhizobiales). Ces enzymes catalysent l'O-carbamoylation de métabolites secondaires tels que des antibiotiques ou des facteurs de nodulation. Ainsi, la protéine TobZ catalyse la carbamoylation de la tobramycine, un antibiotique du type aminoglycoside, afin de produire de la nébramycine. La protéine TobZ de la bactérie *Streptoalloteichus tenebrarius* (*S. tenebrarius*) a fait l'objet d'une étude structure-fonction qui a confirmé une forte similarité avec les protéines YrdC et YgjD ainsi que le rôle respectif de ces domaines (Figure 19A) (Parthier et al., 2012).

Le domaine YrdC de TobZ utilise une molécule de carbamoyl-phosphate et une molécule d'ATP pour produire du PPi et du carbamoyl-AMP (Figure 19B). L'ATP se lie à YrdC au sein d'un site dit cryptique car il est accessible uniquement lorsque le carbamoyl-phosphate y est déjà lié. L'intermédiaire carbamoyl-AMP est ensuite transféré vers le site actif du domaine YgjD via un tunnel formé par l'interface des deux domaines et protégé du solvant. Ceci permet d'éviter la diffusion et/ou la dégradation des ligands, en particulier de l'intermédiaire adénylate labile. Une fois le carbamoyl-AMP et la tobramycine liés, YgjD catalyse le transfert du groupement carbamoyl sur la fonction 6'-OH du substrat. Le domaine YgjD contient un atome de métal nécessaire à cette réaction. L'identification de l'intermédiaire carbamoyl-AMP, synthétisé par le domaine YrdC, a permis de proposer, par homologie avec le système t⁶A qui utilise un groupement thréonyle supplémentaire, que Sua5/YrdC produise une molécule de thréonyl-carbamoyl-AMP (TC-AMP). La révision de la structure de Sua5 de *S. tokodaii* par les auteurs de l'étude sur TobZ est venue conforter cette idée. En effet, le nuage de densité électronique présent au sein du site actif, initialement annoté comme une molécule d'AMP (Agari et al., 2008), s'est avéré être une molécule de TC-AMP co-purifiée (Parthier et al., 2012). Finalement, le TC-AMP, produit par Sua5 de *B. subtilis*, a pu être isolé et caractérisé, confirmant ainsi qu'il s'agit de l'intermédiaire réel de la voie de biosynthèse de t⁶A (Lauhon, 2012).

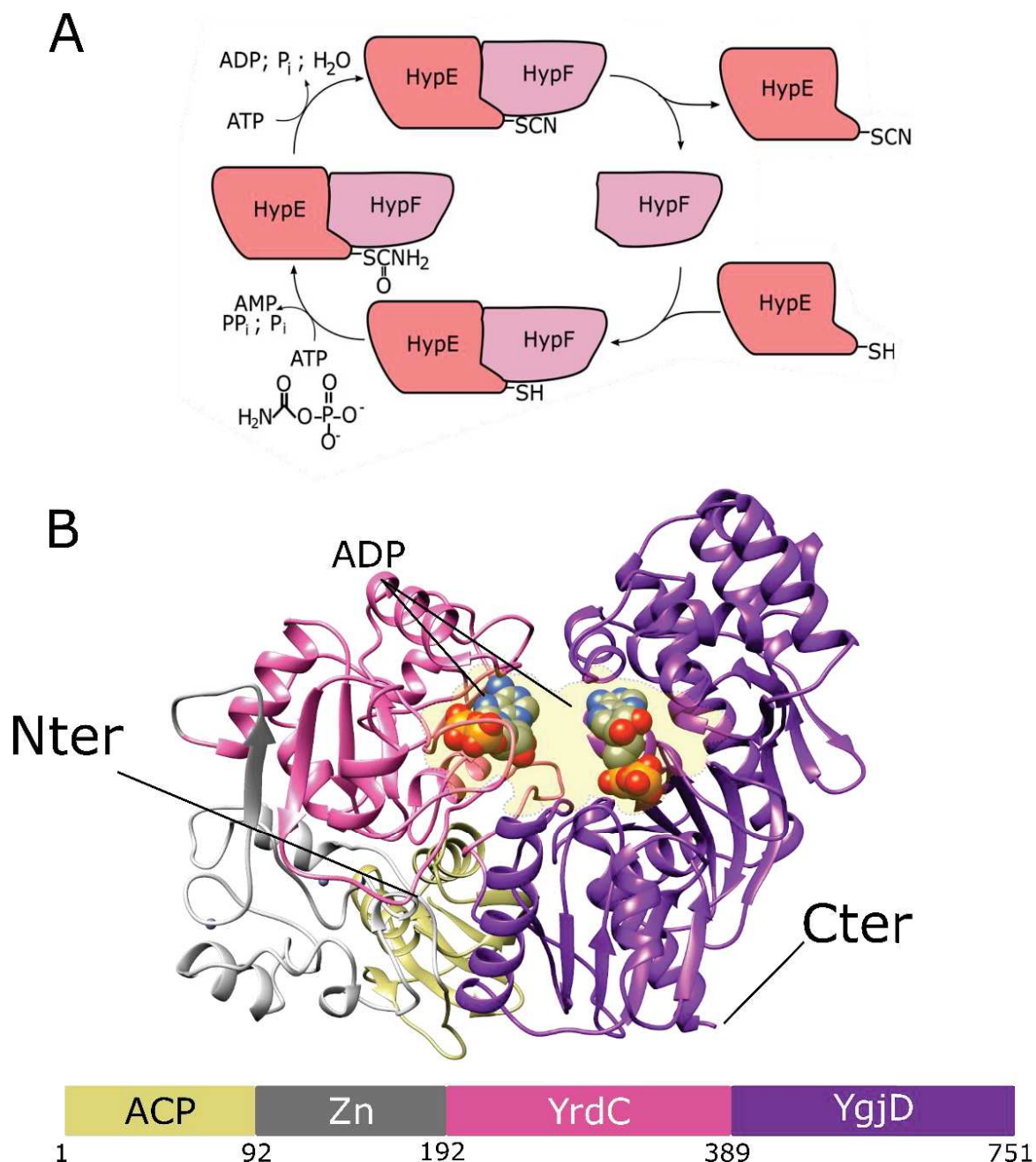


Figure 20. Fonction cellulaire et structure de HypF

(A) Modification de HypE par HypF dans la voie maturation des hydrogénases [NiFe]. HypF consomme un carbamoyl-phosphate (1) et un ATP afin de réaliser la S-carbamoylation du groupement thiol d'une cystéine de HypE. Le groupement thiocarbamate ainsi formé servira de substrat à HypE pour modifier à son tour la grande sous-unité d'une hydrogénase, ceci afin de permettre *in fine* la formation du cluster [NiFe]. Adapté de (Lacasse and Zamble, 2016). **(B)** Structure de HypF de *Caldanaerobacter subterraneus* (code PDB : 3VTI)(Shomura and Higuchi, 2012). Le domaine ACP est en jaune. Le domaine Zn est en gris. Les domaines YrdC et YgjD sont en rose et violet. Une molécule d'ADP est complexée au sein des sites d'YrdC et d'YgjD. Le tunnel formé par les domaines YrdC et YgjD est indiqué en jaune.

La S-carbamoylase HypF

Les hydrogénases catalysent des réactions rédox réversibles impliquant des molécules d'hydrogènes. Les hydrogénases de la classe [NiFe] sont des protéines hétéromériques composées au minimum d'une large sous-unité, qui possède un ion métallique fer Fe et un ion métallique nickel Ni dans son centre catalytique, et d'une petite sous-unité qui possède un ou plusieurs cluster(s) Fe-S. La maturation des hydrogénases [NiFe] implique la liaison de ligands non protéiques sur la grande sous-unité par des protéines accessoires codées par l'opéron *hyp*. HypE est une de ces protéines accessoires. Elle est essentielle et participe à la synthèse de deux groupements thiocyanates sur cette grande sous-unité (Figure 20A). Le nouveau groupement thiocyanate est formé via le transfert et la déshydratation d'un groupement thiocarbamate, présent sur une cystéine modifiée en C-terminal de HypE. Le gène *hypF*, qui ne fait pas partie de l'opéron *hyp*, code pour l'enzyme qui est responsable de la S-carbamoylation du groupement thiol de cette cystéine de HypE (Rangarajan et al., 2008). HypF est retrouvée chez de nombreux procaryotes.

HypF est une protéine monomérique composée d'un domaine N-terminal *acylphosphatase* (ACP) suivi d'un domaine *zinc-finger* puis d'un domaine nommé *middle*, homologue d'YrdC, et enfin d'un domaine C-terminal, homologue d'YgjD (Figure 20B) (Paschos et al., 2002; Shomura and Higuchi, 2012; Tominaga et al., 2012). L'ordre inversé des domaines YgjD et YrdC entre TobZ et HypF suggère qu'elles proviennent de deux événements de fusion indépendants. Comme TobZ, HypF utilise une molécule de carbamoyl-phosphate et un ATP afin de former un intermédiaire carbamoyl-AMP et un PPi. Le groupement carbamoyl de l'intermédiaire est ensuite transféré vers la cystéine C-terminale de HypE afin de former le groupement thiocarbamate final (Blokesch et al., 2004; Paschos et al., 2002; Rangarajan et al., 2008). Les analyses structurales de HypF montrent que, de manière analogue à TobZ, l'organisation des domaines crée une cavité permettant le transfert des ligands d'un domaine au suivant lors de la réaction. Le mécanisme exact de HypF reste à élucider avec certitude. Le modèle actuel suppose que le domaine ACP est impliqué dans l'hydrolyse du carbamoyl-phosphate. Le domaine *middle* (YrdC) assurerait ensuite la formation de carbamoyl-AMP et enfin le domaine YgjD catalyserait le transfert du carbamoyl sur HypE (Lacasse and Zamble, 2016; Petkun et al., 2011; Rangarajan et al., 2008; Shomura and Higuchi, 2012). La fonction du domaine *zinc-finger* est encore inconnue.

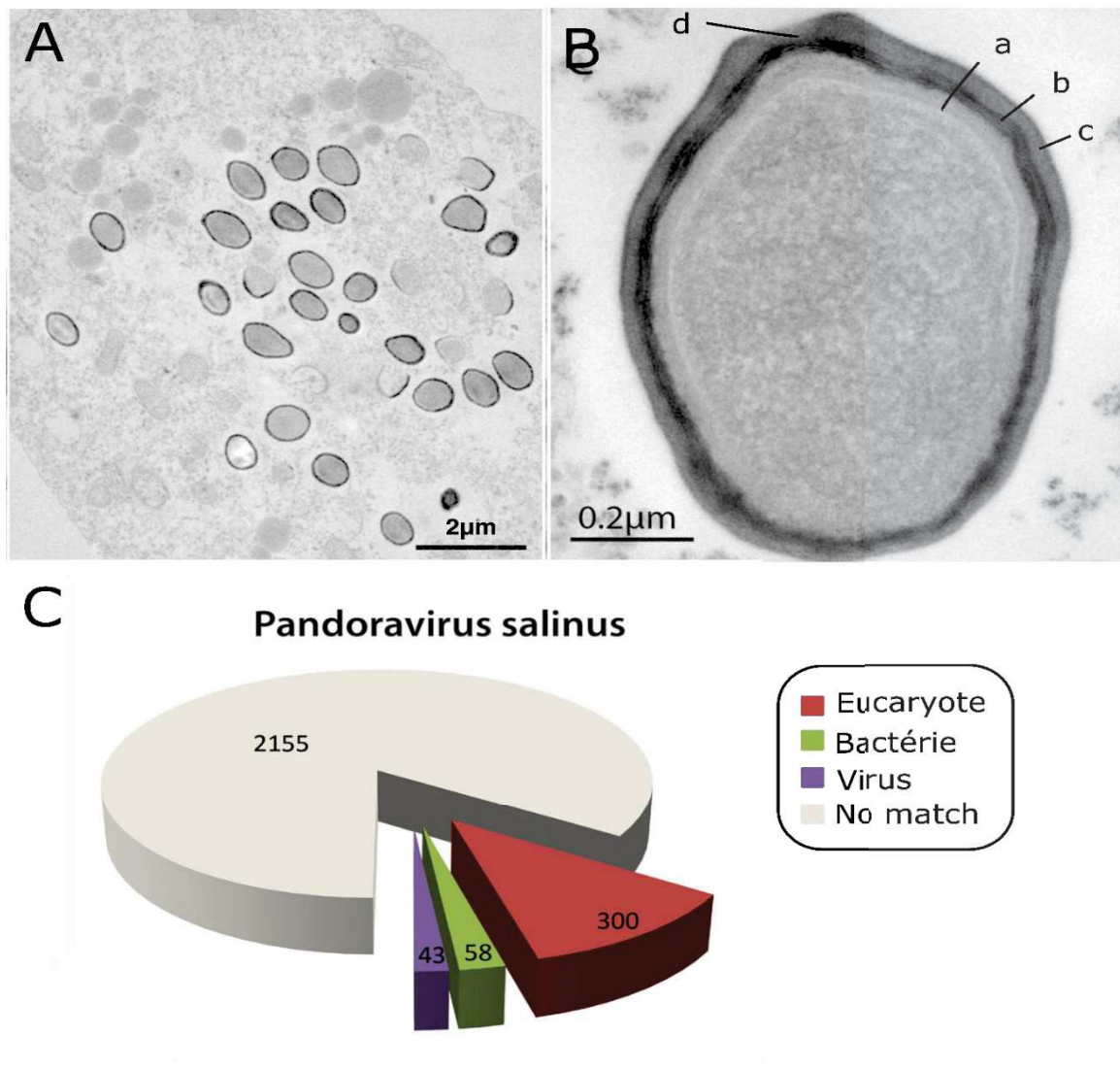


Figure 21. Apparence et génome de *Pandoravirus salinus*

(A) Coupe fine d'une cellule d'*Acanthamoeba* remplie de virus *P. salinus* à des stades variés de leur maturation, vue au microscope électronique. **(B)** Coupe fine de *P. salinus* mature vue au microscope électronique. Le tégument des pandoravirus possède 3 couches, du cytoplasmique vers l'extracellulaire : une couche de faible densité de composition inconnue (a), une couche intermédiaire dense aux électrons composée de mailles fibrilles (b) et une couche externe de densité moyenne proche en composition de la membrane cellulaire de l'hôte (c). Cette dernière couche est responsable de la réponse positive au test de Gram des pandoravirus. Ce tégument unique en son genre est interrompu au sommet par la présence d'un ostiole (d). Cette structure permet la libération du contenu viral dans le cytoplasme de l'hôte, après que la particule virale ait été phagocytée par ce dernier. D'après (Philippe et al., 2013). **(C)** Protéines putatives codées par *P. salinus* et leur plus proches homologues dans les bases de données.

La t⁶A synthétase des virus géants

Les pandoravirus appartiennent à un groupe de virus à ADN double brin, appelés virus géants à cause de leur taille. Leur seul hôte connu actuellement est le genre *Acanthamoeba*, des eucaryotes unicellulaires. Ils apparaissent au microscope optique comme des particules ovoïdes similaire en taille et en apparence à des cellules procaryotes (Figure 21). Ils ont été observés pour la première fois en 2008 (Scheid et al., 2008) et, bien qu'ils n'étaient pas cultivables seuls, leur apparence « cellulaire », ainsi que leur réponse positive au test de Gram, leur a valu d'être considéré comme des bactéries atypiques. Il fallut attendre que leur génome soit séquencé en 2013 pour comprendre qu'il s'agit en réalité de virus (Philippe et al., 2013). En effet, il leur manque des composants essentiels à la vie cellulaire tels qu'une machinerie traductionnelle (ribosome et protéines r), des ATP-synthases ainsi que des éléments de la fission cellulaire ou de la réplication de l'ADN. Ces caractéristiques ne sont pas uniques aux pandoravirus puisque d'autres virus géants, comme les mimivirus, pithovirus et mollivirus, les partagent. Les pandoravirus avaient la particularité d'être les premiers virus géants chez lesquels aucune protéine de capsid n'était clairement identifiée, ce qui expliquerait leur aspect ovoïde. Récemment, un gène candidat a cependant été identifié (Sinclair et al., 2017). Ces dernières années, l'apparence des virus géants et la complexité de leur génome viennent ébranler les barrières, jusque-là évidentes, qui séparent le monde cellulaire et le monde des virus. En particulier, de nombreuses questions quant à leur origine évolutive sont vivement débattues.

Trois souches différentes ont été entièrement séquencées : *Pandoravirus salinus* (*P. salinus*) (2.5 Mb) et *P. dulcis* (1.9 Mb) (Philippe et al., 2013) ainsi que *P. inopinatum* (2.2 Mb) (Scheid, 2016). Les tailles respectives de leur génome en font parmi les plus grands virus actuellement recensés. *P. salinus*, par exemple, code pour plus de 2500 protéines, dont 93 % n'ont aucun homologue dans les bases de données actuelles (Figure 21C). Parmi les 7% restant, un gène particulier a attiré mon attention puisque chez *P. salinus* et *P. dulcis* se trouve une protéine putative correspondant à une fusion de Sua5, YgjD et TcdA (Philippe et al., 2013). Il s'agit du premier cas de t⁶A synthétase identifié chez un virus.

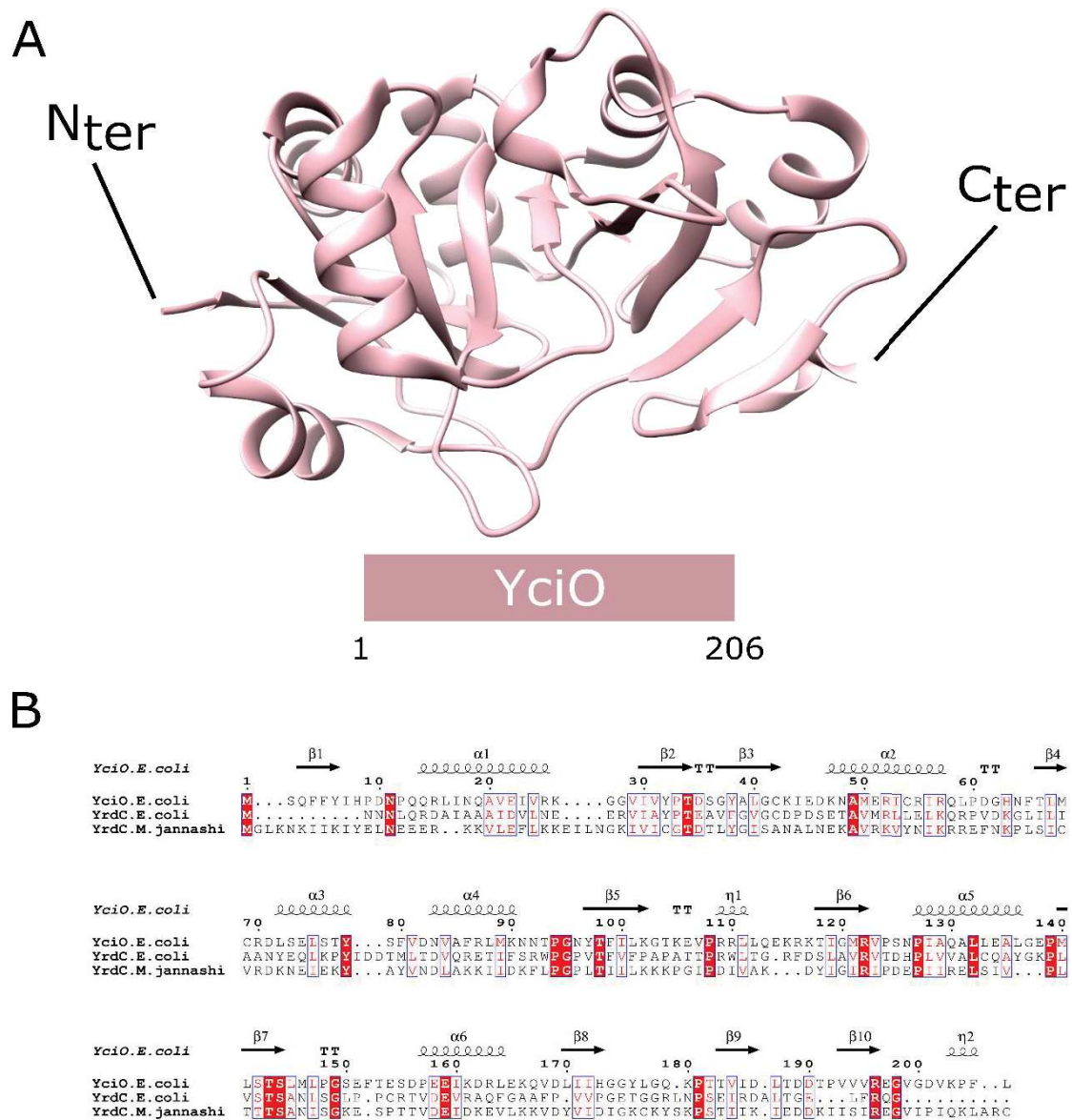


Figure 22. Le paralogue YciO

(A) Structure d'YciO d'*E. coli* (code PDB : 1KK9) (Jia et al., 2002). **(B)** Alignement d'YciO avec YrdC d'*E. coli* et de l'archée *M. jannashii*. Les structures secondaires correspondantes d'après la structure d'YciO sont indiquées au-dessus de l'alignement. Les positions identiques sont surlignées en rouge et les positions similaires sont écrites en rouge.

Les paralogues YciO et Slr0006

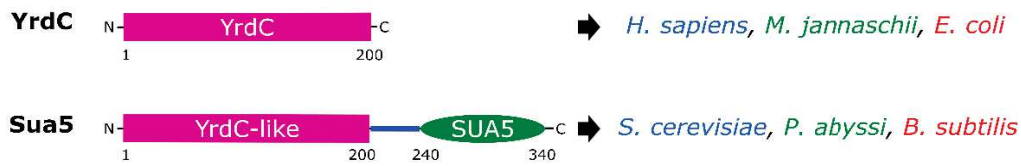
Chez certaines bactéries et certaines plantes, il existe un second membre de la famille Sua5/YrdC, un paralogue d'YrdC, appelé YciO. Ainsi sur les 9200 génomes bactériens analysés par Thiaville et al., plus de la moitié possède YciO (Thiaville et al., 2015a). Malgré une identité de séquence inférieure à 30 %, la structure d'YciO est très similaire à celle d'YrdC (Figure 22) (Jia et al., 2002). YciO ne complémente pas l'absence d'YrdC et n'est pas essentiel chez *E. coli*. (El Yacoubi et al., 2009). La fonction d'YciO semble donc indépendante de la synthèse de t⁶A. Curieusement, *yciO* fait partie des gènes dont la délétion mène à une accumulation de glycogène dans les cellules (Montero et al., 2009). Sa fonction exacte reste cependant inconnue. La présence de Arg¹²² et Ser¹⁸², des résidus conservées de Sua5/YrdC, suggèrent qu'YciO est susceptible de lier la L-thréonine ou un résidu similaire (Kuratani et al., 2011). En revanche la tétrade KxR/SxN, essentielle pour la liaison de l'ATP chez Sua5/YrdC, n'est pas conservée chez YciO (El Yacoubi et al., 2009; Gerdes et al., 2011; Harris et al., 2015).

D'après l'analyse des séquences, un second groupe paralogue plus éloigné, nommé Slr0006, a été identifié chez des *Cyanobactéries* (Carmel et al., 2013; Carvalho et al., 2017). La structure de Slr0006 n'est pas résolue mais les modélisations suggèrent qu'elle possède également un repliement similaire à YrdC. Contrairement à YciO, la délétion de *slr0006* ne mène pas à une accumulation de glycogène chez *E. coli* (Carmel et al., 2013). La fonction de Slr0006 demeure un mystère.

Au bilan, les approches de génomique comparative et de génétique ont démontré le rôle central des familles d'enzymes universelles Sua5/YrdC et Kae1/YgjD/Qri7 dans la synthèse de t⁶A. Cette voie de biosynthèse implique deux étapes séquentielles. Lors de la première étape, Sua5/YrdC utilise comme substrats la L-thréonine, un composé carbonaté ($\text{HCO}_3^-/\text{CO}_2$) et l'ATP pour produire une molécule de L-thréonyl-carbamoyl-adénylate (TC-AMP). Ainsi activé, le groupement L-thréonyl-carbamoyl peut être transféré sur l'ARNt substrat afin de former un t⁶A-ARNt. Cette seconde étape est dépendante de la famille Kae1/YgjD/Qri7 et requiert des protéines accessoires domaine-spécifique. Je vais expliciter ci-dessous les caractéristiques de la famille Sua5/YrdC, qui est au cœur de mon projet de thèse. Une description similaire de la famille Kae1/YgjD/Qri7 et de ses différents partenaires protéiques est disponible en Annexe 1 et peut être lue indépendamment du reste de l'introduction.

A

Eucaryote, Archée, Bactérie



B

	1	200	240	340
<i>E. coli</i>	1	200	240	340
<i>M. jannaschii</i>	1	200	240	340
<i>H. sapiens</i>	1	200	240	340
<i>B. subtilis</i>	1	200	240	340
<i>P. abyssi</i>	1	200	240	340
<i>S. cerevisiae</i>	1	200	240	340

	13	200	240	340
<i>E. coli</i>	13	200	240	340
<i>M. jannaschii</i>	21	200	240	340
<i>H. sapiens</i>	73	200	240	340
<i>B. subtilis</i>	24	200	240	340
<i>P. abyssi</i>	19	200	240	340
<i>S. cerevisiae</i>	55	200	240	340

YrdC-like

	79	200	240	340
<i>E. coli</i>	79	200	240	340
<i>M. jannaschii</i>	86	200	240	340
<i>H. sapiens</i>	135	200	240	340
<i>B. subtilis</i>	86	200	240	340
<i>P. abyssi</i>	81	200	240	340
<i>S. cerevisiae</i>	130	200	240	340

	151	200	240	340
<i>E. coli</i>	151	200	240	340
<i>M. jannaschii</i>	155	200	240	340
<i>H. sapiens</i>	207	200	240	340
<i>B. subtilis</i>	161	200	240	340
<i>P. abyssi</i>	155	200	240	340
<i>S. cerevisiae</i>	208	200	240	340

linker

	207	240	280	340
<i>E. coli</i>	207	240	280	340
<i>M. jannaschii</i>	254	240	280	340
<i>H. sapiens</i>	228	240	280	340
<i>B. subtilis</i>	223	240	280	340
<i>P. abyssi</i>	223	240	280	340
<i>S. cerevisiae</i>	278	240	280	340

SUA5

	279	280	340
<i>E. coli</i>	279	280	340
<i>M. jannaschii</i>	287	280	340
<i>H. sapiens</i>	282	280	340
<i>B. subtilis</i>	352	280	340
<i>P. abyssi</i>	352	280	340
<i>S. cerevisiae</i>	352	280	340

Figure 23. Les enzymes Sua5 et YrdC

(A) Schéma représentatif d'une protéine YrdC (au-dessus) et d'une protéine Sua5 (en-dessous). Des exemples d'espèces des trois domaines du vivant possédant l'une ou l'autre sont donnés à droite. (B) Alignements de trois séquences d'YrdC (au-dessus) et trois séquences de Sua5 (au-dessous). La partie correspondant au domaine YrdC est encadrée en rose, celle correspondant au linker en bleu et celle correspondant au domaine SUA5 en vert. Les positions identiques sont surlignées en rouge. Les positions similaires sont écrites en rouge.

VI- La famille Sua5/YrdC

La première étape de synthèse de t⁶A implique la formation d'un intermédiaire de réaction, le TC-AMP, par les enzymes de la famille universelle Sua5/YrdC. Sua5 et YrdC partagent un domaine catalytique orthologue. La distinction entre les deux se fait actuellement sur la présence ou l'absence d'un second domaine C-terminal. En effet, alors qu'YrdC est constituée uniquement du domaine éponyme, Sua5 possède quant à elle ce second domaine distinct, nommé SUA5, relié via un linker au domaine N-terminal YrdC-like ([Figure 23](#)). La fonction du domaine SUA5 est encore inconnue.

L'analyse de 9200 génomes d'espèces bactériennes a montré que 73 % possédaient YrdC, 31 % possédaient Sua5 et 9 % possédaient les deux ([Thiaville et al., 2015a](#)). En réalité, la majorité des organismes possèdent une seule des deux versions, c'est-à-dire soit Sua5, comme *S. cerevisiae*, *Sulfolobus tokodaii* (*S. tokodaii*) et *B. subtilis*, soit YrdC, comme *H. sapiens*, *H. volcanii* et *E. coli*. Une exception à la règle intéressante est le cas de *Vibrio cholerae* dont le chromosome I code pour les deux versions, YrdC et Sua5. La délétion individuelle des deux gènes n'a pas d'effet phénotypique ce qui confirme que les deux versions sont actives dans la cellule et que chacune d'entre elle produit suffisamment de TC-AMP pour assurer la survie cellulaire ([Chao et al., 2013](#); [Kamp et al., 2013](#)).

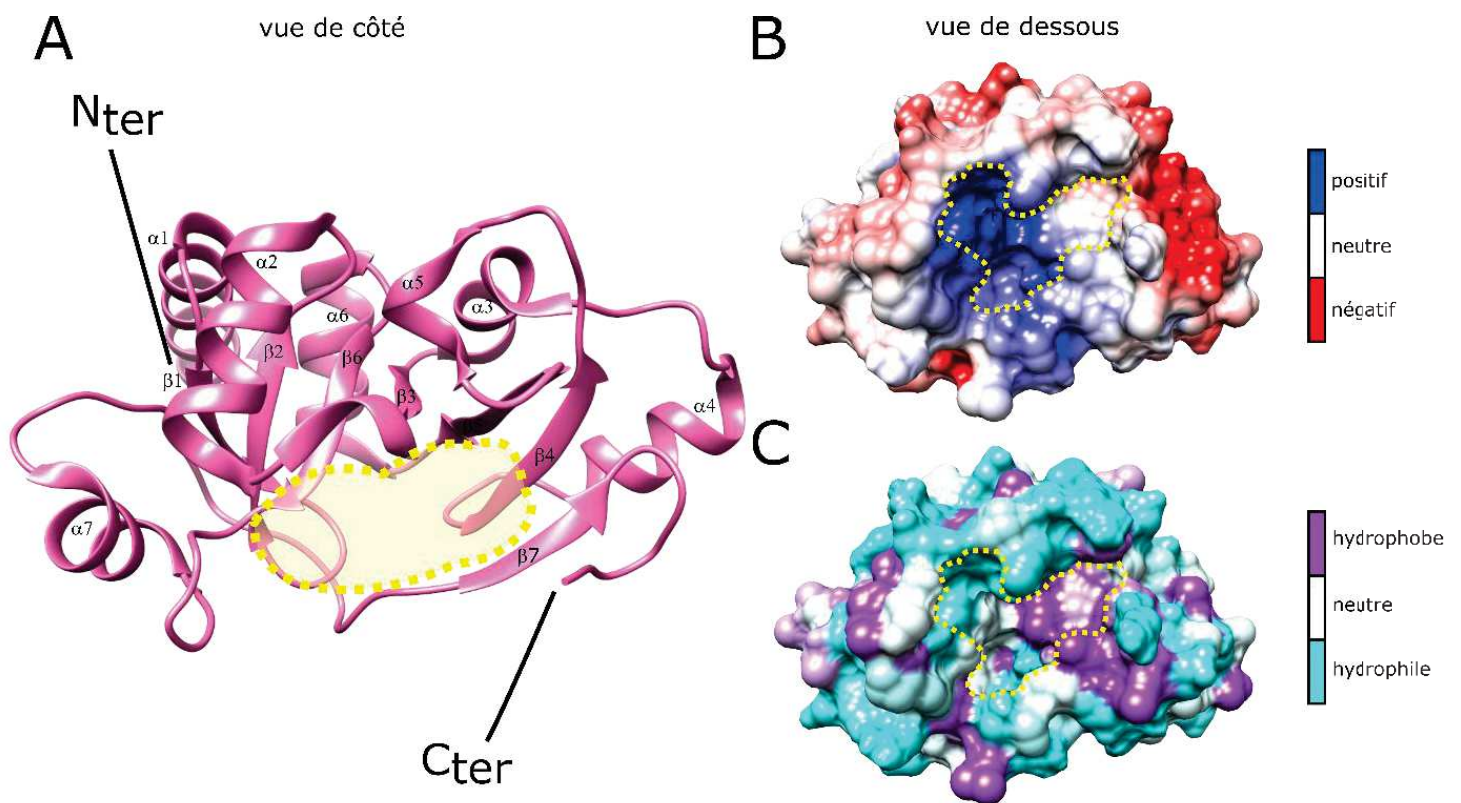


Figure 24. La protéine YrdC d'*E. coli*

(A) Structure d'YrdC d'*E. coli* (code PDB : 1HRU chaîne A) (Teplova et al., 2000) représentée sous forme de rubans. Les numéros des hélices α et brins β , comme décrit dans le texte, sont indiquées en noir. La position du site actif est indiquée en jaune. **(B)** Surface électrostatique de la protéine. La cavité formant le site actif a une charge positive globale importante. **(C)** Surface d'hydrophobicité de la protéine, colorée selon l'échelle d'hydrophobicité de (Kyte and Doolittle, 1982). (B) et (C) correspondent à des vue de dessous (rotation de 90 °C par rapport à A). Le site actif est délimité par un trait jaune en pointillé.

a. Structures d'YrdC et Sua5

La structure d'YrdC d'*E. coli* a été résolue pour la première fois par cristallographie en 2000 (PDB : 1HRU) (Teplova et al., 2000) (Figure 24A). Elle appartient à la superfamille 3,4-dihydroxy-2-butanone-4-phosphate synthase (DHBP synthase – Code CATH : 3.90.870.10). La protéine d'environ 200 acides aminés est composée par 7 hélices α et 7 brins β , organisés en feuillets parallèles et antiparallèles. Elle possède une forme globulaire torsadée. Sa forme générale lui vaut d'être comparée à un gant de baseball. Sa topologie permet de distinguer trois parties : la première est composée par $\alpha 1$, $\beta 1$, $\beta 2$ et $\alpha 2$ et la seconde par $\beta 3$, $\alpha 3$, $\alpha 4$, $\beta 4$, $\alpha 5$, et $\beta 5$. La troisième partie, composée par $\alpha 6$ et $\beta 6$, fait le pont entre les deux précédentes. L'extrémité C-terminale, formée par $\alpha 7$ et $\beta 7$, longe le reste de la protéine. YrdC présente une large cavité formée par $\beta 3$, $\beta 4$, $\beta 5$, $\beta 7$ et $\alpha 5$. La cavité contient de nombreux résidus conservés de la famille Sua5/YrdC. La majeure partie de cette dépression présente un potentiel électrostatique positif (Figure 24B). Le reste de la cavité a globalement un caractère hydrophobe (Figure 24C).

Des mesures de diffusion dynamique de la lumière (*dynamic light scattering* – DLS) montrent qu'YrdC est monomérique en solution (Teplova et al., 2000). Cependant elle forme un homodimère symétrique au sein du cristal. Les contacts entre les deux monomères sont susceptibles d'influencer la structure d'YrdC. En particulier, la structure de l'extrémité C-terminale n'est pas résolue. Pour compléter ces données, la structure du monomère d'YrdC d'*E. coli* en solution a été résolue par RMN (PDB : 2MX1) (Harris et al., 2015). La structure globale est similaire à celle observée précédemment par Teplova et al. L'approche RMN permet cependant de confirmer que l'extrémité C-terminale de la protéine est particulièrement flexible (Carvalho et al., 2017; Harris et al., 2015).

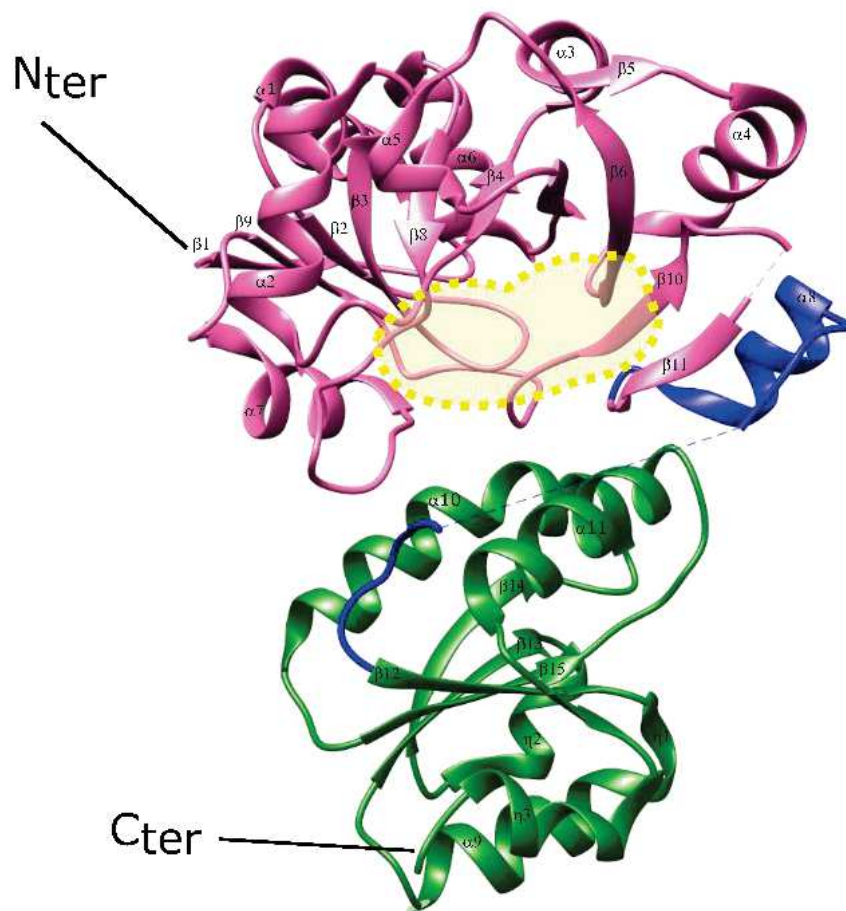


Figure 25. Structure cristallographique de Sua5 de *S. tokodaii*

La protéine Sua5 de *S. tokodaii* (code PDB : 4E1B) est représentée sous forme de rubans (Agari et al., 2008; Parthier et al., 2012). Les couleurs des domaines sont attribuées d'après la Figure 23A. Les numéros des hélices α et brins β , d'après (Kuratani et al., 2011), sont indiquées en noir. La présence de structures supplémentaires par rapport à YrdC d'*E. coli* explique que la correspondance entre numéro d'hélice ou de brin et position dans la structure ne soit pas similaire. La topologie générale du domaine reste cependant la même. Le site actif est indiqué en jaune.

La protéine YrdC d'*E. coli* lie sélectivement l'ATP et la L-thréonine (Cantara et al., 2011; El Yacoubi et al., 2009). Des changements conformationnels au sein de la protéine s'opèrent suite à la liaison des substrats (Harris et al., 2015). Par exemple, la flexibilité des chaînes latérales d'Arg¹⁸⁸ et Asn¹⁸⁹, qui sont particulièrement mobiles par rapport aux autres résidus, est affectée par la liaison de l'ATP. La flexibilité de la région C-terminale Gly¹⁶⁷-Gly¹⁹⁰ diminue également en présence d'ATP. De plus, la présence de deux glycines consécutives Gly¹⁷⁰-Gly¹⁷¹ suggère que cette région peut fonctionner comme une charnière permettant à l'extrémité C-terminale d'agir comme une *gating loop* pendant la liaison des substrats. Les auteurs proposent également l'hypothèse selon laquelle cette région puisse jouer un rôle dans des interactions protéiques avec les autres enzymes de la voie de synthèse de t⁶A et/ou dans l'interaction avec d'autres protéines.

La structure de Sua5 de l'euryarchée hyperthermophile *S. tokodaii* a été résolue par cristallographie (PDB : 3AJE et 4E1B) (Figure 25) (Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012). Sua5 de *S. tokodaii* est également un monomère en solution (Kuratani et al., 2011). La topologie du domaine YrdC-like de Sua5 *S. tokodaii* est similaire à celle de son orthologue YrdC chez *E. coli* avec lequel elle partage 42.9 % de similarité de séquence. Tout comme YrdC, le domaine catalytique de Sua5 présente une cavité électropositive et conservée.

Le domaine C-terminal SUA5, absent chez YrdC, possède un repliement de Rossmann, composé de 3 hélices α (α 9-11) encadrant 5 brins β parallèles (β 12-16). Les hélices 10 et 11 font face au domaine YrdC-like. La structure du domaine SUA5 a été comparée avec les structures connues dans la PDB à l'aide de l'algorithme DALI (Holm and Rosenström, 2010). Le domaine Rossmann étant fréquemment impliqué dans la liaison de nucléotides, les structures les plus proches trouvées par DALI appartenaient à des protéines liant des nucléosides comme par exemple HypB de *M. jannaschii* (PDB code: 2HF9 chain A), une protéine liant le GTP. Cette observation avait mené les auteurs à suggérer que le domaine SUA5 soit impliqué dans la liaison de nucléotide et/ou d'ARN, bien qu'aucune interaction de ce genre n'ait pu être mise en évidence (Agari et al., 2008; Kuratani et al., 2011).

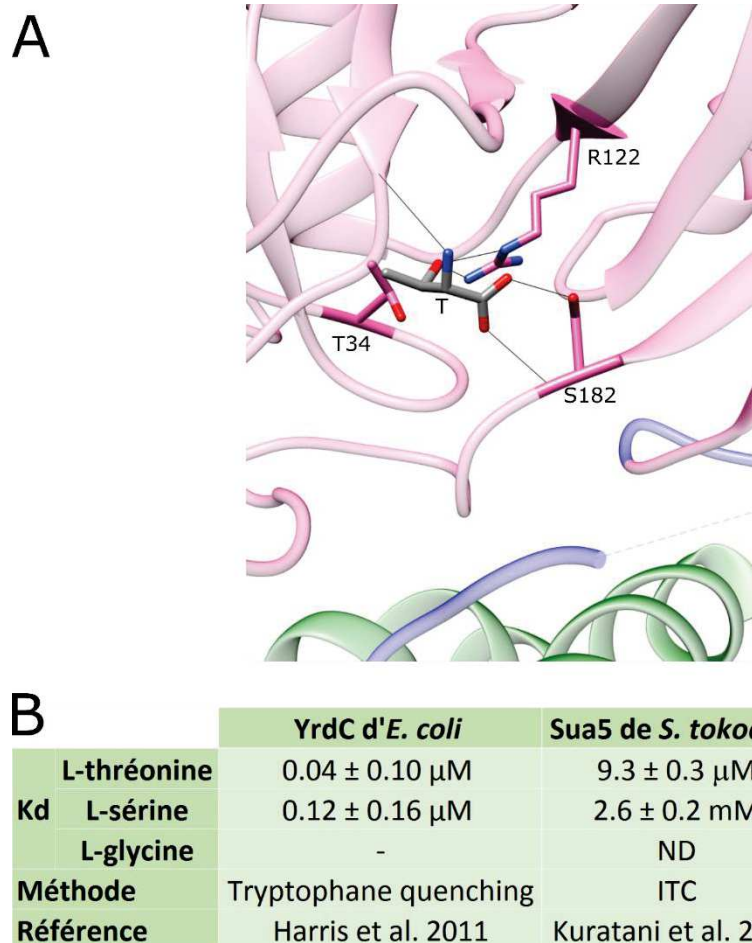


Figure 26. Coordination de la L-thréonine par Sua5/YrdC

(A) La position de la L-thréonine est montrée au sein du site actif de Sua5 de *S. tokodaii* (code PDB : 3AJE) (Kuratani et al., 2011). La chaîne principale est en transparente, à l'exception des résidus impliqués dans la liaison de la L-thréonine, à savoir Thr³⁴, Arg¹²² et Ser¹⁸². **(B)** Résumé des constantes d'affinités pour les acides aminés de Sua5 et YrdC.

b. Les modes de liaison des substrats

Liaison de $\text{HCO}_3^-/\text{CO}_2$

La nature exacte du substrat carbonaté nécessaire pour la formation de TC-AMP n'est pour le moment pas déterminée avec certitude. En effet HCO_3^- et CO_2 coexistent dans un environnement aqueux et les données actuelles ne permettent de savoir lequel est le substrat réel de Sua5/YrdC. La réaction *in vitro* de synthèse de $t^6\text{A}$ peut donc se faire en ajoutant l'un ou l'autre à la solution de réaction.

Liaison de la L-thréonine

Des mesures ITC évaluent la constante d'affinité de Sua5 de *S. tokodaii* à $9.3 \pm 0.3 \mu\text{M}$ pour la L-thréonine contre $2.6 \pm 0.2 \text{ mM}$ pour la L-sérine (Kuratani et al., 2011). Aucun signal n'est détecté pour la L-glycine. De même la liaison de la L-thréonine à YrdC d'*E. coli* a été suivie en mesurant la fluorescence des tryptophanes lors de la titration par l'acide aminé. YrdC d'*E. coli* lie la L-thréonine avec un K_d de $0.04 \pm 0.10 \mu\text{M}$ et la L-sérine avec un K_d de $0.12 \pm 0.16 \mu\text{M}$ (Harris et al., 2011). Ces données montrent une grande spécificité de Sua5 et YrdC pour la L-thréonine. La structure de Sua5 de *S. tokodaii* en complexe avec la L-thréonine et l'AMPPNP, un analogue non hydrolysable de l'ATP, montre que le site de liaison de la L-thréonine se situe dans la partie profonde de la cavité (Kuratani et al., 2011). Elle interagit avec plusieurs résidus dont Thr³⁴, Arg¹²² et Ser¹⁸² qui sont conservés au sein de la famille Sua5/YrdC (Figure 26) (Kuratani et al., 2011). Thr³⁴ procure un environnement hydrophobe pour le carbone γ de la thréonine, Arg¹²² forme une liaison hydrogène avec la fonction hydroxyle de l'acide aminé et Ser¹⁸² forme deux liaisons hydrogènes avec les oxygènes de la fonction carboxyle. La mutation des résidus correspondants Thr⁷⁰, Arg¹⁷⁴ et Ser²⁴³ chez *S. cerevisiae* diminue de 75 % la synthèse de $t^6\text{A}$ *in vitro* pour le premier et l'abolit complètement pour les autres (Wan et al., 2013). Des analyses RMN d'YrdC d'*E. coli* détectent une liaison avec la L-thréonine pour les résidus correspondants Thr²⁷ et Ser¹⁷⁶, mais pas Arg¹¹⁷ (Harris et al., 2015). Le *docking* moléculaire de la L-thréonine sur YrdC d'*E. coli* suggère que Ser¹⁷⁶ coordonne la L-thréonine de manière similaire chez Sua5 de *S. tokodaii* et YrdC d'*E. coli*. En revanche Thr²⁷ semble agir différemment puisqu'elle forme des liaisons hydrogènes avec la fonction hydroxyle du substrat (Harris et al., 2015). D'après les auteurs, cette différence pourrait être une authentique distinction entre le mode de liaison de la L-thréonine de Sua5 et YrdC ou un artefact du *docking* moléculaire.

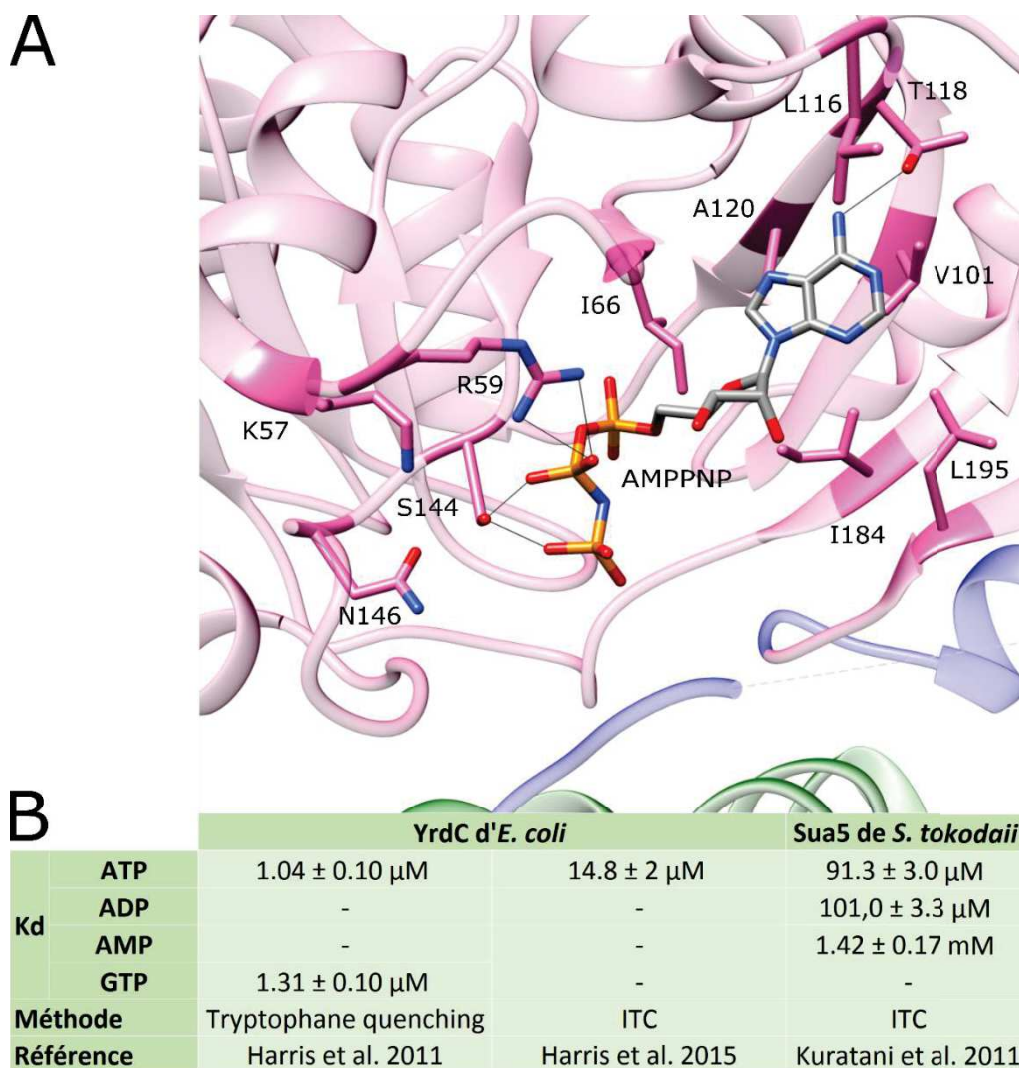


Figure 27. Coordination de l'ATP par Sua5/YrdC

(A) La structure de Sua5 de *S. tokodaii* a été résolue en complexe avec une molécule d'AMPPNP et une molécule de L-thréonine (non montrée) (code PDB : 3AJE) (Kuratani et al., 2011). Le groupement adénine occupe une surface hydrophobe possédant de nombreux résidus conservés. Les phosphates sont coordonnés par la tétrade K⁵⁷XR⁵⁹/S¹⁴⁴YN¹⁴⁶. La chaîne principale est en transparente, à l'exception des résidus impliqués dans la liaison de l'ATP. **(B)** Résumé des constantes d'affinités pour les nucléotides de Sua5 et YrdC.

Liaison de l'ATP

La structure de Sua5 de *S. tokodaii* en complexe avec L-thréonine et AMPPNP révèle que le nucléoside se positionne également dans la cavité conservée du domaine catalytique (Kuratani et al., 2011). L'affinité de Sua5 de *S. tokodaii* pour l'ATP, l'ADP et l'AMP, sans rajout d'autres ligands, a été déterminée par ITC respectivement à $61.3 \pm 3.0 \mu\text{M}$, $101 \pm 3.3 \mu\text{M}$ et $1.42 \pm 0.17 \text{ mM}$ (Kuratani et al., 2011). Sua5 possède donc une spécificité significative pour l'ATP, bien qu'elle semble pouvoir également se lier à l'ADP voire même l'AMP *in vitro*. Plus récemment l'affinité d'YrdC d'*E.coli* pour l'ATP a été mesurée à $14.8 \pm 2.0 \mu\text{M}$ par ITC (Harris et al., 2015).

L'adénine se positionne dans la région hydrophobe de la cavité où elle est coordonnée par les résidus Ile⁶⁶, Val¹⁰¹, Leu¹¹⁶, Ala¹²⁰, Ile¹⁸⁴ et Leu¹⁹⁵ (Figure 27). La liaison de l'ATP à YrdC d'*E. coli* a été analysée par ¹H-¹⁵N-HSQC et par *docking* moléculaire grâce au logiciel HADDOCK (van Zundert et al., 2016) (Harris et al., 2015). Le spectre ¹H-¹⁵N-HSQC est une méthode RMN qui permet d'identifier les décalages chimiques des couples H-N, chacun des couples étant généralement associé à un seul résidu au sein de la protéine. Ces décalages indiquent un changement dans l'environnement du résidu. Dans les deux approches, les résidus qui présentent les décalages chimiques les plus significatifs sont Ile⁵⁹, Leu¹¹⁴ et Ala¹¹⁵ (respectivement Ile⁶⁶, Leu¹¹⁶ et Ala¹²⁰ chez Sua5 de *S. tokodaii*). Tous ces résidus sont situés dans la région de liaison de l'adénine.

L'approche RMN a également détecté un déplacement chimique significatif pour le résidu Gly¹⁰⁹, bien qu'aucun changement au niveau de ce résidu n'ait été observé avec le *docking* moléculaire. Ce résidu est localisé dans la boucle entre $\beta 4$ et $\beta 5$, une région de la protéine qui n'a aucun contact direct avec l'ATP. Une possibilité est que la liaison de l'ATP provoque un changement de conformation de cette boucle chez YrdC (Harris et al., 2015).

Les phosphates de l'ATP sont liés au niveau de la dépression chargée positivement. Le motif LysxArg/SerxAsn (KxR/SxN) est un motif hautement conservé dans la famille Sua5/YrdC impliqué dans la liaison de l'ATP (El Yacoubi et al., 2009). La mutation de ces résidus chez Sua5 de *S. cerevisiae* inactive l'enzyme (Wan et al., 2013). Le motif K⁵⁷xR⁵⁹/S¹⁴⁴xN¹⁴⁶ est important pour la coordination des phosphates de l'AMPPNP dans la structure de Sua5 (Kuratani et al., 2011) pourtant l'approche ¹H-¹⁵N-HSQC ne détecte aucun *chemical shift* pour les résidus correspondants K⁵⁰, R⁵² ou N¹⁴¹ (S¹³⁹ n'est pas observable) lors de la liaison de l'ATP à YrdC, en absence d'autres substrats (Harris et al., 2015). La modélisation de de l'ATP au sein d'YrdC montre également que les phosphates ne sont en contact avec aucune région de la protéine (Harris et al., 2015).

L'affinité de Sua5 de *S. tokodaii* pour l'ATP *in vitro* est plus de trois fois inférieure à celle pour la L-thréonine. Dans la structure de Sua5, l'AMPPNP se positionne au-dessus du ligand carbonaté et de la L-thréonine et empêche donc leur passage. Ceci suggère que, pour que le TC-AMP puisse être formé, l'ATP doit être lié en dernier par l'enzyme. La modélisation d'YrdC d'*E. coli* liée à la L-thréonine et à l'ATP montre que, dans ce cas de figure, le motif K⁵⁰xR⁵²/S¹³⁹xN¹⁴¹ devient suffisamment proche des phosphates de l'ATP pour former des liaisons hydrogènes avec ces derniers (Harris et al., 2015). Ces données sont cohérentes avec le fait qu'YrdC ne produit de l'AMP qu'en présence de L-thréonine et HCO₃⁻/CO₂ (Deutsch et al., 2012). Cependant les modalités exactes de liaison des substrats et les constantes cinétiques de cette réaction demeurent inconnues pour le moment.

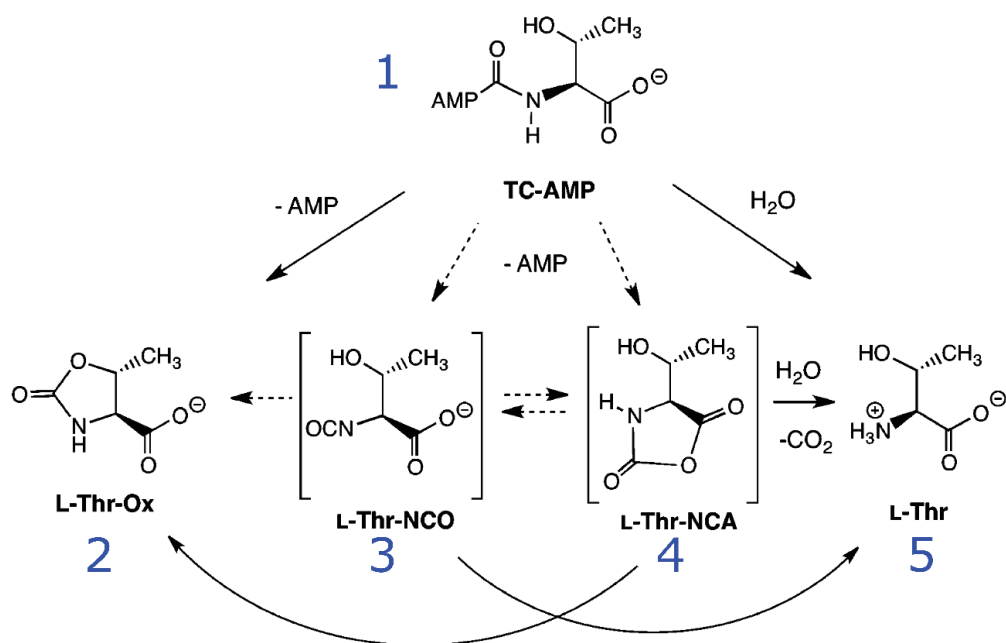


Figure 28. La dégradation du TC-AMP

En milieu aqueux, le TC-AMP (1) se dégrade spontanément pour donner comme produits finaux de l'AMP et un mélange équimolaire de L-thréonine (5) et de 5-méthyl-2-oxazolidinone-4-carboxylate (2). Les analyses HPLC et H¹-RMN du processus de dégradation suggèrent l'existence d'un ou plusieurs intermédiaires de dégradation, susceptible d'être un isocyanate (3) et/ou un N-carboxyanhydride (4). Ces composés, plus réactifs, se dégraderaient rapidement soit en 2 soit en 5. D'après (Lauhon, 2012).

c. Le L-thréonyl-carbamoyl-AMP

La structure de Sua5 en complexe avec une molécule de TC-AMP (code PDB : 4E1B) révèle la position de ce dernier au sein du site actif (Agari et al., 2008; Parthier et al., 2012). Les groupements L-thréonyl et adényle occupent la même position que celle des substrats L-thréonine et AMPPNP, respectivement, et sont coordonnés de la même manière. Un ion magnésium est impliqué dans la coordination du phosphate α . En outre, le magnésium est essentielle à l'activité *in vitro* d'YrdC/Sua5 (Agari et al., 2008; Parthier et al., 2012). Comme dans la plupart des protéines liant un NTP, il est probablement important pour la coordination des phosphates de l'ATP et/ou du TC-AMP au sein du site actif.

YrdC et Sua5 catalysent toutes les deux la formation du précurseur activé L-thréonylcarbamoyl-AMP (TC-AMP) à partir de L-thréonine, ATP, et $\text{HCO}_3^-/\text{CO}_2$ (Lauhon, 2012). Cet intermédiaire a été isolé et caractérisé par HPLC et RMN (Lauhon, 2012). D'un point de vue structural, le TC-AMP correspond à l'addition d'un groupement thréonyl-carbamoyl lié au phosphate α de l'AMP (Figure 28). Le TC-AMP se décompose spontanément en solution aqueuse pour produire de l'AMP et un mélange quasi-équimolaire de L-thréonine et 5-méthyl-2-oxazolidinone-4-carboxylate (L-Thr-ox), un dérivé cyclique de la L-thréonine. La durée de demi-vie de TC-AMP est de 3.5 min à pH 7.5, 37 °C, 2 mM MgCl_2 . Cette durée de demi-vie décroît rapidement lorsque le pH, la concentration en MgCl_2 ou la température augmente. La présence de Sua5 de *B. subtilis* ou d'YrdC d'*E. coli* stabilise faiblement mais significativement le TC-AMP, probablement en le séquestrant et/ou en le liant dans une conformation stable. La réaction de synthèse de TC-AMP est réversible puisque Sua5 de *B. subtilis* ou YrdC d'*E. coli* sont capables de catalyser la formation d'ATP à partir de TC-AMP et PPi (Lauhon, 2012).

Les *adenylate-forming enzymes* utilisent une chimie similaire à celle de la synthèse de t⁶A (Schmelz and Naismith, 2009). Ces enzymes activent un substrat carboxylate en le condensant avec un ATP pour un carboxylate-adénylate. Ce dérivé est très réactif et se décompose spontanément dans l'eau. Pour limiter cette hydrolyse, les *adenylate-forming enzymes* catalysent généralement elle-même une seconde étape lors de laquelle un nucléophile (une amine, un alcool ou un thiol) réagit avec l'intermédiaire pour former le produit désiré. C'est le cas par exemple des aaRS qui catalysent la formation de l'intermédiaire aa-AMP avant de former les produits aa-ARNt et AMP. C'est aussi le cas des enzymes TobZ et HypF qui utilise le carbamoyl-AMP comme intermédiaire activé. Dans le cas de la t⁶A, la première étape, impliquant la synthèse de l'intermédiaire TC-AMP, est catalysée par les enzymes Sua5/YrdC mais la seconde étape, impliquant l'attaque du TC-AMP par la fonction amine de l'A₃₇, n'est pas réalisée au sein de la même enzyme. Elle nécessite un transfert vers les complexes KEOPS/DEZ ou vers l'enzyme Qri7. Ceci rend la réaction de synthèse de t⁶A encore plus complexe à comprendre.

Le taux de conversion en t⁶A par le complexe DEZ de *B. subtilis* est plus lent lorsqu'on lui fournit le TC-AMP déjà formé que lorsqu'on ajoute Sua5 et les précurseurs initiaux de la réaction (c'est à dire L-thréonine, bicarbonate and ATP) (Lauhon, 2012). Cette différence de cinétique et l'instabilité évidente du TC-AMP sont des arguments qui suggèrent que le transfert de TC-AMP entre Sua5/YrdC et Kae1/YgjD/Qri7 ne se fasse pas par simple diffusion de l'intermédiaire. Bien qu'une interaction entre YrdC et YgjD-YeaZ d'*E. coli* ait été observée par la méthode de *pull-down* (Deutsch et al., 2012), une telle interaction n'a jamais été confirmée par une autre méthode chez *E. coli* et n'a jamais été observée pour d'autres systèmes. De plus il a été montré que lorsque Sua5 et Qri7 de *S. cerevisiae* sont séparés par une membrane ne permettant la diffusion que des molécules de petite taille, comme le TC-AMP, la synthèse de t⁶A est toujours réalisée (Wan et al., 2013).

Au bilan, le TC-AMP est une molécule réactive et instable. Le procédé par lequel elle est transférée de Sua5/YrdC vers Kae1/YgjD/Qri7 est incertain. Bien qu'une interaction directe ne soit apparemment pas essentielle, un transfert par simple diffusion paraît insuffisant pour expliquer la synthèse de t⁶A dans l'environnement cellulaire, encombré et dynamique. En particulier dans le cas des organismes hyperthermophiles, soumis à des températures favorisant la dégradation de TC-AMP.

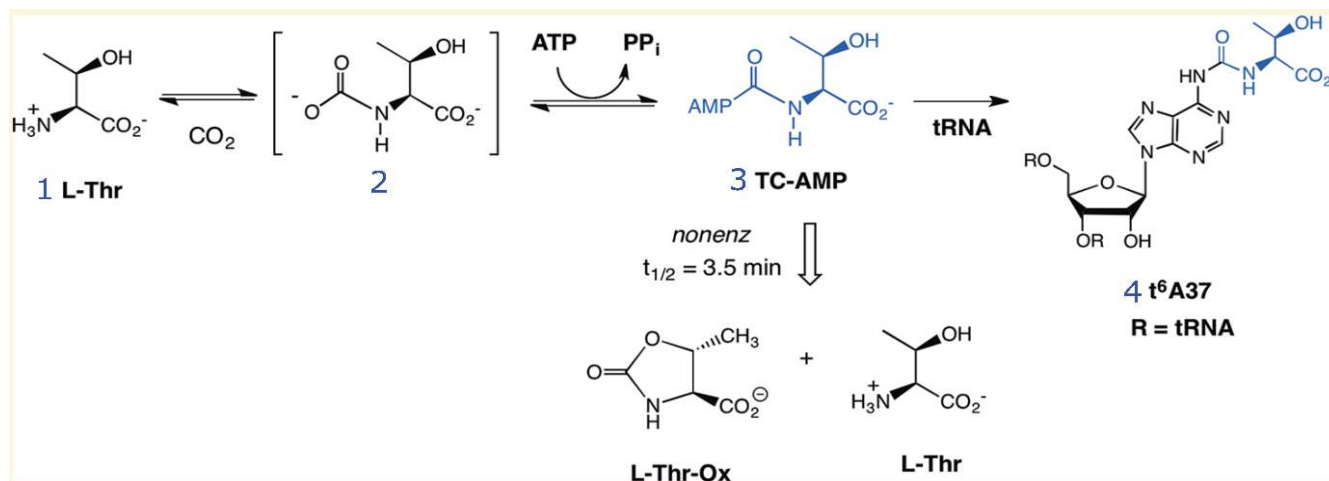


Figure 29. Mécanisme proposé pour la synthèse de t⁶A-ARNt

Dans ce modèle, les enzymes Sua5/YrdC utilisent la L-thréonine (1) et le CO₂ (ou le HCO₃⁻) pour synthétiser un intermédiaire de réaction, la N-carboxy-L-thréonine (2). Cette intermédiaire est ensuite capable d'attaquer le phosphate α de l'ATP afin de former une molécule de TC-AMP (3) et une molécule de PP_i. Le TC-AMP est transmis aux complexes DEZ/KEOPS ou à Qri7 qui, lors de la seconde étape de synthèse, transfère le groupement L-thréonyl-carbamoyl sur l'ARNt substrat pour former un ARNt-t⁶A (4). Les modalités de transfert du TC-AMP de Sua5/YrdC vers DEZ/KEOPS/Qri7 restent encore incompréhensibles, d'autant plus que le TC-AMP est une molécule hautement instable *in vitro*. D'après (Lauhon, 2012).

d. Le mécanisme enzymatique de formation du TC-AMP

Le mécanisme derrière la synthèse de TC-AMP par Sua5 et YrdC est encore mal compris. Le modèle proposé actuellement implique la formation d'un intermédiaire N-carboxy-L-thréonine, formé à partir de $\text{HCO}_3^-/\text{CO}_2$ et de L-thréonine (Figure 29) (Lauhon, 2012; Parthier et al., 2012). Les amines aliphatiques, y compris les acides aminés tels que la thréonine, sont connues pour former des carbamates stables à pH alcalin ($\sim 8.5-12$) lorsque les substrats, ici $\text{HCO}_3^-/\text{CO}_2$ et L-thréonine, sont en concentrations élevées (Yamamoto et al., 2012). L'enzyme aurait pour fonction de concentrer localement les deux substrats et d'augmenter le pH du site actif pour permettre la formation de la N-carboxy-L-thréonine. Dans ce cas, le substrat carbonaté de Sua5/YrdC est plus probablement CO_2 , et non HCO_3^- puisque CO_2 est plus électrophile et donc favorise d'avantages la formation de N-carboxy-L-thréonine. Une fois formé, le carbamate, de par sa nature chimique et sa position dans le site actif, est susceptible d'attaquer le phosphate α de l'ATP pour former le TC-AMP.

La formation du carbamate reste cependant une réaction énergétiquement défavorable. C'est pourquoi une version différente de ce mécanisme a été proposée, dans laquelle la formation de la N-carboxy-L-thréonine se fait entre la L-thréonine et un intermédiaire actif carboxy-AMP, formé à partir de $\text{HCO}_3^-/\text{CO}_2$ et d'un ATP (Perrochia et al., 2013b). Dans ce cas la formation de TC-AMP consomme deux molécules d'ATP. L'ensemble des données structurales et le mécanisme proposé semblent indiquer que tous les groupes fonctionnels impliqués dans la synthèse du TC-AMP sont fournis par les substrats eux-mêmes. Dans ce cas les protéines YrdC/Sua5 auraient un rôle passif dans cette réaction qui serait alors une catalyse entièrement assistée par les substrats (Parthier et al., 2012).

Référence	Substrat	Kd
Teplova et al., 2000	ADN simple ou double brin	ND
	ARN dodecamer simple brin	ND
	ARN dodecamer double brin (CGCGAAUUCGCG)	$0.18 \pm 0.04 \mu\text{M}$
	Mélange commercial ARNt <i>E. coli</i>	$0.68 \pm 0.15 \mu\text{M}$
El Yacoubi et al., 2009	ARNt ^{Thr(CGU)} complètement modifié	$0.6 \pm 0.13 \mu\text{M}$
	ARNt ^{Thr(CGU)} hypomodifié (no t ⁶ A)	$0.11 \pm 0.05 \mu\text{M}$
	ARNt ^{Thr(CGU)} non modifié	ND
Harris et al., 2011	ASL ^{Lys(UUU)} -t ⁶ A	$0.27 \pm 0.2 \mu\text{M}$
	ASL ^{Lys(UUU)} non modifié	$1.36 \pm 0.39 \mu\text{M}$

Table 4. Résumé des affinités d'YrdC d'*E. coli* pour les acides nucléiques

e. D'autres fonctions pour Sua5/YrdC

La taille et le potentiel électrostatique positif de la cavité du site actif d'YrdC suggèrent que cette protéine puisse interagir avec des acides nucléiques (Harris et al., 2011; Teplova et al., 2000). Une telle interaction est décelée pour YrdC d'*E. coli* en utilisant la technique d'extinction de la fluorescence des tryptophanes (en anglais *fluorescence quenching*) (Table 4). Elle est capable de fixer spécifiquement *in vitro* l'ARN double brin, en particuliers les ARNt hypomodifié, avec une constante d'affinité de l'ordre de 0,1-1.4 μ M mais ne fixe ni l'ADN double brin ni l'ARN simple brin (El Yacoubi et al., 2009; Harris et al., 2011; Teplova et al., 2000). En revanche aucune interaction entre *Pa*-Sua5 ou *Sc*-Sua5 et un ARN n'a pu être mise en évidence par la technique de EMSA (Perrochia et al., 2013b). Ces résultats suggèrent qu'YrdC puisse avoir un rôle différent et/ou une autre fonction que celle de Sua5 au sein des cellules.

La protéine RS1 de souris a été détectée comme un interactant d'IRIP (La protéine YrdC de souris) par double hybride puis validée par co-immunoprécipitation (Jiang et al., 2005). Il s'agit du seul interactant de la famille Sua5/YrdC décrit à ce jour. RS1 est une protéine qui régule l'activité des transporteurs Na⁺-D-glucose au niveau transcriptionnel et/ou post-transcriptionnel (Veyhl et al., 1993, 2003). La surexpression d'YrdC chez des lignées cellulaires HeLa réduit significativement l'activité des transporteurs type Na⁺-D-glucose (LI et al., 2013). Leur activité catalytique intrinsèque n'est cependant pas modifiée ce qui mène les auteurs à suggérer que l'effet d'YrdC n'est pas direct mais serait lié au système global de régulation des systèmes de transporteurs. Le lien entre RS1/YrdC et régulation des transporteurs n'est cependant encore ni avéré ni compris.

L'absence de Sua5 chez *S. cerevisiae* est corrélée à une réduction progressive de la taille des télomères (Meng et al., 2010). Des expériences de retard sur gel *in vitro* montrent que Sua5 se lie de manière spécifique à l'ADN télomérique. Les auteurs proposent donc que Sua5 ait un rôle dans l'homéostasie des télomères chez *S. cerevisiae*.

Disease category	Disease	Affected tRNA modification
Neurological	Galoway-Mowat syndrome (GAMOS)	t ⁶ A
	Intellectual disability	2'O-methylribose
		m ² ₂ G
		m ⁵ C
		m ⁷ G
		A-to-I editing
	Familial dysautonomia	mcm ⁵ s ² U
	Amyotrophic lateral sclerosis	mcm ⁵ s ² U
Cardiac	Rolandic epilepsy	mcm ⁵ s ² U
	Dubowitz-like syndrome	m ⁵ C
Respiratory	Noonan-like syndrome ^d	m ⁵ C
Cancer	Bronchial asthma	mcm ⁵ s ² U
	Skin, breast, and colorectal	m ⁵ C
		wybutosine
	Breast	m ⁵ U
	Colorectal	m ¹ G
	Urothelial	mcm ⁵ U
	Breast, bladder, colorectal, cervix, testicular	mcm ⁵ U
	Epigenetic cancer treatment	m ⁵ C
Metabolic	Type 2 diabetes	ms ² t ⁶ A
Mitochondrial-linked	MELAS	τm ⁵ U
	MERRF	τm ⁵ s ² U
	Infantile liver failure	s ² U
	Deafness associated with mutations in mitochondrial 12S ribosomal RNA	s ² U

Table 5. Les maladies humaines liées à des défauts de modifications des ARNt

Adapté de (Powell et al., 2015)

VII- Modifications des ARNt et maladies humaines.

Les modifications des ARNt affectent aussi bien leur stabilité générale que l'efficacité et la fidélité de la traduction. Selon la modification concernée, une hypo-modification des ARNt peut avoir soit des effets globaux sur la synthèse protéique soit des effets ne concernant qu'une partie des transcrits contenant des codons d'un certain type (Novoa and Ribas de Pouplana, 2012). Dans les deux cas tous les organismes sont susceptibles d'être affectés. L'humain ne fait pas exception à la règle. Ces dernières années, un lien fort entre des dysfonctionnement du système de modifications des ARNt et des pathologies humaines complexes a été mis en évidence (Table 5) (Powell et al., 2015).

La plupart de ces pathologies présentent un symptôme de déficience neurologique. Par exemple le gène *TRM1* code pour une méthyltransférase catalysant la diméthylation de la G en position 26 (m^2G_{26}) des ARNt (Liu and Strâby, 2000). Une mutation dans ce gène provoquant un décalage du cadre de lecture, et par conséquent inactivant ce gène, est associé à des troubles cognitifs autosomiques récessifs sévères (Najmabadi et al., 2011).

Par ailleurs plusieurs études suggèrent un lien entre différents types de cancers et un comportement aberrant dans le système de modifications d'ARNt (Frye and Watt, 2006; Rodriguez et al., 2007). Ainsi, l'outil de diagnostic Mammostrat utilise le gène *TRMT2A*, codant pour l'enzyme qui catalyse la formation de 5-méthyluridine (m^5U) en position 42 et 54 des ARNt, comme l'un des 5 biomarqueurs permettant de détecter le cancer du sein (Bartlett et al., 2010).

La t^6A et ses dérivées sont elles aussi impliquées dans des maladies humaines aux phénotypes variés et sévères. Le diabète de type 2, qui se caractérise par un défaut de synthèse de l'insuline au niveau du pancréas et une résistance à l'insuline, possède également un lien avec une modification des ARNt : la ms^2t^6A des $ARNt^{Lys(UUU)}$. L'enzyme catalysant cette réaction est codée par le gène *CDKAL1* chez l'humain. Les données suggèrent qu'un manque en ms^2t^6A provoque des erreurs lors de la traduction de diverses protéines dont la pro-insuline. La pro-insuline anormale ne pouvant pas être convertie en insuline, elle s'accumulerait et engendrerait une intolérance au glucose (Wei and Tomizawa, 2011; Wei et al., 2011).

En 2017, une mutation ponctuelle récessive dans Kae1 a été associée à un retard de développement général, un retard staturo-pondéral et une anomalie tubulaire rénale observées au sein d'une même famille (Edvardson et al., 2017). Plus récemment encore, un défaut de synthèse de t⁶A, associée à diverses mutations dans le complexe KEOPS humain, a été identifié comme étant une cause du syndrome de Galloway-Mowat (GAMOS) (Braun et al., 2017). Le GAMOS est une maladie autosomique récessive rare dont le déterminisme génétique reste encore à élucider. Il se caractérise par un syndrome néphrotique cortico-résistant, des anomalies du système nerveux central et une microcéphalie précoce. Les patients décèdent à un jeune âge.

Au bilan, l'importance des modifications des ARNt chez l'humain, y compris celle de la t⁶A, commence tout juste à être appréhendée. Les mécanismes moléculaires derrière ces connections restent cependant encore en grande partie inconnus et la plupart de ces pathologies n'ont actuellement aucun traitement. L'étude de ces voies de modifications possède donc un intérêt majeur puisqu'elle permettra, à long terme, de développer des approches thérapeutiques visant à corriger les hypo-modifications des ARNt ou à moduler l'expression des enzymes de modification des ARNt.

Problématique

Le nucléoside modifié t⁶A est universellement trouvé en position 37 de tous les ARNt décodant des codons ANN. Cette position stratégique et sa structure chimique lui permettent de stabiliser la boucle AC et de renforcer l'appariement codon-anticodon lors de la traduction. Son absence est létale pour de nombreux procaryotes et provoque des phénotypes sévères chez les eucaryotes. Bien que la t⁶A ait été identifiée en 1969, les enzymes responsables de sa synthèse sont restées mystérieuses près de 40 ans. L'avènement de la génomique comparative a permis d'identifier les familles universelles Sua5/YrdC et Kae1/YgjD/Qri7 comme étant des éléments conservés et essentiels pour la synthèse de la t⁶A. Dans les années qui suivirent ces découvertes, les travaux de plusieurs équipes à travers le monde, dont les deux laboratoires qui m'accueillent, ont permis de faire d'importants progrès quant à la compréhension de ces enzymes. Cependant, de nombreuses questions relatives au mécanisme enzymatique de synthèse de t⁶A restent encore inconnues. Notamment en ce qui concerne la première réaction complexe catalysée par les enzymes Sua5/YrdC.

La protéine YrdC fonctionne en tant que protéine à domaine unique et ne nécessite aucun partenaire pour être active. La protéine Sua5 possède quant à elle, en plus du domaine catalytique orthologue à YrdC, un domaine SUA5 C-terminal. Les deux versions sont capables de synthétiser du TC-AMP par un mécanisme qui reste à être validé. En particulier, on ignore actuellement s'il existe des différences fonctionnelles majeures entre les domaines catalytiques d'YrdC et Sua5 et, si oui, lesquels. Le domaine SUA5 ne possède aucun homologue de séquence dans les bases de données actuelles. En d'autres termes, le domaine SUA5 est retrouvé exclusivement au sein des protéines Sua5. La fonction de ce domaine, ainsi que celle du linker, restent donc mystérieuse, d'autant plus que la version YrdC est fonctionnelle sans ces derniers.

La vaste majorité des organismes codent exclusivement pour YrdC ou pour Sua5. Cependant les deux formes sont présentes dans les trois domaines du vivant. C'est-à-dire que certaines archées, bactéries et eucaryotes utilisent Sua5 quand d'autres utilisent YrdC. Cette répartition déroutante ne permet pas de comprendre facilement l'histoire évolutive de cette famille et de savoir quelle forme était présente chez LUCA.

Objectifs

Au vu des données présentées dans l'introduction, cette thèse a pour objectif de répondre aux réponses suivantes - selon leur ordre de priorité :

- Déterminer le rôle du linker et du domaine SUA5 chez les protéines Sua5 (Article I).
- Comprendre les différences fonctionnelles entre Sua5 et YrdC (Article I et II).
- Proposer un modèle évolutif expliquant l'existence des deux formes et leur répartition Sua5/YrdC au sein du vivant (Article II).
- Caractériser la t⁶A synthétase de *P. salinus* (Article III).

Choix de l'approche

Le point de départ de ma thèse était de déterminer le rôle du domaine SUA5 et du linker. Aucune étude de ce genre n'a été décrite pour le moment. Afin de répondre à cette question, je me suis basée sur les structures résolues de Sua5 de *P. abyssi* par Wenhua Zhang, étudiant en thèse de 2011 à 2015 dans l'équipe Fonction et Architecture des Assemblages Macromoléculaire (FAAM). Des résidus présents au niveau du linker et sur le domaine SUA5 potentiellement important pour Sua5 ont été identifiés puis ciblés par mutagénèse. J'ai utilisé une approche biochimique *in vitro* couplée à une approche *in vivo* de complémentation chez *S. cerevisiae* afin de caractériser ces mutants. La double approche *in vivo/in vitro* permet d'obtenir des données complémentaires susceptibles de nous aider à mieux répondre à la question biologique posée. Une approche *in silico*, utilisant les bases de données actuelles disponibles, a permis de compléter les données *in vitro/in vivo*.

Les protéines décrites et purifiées dans cette thèse ont été produites en système hétérologue par surexpression chez *E. coli*. Cette méthode est désormais bien connue et utilisée en routine avec succès dans un grand nombre de laboratoires. L'obtention de protéines solubles et actives par cette méthode reste cependant empirique et peu prévisible. De plus les protéines obtenues ne portent pas les modifications post-traductionnelles éventuelles qui peuvent exister *in vivo* chez l'organisme d'origine et qui sont susceptible d'agir sur leur activité.

Pour faciliter les étapes de purification, chaque protéine est fusionnée à une étiquette peptidique (hexahistidine ou streptavidine) placée en C-terminal pour les protéines entières, et en C-terminal ou en N-terminal pour les mutants tronqués selon le site de la troncature. En raison de sa faible taille (6 ou 7 acide aminés), l'étiquette n'est pas enlevée. La présence d'une telle étiquette peut dans certains cas perturber le repliement, la stabilité et/ou l'activité de la protéine. Il faut donc garder à l'esprit l'existence de ce paramètre dans l'analyse des données *in vitro*.

Choix des modèles d'études

Le modèle d'étude principal de ma thèse est la protéine Sua5 de *P. abyssi*, une euryarchée hyperthermophile ayant une température optimale de croissance de 95 °C en laboratoire. Il s'agit d'un des modèles de prédilection du laboratoire Biologie Cellulaire des Archées (BCA) qui m'accueille. Découverte en 1992 au niveau d'un fumeur noir à plus de 3500 m de profondeur dans l'océan Pacifique, son génome a été séquencé en 2001 (Lecompte et al., 2001) avant d'être ré-annoté en 2003 (Cohen et al., 2003) - un travail auquel a participé l'équipe BCA. Le génome de *P. abyssi* est composé d'un chromosome circulaire de 1,76 Mb et d'un plasmide de 3,5 kb, nommé pGT5. *P. abyssi* a servi de modèle à de nombreuses études sur la biologie des archées depuis. Elle a par exemple permis d'identifier l'origine de répllication de ces dernières par l'équipe BCA (Myllykallio et al., 2000). *Pyrococcus abyssi* porte un seul gène codant pour Sua5 (Gene ID: 1495884). Le complexe KEOPS, responsable de la seconde étape de synthèse de t⁶A après Sua5/YrdC, utilisé lors de ma thèse, provient lui aussi de *P. abyssi*. Le travail de thèse de Ludovic Perrochia, de 2010 à 2013, au sein de l'équipe BCA, a permis d'établir, entre autres, que la protéine sauvage de Sua5 de *P. abyssi*, portant un His-tag en C-terminal, ainsi que le complexe KEOPS de *P. abyssi*, portant un His-tag en N-terminal de sa sous-unité Pcc1, s'expriment tous les deux de façon soluble chez *E. coli*, se purifient correctement et sont actifs *in vitro*.

Les alignements de séquences des protéines Sua5/YrdC d'archées, réalisés dans le but d'identifier des caractéristiques propres à chaque forme, ont mis en évidence chez *A. profundus* l'existence un orthologue de Sua5 atypique (Gene ID: 8738752) qui n'a fait l'objet d'aucune étude à ce jour. La Sua5 d'*A. profundus* possède des caractéristiques intermédiaires entre Sua5 et YrdC, ce qui en fait un modèle d'étude intéressant pour comprendre les différences fonctionnelles de chaque forme et l'évolution de ces enzymes. *Archaeoglobus profundus* (*A. profundus*) est également une euryarchée hyperthermophile (température optimale de croissance en laboratoire de 85 °C). Cet organisme a été isolé à partir des cheminées hydrothermiques des grands fonds en océan Pacifique (profondeur: 2000 m) à Guaymas, Mexique en 1990 (Burggraf et al., 1990) puis son génome a été séquencé en 2010 (von Jan et al., 2010). Le génome d'*A. profundus* est composé d'un chromosome circulaire de 1,56 Mb et d'un plasmide de 2,8 kb, nommé pArcpr01.

Au niveau biochimique, l'étude de protéines provenant d'organismes hyperthermophiles présente de nombreux avantages qui n'ont pas échappé aux biochimistes : facilité de purification, meilleure stabilité, cristallogenèse améliorée. Ainsi le protocole de purification des protéines Sua5 hyperthermophiles implique typiquement une étape de chauffe à 65 °C pendant 30 min et les tests d'activité *in vitro* se font à 50 °C. La probabilité qu'un contaminant issu d'*E. coli* soit co-purifié et actif malgré ces conditions est très faible.

Pandoravirus salinus (*P. salinus*) est un virus géant dont le seul hôte connu actuellement est l'amibe *Acanthamoeba*. Son génome a été séquencé en 2013 (Philippe et al., 2013). L'annotation de ce génome a mis en évidence un gène codant pour une protéine putative correspondant à une fusion entre Sua5, YgjD et TcdA (l'enzyme responsable de la cyclisation en ct^6A). Cette protéine et sa fonction n'ont encore jamais été décrites. Il s'agit du seul cas répertorié d'un système de synthèse de t^6A sous forme d'un seul polypeptide chez un virus. L'étude de cette protéine a deux intérêts majeurs évidents. Tout d'abord un intérêt biochimique puisque que l'on est susceptible d'en apprendre d'avantages sur les mécanismes de synthèse et de transfert de TC-AMP entre Sua5 et YgjD ainsi que sur la fixation de l'ARNt à ce système, entre autres. Ensuite un intérêt évolutif puisque, si le système t^6A synthétase de *P. salinus* fonctionne sans protéine partenaire, il s'agirait d'un système minimal dont la fonction pour le virus est totalement inconnue.

ARTICLE I

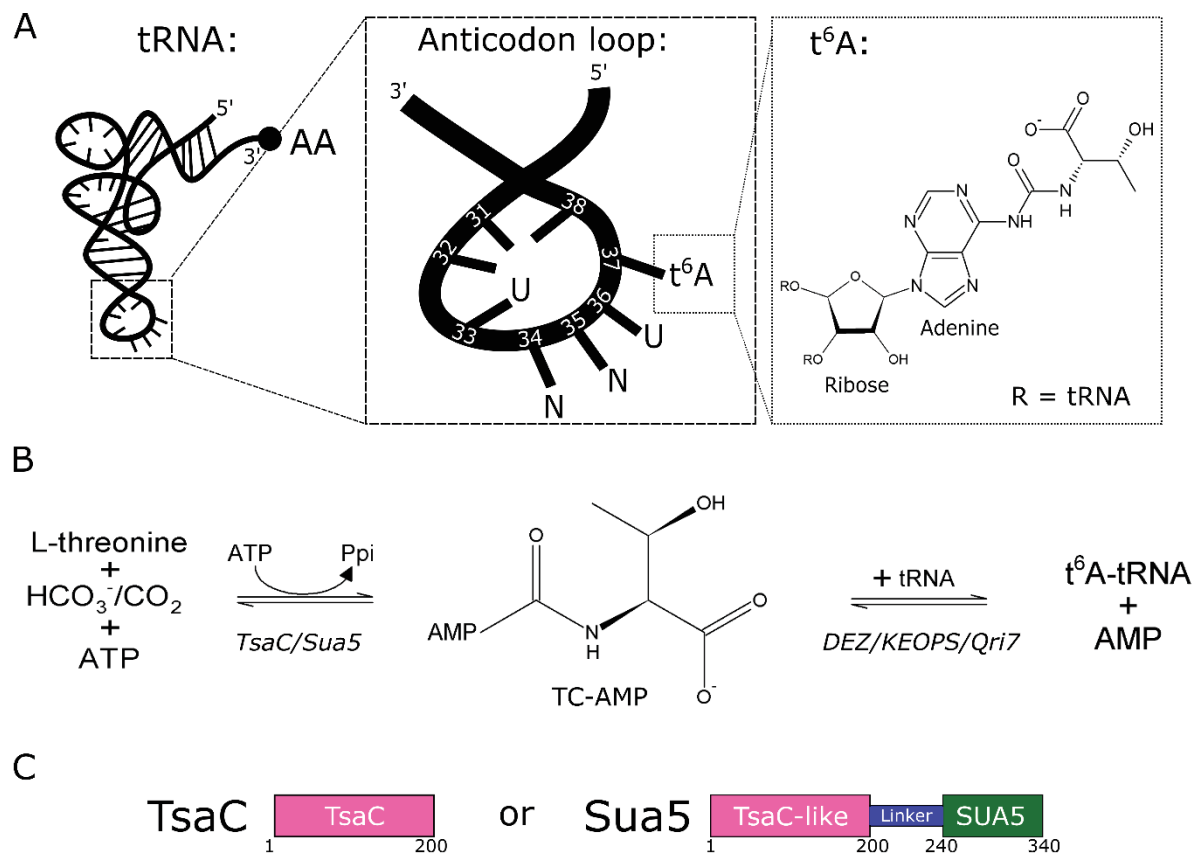


Figure 30. Structure and biosynthesis pathway of t^6A

(A) N^6 -threonyl-carbamoyl adenosine (t^6A) is depicted at position 37 of the anticodon loop (enlarged) of tRNA carrying an NNU anticodon (where N = A, U, G, or C). In absence of t^6A , the invariable U_{33} nucleoside can form Watson-Crick hydrogen bonds with A_{37} which results in a conformational change of the anticodon loop and less efficient codon-anticodon interaction. The black bars represent the orientation of the nucleoside bases against the phosphodiester backbone. The right panel shows the chemical structure of t^6A . **(B)** The t^6A biosynthesis pathway. In the first step, condensation of L-threonine, HCO_3^-/CO_2 and ATP to form L-threonyl-carbamoyl adenylate (TC-AMP) and inorganic pyrophosphate (PPi) is catalyzed by two distinct enzyme variants: Sua5 and YrdC. In the second step of the synthesis TC-AMP is bound by the KEOPS complex (in archaea and in the cytoplasm of eukaryotes), the Qri7 protein (in mitochondria) or the DEZ complex (in bacteria) that catalyze the transfer of the L-threonyl-carbamoyl moiety to A_{37} of substrate tRNA to form t^6A -tRNA and one molecule of AMP. **(C)** Schematic representation of YrdC and Sua5. YrdC is a single domain protein of about 200 amino acids (in pink) while Sua5 consists of a YrdC-like domain connected via a linker of about 40 residues (in blue) to a globular C-terminal domain of about 100 amino acids (in green).

Structure function analysis of Sua5 protein reveals novel functional motifs required for the synthesis of the universal t⁶A tRNA modification

Adeline Pichard-Kostuch*, Wenhua Zhang*, Dominique Liger, Marie-Claire Daugeron, Juliette Letoquart, Patrick Forterre, Bruno Collinet, Herman van Tilbeurgh#, Tamara Basta#

*first co-author, # corresponding authors

Introduction

To be functional in the cell, tRNA molecules need to be post-transcriptionally modified by addition of a variety of chemical groups (Grosjean et al., 1995). One of the most complex modification is the N⁶-threonyl-carbamoyl adenosine (t⁶A) found exclusively at position 37 (3' adjacent to anticodon) of tRNA that decode codons of the type ANN (where N = A, U, G or C) (Chheda et al., 1969; Powers and Peterkofsky, 1972a; Takemura et al., 1969) (Figure 30). Except in some insect symbionts with highly reduced genomes, the t⁶A modification is ubiquitously distributed among cellular organisms as well as in organelles such as mitochondria and chloroplasts (Machnicka et al., 2013). t⁶A prevents frameshifting during translation by stabilizing the pairing between the A₁ from codon and the U₃₆ from anticodon (Sonawane and Sambhare, 2015; Weissenbach and Grosjean, 1981). In addition, its presence was reported to be critical for the binding of the anticodon stem loop to mRNA-ribosome complex *in vitro* (Agris, 2008; Lescrinier et al., 2006; Murphy et al., 2004; Stuart et al., 2000; Yarian et al., 2002). Abolition of t⁶A biosynthesis is lethal in *Escherichia coli* (*E. coli*) (El Yacoubi et al., 2009; Thiaville et al., 2015b). In *Saccharomyces cerevisiae* (*S. cerevisiae*), the absence of t⁶A increases the occurrence of frame shifting, leaky scanning of start codons and read-through of stop codons and leads to a dramatically reduced fitness (Daugeron et al., 2011; El Yacoubi et al., 2011; Lin et al., 2010; Thiaville et al., 2016). Recently, mutations in the t⁶A synthetase genes was linked to a severe neurological and renal genetic disease in humans (Braun et al., 2017; Edvardson et al., 2017).

The t⁶A biosynthesis reaction proceeds in two consecutive steps ([Figure 30B](#)) catalyzed by two universal enzyme families: Sua5/YrdC* (* recently renamed TsaC) and Kae1/YgjD‡/Qri7 (‡ recently renamed TsaD). Sua5/YrdC catalyzes the condensation of L-threonine, bicarbonate (HCO₃⁻) or carbon dioxide (CO₂), and ATP to form inorganic pyrophosphate (PPi) and L-threonyl-carbamoyl adenylate (TC-AMP). The exact nature of the carbonaceous substrate used, HCO₃⁻ or CO₂, is not yet unambiguously identified. The TC-AMP intermediate is then taken up by Kae1/YgjD/Qri7 that transfers the L-threonyl-carbamoyl moiety onto the substrate tRNA to form t⁶A-tRNA (Deutsch et al., 2012; El Yacoubi et al., 2009, 2011; Lauhon, 2012; Perrochia et al., 2013b; Srinivasan et al., 2011; Wan et al., 2013).

Except in mitochondria, where Qri7 works alone as a homodimer (Thiaville et al., 2014; Wan et al., 2013), Kae1/YgjD require the presence of essential accessory proteins to be active. In bacteria, YgjD forms with two bacteria-specific proteins, TsaB (previously YeaZ) and TsaE (previously YjeE), a complex called DEZ (Deutsch et al., 2012; Lauhon, 2012; Nichols et al., 2013; Zhang et al., 2015a). In archaea and in the cytoplasm of eukaryotes, Kae1 associates with Bud32, Pcc1 and Cgi121 into a complex called KEOPS or EKC (recently renamed TCTC, for nomenclature of t⁶A synthetic genes see (Thiaville et al., 2014)) thereafter referred to as KEOPS for simplicity (Downey et al., 2006; Kisseleva-Romanova et al., 2006; Mao et al., 2008; Perrochia et al., 2013a; Thiaville et al., 2014; Zhang et al., 2015a). A fifth member of KEOPS, Gon7, is found exclusively in fungi and its counterpart, C14ORF142, was recently reported in humans and several other metazoans (Wan et al., 2016). The exact function of accessory proteins is unknown.

Sua5/YrdC orthologs require no partners to catalyze the formation of TC-AMP. YrdC and Sua5 both use an orthologous catalytic domain of about 200 amino acids. Yet YrdC is a single domain protein while Sua5 has a supplementary C-terminal domain of about 100 residues, named SUA5 domain ([Figure 30C](#)). The latter is connected to the YrdC-like domain via a linker containing about 40 residues. Linker and SUA5 domain are specific of Sua5 proteins.

The YrdC protein from *E. coli* (*Ec*-YrdC) folds in a compact α/β twisted open-sheet structure composed of 7 α helices and 7 adjacent β strands in parallel and antiparallel orientations (Fu et al., 2010; Harris et al., 2015; Teplova et al., 2000). The overall structure is reminiscent of a baseball glove with a large central concave cavity. This cavity has a positive global charge and was suggested to act as nucleic acid binding surface (Teplova et al., 2000). Indeed *Ec*-YrdC protein binds tRNA or the anticodon arm structure with high affinity *in vitro* (El Yacoubi et al., 2009; Harris et al., 2011, 2013).

The active site of Sua5/YrdC is found in this cavity and involves several highly conserved residues among which the tetrad K⁵⁰xR⁵²/S¹³⁹xN¹⁴¹ (*Ec*-YrdC numbering) that plays a role in the binding of ATP and is essential for activity both in YrdC and Sua5 proteins (El Yacoubi et al., 2009; Harris et al., 2015; Kuratani et al., 2011; Teplova et al., 2000; Wan et al., 2013). The C-terminal extremity of *Ec*-YrdC protein was shown to have an abnormally high flexibility and was proposed to be important for the enzymatic activity by acting like a gating loop (Harris et al., 2015).

The structure of Sua5 from the archaeon *Sulfolobus tokodaii* (*St*-Sua5) was resolved in complexes with L-threonine and AMPPNP (Kuratani et al., 2011) and with TC-AMP (Agari et al., 2008; Parthier et al., 2012). The overall fold of the YrdC-like domain of *St*-Sua5 is very similar to the one of *Ec*-YrdC while the SUA5 domain contains a Rossmann fold (Agari et al., 2008; Kuratani et al., 2011). About twenty residues in the linker connecting the two domains were disordered and could not be resolved in the published structures of *St*-Sua5. Despite marked structural differences, Sua5 and YrdC enzymes catalyze the formation of TC-AMP via a complex mechanism that is not yet understood.

TC-AMP is an unstable molecule with a half-life of 3.5 min under standard physiological conditions i.e pH 7.5, 37 °C, 2 mM MgCl₂ (Lauhon, 2012). Decomposition of TC-AMP in aqueous media produce AMP and a nearly equimolar mixture of L-threonine and its cyclic derivate 5-methyl-2-oxazolidinone-4-carboxylate as final products (Lauhon, 2012). Despite this instability, direct contact between Sua5 and Qri7 from *S. cerevisiae* is not required for *in vitro* t⁶A synthesis suggesting that TC-AMP is a diffusible reaction intermediate (Wan et al., 2013). Still, the exact mechanism of transfer of TC-AMP during t⁶A synthesis remains to be fully elucidated. Domains homolog to YrdC are also found in the [NiFe]-hydrogenase maturation factors HypF and the O-carbamoyl transferases CmcH/NodU/TobZ family (Parthier et al., 2012; Petkun et al., 2011) where they catalyze carbamoylation reactions similar to those of Sua5/YrdC.

Both Sua5 and YrdC are found in the three domains of life, Archaea, Eukarya and Bacteria, but most species encodes only one of the two homologs (Thiaville et al., 2015a). For instance *Homo sapiens* carries a YrdC protein while *S. cerevisiae* and *Plasmodium falciparum* carry a Sua5 protein. This puzzling phylogenetic repartition and the successful *in vitro* and *in vivo* complementation assays between *Ec*-YrdC and Sua5 from *S. cerevisiae* (*Sc*-Sua5) (El Yacoubi et al., 2009) as well as the lack of data concerning the SUA5 domain provide little information about the function of this domain. To better understand the role of the linker and the SUA5 domain and gain insight into the functional differences between YrdC and Sua5 enzymes versions, we performed structure – function study of Sua5 (*Pa*-Sua5) from the hyperthermophilic archaeon *Pyrococcus abyssi* (*P. abyssi*).

We present here the crystal structure of *Pa*-Sua5 including for the first time the linker region structured between the YrdC-like and SUA5 domains. The linker is positioned in front of the catalytic cavity. Using structure-guided mutational analysis we established that conserved residues in the linker and in the SUA5 domain are essential for the TC-AMP formation. The data further suggest that the loop may control the access of the substrates/products to the active site and that tight interaction between the YrdC-like domain and SUA5 domain is required for this function. The identification of novel functional motifs specific to Sua5 proteins is an asset for further studies of the complex catalytic mechanism taking place during the biosynthesis of the universal t⁶A-tRNA modification.

Materials and methods

Cloning procedures and mutagenesis

The recombinant plasmid pET26b(+)-PAB1302-(wt)_{6H} (Perrochia et al., 2013b), was used as template for site-directed mutagenesis. The mutants were constructed using Phusion polymerase (ThermoFisher). The oligonucleotides used were synthesized by Genewiz and are listed in Annex 3. For generating point mutations, the QuickChange Kit II (ThermoFisher) protocol was used. For construction of truncated proteins, the sequences encoding the isolated protein domains were amplified using PCR. A hexahistidine tag or a Strep-tag II was added in 5' or 3' of the gene which was subsequently cloned in pET26b(+) vector (Novagen) using *NdeI/XhoI* restriction sites. The sequences of the constructs are available on request. Recombinant plasmids were introduced into chemically competent *E. coli* strain TOP10 (Invitrogen). The recombinant plasmid used for heterologous expression of KEOPS complex of *P. abyssi* was previously described (Perrochia et al., 2013b).

Recombinant gene expression and protein purification

Recombinant genes were expressed in *E. coli* strains Rosetta2 (DE3) pLysS (Novagen) or BL21-Gol (Agilent). Protein overproduction was done in Auto Induction Media Terrific Broth Base including trace elements (Formedium) prepared according to manufacturer's protocol. Cells were collected by centrifugation, resuspended in lysis buffer 50 mM Tris-HCl pH 8, 200 mM NaCl, 10 % glycerol. Lysozyme (0.1 mg/mL final) and anti-protease Roche cOMplete ULTRA tablets EDTA-free were added. Cell suspension was homogenized using the high pressure cell disrupter One Shot (Constant Systems Ltd). After centrifugation at 30 000 g for 30 min, the supernatant was heated at 65 °C for 30 min, except for the truncated mutants, to precipitate bacterial proteins. Protein precipitate was removed by centrifugation at 30 000 g for 60 min. His-tagged proteins from the soluble fraction were purified on a HisTrap HP using an AKTA FPLC system (GE Healthcare). Strep-tagged proteins from soluble fraction were purified on Strep-Tactin resin (Qiagen). Fractions of interest were pooled and injected on the Superdex 75 HiLoad 16/60 (or 10/300) (GE Healthcare). Fractions containing pure proteins were concentrated, flash-frozen in liquid nitrogen and stored at -80 °C in lysis buffer containing 10-20 % glycerol.

Optimization attempts to increase the solubility of *Pa*-Sua5 truncated mutants included: (i) using classical LB media and induction with IPTG for 1-2 hour at OD₆₀₀ 0.6, (ii) using heat-shock at 42 °C before induction to induce chaperon expression, (iii) growing culture at 16 and 22 °C, (iv) using several purification buffers between pH 7 and 8,5 with 200-500 mM NaCl and 0-20 % glycerol. In hope to enhance the stability of the purified proteins, storage conditions at 4°C, -20 °C and -80°C in buffer with glycerol between 10 and 30 % were tested.

tRNA substrate production and purification

The template sequences for *in vitro* transcription were synthesized and cloned between *HindIII* and *EcoRI* restriction sites in pUC18 vector (GeneScript). The gene coding the *Pa*-tRNA^{Lys(UUU)} is surrounded by two sequences coding hammerhead ribozyme in 5' and glmS ribozyme followed by a *BbsI* site in 3' (Batey and Kieft, 2007; Price et al., 1995). Plasmids were linearized by the restriction endonuclease *BbsI* overnight at 37°C. T7 RNA polymerase was purified as described by (Grodberg and Dunn, 1988). Transcription assay contained 80 mM Na-HEPES pH 7.5, 2 mM spermidine, 40 mM DTT, 20 mM MgCl₂, 2.5 mM NTP, 200 µg of linearized plasmid, 31 µg/mL T7 RNA polymerase in a 10 mL reaction. The reaction was done for 1 hour at 37 °C. The hammerhead ribozyme cleavage occurs simultaneously during the transcription. The precipitated magnesium PPi is pelleted by centrifugation at 10,000 g for 10 min. The glmS ribozyme cleavage is performed by incubating the supernatant with 1 mM glucosamine-6-phosphate for 5 min at room temperature. The reaction was heated at 80 °C for 10 min, to inactivate the enzyme, then centrifuged at 10 000 g for 10 min. The supernatant was injected on a MonoQ 5 mL column using an AKTA FPLC system (GE Healthcare). The elution was performed by a gradient of NaCl from 0.3 mM to 1 M in 50 mM Na-HEPES pH 7.5, 5mM MgCl₂. After analysis on a 12% (w/v) polyacrylamide gel containing 8 M urea, the fractions containing the pure tRNA were pooled, ethanol precipitated and solubilized in 50 mM Na-HEPES pH 7.5, 5mM MgCl₂. The samples were stored at -20°C.

Dataset		<i>Pa</i> -Sua5 ^{wt}		<i>Pa</i> -Sua5 ^{H234A/Y235A}
		threonine-PPi complex	threonine complex	mutant
Wavelength (Å)		0.98011	0.97857	1.0023
Composition		<i>Pa</i> -Sua5 ^{wt} -threonine-PPi	<i>Pa</i> -Sua5 ^{wt} -threonine	<i>Pa</i> -Sua5 ^{H234A/Y235A} -threonine- HCO ₃ ⁻
Ligands		L-threonine Inorganic pyrophosphate	L-threonine	L-threonine HCO ₃ ⁻
Resolution range (Å)		49.03 - 2.62 (2.71 - 2.62)	48.78 - 2.86 (2.96 - 2.86)	47.92 – 2.81 (2.88 – 2.81)
Space group		C2	C2	C2
Cell dimensions	a, b, c (Å)	150.4, 144.1, 103.7	152.2, 142.4, 103.9	177.4, 125.3, 97.9
	α, β, γ (°)	90, 103.1, 90	90, 101.8, 90	90, 91.8, 90
Unique reflections		64448 (6230)	49930 (4795)	24978(3990)
Completeness (%)		99.44 (96.35)	99.29 (95.23)	97.8 (97.3)
Mean I/sigma(I)		14.29 (2.36)	10.82 (2.21)	15.59 (3.26)
R-factor/ R-free		0.18/0.22	0.19/ 0.26	0.16/0.21
Protein residues		1367	1365	807
RMS (bonds) (Å)		0.017	0.024	0.014
RMS (angles) (°)		1.8	1.8	1.8
Ramachandran plot (%)	avored	97	98	97
	allowed	2.7	1.5	2.7
	outliers	0.3	0.5	0.3
Average B-factor		52.4	68.5	58.74

Table 6. The X-ray data collection and structural model refinement statistics

Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) experiments were performed using a VP-DSC calorimeter (MicroCal). *Pa*-Sua5 proteins were used at a concentration around 13 μ M in 50 mM K-HEPES pH 8, and 35 mM KCl. Scanning from 20 to 100 °C at a heating speed of 1 °C/min was done for each experiment. Data were analyzed using the MicroCal Origin software provided by the manufacturer.

Size-Exclusion-Chromatography coupled with Multi-Angle Laser Light Scattering

The molecular weight of the wild-type *Pa*-Sua5 was determined using Size Exclusion Chromatography coupled to Multi-Angle Laser Light Scattering (SEC-MALLS) with the SEC-TDA equipment (Malvern, France). The Superdex™ 200 (75) HR 10/300 columns (GE Healthcare) or Bio-SEC3 column (Agilent) for SEC was equilibrated with buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 5 mM 2- β -mercaptoethanol. 80 μ l of *Pa*-Sua5 at a concentration of 5 mg/ml were injected at a flow rate of 0.5 ml min⁻¹ onto the SEC column and eluted with the equilibration buffer at a flow rate of 0.3 ml min⁻¹. Elution was followed by a UV-visible spectrophotometer, a differential refractometer, a 7° and a 90° angle light scattering detector and a differential pressure viscometer. The OmniSEC software program was used for the acquisition and analysis of the data. BSA was used as standard reference protein of known molecular weight, concentration, refractive index increment ($dn/dc = 0.185 \text{ ml}\cdot\text{g}^{-1}$), and intrinsic viscosity to calibrate the instrument.

Crystallization, data acquisition and structure determination

Pa-Sua5 at a final concentration of 9 mg/ml was either crystallized alone or co-crystallized with ATP/MgCl₂/L-threonine/HCO₃⁻ (all ligands were added at 5 mM) by the sitting-drop vapor diffusion method. Crystals were obtained within 3 days from a crystallization reservoir containing 0.1 M Tris pH 8.5-8.8, 3.5 M ammonium sulfate. Crystals were cryo-protected in crystallization reservoir solution supplemented with 30 % glycerol and flash-frozen in liquid nitrogen. The X-ray diffraction data were collected at 100K at the Soleil Synchrotron Light Source Proxima 1 (wavelength 0.98011 Å, PILATUS 6M detector). The images were successfully processed and scaled with the XDS package (Kabsch, 2010). The data processing and refinement statistics are summarized in [Table 6](#).

The structure was solved by molecular replacement with the program MOLREP implemented in the CCP4 suite (Winn et al., 2011) using the structure of *St*-Sua5 as the search model (PDB code: 2EQA). The model building was carried out with the program Coot (Emsley and Cowtan, 2004) and the structure refinement was performed with Refmac5 (Winn et al., 2003).

Sequence Alignment and structure comparison

The sequence conservation was analyzed using the program ConSurf (Ashkenazy et al., 2016) with the structure of *Pa*-Sua5 as the query structure. Based on the default criteria, the program ConSurf employed CS-BLAST to identify the 500 closest homologs by BLOSUM-60 on the Uniref-90 Protein database and aligned them using MAFFT (Yamada et al., 2016). The conservation rate of each position was calculated with Bayesian algorithms using the 150 homologous sequences with the lowest E-value and mapped on the structure of *Pa*-Sua5. The PDB codes of the structures are 1HRU and 2MX1 for *Ec*-YrdC, 2EQA, 3AJE and 4E1B for *St*-Sua5, 3VEZ for *St*-TobZ. The structural comparison and graphic representations of the protein structures were produced with UCSF Chimera (Pettersen et al., 2004). The electrostatic potential representation were rendered and produced using APBS and PDB2PQR (Unni et al., 2011). Ligands coordination was analyzed with LigPlot+ (Laskowski and Swindells, 2011).

***In vitro* assay for the synthesis of t⁶A-modified tRNA**

The reaction was performed using *Pa*-KEOPS and *Pa*-Sua5 proteins (2 µM each), purified *Pa*-tRNA^{Lys(UUU)} (5 µM) in reaction buffer composed of 50 mM K-HEPES pH 8, 35 mM KCl, 10 mM MgCl₂, 1 mM MnCl₂, 5 mM DTT, 1 mM ATP, 10 mM NaHCO₃⁻, supplemented with 182 µM L-[1-¹⁴C]-threonine (0,05 µCi, 55 Ci/mol) (American Radiolabeled Chemicals). After incubation at 50 °C for 1 hour, macro-molecules were precipitated by addition of 1 mL of Trichloroacetic acid 15% (TCA), and incubated on ice for 1 hour. Precipitated material was applied to glass microfiber filters GF/F (Whatmann), pre-wet with TCA 15%, using a vacuum apparatus (Millipore).

Assay tubes were rinsed with 1 mL of TCA 15%, and filters were washed three times with 1 mL of TCA 5% and three times with 1mL of Ethanol 95%. After extensive drying, filters were placed into scintillation vials and soaked with 3 mL of Emulsifier Safe scintillation cocktail (Zinsser analytic). Radioactivity was recorded as average counts per minute (CPM) during 5 min, with Packard Liquid Scintillation Analyzer.

***In vitro* ATP hydrolysis assay**

The ATP hydrolysis reactions (conversion into AMP and PPi) were carried out using *Pa*-Sua5 (1 or 2 μ M) and, when indicated, with *Pa*-KEOPS (2 μ M) and *Pa*-tRNA^{Lys(UUU)} (5 μ M) in the same buffer as for the t⁶A assay, supplemented with [α^{32} -P] ATP (3 Ci/ μ mol) (Perkins Elmer). The reaction was performed at 50 °C and stopped on ice. Radioactive nucleotides were separated by thin layer chromatography using 0.4 M KH₂PO₄- pH 3.4 as mobile phase. 1 μ L of the mix was spotted on a 10 x 10 cm PEI-Cellulose plates (Merck), pre-run in distilled water. Plates were dried and radioactivity was revealed by phosphorimaging using the scanner Typhoon Trio (GE Healthcare). Quantification was done with ImageQuantTL software.

For the determination of enzymatic constants, L-threonine and cold ATP were added at different concentrations to measure the initial velocity of ATP hydrolysis. The reaction was performed with an incubation time up to 6 min for L-threonine titration and up to 8 min for ATP titration. We could not detect any significant background of ATPase activity in absence of enzyme. The data were then fitted using non-linear least square regression applied to the Michaelis-Menten equation using Graph Pad Prism. All the experiments were performed at least in quadruplicates.

Functional complementation assay in *S. cerevisiae* cells

Yeast cells were grown at 28 °C in standard rich medium YEPD (1% yeast extract, 2% peptone, 2% glucose) or selective minimal media minus leucine with 2% galactose (GAL-LEU) or glucose (GLU-LEU) (Sigma or Euromedex). Cells were transformed using lithium acetate method as previously described (Gietz and Schiestl, 2007).

For the construction of the YcpLac33-*SUA5* plasmid, the wild-type *SUA5* gene was amplified by PCR using genomic DNA from strain BY496 as template (see Table [Annex 3](#) for oligonucleotide sequences). The insert was cloned into the YcpLac33 vector using *Sall/SacI* restriction sites. The resulting construct was introduced in the $\Delta sua5::KanMX$ strain, derived from the W303 reference strain. The genes encoding wild-type and point mutant *Pa-Sua5* proteins were amplified by PCR using the recombinant pET26b(+) plasmids (Cloning procedures section) as templates. The PCR products were cloned into pESC-LEU vector between *BglII/SacI* or *BamHI/Sall* and the genes were expressed under the control of respectively pGAL10, producing FLAG- tagged proteins, or pGAL1 promoters, producing myc-tagged proteins. The *S. cerevisiae* strain $\Delta sua5::KanMX$ – (YcpLac33-*SUA5*) was transformed with the recombinant pESC-LEU vectors and the transformants were selected on GLU-LEU. For each construction, three transformed clones were selected and grown on GLU-LEU petri dishes. Cells were serially diluted and spotted in parallel on GAL-LEU to induce the plasmid driven expression of *Pa-Sua5* variants and on GAL-LEU supplemented with 1 g.L⁻¹ 5-fluoroorotic acid (5-FOA) to counter select YcpLac33-*SUA5* plasmid. The plates were incubated at 28 °C for the time indicated in the main text.

Colony PCR

To validate the genetic constructions, several colonies per clone were selected from the GAL-LEU 5FOA plates and suspended in 20 to 60 µL NaOH 20 mM. After incubation at 95 °C for 45 min and 300 rpm, the extract was centrifuged at 11 000 g for 10 min to get rid of cell debris. Supernatant was diluted and used with the adequate primers for PCR reaction using Phusion (ThermoFisher). PCR products were sequenced for validation (Eurofins).

Western blot

Protein extraction was done using the standard protocol as previously described (Liger et al., 2011) for FLAG-tagged proteins or the Kushnirov extraction method (Kushnirov, 2000) for myc-tagged proteins. Proteins from soluble fraction were resolved by SDS-PAGE on a 15% polyacrylamide gel and transferred onto Protran nitrocellulose membrane (Whatman). Detection of myc-tagged proteins was performed using mouse 9E10 anti-myc as primary antibody (1/3000 dilution, Sigma-Aldrich) and goat anti-mouse HRP-conjugated IgG as secondary antibody (1/5000 dilution) (Sigma-Aldrich). Detection of FLAG-tagged proteins was performed using mouse anti-FLAG M2-peroxidase (HRP) (1/10 000 dilution) (Sigma-Aldrich). Immunoblots were revealed using ECL kit SuperSignal™ West Pico Chemiluminescent Substrate (ThermoFisher).

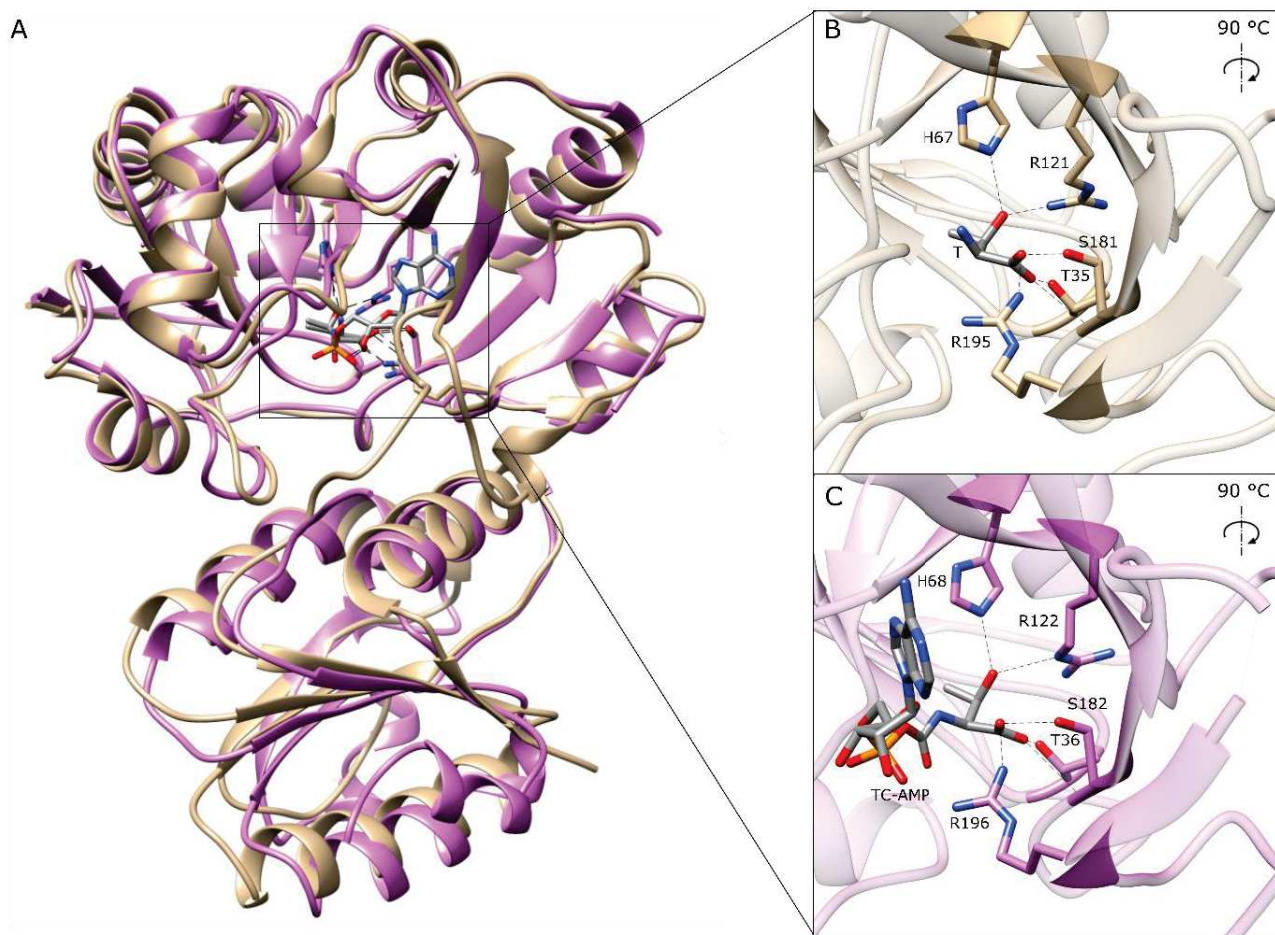


Figure 31. Comparison of the *Pa*- and *St*-Sua5 structures.

(A) Superposition of the structure of *Pa*-Sua5 bound to L-threonine (in tan) and the structure of *St*-Sua5 bound to co-purified TC-AMP (PDB: 4E1B) (in purple). **(B)** Zoom in the active site of *Pa*-Sua5 containing L-threonine. The L-threonine occupies the same position in all published Sua5 structures and is bound via H-bonds (dashed black lines) by four residues conserved among Sua5/YrdC proteins. **(C)** Zoom in the active site of *St*-Sua5 bound to TC-AMP. The threonyl-moiety from TC-AMP is coordinated by four conserved residues as for the L-threonine. Except for the ligands and the four conserved residues, the protein main chain is shown as transparent.

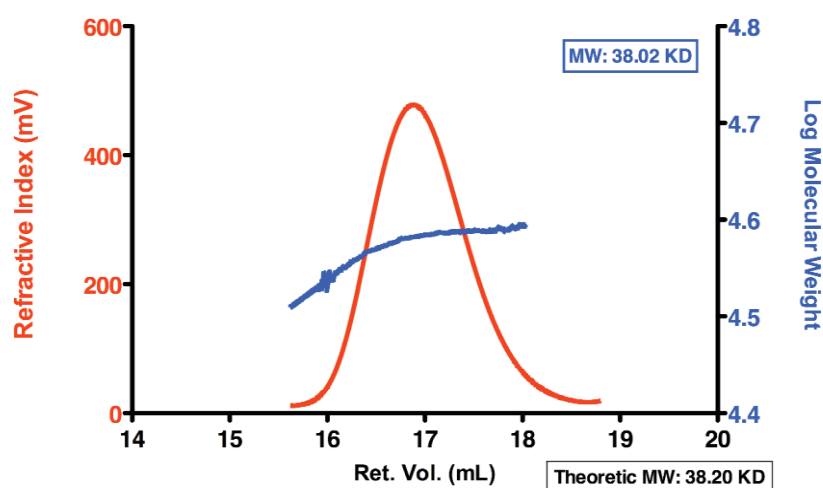
Results

Crystallographic structure of Sua5 from P. abyssi reveals a novel linker conformation

We first determined the apo-structure of *Pa*-Sua5 at 2.8 Å resolution. The asymmetric unit contains four copies with an average root-mean square-deviation (RMSD) of 0.6 Å over 341 Cα atoms. *Pa*-Sua5 and *St*-Sua5 share 80 % sequence similarity and their overall structure is very similar with an RMSD of 0.67 Å over 318 Cα atoms (Figure 31A). The molecular mass of *Pa*-Sua5 determined using SEC-MALLS was 38.02 kDa, which is consistent with *Pa*-Sua5 being a monomer in solution (Figure 32) as reported for *St*-Sua5 (Agari et al., 2008; Kuratani et al., 2011).

A residual electron density cloud at the bottom of the catalytic pocket could be fitted by a L-threonine molecule, which co-purified with the enzyme. L-threonine sits in a pocket that is covered by similar bound nucleotides as the threonyl-moiety of TC-AMP in the *St*-Sua5 structure (Figure 31B and 31C). It forms hydrogen bonds (H-bonds) with five highly conserved residues of the catalytic domain: Thr³⁵, His⁶⁷, Arg¹²¹, Ser¹⁸¹ and Arg¹⁹⁵ (Figure 31B). Atop the active site, two H₂O molecules (Figure 34B) are bound tightly via H-bonds with the conserved residues Arg⁵⁸, Asn⁶² and His²³⁴ for one and Arg⁵⁸ and Ser¹⁴⁷ for the other.

Figure 32. SEC-MALLS analysis of Pa-Sua5



The monomeric state of *Pa*-Sua5 in solution was determined by SEC-MALLS. The molecular weight of *Pa*-Sua5 in solution was determined to be 38.02 kDa, which is in proximity to a calculated molecular weight of 38.20 kDa from the sequence. See Materials and Methods for further details.

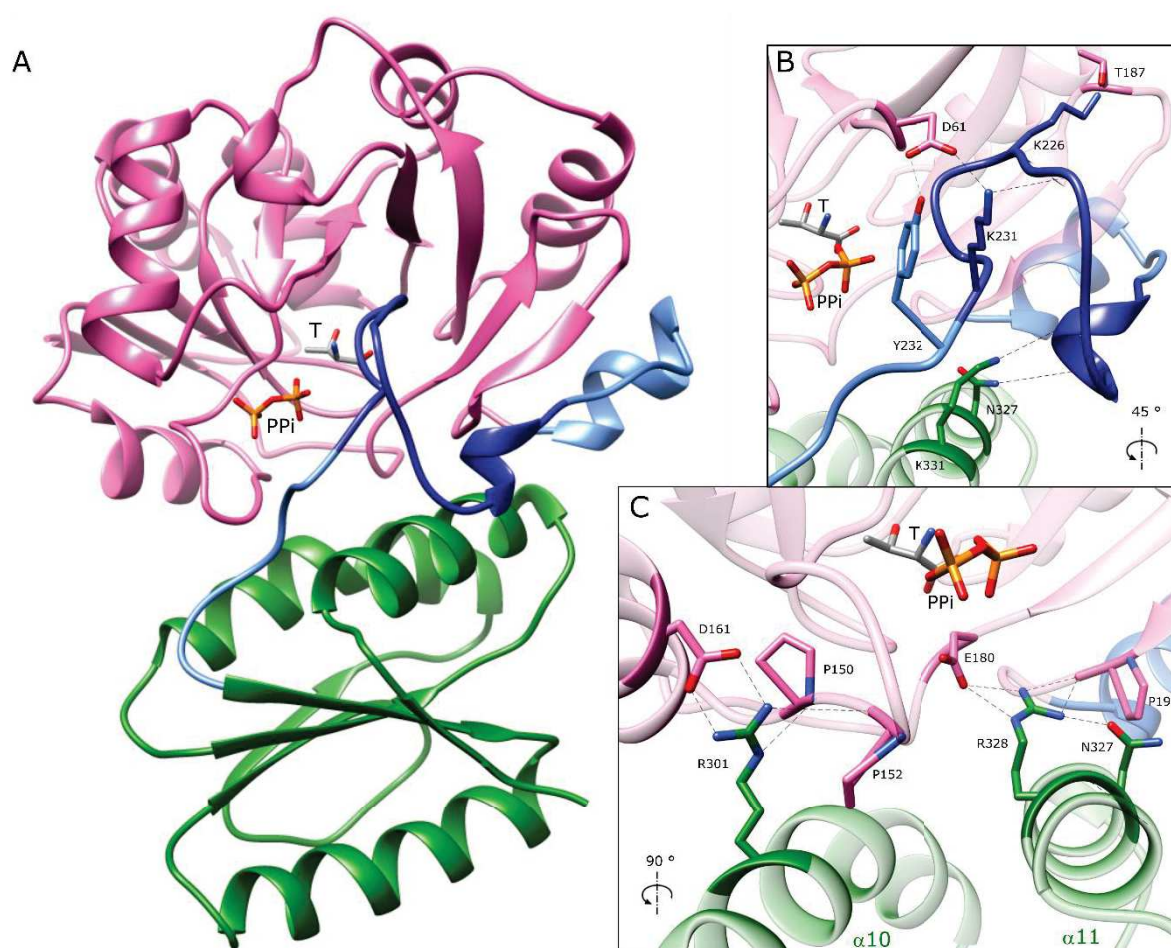


Figure 33. The structure of Sua5 from *Pyrococcus abyssi*

(A) The overall structure of *Pa*-Sua5 bound to PPi and L-threonine (denoted as T). The YrdC-like domain (in pink) is similar to the *Ec*-YrdC structure. The SUA5 domain (in green) is a Rossmann fold. The linker (in blue) is fully ordered in the *Pa*-Sua5 structures. The region that was not resolved in the previous *St*-Sua5 structures is colored in dark blue. **(B)** Zoom in the active site of *Pa*-Sua5. The linker is stabilized by several H-bonds (indicated as black dashed lines) with conserved residues of both the YrdC-like and the SUA5 domains. Its hairpin structure is twisted and slightly enters the catalytic cavity. Regions apart from the linker and the residues involved in its stabilization are shown in transparent. **(C)** Zoom in the inter-domain surface of *Pa*-Sua5. The conserved residues Arg³⁰¹ and Arg³²⁸ interact with several residues of the YrdC-like domain that are highly conserved in Sua5 proteins.

We compared the structure of *Pa*-Sua5 with the structures of *St*-Sua5 in complex with AMPPNP and L-threonine (PDB: 3AJE) or with TC-AMP (PDB: 4E1B) (Figure 31A). The main difference between the structures resides in the linker peptide connecting the YrdC-like and the SUA5 domains. This linker region is disordered in the *St*-Sua5 structures, respectively from Pro²¹⁵ to Lys²³³ (*St*-Sua5 numbering) for the TC-AMP bound form and from Glu²¹⁶ to Lys²³⁶ for the AMPPNP bound form. In *Pa*-Sua5, this region is fully ordered (Figure 33A). It starts with one short α -helical turn from Pro²¹⁵ to Tyr²¹⁸, continues by a twisted hairpin from Asp²²³ to Lys²³¹ that slightly enters the active site pocket and continues by a ω -loop that folds up against the active site cavity before joining the SUA5 domain.

This linker conformation is stabilized by several interactions with conserved residues in Sua5 proteins (Figure 33B). The α -helical turn is stabilized through H-bond between the backbones of Ala²¹⁶ and Val²¹⁷ and the amine functions of, respectively, Lys³³¹ and Asn³²⁷ in the SUA5 domain. The hairpin is stabilized through interaction between Lys²²⁶ and the residue Thr¹⁸⁷ in the YrdC-like domain. The residues Lys²³¹ and Tyr²³² further stabilize the linker by forming H-bonds with Asp⁶¹ in the YrdC-like domain and, for Lys²³¹, with the backbone of Asp²²³.

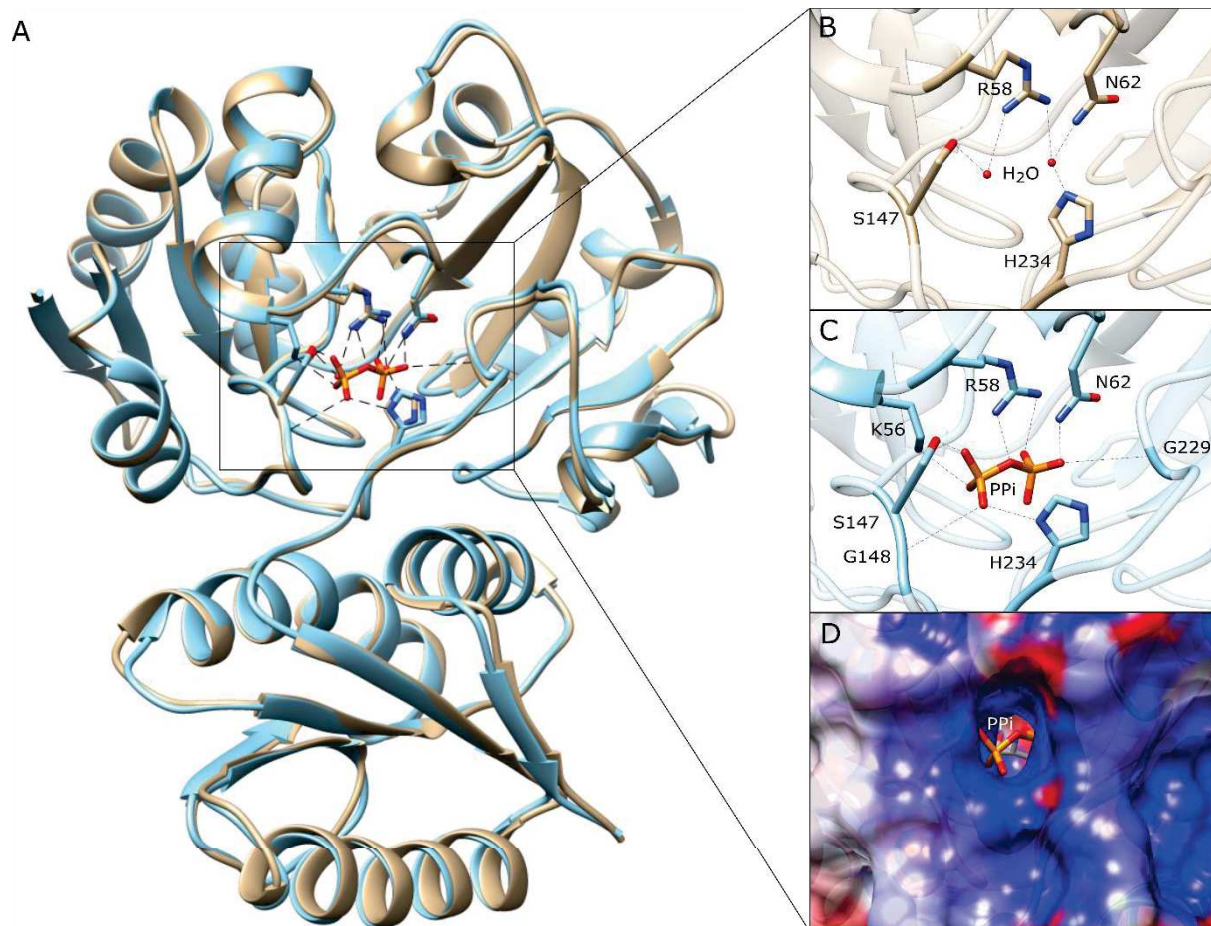


Figure 34. Comparison of the *Pa-Sua5* structures bound to PPi or water molecules

(A) Superposition of the structures of *Pa-Sua5* bound to co-purified L-threonine (in tan) and to L-threonine and PPi (in blue). The L-threonine is not shown (see [Figure 31](#) and [36](#)). In absence of added ligands, *Pa-Sua5* crystallized in complex with L-threonine and two H₂O molecules. When crystallized in presence of all its substrates, the product PPi was observed in the active site instead of the H₂O molecules. PPi and the H₂O molecules as well as the side chains of the residues they form H-bonds with are shown in the active site. Dashed lines indicate H-bonds. **(B)** Coordination of the H₂O molecules in the active site, only the side chains of residues forming H-bonds with H₂O molecules are indicated. **(C)** Coordination of PPi in the active site only the side chains of residues forming H-bonds with PPi molecules are indicated. **(D)** Electronic potential surface of *Pa-Sua5* (same view as (C)) reveals that the PPi is bound in a negative charged patch and that the structuration of the linker leaves an opening atop its binding site large enough for its passage. Blue and red color correspond to positively or negatively charged surfaces, respectively.

To map the substrate/product binding sites of *Pa*-Sua5, we co-crystallized the protein with ATP, L-threonine, HCO_3^- and MgCl_2 . Crystals were obtained in the same conditions as for the L-threonine bound form and diffracted at 2.6 Å. The asymmetric unit contains four copies of *Pa*-Sua5 with average RMSD of 0.53 Å over 340 Cα atoms. There was no indication for bound nucleotide but we observed residual electron density in the active site cavity, which could be satisfactorily fitted by a PPi ion and an L-threonine (Figure 33A).

The L-threonine bound and the PPi bound *Pa*-Sua5 structures superpose with a RMSD of 0.63 Å over 340 Cα atoms (Figure 34A). The L-threonine occupies the same position in both (Figure 36C and 36D). The PPi was likely generated by the hydrolysis of ATP during the crystallization and displaces the two water molecules present in the L-threonine bound structure (Figure 34C). It is bound slightly at the periphery of the active site where it occupies a positively charged electrostatic surface potential patch (Figure 34D). The oxygens of the phospho-moieties are coordinated by H-bonds respectively with Arg⁵⁸, Asn⁶² and Gly²²⁹ and with Lys⁵⁶, Arg⁵⁸, Ser¹⁴⁷, Gly¹⁴⁸ and His²³⁴ (Figure 34C), all highly conserved residues. Pro²²⁸ and Lys⁵⁶ also participate by making hydrophobic contacts. Lys⁵⁶, Arg⁵⁸ and Ser¹⁴³ belong to the conserved tetrad K⁵⁶xR⁵⁸/S¹⁴³xN¹⁴⁵, shown to be essential for ATP binding.

Notably the PPi ion is stabilized through interaction with two conserved residues of the linker, Gly²²⁹ and His²³⁴ (Figure 34C). This is a distinctive feature of Sua5 proteins since the YrdC proteins have no linker. At the top of the active site is an opening that could serve as a path for the PPi release (Figure 34D). PPi does not occupy the same position as the β and γ-phosphate groups of AMPPNP in *St*-Sua5 (Figure 36B and 36D). Notably, the PPi clashes with the β phosphate of AMPPNP and with the phospho-moiety of TC-AMP (Figure 35C) from *St*-Sua5 structures. This suggests that the presence of PPi in the active site of *Pa*-Sua5 may prevent the simultaneous binding of a nucleotide.

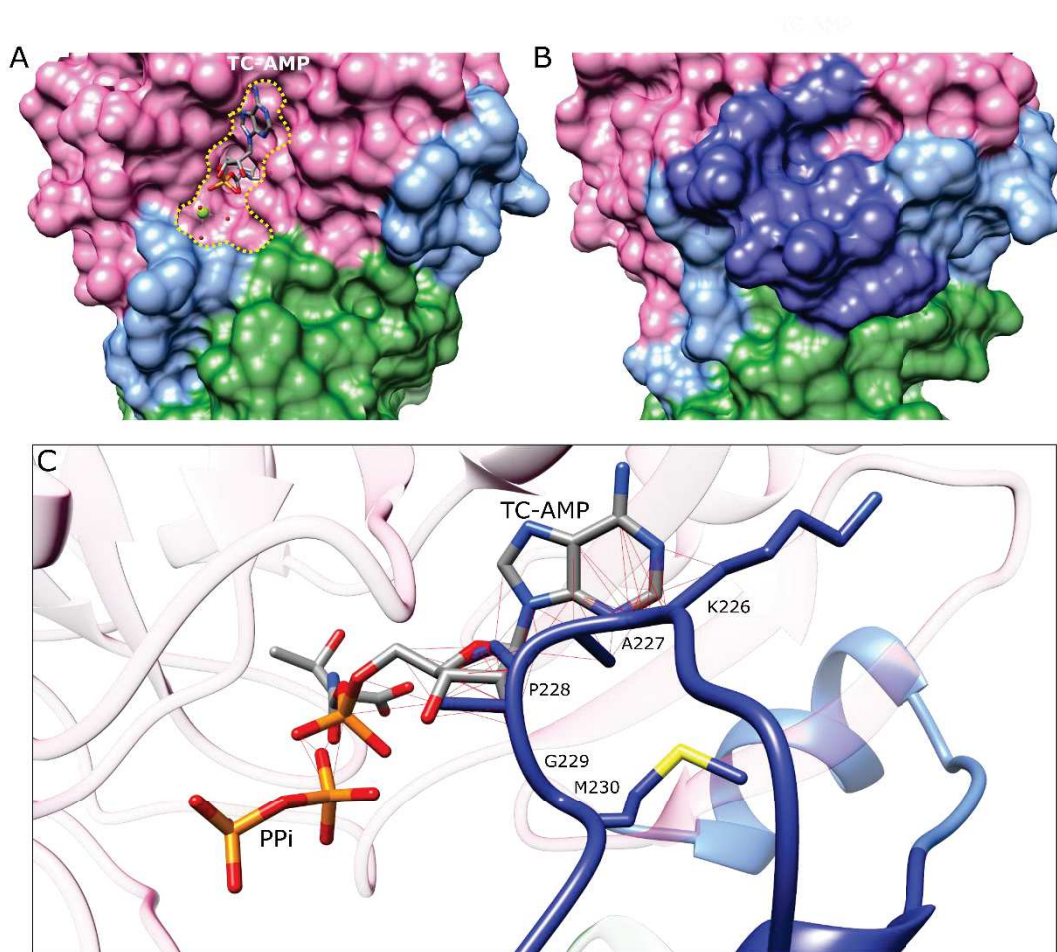


Figure 35. Linker acts as a lid closing the active site pocket of Sua5 proteins

(A) Surface representation of *St*-Sua5 bound to TC-AMP (PDB: 4E1B). The active site pocket is bordered by yellow dots. The surface of the structured part of the linker is colored in blue **(B)** Surface representation of *Pa*-Sua5. The surface of the linker is shown in blue and in dark blue. The linker closes completely the active site rendering it inaccessible for ligands or solvent. **(C)** Superposition of *St*-Sua5 bound to TC-AMP and *Pa*-Sua5 bound to PPi. Ligands and the side chains of the linker are indicated in color while the protein main chain is shown as transparent. Clashes between the linker hairpin residues and the adenylyl-moiety of TC-AMP and between PPi and the phospho-moiety of TC-AMP are shown by red lines. Clashes were determined using UCSF Chimera (Pettersen et al., 2004).

The linker is fully resolved and adopts a similar position in both structures, indicating that PPi is not required for its structuration (Figure 34A). The linker of *Pa*-Sua5 lies over the active site gorge and blocks its access (Figure 35A). The adenylate parts of AMPPNP and TC-AMP occupy the same position in the *St*-Sua5 structures (Figure 36A and 36B). Structure superposition of *Pa*-Sua5 with *St*-Sua5 reveals a clash between the linker hairpin residues Lys²²⁶-Ala²²⁷-Pro²²⁸ and the adenylyl-moiety of both AMPPNP and TC-AMP (Figure 35C). This linker structure is therefore not compatible with the presence of a nucleotide in the active site.

In *Pa*-Sua5, the conserved His²³⁴ of the linker binds to PPi while in the TC-AMP bound form of *St*-Sua5, its side chain points outwards of the active site (Figure 36). Its neighboring residue Tyr²³⁵ also has different orientation depending on the conformation adopted by the linker. Its side chain is oriented deeper in the active site in the AMPPNP *St*-Sua5 structure and is able to bind the β phosphate of AMPPNP and Ser¹⁴⁴ of the KxR/SxN tetrad (Figure 36B). The orientation of these residues is related to the flexibility of the linker and may influence the binding of ligands. The highly conserved Pro²³⁷ is found next to the bend of the linker C-terminal extremity (Figure 36).

As in the previous Sua5 structures, the *Pa*-SUA5 domain adopts a Rossmann fold with an inner β sheet composed of 5 strands and framed by 3 α helices. Helices α 10 and α 11, respectively from Val²⁸⁹ to Arg³⁰⁵ and from Gly³²⁰ to Ser³³³, interact with each other in a typical α/α coil *via* a hydrophobic zipper formed by the conserved Leu²⁹⁶, Leu³⁰⁰, Leu³²⁹ and Val³²⁵. Several conserved residues of α 10 and α 11 form a network of non-covalent interactions with residues from the catalytic domain conserved among Sua5 proteins. Notably Arg³⁰¹ forms H-bonds with the side-chain of Asp¹⁶¹ and with the Pro¹⁵⁰ and Pro¹⁵² backbones and, in a similar way, Arg³²⁸ interacts with Glu¹⁸⁰ and Pro¹⁹⁶ (Figure 33C). Arg³⁰¹ and Arg³²⁸ appear to act like two arms holding together the two domains of Sua5-proteins.

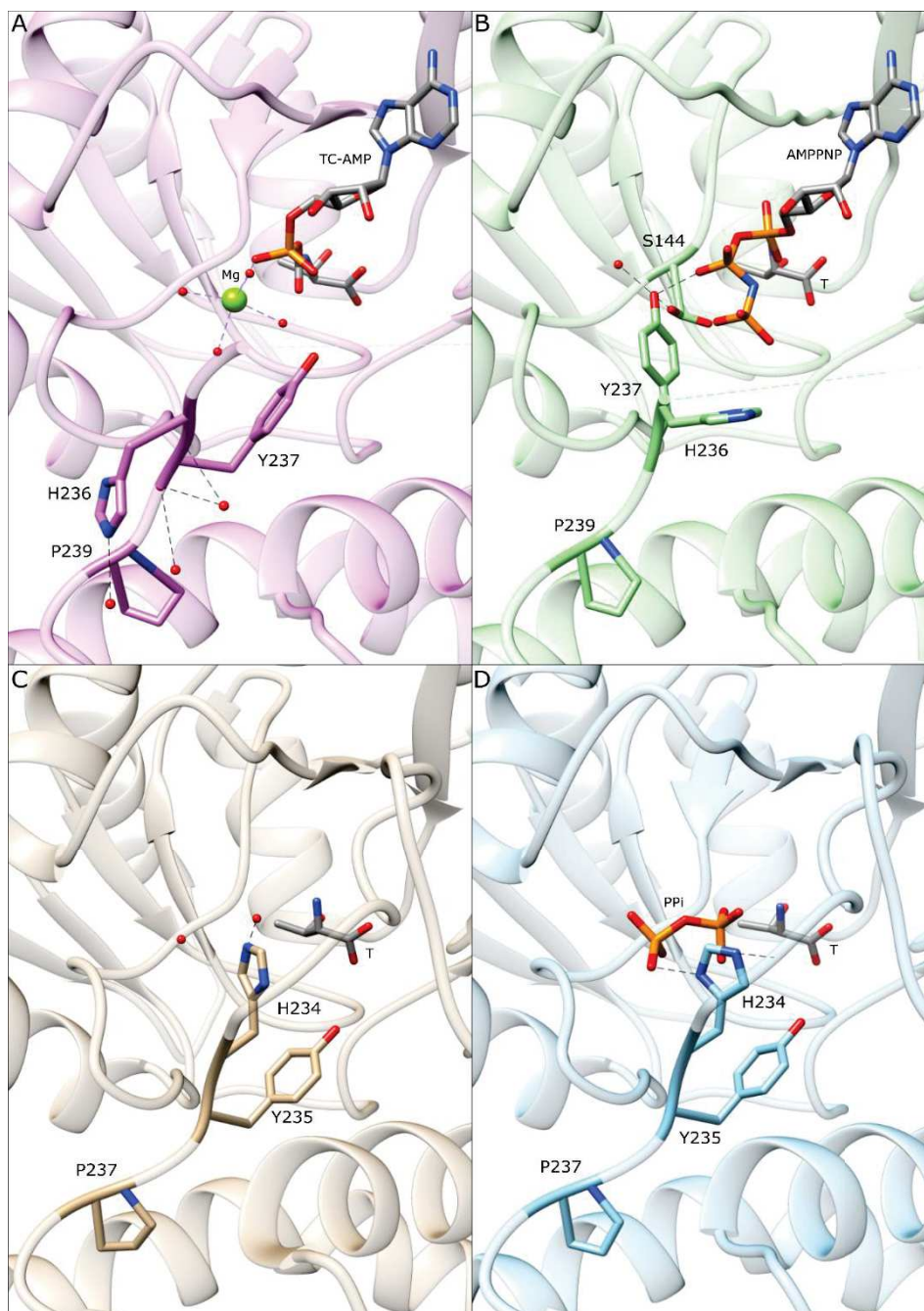


Figure 36. Orientation of the His²³⁴-Tyr²³⁵ motif side chains in the Sua5 structures

Zoom in the active site of *St*-Sua5 bound to TC-AMP **(A)** or to L-threonine and AMPPNP **(B)** and of *Pa*-Sua5 bound to L-threonine **(C)** or to L-threonine and PPi **(D)**. Proteins are colored as in [Figure 31](#) and [34](#). The main chain ribbons are transparent. The H-bonds formed by the His²³⁴-Tyr²³⁵ motif are shown by black dashed lines. The highly conserved Pro²³⁷ in *Pa*-Sua5 (Pro²³⁹ in *St*-Sua5) found at the C-terminal extremity of the linker is shown. The residues His²³⁴-Tyr²³⁵ facing the active site are highly conserved among Sua5 proteins. Their side-chains orientation, and thus their ability to interact with nucleotides or PPi, varies with the linker conformation and/or bound ligand

***In vitro* kinetics of ATP hydrolysis by Pa-Sua5^{wt}**

Sua5/YrdC proteins require L-threonine, ATP and HCO₃⁻/CO₂ to synthesize TC-AMP and PPi via a complex mechanism that involves the formation of two covalent bonds (Agari et al., 2008; El Yacoubi et al., 2009; Harris et al., 2015; Lauhon, 2012). The precise reaction mechanism of the enzymes is still unknown. As an initial step towards the elucidation of the catalytic mechanism of Sua5 proteins, we determined the kinetic constants for the hydrolysis reaction of ATP into AMP catalyzed by *Pa*-Sua5^{wt}. We used a radioactive assay to measure the initial velocity of the reaction at varying concentrations of either L-threonine or ATP. The concentration of HCO₃⁻/CO₂ was presumed to be constant as it is always present in solution if no precautions are undertaken. The best non-linear least square fit of the data produced a hyperbolic curve for both substrates indicating that the reaction obeys the classical Michaelis-Menten kinetics (Figure 37). We found K_m and k_{cat} values of $11.5 \pm 1.3 \mu\text{M}$ and $0.05 \pm 0.0017 \text{ s}^{-1}$ for L-threonine and $5.5 \pm 1.3 \mu\text{M}$ and $0.03 \pm 0.0013 \text{ s}^{-1}$ for ATP under our experimental conditions. The unexpected difference in k_{cat} values may be a consequence of large variations in quantification of radioactive AMP at low concentrations of ATP (Figure 37B). Our data are comparable to the apparent rate constants for the ATPase activity of *St*-Sua5 of 0.04 and 0.1 s⁻¹ at 37 and 60 °C, respectively, in the presence of 1 mM ATP, 5 mM MgCl₂, 50 mM K-HEPES, and 100 mM KCl, pH 7.5 (Agari et al., 2008).

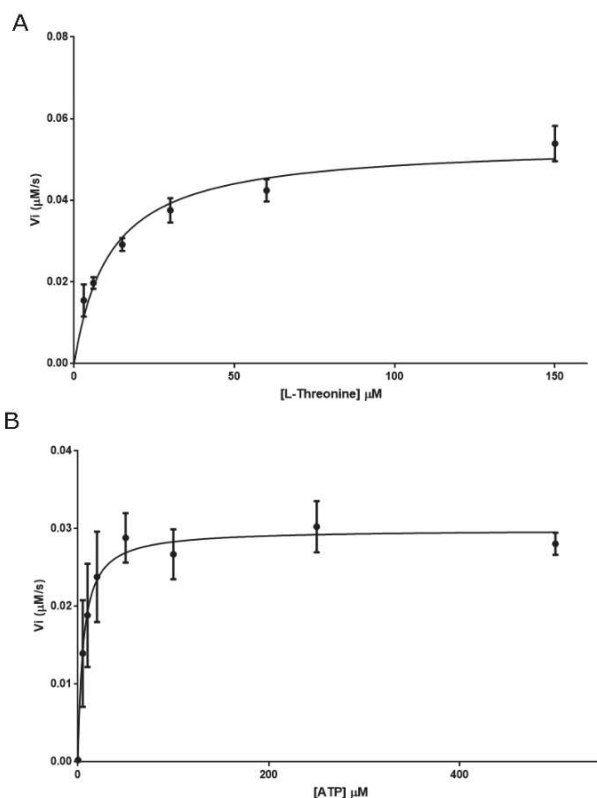


Figure 37. Kinetics of the ATP hydrolysis by *Pa*-Sua5^{wt}

Initial velocity of the AMP formation catalyzed by *Pa*-Sua5^{wt} was measured for varied concentrations of L-threonine (A) or ATP (B). For each concentration of L-threonine or ATP the initial velocity was determined as the slope of the linear curve obtained by plotting the amount of produced AMP versus time for at least four time points. The initial velocity values correspond to the amount of ATP converted to AMP and pyrophosphate per second. Error bars correspond to standard deviation from the mean value.

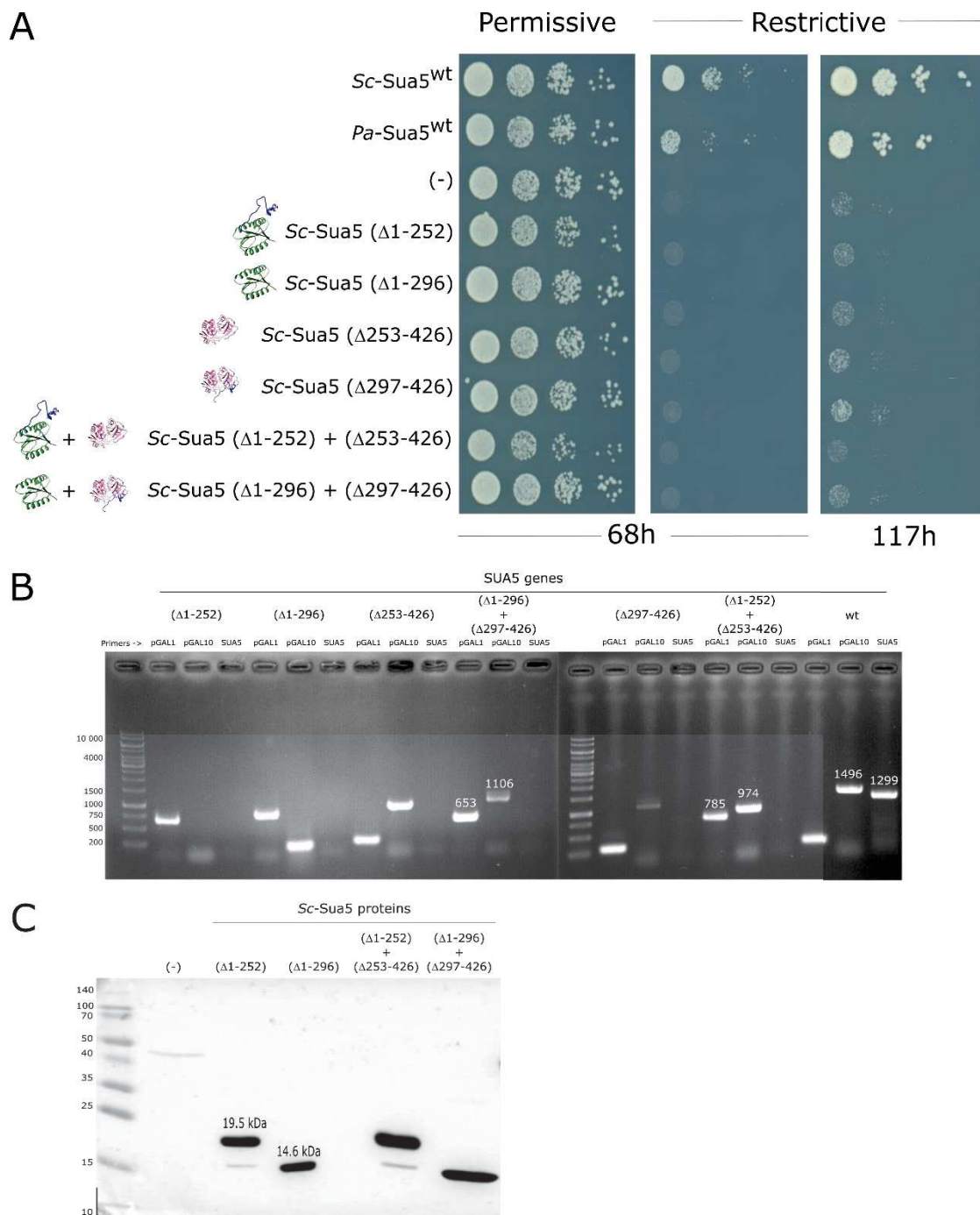


Figure 38. Complementation of Δ sua5 *S. cerevisiae* by truncated *Sc*-Sua5 proteins

(A) Complementation assays were realized with truncated mutants of *Sc*-SUA5. As for *Pa*-Sua5 truncated mutants, the YrdC-like domain of *Sc*-Sua5 was expressed with (Δ297-436) or without (Δ253-426) the linker. Similarly, the SUA5 domain was expressed with (Δ1-252) or without (Δ1-296) linker. **(B)** The genetic constructions were validated by colony PCR (see Figure 42). **(C)** Expression of recombinant proteins in *S. cerevisiae* was analyzed by western blot. The FLAG-tagged YrdC-like domains with or without linker (Δ253-426 or Δ297-426), expressed under pGAL10, were not detectable. See Table Annex 3 for the gene and protein nomenclature.

The YrdC-like domain of Pa-Sua5 is highly unstable in absence of the SUA5 domain

The structures of the catalytic domains of *Ec*-YrdC and of *Pa*- and *St*-Sua5 enzymes are very similar. To find out if the SUA5 domain is required for TC-AMP synthesis, we tested truncated variants of *Pa*-Sua5. Using the PROSITE-ProRule annotation and structure superposition, we made three constructs. *Pa*-Sua5 $\Delta(200-340)$ and *Pa*-Sua5 $\Delta(240-340)$ correspond to the YrdC-like domain, respectively with or without the linker, and *Pa*-Sua5 $\Delta(1-239)$ corresponds to the SUA5 domain. After expression in *E. coli*, *Pa*-Sua5 $\Delta(200-340)$ and *Pa*-Sua5 $\Delta(240-340)$ could not be detected in the soluble fraction and *Pa*-Sua5 $\Delta(1-239)$ was significantly less soluble than *Pa*-Sua5^{wt} under different expression conditions tested (see materials and methods).

Nevertheless, small amounts of *Pa*-Sua5 $\Delta(200-340)$, *Pa*-Sua5 $\Delta(240-340)$ and *Pa*-Sua5 $\Delta(1-239)$ could be obtained in soluble and pure form but none of them exhibited detectable ATPase activity *in vitro* (data not shown). We noticed that these truncated proteins were rapidly degraded under several storage conditions tested indicating that they are highly unstable. In agreement with these observations we were unable to detect soluble expression of truncated mutants of Sua5 from *S. cerevisiae* using *in vivo* complementation assays (Figure 38). Linker and SUA5 domain appear to be essential for the stability of the YrdC-like domain of *Pa*-Sua5 and *Sc*-Sua5.

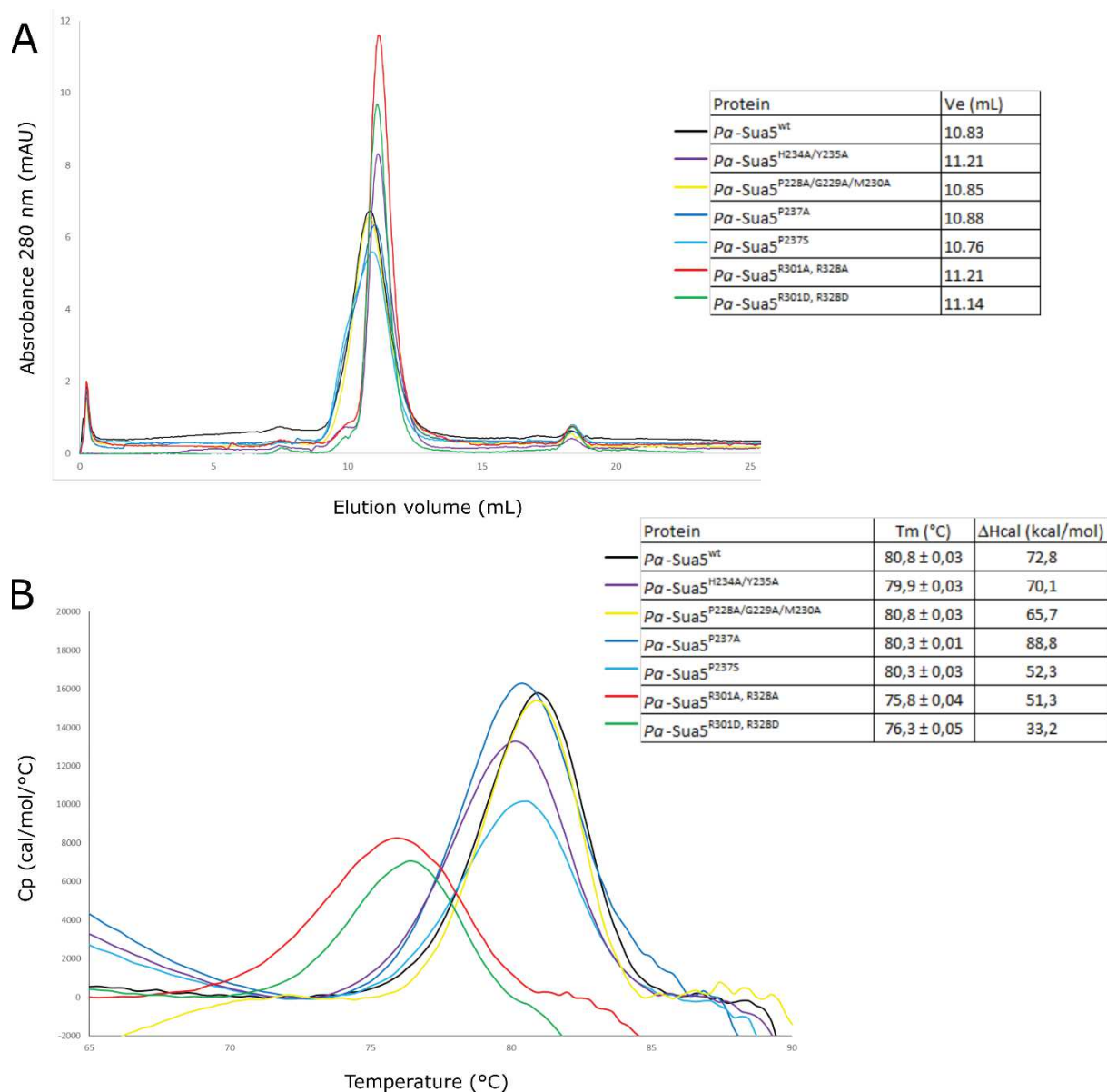


Figure 39. Analysis of recombinant *Pa*-Sua5 proteins stability

(A) Gel filtration profiles: each protein was injected at 8 μ M concentration and the gel filtration was performed under identical buffer and run conditions (see materials and methods). The recorded individual chromatograms were superposed. **(B)** Differential Scanning Calorimetry profiles temperature scan, converted to partial molar heat capacity as a function of temperature. Proteins were between 0.4 and 0.7 mg/ml in buffer containing 50 mM HEPES pH 8 KCl 35 mM. Data were obtained at a scan rate of 1 $^{\circ}$ C/min using a VP-DSC calorimeter (MicroCal).

Mutation of conserved residues of linker and SUA5 domain produce stable proteins

The structure of *Pa*-Sua5 suggests that the linker and the SUA5 domain could interfere with enzymatic activity. We set out to test this hypothesis by mutating a few conserved residues in the linker and at the domain interface. The sequence conservation score of each residue was calculated using 500 orthologs and mapped onto the *Pa*-Sua5 structure using Consurf (Ashkenazy et al., 2016) (Figure 40). The linker region from Ala²²⁵ to Ala²³⁹, facing the active site, contains several highly-conserved residues. Notably the motifs Pro²²⁸-Gly²²⁹-Met²³⁰ (Figure 35C) and His²³⁴-Tyr²³⁵ (Figure 36) are positioned respectively near the adenylyl-moiety and the phosphate groups of PPi/ATP. To test the importance of these motifs for the activity we generated the *Pa*-Sua5^{P228A/G229A/M230A} and *Pa*-Sua5^{H234A/Y235A} mutant variants. We also tested the *Pa*-Sua5^{P237A} and *Pa*-Sua5^{P237S} mutants reasoning that this proline could influence the flexibility of the linker region. In addition, we constructed the *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D} mutants to test if the integrity of the domain-interface affects the activity of Sua5 proteins.

We first evaluated the effect of the mutations on protein stability using gel filtration experiments and DSC analysis (Figure 39). The gel filtration elution profiles of the mutant proteins were similar to that of the wild-type protein indicating that the mutations did not modify significantly the Stokes radius of *Pa*-Sua5 (Figure 39A). The DSC analysis produced single relatively narrow and symmetric peaks for all proteins corresponding to a highly cooperative unfolding process typical for compact globular proteins (Figure 39B). The sharp drop at the end of the curve indicated a non-reversible denaturation for all proteins. *Pa*-Sua5^{wt} has a melting temperature of 80.8 ± 0.3 °C in agreement with the hyperthermophilic nature of *P. abyssi*. The linker mutants *Pa*-Sua5^{P228A/G229A/M230A}, *Pa*-Sua5^{H234A/Y235A}, *Pa*-Sua5^{P237A} and *Pa*-Sua5^{P237S} unfolded over a temperature range similar to that of *Pa*-Sua5^{wt} (75 °C – 85 °C), with no major impact on melting temperatures (78.85 – 80.86 °C). In contrast, the endothermic peaks recorded for *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D} were broader and their melting temperatures were lowered by 5°C suggesting that mutations of the inter-domain interface affect the stability of *Pa*-Sua5.

A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
	T	I	I	I	N	V	R	E	R	I	E	E	W	K	I	R	I	A	A	G	F	I	R	E	G	K	L	V	A	F	P	T	E	T	V	Y	G	L	G
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
A	N	A	L	D	E	N	A	V	K	R	I	F	E	A	K	G	R	P	A	D	N	P	L	I	I	H	I	A	S	F	E	Q	L	E	V	L	A	K	E
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
I	P	E	E	A	E	M	L	A	K	R	F	W	P	G	P	L	T	L	V	L	P	K	S	E	V	V	P	R	V	I	T	G	G	L	D	T	V	A	V
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
R	M	P	A	H	E	I	A	L	K	L	I	E	L	S	E	R	P	I	A	A	P	S	A	N	I	S	G	K	P	S	P	T	S	A	H	H	V	A	E
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
D	F	Y	G	K	I	E	C	I	I	D	G	G	E	T	R	I	G	V	E	S	T	V	I	D	L	T	E	W	P	P	V	L	L	R	P	G	L	P	
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
L	E	E	I	E	K	V	I	G	E	I	R	I	H	P	A	V	Y	G	K	S	V	D	T	A	K	A	P	G	M	K	Y	R	H	Y	A	P	S	A	E
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280
V	I	V	V	E	G	P	R	D	K	V	R	K	I	E	E	L	I	A	K	F	K	E	E	G	K	K	V	G	V	I	I	G	S	G	S	Y	D	A	D
281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320
E	V	F	Y	L	G	D	T	V	E	E	I	A	R	N	L	F	K	A	L	R	H	M	D	R	T	G	V	D	V	I	L	A	E	G	V	E	E	K	G
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339																					
L	G	L	A	V	M	N	R	L	R	K	A	S	G	Y	R	I	I	K																					

B

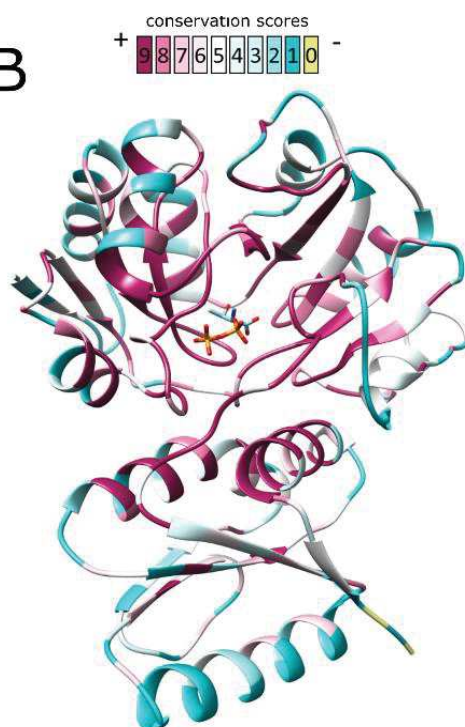


Figure 40. Sequence conservation analysis of *Pa*-Sua5

Conservation score mapped on the sequence **(A)** and on the structure **(B)** of *Pa*-Sua5. Sequence conservation levels were calculated with Consurf program (Ashkenazy et al., 2016). 500 homologous sequences to *Pa*-Sua5 sequence were retrieved with CS-BLAST (Biegert and Söding, 2009) from the Uniref 90 database and multiple sequence alignment was generated. The position-specific conservation scores were computed using the empirical Bayesian algorithms. The continuous conservation scores were divided into nine grades, from grade 1 for the most variable positions (in turquoise), through grade 5 for intermediately conserved positions (in white), to grade 9 for the most conserved (in purple).

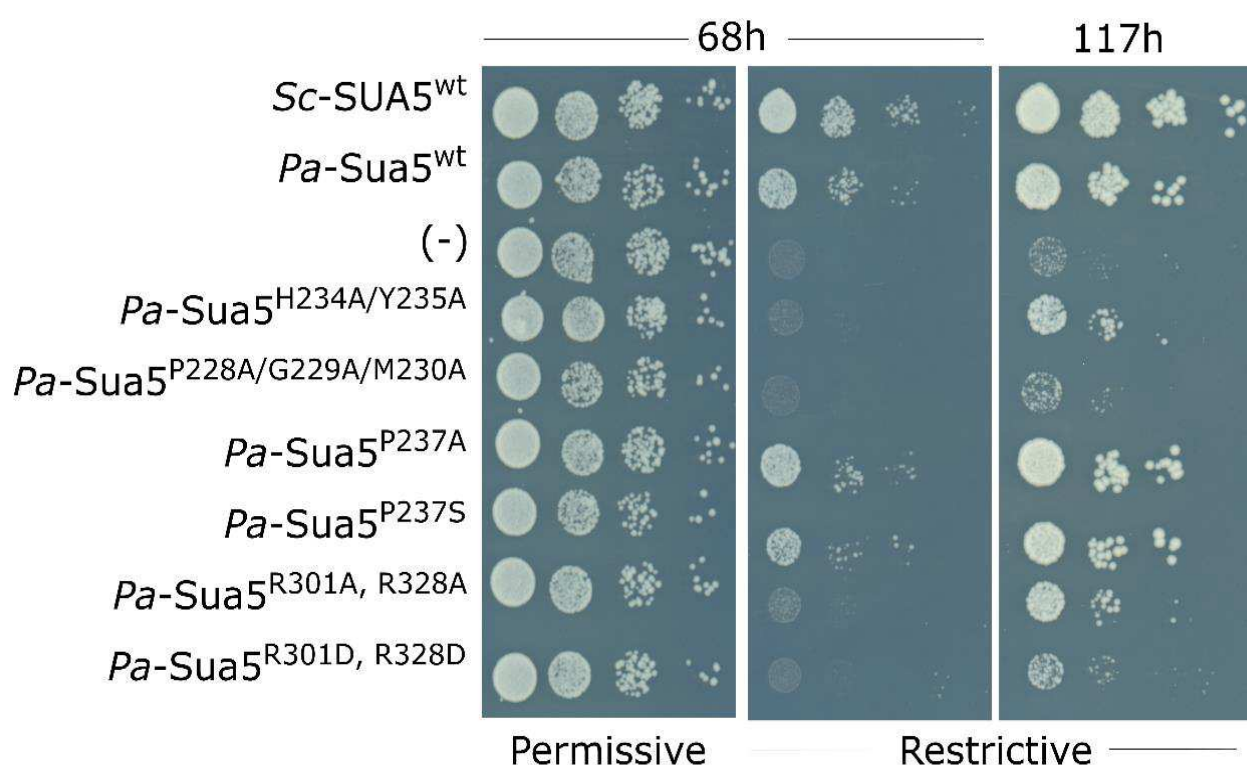


Figure 41. Functional complementation of Δ sua5 *S. cerevisiae* by *Pa*-Sua5 proteins

The function of the *Pa*-Sua5 mutant alleles was tested in a complementation assay of a Δ sua5 *S. cerevisiae* strain carrying a URA3-plasmid expressing SUA5 in *trans*. pESC-LEU plasmids carrying *Pa*-Sua5 alleles under the control the strong inducible yeast promoters pGAL1 or pGAL10 were transformed into the yeast strain. Transformants were grown in GAL-LEU medium and serial dilutions of the cultures were spotted in parallel on GAL-LEU (permissive) and GAL-LEU+5FOA (restrictive) plates and incubated at 28 °C to monitor the complementation. The control experiment under permissive conditions (left panel) showed that all the constructs grew similarly.

The conserved residues in the linker and SUA5 domain are functionally important

To test if the identified conserved residues are functionally important, we used *in vivo* complementation assay in *S. cerevisiae*. The absence of the *SUA5* gene in yeast results in an extremely slow growth phenotype. Thus, we used as host a *S. cerevisiae* strain that carried chromosomal deletion of the *SUA5* gene and a plasmid YcpLac33 expressing the endogenous *SUA5* gene in *trans*. The genes encoding mutant *Pa*-Sua5 proteins were then introduced on a second plasmid pESC-LEU. Plasmid shuffling i.e. selection of the strain which lost the YcpLac33-*SUA5* plasmid was done in restrictive condition by addition of 5-FOA into the growth medium. For all constructs, the presence of the pESC-LEU plasmid and the loss of the YcpLac33-*SUA5* plasmid after plasmid shuffling were validated by colony PCR (Figure 42A). In addition, the heterologous expression of all recombinant *Pa*-Sua5 proteins in yeast cells was confirmed using western blot (Figure 42B).

After plasmid shuffling, *Pa*-Sua5^{wt} robustly complemented the absence of *Sc*-SUA5 despite the incubation temperature (28 °C) much below the optimal growth temperature of *P. abyssi* (96 °C) (Figure 41). A similar phenotype was observed for *Pa*-Sua5^{P237A}, but only partial growth complementation was observed for *Pa*-Sua5^{H234A/Y235A}, *Pa*-Sua5^{P237S}, *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D}. Of note, the *Sua5*^{R301D, R328D} mutant was more affected than the *Pa*-Sua5^{R301A, R328A} mutant. Expression of *Pa*-Sua5^{P228A/G229A/M230A} mutant resulted in an amorphic phenotype (similar to that of Δ *sua5* strain) indicating that this mutant has no significant activity. Together, the data indicate that the conserved residues in the linker and in the SUA5 domain are functionally important *in vivo*.

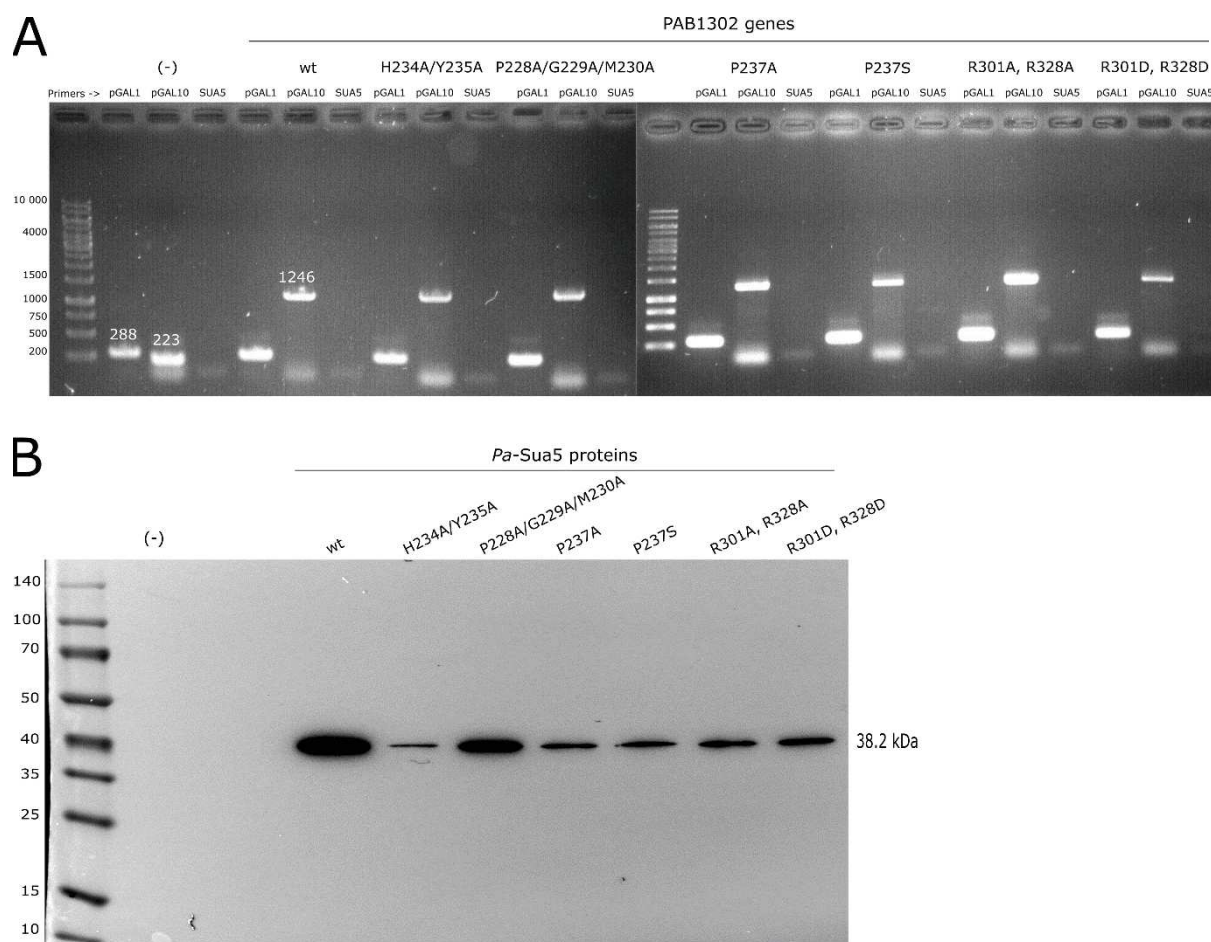


Figure 42. Validation of the *in vivo* complementation assays for *Pa-Sua5* proteins

(A) Colony PCR was done after plasmid shuffling to test the presence of the recombinant pESC-LEU plasmid and the loss of the YcpLac33-URA-*SUA5* plasmid. Oligonucleotides were designed to bind to the pGAL1 or pGAL10 promoters and gene coding for *Pa-Sua5* proteins or to the yeast *SUA5* gene. The expected size of PCR generated fragments is indicated in bp. All the *Pa-Sua5* proteins were expressed under pGAL10 control i.e with a FLAG-tag in N-terminal. **(B)** Western blot was done to confirm the recombinant expression of *Pa-Sua5* proteins in *S. cerevisiae* cells. For each construct, the total soluble proteins were extracted and probed with anti-FLAG primary antibody. See [Annex 3](#) for the gene and protein nomenclature.

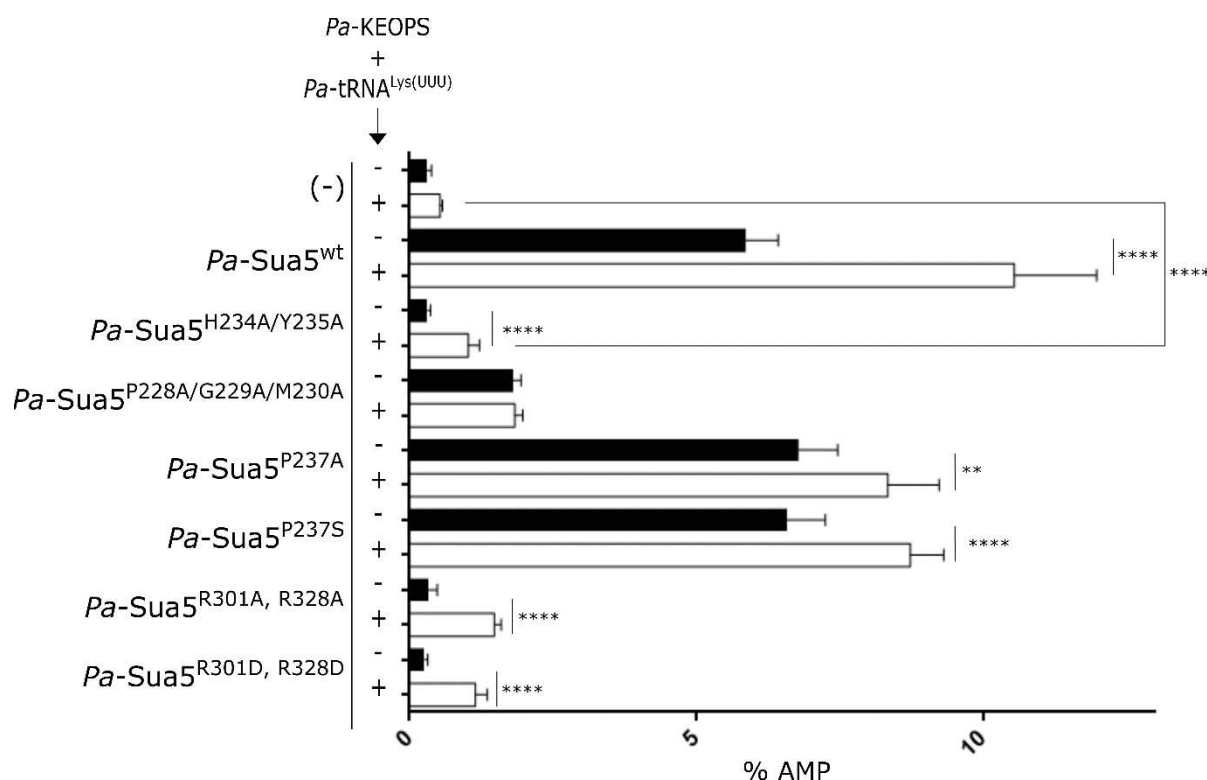


Figure 43. *In vitro* ATPase activity of *Pa*-Sua5 proteins

ATPase activity assays were performed using 2 μ M of *Pa*-Sua5 enzymes. Where indicated, the mixture was incubated in presence of *Pa*-KEOPS at 2 μ M and *Pa*-tRNA^{Lys(UUU)} at 5 μ M (white bars). After incubation for 1 h at 50 °C in presence of radioactive [α^{32} -P] ATP, the nucleotides were separated using TLC (see Figure 44 and materials and methods section). Radioactive spots corresponding to ATP and AMP were recorded and quantified. All experiments were done at least in technical triplicates and biological duplicate. The bars correspond to standard error deviation. The data were analyzed using a t test (p value < 0.05). Significant differences are indicated by **** where p value < 0.0001 and ** where p value < 0.01).

To obtain complementary data, we next measured the *in vitro* ATPase activity of *Pa*-Sua5 mutants corresponding to the amount of AMP detected after 1 h of incubation at 50 °C (Figure 43 - black bars). The activities of the *Pa*-Sua5^{P237A} and *Pa*-Sua5^{P237S} mutants were similar to that of *Pa*-Sua5^{wt}. These data are in line with the *in vivo* data and suggest that the strictly conserved Pro²³⁷ at the base of the linker is not essential for the function in the conditions tested.

Interestingly, *Pa*-Sua5^{H234A/Y235A}, *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D} which partially complemented the yeast growth phenotype, were totally inactive *in vitro* suggesting that the activity of these mutants can be rescued in the *in vivo* conditions. Combined with the *in vivo* hypomorphic and amorphic phenotypes of these mutations this indicates that the His²³⁴-Tyr²³⁵ motif as well as the interface residues Arg³⁰¹ and Arg³²⁸ are important for *Pa*-Sua5 function. In contrast, *Pa*-Sua5^{P228A/G229A/M230A} exhibited significant ATPase activity (an approximate two-fold decrease of AMP production as compared to *Pa*-Sua5^{wt}) even though it did not complement the growth of yeast. This suggest that *Pa*-Sua5^{P228A/G229A/M230A} does not produce enough TC-AMP to sustain growth or that the observed *in vitro* activity does not correspond to production of TC-AMP.

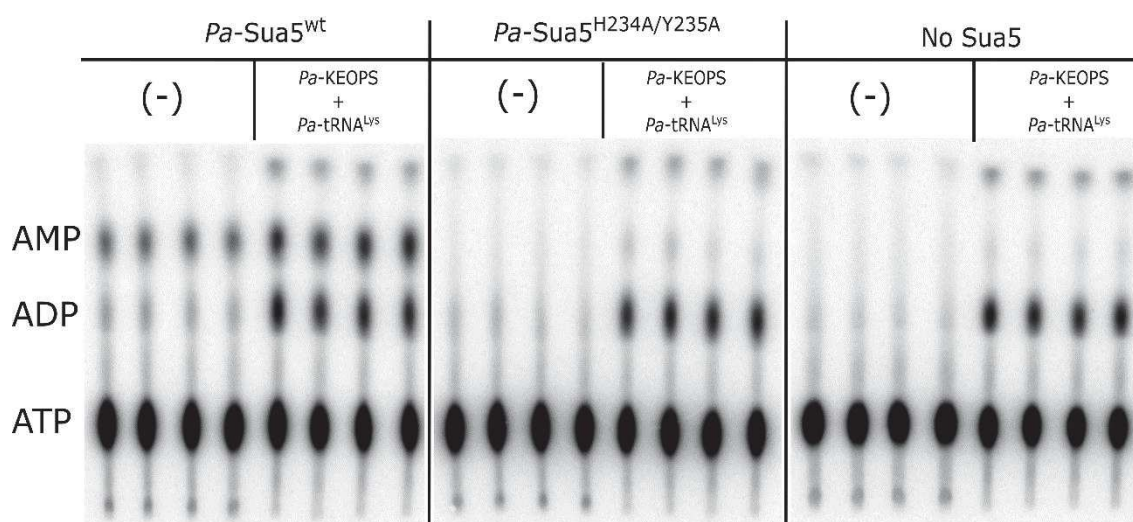


Figure 44. Nucleotide separation by TLC for *in vitro* ATPase assay

ATPase activity assays were performed using 2 μ M of *Pa*-Sua5 enzymes. Where indicated, the mixture was incubated in presence of *Pa*-KEOPS at 2 μ M and *Pa*-tRNA^{Lys(UUU)} at 5 μ M. The reaction mixtures were incubated for 1 h at 50 °C in presence of radioactive [α^{32} -P] ATP and then the nucleotides were separated using TLC. Radioactive spots corresponding to ATP, ADP and AMP were recorded using phosphorimager and quantified using ImageQuant software. The ADP detected is produced hydrolysis of ATP by the KEOPS complex during t⁶A-tRNA synthesis (Perrochia et al., 2013a).

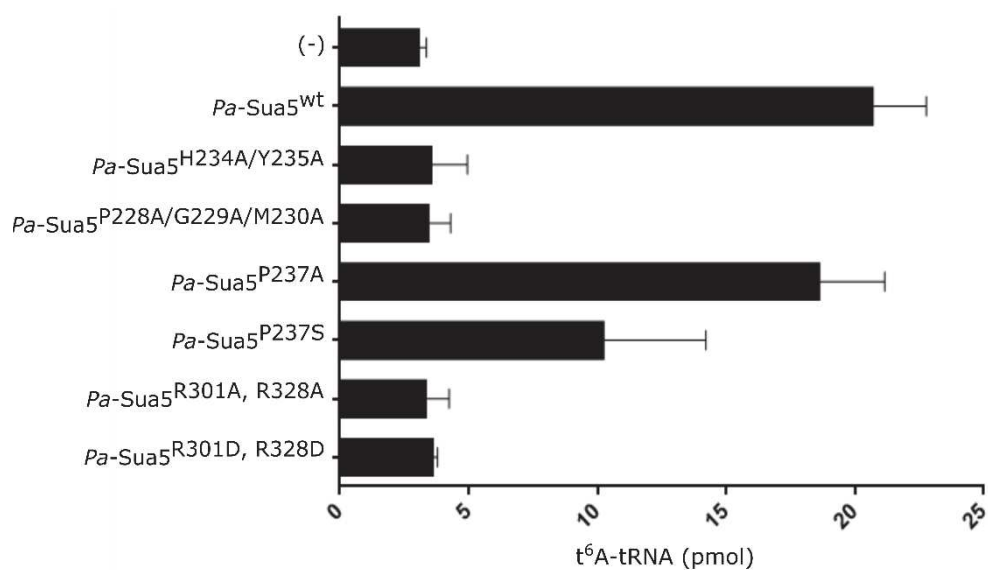


Figure 45. *In vitro* t⁶A synthesis activity of *Pa-Sua5* proteins

t⁶A activity assays were performed using 2 μM of *Pa-Sua5* and *Pa-KEOPS*, 5 μM of *Pa-tRNA*^{Lys(UUU)} and in presence of L-[1-¹⁴C]-threonine. The reaction mixtures were incubated for 1 h at 50 °C. The reactions were stopped on ice and macromolecules were precipitated by addition of 15% TCA. The tRNA were recovered on filters and the amount of radioactive tRNA, corresponding to L-[1-¹⁴C]-t⁶A-tRNA, was determined by scintillation counting. The bars correspond to standard deviation.

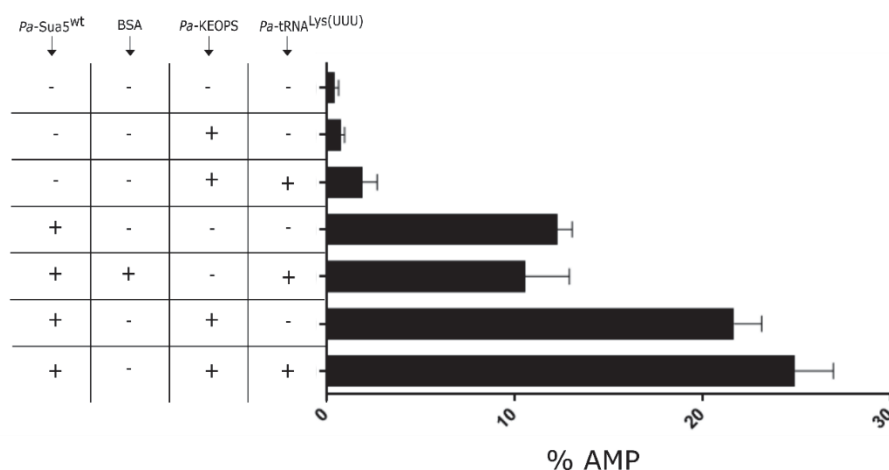


Figure 46. *In vitro* ATPase activity of *Pa-Sua5* in presence of *Pa-KEOPS*

The ATPase activity of 2 μM *Pa-Sua5*^{wt} was measured in the presence of 2 μM of *Pa-KEOPS* (or BSA) and 5 μM *tRNA*^{Lys(UUU)}. The amount of produced radioactive AMP was quantified using phosphoimager. The E-values correspond to the average E-value obtained for at least four replicates. The error bars correspond to standard deviation.

Effect of the presence of Pa-KEOPS on Pa-Sua5 in vitro ATPase activity

Sua5 proteins together with the KEOPS complex constitute the t⁶A pathway in Eukarya and Archaea. To test whether KEOPS influences the activity of Sua5, we added in our ATPase assay the *Pa*-KEOPS complex together with cognate substrate *Pa*-tRNA^{Lys} (UUU), creating reaction conditions that allow the synthesis of t⁶A-tRNA. The quantity of AMP produced by *Pa*-Sua5^{wt} increased approximately two-fold in presence of KEOPS or KEOPS and tRNA but not when BSA was added (Figure 46). *Pa*-Sua5^{P237A} and *Pa*-Sua5^{P237S} mutants also showed a small but significant increase of activity (*p* value < 0.005) (Figure 43 – white bars).

Unexpectedly, significant amount of AMP was detected for the inactive mutants *Pa*-Sua5^{H234A/Y235A}, *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D}, respectively at 9.7, 14 and 10.9 % of the wild-type level.. In contrast, the quantity of AMP produced by *Pa*-Sua5^{P228A/G229A/M230A} was not affected by the presence of KEOPS and/or tRNA. We verified indirectly whether the AMP detected corresponded to actual production of TC-AMP by measuring the t⁶A-tRNA produced in the same reaction conditions. No significant amounts of t⁶A could be detected for *Pa*-Sua5^{H234A/Y235A}, *Pa*-Sua5^{P228A/G229A/M230A}, *Pa*-Sua5^{R301A, R328A} and *Pa*-Sua5^{R301D, R328D} (Figure 45). This suggests that these mutants hydrolyze ATP but do not form TC-AMP. The mechanism behind the stimulation effect of KEOPS and its absence in the case of the *Pa*-Sua5^{P228A/G229A/M230A} mutant remains unknown.

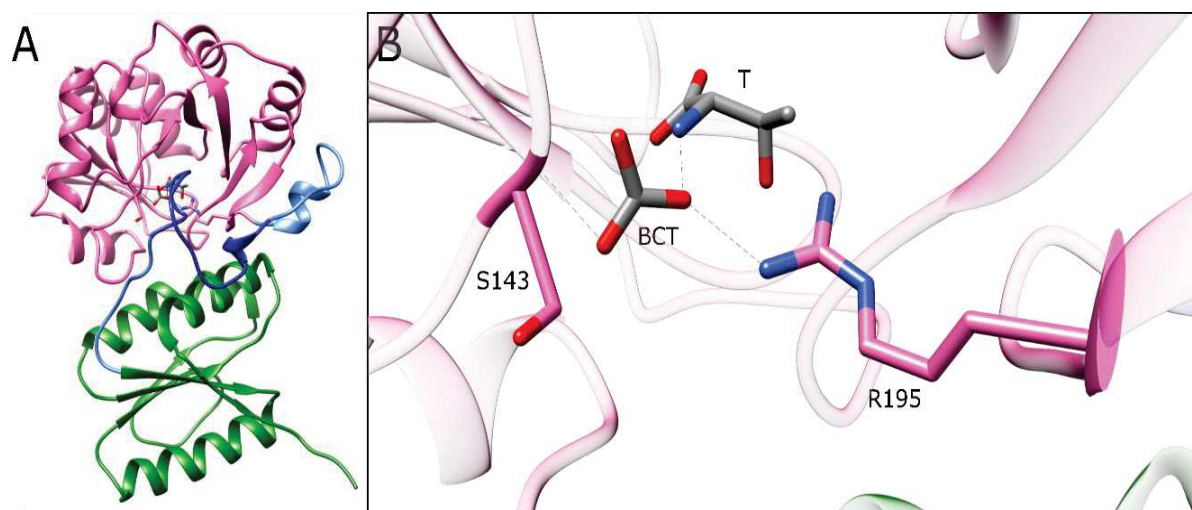


Figure 47. Structure of *Pa*-Sua5^{H234A/Y235A}

(A) The overall structure of *Pa*-Sua5^{H234A/Y235A} is similar to the wt structure. Only the monomer A is shown and is colored by domain as in [Figure 33](#). **(B)** An electronic density cloud next to the substrate L-threonine in the active site could be fitted by a HCO₃⁻ ion (BCT). It faces the amine function of the L-threonine. Except for the ligands and the binding residues of HCO₃⁻, the protein main chain is shown as transparent.

Crystallographic structure of Pa-Sua5^{H234A/Y235A}

The mutant *Pa-Sua5^{H234A/Y235A}* was co-crystallized with ATP, L-threonine, HCO_3^- and MgCl_2 in the same conditions as for *Pa-Sua5^{wt}*. Crystals belonged also to the C2 space group but with different cell edges from the crystals of the wild type protein. Data were obtained at a resolution of 2.8 Å. The asymmetric unit contains two copies of *Pa-Sua5*. The overall structure of *Pa-Sua5^{H234A/Y235A}* is almost identical to that of the wild type with an RMSD of 0.53 Å over 331 Cα atoms for monomer A and of 0.565 Å over 311 Cα atoms for monomer B. The linker was clearly defined up to residue 233 in one copy of *Pa-Sua5* but the two mutated residues were disordered.

The linker is completely disordered in the second copy of *Pa-Sua5*. The structured linker region superposes well with the *Pa-Sua5^{wt}* linker and is probably stabilized by crystal contacts. Although no ATP, TC-AMP or PPi could be detected in the residual electron density, L-threonine and an electronic density cloud that could be fitted by a HCO_3^- ion were present in the active site (Figure 47). The HCO_3^- ion is facing the nitrogen from L-threonine (C-N distance is 3.1 Å), which occupies the same position as in all other Sua5 structures, and is stabilized by ionic interactions with Arg¹⁹⁵ and the backbone of Ser¹⁴³, both of them conserved residues in the catalytic site of Sua5/YrdC proteins (Figure 47B). The HCO_3^- ion superposes with the carbamoyl group of the TC-AMP reaction product.

Discussion

t⁶A is a universal nucleoside found at position 37, next to the anticodon, of all tRNA decoding ANN codons. Its presence is critical for translation accuracy. The first step of the t⁶A-tRNA reaction pathway consists in the formation of TC-AMP. This intermediate is highly instable and decomposes rapidly to give AMP and a nearly equimolar mixture of L-threonine and a cyclic derivative as final products (Lauhon, 2012). Remarkably, the synthesis of TC-AMP can be catalyzed by two related enzymes, YrdC and Sua5. Both share an orthologous catalytic domain but Sua5 proteins have an extra C-terminal domain, called SUA5 domain, which has no attributed function so far. The previous structures of Sua5 from *S. tokodaii* (Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012) revealed that the SUA5 domain contains a Rossmann fold. Its closest structural analogs found using the DALI server (Holm and Laakso, 2016) are nucleotide binding proteins such as the Signal recognition particle protein Ffh (PDB: 5GAD - chain i). As a consequence, it had been suggested that this domain may have a similar function (Agari et al., 2008; Kuratani et al., 2011; Lauhon, 2012). Yet the structure comparison of the SUA5 domain and its structural analogs shows that it is lacking residues or region involved in nucleotide binding (data not shown). In line with these data, we could not detect any *in vitro* ATPase activity for SUA5 domain alone. In addition, the SUA5 domain has no significant sequence conservation, except for residues involved in the interaction with the YrdC-like domain. Mutations disrupting this interaction in Sua5 from *P. abyssi* provoked a significant loss in activity indicating that the integrity of this interface is important for the protein function. Taken together these data suggest that the formation of inter-domain interface is the main function of the SUA5 domain.

We resolved the structure of *Pa*-Sua5 and its overall fold is similar to the one of *St*-Sua5 (Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012). The linker was partly disordered in the *St*-Sua5 structures while it has a well-defined electron density in the *Pa*-Sua5 structures. Notably, the residues from Asp²²³ to Lys²³¹ folds as a twisted hairpin that enters the active site. These residues overlap with the position of AMPPNP or TC-AMP in the *St*-Sua5 structures (Figure 35C). These data reveal that, when the linker is folded, it completely closes the active site preventing binding of nucleotides (Figure 35B).

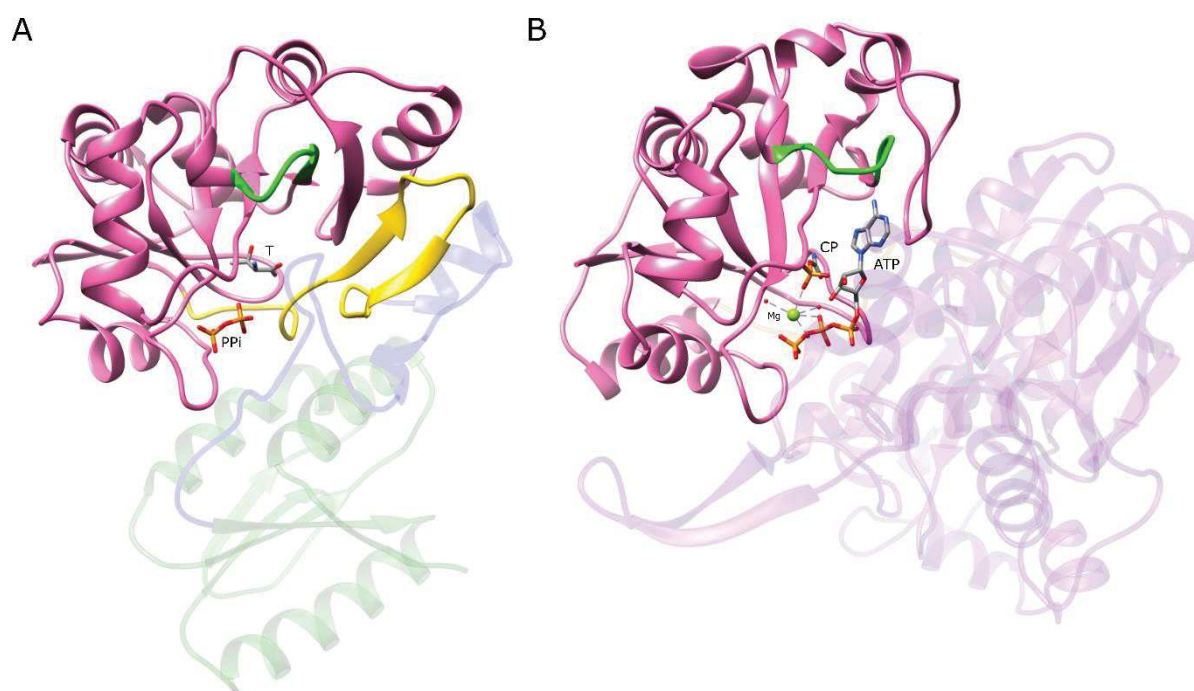


Figure 48. Comparison of the YrdC domains of TobZ and Sua5

(A) The structure of *Pa*-Sua5 bound to L-threonine and PPi (colored as in Figure 33 – linker and SUA5 domain in transparent). The residues from Gly¹⁷³ to Leu¹⁹⁹ in *Pa*-Sua5 (in yellow) are absent in the TobZ ortholog domain. **(B)** The structure of TobZ from *Streptoalloteichus tenebrarius* (PDB: 3VEZ) bound to ATP and carbamoyl-phosphate (CP). The YrdC domain is in pink and is shown in the same orientation as for *Pa*-Sua5 protein. The Kae1 domain is in transparent purple. Carbamoyl-phosphate binds in the same region as the carbamoyl-threonyl moiety in Sua5 proteins but does not involve the same residues. The residues from Val⁴⁹⁰ and Thr⁴⁹⁷ (in green) in TobZ acts as a gating loop that closes the active site of the YrdC-like domain when an ATP is bound. The corresponding region in *Pa*-Sua5, from residues Thr¹¹² to Val¹¹⁸ (also in green), was inferred from the structures superposition. It has no sequence similarity with TobZ and no flexibility was observed in this region. This suggests that the gating loop function of this region is specific to TobZ.

Based on our structural data we propose that the linker acts as a lid that undergoes conformational changes during the reaction cycle, controlling the access of substrates/products to the active site. The SUA5 domain would therefore serve as an anchoring platform to hold the C-terminal extremity of the linker and limit its flexibility. Closure of the active site by surface loops is a recurrent mechanism used by enzymes to create favorable reaction conditions. Although this reaction mechanism is plausible for Sua5, it is not clear why YrdC is able to catalyze the same reaction with no linker and SUA5 domain. In the case of the *Ec*-YrdC, it was reported that the last 20 residues are flexible and that their conformation is affected by ATP binding (Harris et al., 2015; Teplova et al., 2000). This prompted Harris and colleagues to hypothesize that the C-terminus acts as a flexible arm that may participate in the binding of ATP (Harris et al., 2015). It is therefore tempting to speculate that the C-terminal part of YrdC and the linker of Sua5 play similar roles.

The enzyme TobZ catalyzes the O-carbamoylation of the antibiotic tobramycin. The structure of TobZ from *Streptoalloteichus tenebrarius* (PDB: 3VEZ) shows that its N-terminal domain is homologous to Kae1/YgjD/Qri7 while its C-terminal domain is homologous to YrdC (Parthier et al., 2012) (Figure 48B). Both domains closely interact creating a wide internal tunnel that connects their respective active sites. In TobZ, the YrdC-like domain catalyzes the formation of a carbamoyl-AMP intermediate using carbamoyl-phosphate and ATP as substrates. After binding to TobZ, carbamoyl-phosphate is dephosphorylated and the carbamoyl group performs a nucleophilic attack upon the α -phosphate of ATP giving PPI and carbamoyl-AMP. Interestingly, the region between Val⁴⁹⁰ and Thr⁴⁹⁷ acts as a gating loop that closes the active site of the YrdC-like domain after ATP binding (Figure 48). This loop has no equivalent in Sua5/YrdC. Still, it appears that restricting the diffusion of substrates, by the use of a gating loop or a flexible linker, is a general requirement for the enzymatic reaction of the YrdC domains. Parthier *et al.* proposed that the synthesis of carbamoyl-AMP by TobZ proceeds through substrate-assisted catalysis (Parthier et al., 2012). This hypothesis is motivated by the fact that no clear catalytic side chains are present near the carbamoyl moiety of the substrate. Similarly, all the functional groups involved in TC-AMP synthesis are furnished by the substrates suggesting that Sua5 proteins also use a substrate-assisted catalysis for its formation.

After crystallization of *Pa*-Sua5 with all substrates, we identified in the active site an L-threonine and a PPi molecule, one of the reaction products. The L-threonine occupies the same deeply buried site in all the Sua5 structures. This cavity is not accessible when nucleotide analogues are occupying the active site and therefore L-threonine and $\text{HCO}_3^-/\text{CO}_2$ must bind first to ensure the synthesis of TC-AMP. Similarly, the proper coordination of ATP in the binding pocket in *Ec*-YrdC seems to require that the binding of L-threonine occurs first (Harris et al., 2015). In TobZ, ATP is proposed to bind via a cryptic site that is restored when the carbamoyl-phosphate binds first (Parthier et al., 2012). In all these orthologs there exists a constraint on the substrates binding order. In *Pa*-Sua5, the binding of PPi is stabilized by interactions with residues from the linker, in particular with the conserved motif His²³⁴-Tyr²³⁵ whose side chains are pivoted inwards when the linker is structured (Figure 36). But the closed conformation of the linker is also observed in absence of bound PPi (Figure 47).

Mutation of His²³⁴-Tyr²³⁵ shows that this motif is important for TC-AMP synthesis, likely by playing a role in the binding of ATP and/or PPi. This may be a significant functional difference with YrdC. Lauhon and coworkers observed that the TC-AMP synthesis activity of Sua5 from *Bacillus subtilis* (*Bs*-Sua5) is strongly activated in the presence of pyrophosphatase, an enzyme that hydrolyzes PPi into phosphate (Lauhon, 2012). In our structure, PPi partially overlaps with the position of the β -phosphate moiety of the nucleotide compounds in *St*-Sua5 AMPPNP and TC-AMP complexes (Figure 35C). Taken together these data suggest that PPi may be an inhibitor of the Sua5 enzymes. Yet it was also observed that the enzyme can synthesize ATP when given TC-AMP and PPi suggesting that both may bind concomitantly in the active site (Lauhon, 2012).

The current proposed mechanism for TC-AMP synthesis by Sua5/YrdC proteins involves the formation of the intermediate N-carboxy-L-threonine from $\text{HCO}_3^-/\text{CO}_2$ and L-threonine. The carbamate could then attack the α -phosphate of ATP to give TC-AMP and PPi (Lauhon, 2012; Perrochia et al., 2013b). Unexpectedly, the structure of the *Pa*-Sua5^{H234A/Y235A} mutant revealed the presence of an L-threonine and a bicarbonate (HCO_3^-) molecule in the active site (Figure 47) suggesting that the latter molecule may be the actual substrate for TC-AMP synthesis. The L-threonine molecule is pointing its NH_2 group towards the HCO_3^- ion. This likely presents the Michaelis-complex preceding the synthesis of N-carboxy-L-threonine, the hypothesized first intermediate of the reaction.

The *Pa*-Sua5^{P228A/G229A/M230A} does not complement SUA5 deletion *in vivo* suggesting it is not capable of producing TC-AMP. The conserved Pro²²⁸-Gly²²⁹-Met²³⁰ motif in the linker is part of the hairpin blocking the nucleotide binding site and closing the active site cavity. Its absence is likely to influence the flexibility of the linker and may therefore affect the synthesis of N-carboxy-L-threonine. We therefrom propose the following mechanistic model for TC-AMP synthesis by Sua5 proteins. L-threonine and $\text{HCO}_3^-/\text{CO}_2$ bind first to the narrow cavity at the bottom of the active site. The linker-lid likely closes the active site which may favor the synthesis and stabilization of N-carboxy-L-threonine by excluding solvent and increasing the local concentration of the substrates. The subsequent opening of the lid allows ATP binding followed by nucleophilic attack of the carboxyl group of N-carboxy-L-threonine leading to TC-AMP and PPi.

In Eukarya and Archaea, the KEOPS complex is responsible for the transfer of the L-threonyl-moiety from TC-AMP onto the substrate tRNA to form t⁶A-tRNA. The KEOPS complex is known to hydrolyze ATP into ADP but not to AMP (Perrochia et al., 2013a). Unexpectedly, we detected an increase in AMP production specifically when *Pa*-KEOPS was added to *Pa*-Sua5^{wt}. Furthermore, AMP formation was observed with inactive mutants upon the addition of KEOPS. However, this ATPase activity was not coupled to detectable t⁶A formation *in vitro*. Notably, the ATPase activity of *Pa*-Sua5^{P228A/G229A/M230A} mutant was not affected suggesting that this motif is required for the stimulation effect via *Pa*-KEOPS.

Pa-Sua5^{P228A/G229A/M230A} mutant was neither able to complement the Δ *sua5* yeast strain nor produce t⁶A *in vitro* but was able to produce AMP *in vitro*. This suggests that mutation of the Pro²²⁸-Gly²²⁹-Met²³⁰ motif results in decoupling of TC-AMP synthesis and ATP hydrolysis. Interestingly, Lauhon and colleagues detected the formation of AMP in the presence of L-threonine for *Bs*-Sua5 and, in significantly smaller amount, for *Ec*-YrdC (Lauhon, 2012). Still it was unclear whether AMP was produced by the direct hydrolysis of ATP or via the degradation of TC-AMP. The *Pa*-Sua5^{P228A/G229A/M230A} mutant suggests that Sua5 may have an actual ATPase activity. Further studies, especially of the *Pa*-Sua5^{P228A/G229A/M230A} mutant, are required to understand these observations. Hypothetically, this activity could serve to the Sua5 enzymes to hydrolyze ATP bound first in absence of L-threonine and/or HCO₃⁻/CO₂. *Ec*-YrdC is less prone to bind ATP in absence of the others substrates (Harris et al., 2015) and thus may have less need for this mechanism. It was also suggested that the SUA5 domain may serve to hydrolyze ATP into AMP (Lauhon, 2012). Our structure-function analysis of *Pa*-Sua5 does not suggest such a role for this domain. In particular we could not detect any *in vitro* ATPase activity for the SUA5 domain truncated mutant (Data not shown).

In conclusion, we identified two novel binding sites in Sua5 proteins: one for the bicarbonate substrate and one for the PPi product. The structure and the functional properties of the linker suggest it plays an active role during the reaction cycle, probably related to the stabilization/synthesis of the N-carboxy-L-threonine intermediate and to the binding of ATP/PPi. The absence of linker and SUA5 domain in the YrdC proteins demonstrates that these proteins use different mechanism to synthesize TC-AMP. YrdC and Sua5 are therefore closely related but functionally distinct orthologs. The identified functional differences provide clues to help understand the complex evolutionary history of these universal proteins.

Acknowledgments

We thank Domenico Libri for the generous gift of the *S. cerevisiae* [Δ *sua5*::Kan] strain. This work was supported by Lidex BIG grant of the Paris-Saclay University to TBL and HvT. This work has benefited from the platform and expertise of the Macromolecular interactions measurements Platform of I2BC

ARTICLE II

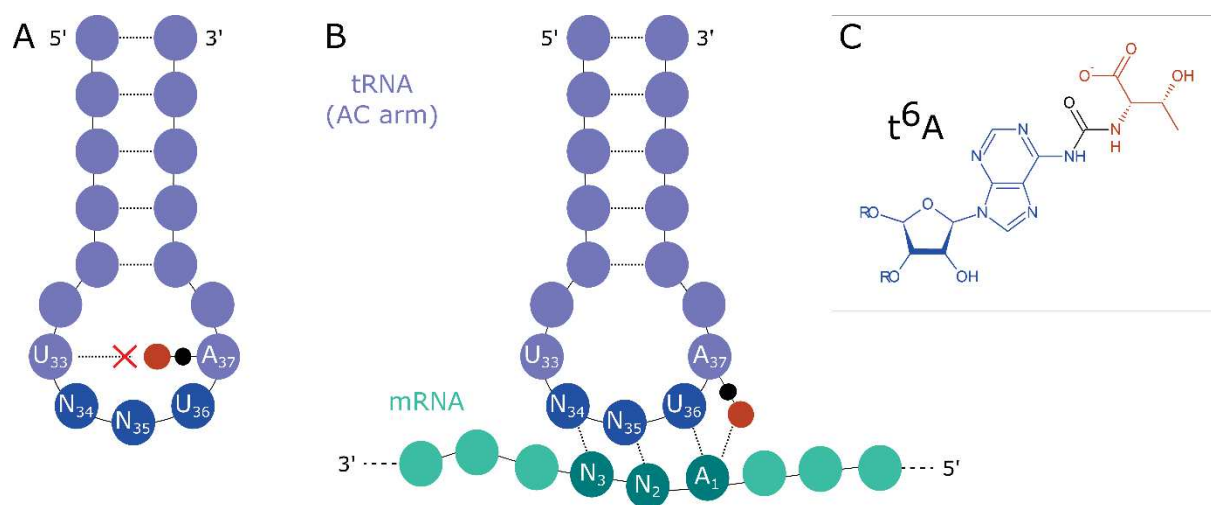


Figure 49. Structure and properties of the universal modified nucleoside t⁶A

The tRNA anticodon (AC) stem is shown in blue, anticodon bases a darker tint. The mRNA is shown in green, codon bases in a darker tint. Watson-Crick pairings are indicated by dashed lines. The t⁶A modification is symbolized by two supplementary circles on A₃₇ that are colored as in (C). **(A)** When not bound to mRNA, the side-chain of t⁶A is turned inside the AC loop and prevents the pairing between the canonical U₃₃ and A₃₇. **(B)** When bound to mRNA, the base N₃₇ undergoes a change in orientation that allows the side chain of t⁶A to form H-bonds with A₁ and to produce staking effects with N₁ and N₃₈ (not shown). As a consequence the codon-anticodon pairing strength and accuracy are increased. This prevents frameshifting and participates in the fidelity of translation. **(C)** Chemically, t⁶A is formed by the addition of a L-threonyl moiety (in orange) bound via a carbamoyl moiety (in black) to the N⁶ of A₃₇ (in blue).

Evolution of the universal Sua5/YrdC enzymes: when simplification is the ultimate sophistication.

Adeline Pichard-Kostuch, Violette Da Cunha, Jacques Oberto, Herman van Tilbeurgh, Patrick Forterre, Tamara Basta

Introduction

tRNA requires maturation to be fully efficient during the translation process. This involves *inter alia* the enzymatic modification of the canonical A, U, C and G bases to form modified nucleosides by addition of a variety of chemical groups (for a recent review about modified tRNA nucleosides see (Väre et al., 2017)). Around one hundred types of modified nucleosides are presently identified in tRNA, twenty of which are shared by the three domains of life (Cantara et al., 2011; Machnicka et al., 2013). These latter modifications were possibly inherited from the Last Universal Common Ancestor (LUCA) that was thus already using a set of modified nucleosides to enhance translation efficiency and/or accuracy. The universal modified nucleoside N⁶-threonyl-carbamoyl-adenosine (t⁶A) is found on the adenosine in position 37, i.e the 3' of the anticodon, of all tRNA that recognize ANN codons (Ishikura et al., 1969; Powers and Peterkofsky, 1972a, 1972b; Takemura et al., 1969) ([Figure 49C](#)).

The physico-chemical properties of t⁶A prevent the base-pairing between A₃₇ and the conserved U₃₃ (Stuart et al., 2000; Yarian et al., 2002), maintaining the anticodon loop in an open conformation, and preserve the canonical U-turn motif of this loop (Bénas et al., 2000; Sundaram et al., 2000) ([Figure 49A](#)). t⁶A also stabilizes the codon-anticodon pairing by strengthening the interaction between the first base of the codon A₁ and the last base of the anticodon U₃₆ (Murphy et al., 2004) ([Figure 49B](#)).

At the cellular level, the presence of t⁶A has pleiotropic consequences, not yet fully understood, which depend on the nature of the substrate tRNA, its molecular context as well as the organism. For example t⁶A was found to be required for translation accuracy (Daugeron et al., 2011; El Yacoubi et al., 2011; Freier and Tinoco, 1975; Högenauer et al., 1972; Lin et al., 2010; Weissenbach and Grosjean, 1981), aminoacylation of some tRNA species (Nureki et al., 1994) and translocation in the ribosome (Phelps et al., 2004). It is strictly essential in most model prokaryotes such as *Escherichia coli* (*E. coli*) and the archaeon *Haloferax volcanii* (*H. volcanii*) but not in *Saccharomyces cerevisiae* (*S. cerevisiae*), which survives in its absence, despite being significantly affected (for a review about the essentiality of t⁶A see (Thiaville et al., 2015b)).

t⁶A-tRNA biosynthesis is a two-step mechanism. During the first step, the universal enzyme family Sua5/YrdC (COG0009) uses L-threonine, ATP and bicarbonate/carbon dioxide to synthesize an activated intermediate called threonyl-carbamoyl-adenylate (TC-AMP) (Figure 50C) (El Yacoubi et al., 2009; Harris et al., 2015; Lauhon, 2012; Parthier et al., 2012; Perrochia et al., 2013b). In the second step, t⁶A-tRNA is formed by the transfer of the threonyl-carbamoyl moiety from the TC-AMP onto the substrate tRNA. This reaction is catalyzed by the complex DEZ in Bacteria (Deutsch et al., 2012; El Yacoubi et al., 2011; Zhang et al., 2015a) and by the complex KEOPS in Archaea and Eukarya (El Yacoubi et al., 2011; Perrochia et al., 2013a; Srinivasan et al., 2011). DEZ and KEOPS complexes both share a catalytic subunit belonging to the universal family Kae1/TsaD/Qri7. However, their respective accessory proteins show no homology (for a review about t⁶A synthetases diversity see (Thiaville et al., 2015a)). Notably, the mitochondrial ortholog Qri7 is able to catalyze the reaction without accessory proteins (Wan et al., 2013).

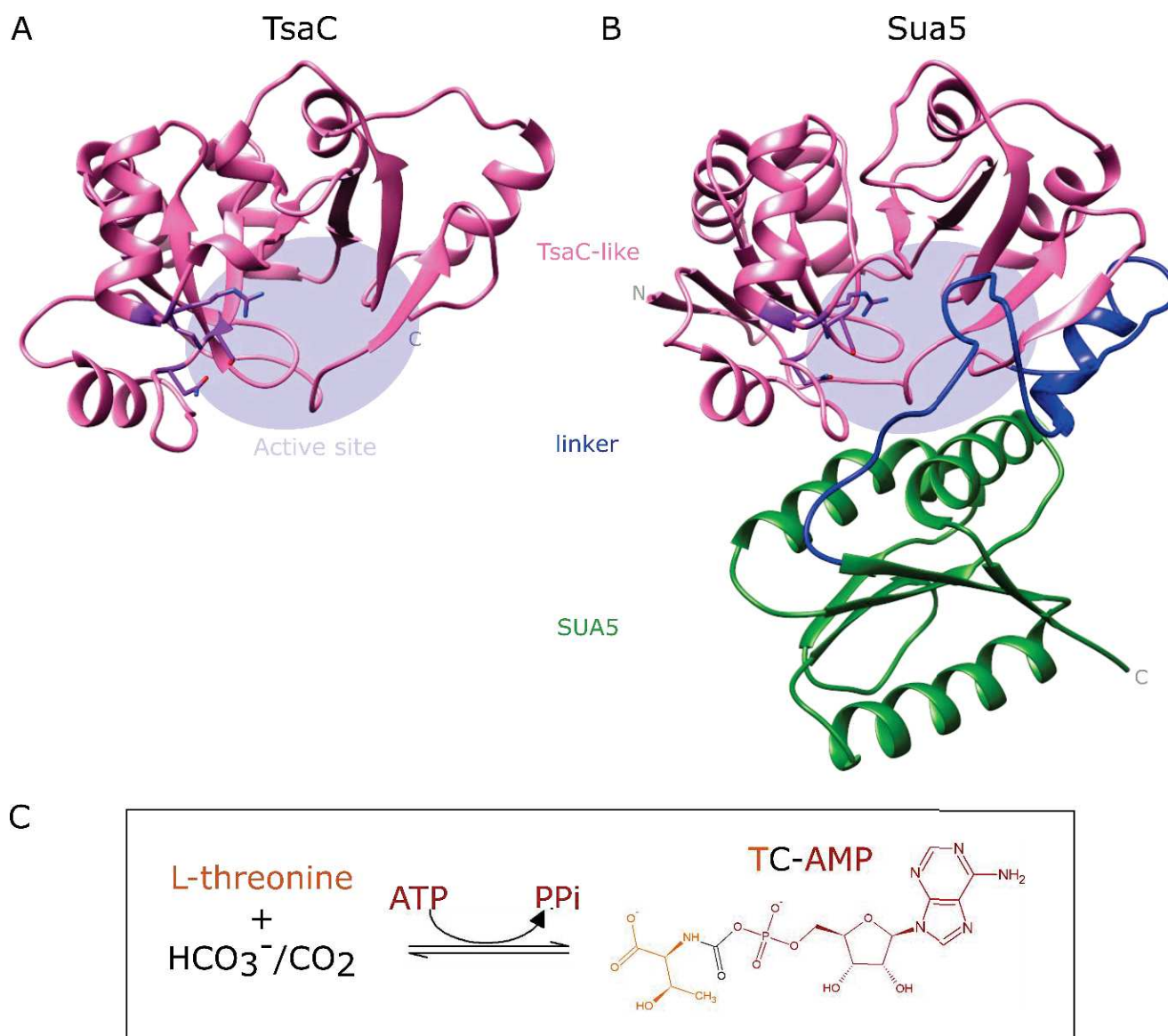


Figure 50. The Sua5/YrdC universal enzyme family

(A) The structure of YrdC from *E. coli* (Code PDB: 1HRU) (Teplova et al., 2000). **(B)** The structure of Sua5 from *P. abyssi* (Pichard-Kostuch et al., Article I). YrdC and Sua5 variants have an orthologous catalytic domain (in pink) that contains a highly conserved active site. Notably the KxR/SxN tetrad (side-chains indicated in purple) is involved in the binding of ATP and is essential for the activity in both variants. Sua5 has an additional C-terminal domain called SUA5 (in green), linked to the YrdC-like domain via a linker (in blue) that is able to change conformation and acts as a gating loop. Conserved residues in the linker and SUA5 domain were shown to be important for activity of Sua5 from *P. abyssi*. **(C)** Despite their structural heterogeneity, both YrdC and Sua5 catalyze the formation of threonyl-carbamoyl-AMP (TC-AMP) from L-threonine, bicarbonate/carbon dioxide and ATP.

The Sua5/YrdC family is found in all current complete sequenced genomes, with the exception of the reduced genomes of some insect symbionts (Thiaville et al., 2014). It has thus been presumed to have been present in the LUCA (Koonin, 2003; Ranea et al., 2006). The Sua5/YrdC family is composed of two variants, YrdC and Sua5, which both catalyze the synthesis of TC-AMP and share an orthologous catalytic domain. The structures of YrdC from *E. coli* (*Ec*-YrdC) (Harris et al., 2015; Teplova et al., 2000) and Sua5 from the *Sulfolobus tokodaii* (*St*-Sua5) (Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012) and *Pyrococcus abyssi* (*Pa*-Sua5) (Pichard-Kostuch et al., Article I) revealed that this domain adopts a globular twisted fold made of 7 parallel and anti-parallel β strands bordered by 7 α helices (Figure 50). The active site is positioned in the cavity formed by this domain and is highly conserved. Notably, the residues of the strictly conserved motif K⁵⁶xR⁵⁸/S¹⁴³xN¹⁴⁵ (*Pa*-Sua5 numbering) were shown to be required for TC-AMP synthesis in Sua5 from *S. cerevisiae* (Wan et al., 2013) and *Ec*-YrdC (Harris et al., 2015) likely by playing a role in ATP binding.

While YrdC is found as a single domain protein, Sua5 contains an additional domain, named SUA5. The SUA5 domain has a typical Rossmann fold made of 3 α helices and 5 parallel β strands (Figure 50B). There is no significant sequence conservation among SUA5 domains except in the two α helices involved in the interaction with the YrdC-like domain, namely α 10 and α 11. Most species encode either a YrdC or Sua5 protein (Thiaville et al., 2015a) though the distribution of these two variants does not follow a clear phylogenetic pattern, with both variants being found distributed among Archaea, Bacteria and Eukarya. For example the genomes of *P. abyssi*, *Bacillus subtilis* and *S. cerevisiae* (an archaeon, bacterium and eukaryote, respectively) encode Sua5 while *H. volcanii*, *E. coli* and *Homo sapiens* (also an archaeon, bacterium and eukaryote, respectively) encode YrdC. Additionally, *in vivo* and *in vitro* assays done between YrdC and Sua5 variants show functional complementation of the two variants (El Yacoubi et al., 2009), indicating that the SUA5-domain has no obvious role in other cellular activities.

The inter-domain linker of Sua5 proteins also contains conserved residues, notably the highly conserved motifs Pro²²⁸-Gly²²⁹-Met²³⁰ and His²³⁴-Tyr²³⁵ (*Pa*-Sua5 numbering). Using structure-function analysis of *Pa*-Sua5, we have shown recently that the linker may change conformation and that conserved motifs in the inter-domain interface and the linker are important for the protein activity (Pichard-Kostuch et al., Article I). We therefore proposed that the linker acts as a gating loop that influences the entry and exit of ligands during enzymatic activity by switching between a closed ordered state and an open disordered state. The SUA5 domain acts as an anchoring platform to fix the C-terminal extremity of the linker.

The high degree of sequence conservation of the linker and the inter-domain interface suggests that this mechanistic model is conserved among Sua5 proteins. Why the catalytic YrdC-like domain in Sua5 proteins is not sufficient for TC-AMP synthesis when its counterpart YrdC is fully active alone in many organisms and how this family of proteins evolved to yield two variants remind open questions.

To address these questions, we performed phylogenetic analysis of Sua5/YrdC proteins across the three domains of life. This analysis revealed that the phylogenetic signal contained in the sequences is not sufficient to yield robust phylogenetic trees. The distribution of Sua5 and YrdC across the universal phylogenetic tree based on RNA polymerase sequences shows that specificity for either variant is found within groups at different taxonomic levels. Using sequence comparisons, we identified atypical Sua5 proteins in Archaeoglobales archaea which lack key residues in the linker and SUA5 domain. Biochemical and mutational analysis of Sua5 from *Archaeoglobus profundus* revealed that, unlike canonical Sua5 proteins, this YrdC-like domain remains active in absence of the SUA5 domain. Together the data indicate that Sua5 from *A. profundus* is in the process of shortening to become YrdC variant. This corroborates the scenario in which Sua5 variant is the ancestral version that was present in the LUCA while YrdC arose through loss of SUA5 domain in the course of evolution.

Materials and methods

Phylogenetic analysis

The Sua5/YrdC sequences of the species dataset previously used in (Da Cunha et al., 2017) were extracted from the Uniprot database and used to construct several phylogenetic analyses per domain, but also at the level of the entire tree of Life. The selected sequences were first aligned using MAFFT v7 with auto settings (Yamada et al., 2016) and then trimmed with BMGE (Criscuolo and Gribaldo, 2010) with a BLOSUM30 matrix. The maximum likelihood (ML) phylogenetic analyses were performed using IQ-TREE v1.4.3 (<http://www.iqtree.org/>) with the TESTNEW option for model selection, and with both nonparametric bootstrap (100 replicates) and/or ultrafast bootstrap approximation (1,000 replicates) (Kalyaanamoorthy et al., 2017; Minh et al., 2013).

Cloning procedures and mutagenesis

The gene ARCPR_RS00515 encoding *Ap*-Sua5^{wt} was synthesized (Genscript) and cloned in the recombinant plasmid pET26b(+) using *NdeI/XhoI* sites. This plasmid was used as template for mutagenesis. The mutants were constructed using Phusion polymerase (ThermoFisher). The oligonucleotides used were synthesized (Genewiz) and are listed in Annex 3. For generating point mutations, the QuickChange Kit II (Thermofisher) protocol was used. For construction of the truncated protein, the sequences encoding the isolated protein domains were amplified using PCR. A hexahistidine tag was added in 5' of the gene which was subsequently cloned in pET26b(+) vector (Novagen) using *NdeI/XhoI* restriction sites. The sequences of the constructs are available on request. Recombinant plasmids were introduced into chemically competent *E. coli* strain TOP10 (Invitrogen). The recombinant plasmid used for heterologous expression of KEOPS complex of *P. abyssi* was previously described (Perrochia et al., 2013b).

Recombinant gene expression and protein purification

Recombinant genes were expressed in *E. coli* strain BL21-Gold (Agilent). Protein overproduction was carried out in Auto Induction Media Terrific Broth Base including trace elements (Formedium) prepared according to manufacturer's protocol. Cells were collected by centrifugation, resuspended in lysis buffer 50 mM Tris-HCl pH 8, 200 mM NaCl, 10 % glycerol. Lysozyme (0.1 mg/mL final) and anti-protease Roche cOmplete ULTRA tablets EDTA-free were added. Cell suspension was homogenized using the high pressure cell disrupter One Shot (Constant Systems Ltd). After centrifugation at 30 000 g for 30 min, the supernatant was heated at 65 °C for 30 min to precipitate bacterial proteins. Protein precipitate was removed by centrifugation at 30 000 g for 60 min. His-tagged proteins from the soluble fraction were purified on a HisTrap HP using an AKTA FPLC system (GE Healthcare). Strep-tagged proteins from soluble fraction were purified on Strep-Tactin resin (Qiagen). Fractions of interest were pooled and injected on the Superdex 75 HiLoad 16/60 (or 10/300) (GE Healthcare). Fractions containing pure proteins were concentrated, flash-frozen in liquid nitrogen and stored at -80 °C in lysis buffer containing 10-20 % glycerol.

tRNA substrate production and purification

tRNA production and purification were done following the protocol described in (Pichard-Kostuch et al., Article I).

Differential Scanning Calorimetry

DSC experiments were performed using a VP-DSC calorimeter (MicroCal). *Ap-Sua5* proteins were used at a concentration of 18 or 24 μ M in 50 mM K-HEPES pH 8, and 35 mM KCl. Scanning from 20 to 120 °C at a heating speed of 1 °C/min was done for each experiment. Data were analyzed using the MicroCal Origin software provided by the manufacturer.

Structure prediction and analysis

The tertiary structure of *Ap*-Sua5^{wt} was modeled using the program Phyre 2 in intensive mode and standard settings (Kelley et al., 2015). The structural comparison and graphic representations of the protein structures were produced with UCSF Chimera (Pettersen et al., 2004). The electrostatic potential representation were rendered and produced using APBS and PDB2PQR (Unni et al., 2011). The PDB codes of the structures are 1HRU and 2MX1 for *Ec*-YrdC, 2EQA, 3AJE and 4E1B for *St*-Sua5, 3VEZ for *St*-TobZ and 3TTC for *Ec*-HypF. Ligands coordination was analyzed with LigPlot+ (Laskowski and Swindells, 2011).

In vitro activity assays

The t⁶A synthesis reaction was performed using *Pa*-KEOPS and *Ap*-Sua5 proteins (2 μM each), *Pa*-tRNA^{Lys}(UUU) (5 μM) in reaction buffer (50 mM K-HEPES pH 8, 35 mM KCl, 10 mM MgCl₂, 1 mM MnCl₂, 5 mM DTT, 1 mM ATP, 10 mM NaHCO₃⁻) supplemented with 182 μM L-[1-¹⁴C]-threonine (0,05 μCi, 55 Ci/mol) (American Radiolabeled Chemicals). After incubation at 50 °C for 1 hour, macro-molecules were precipitated by addition of 1 mL of Trichloroacetic acid 15% (TCA), and incubated on ice for 1 hour. Precipitated material was applied to glass microfiber filters GF/F (Whatmann), pre-wet with TCA 15%, using a vacuum apparatus (Millipore). Assay tubes were rinsed with 1 mL of TCA 15%, and filters were washed three times with 1 mL of TCA 5% and three times with 1mL of Ethanol 95%. After extensive drying, filters were placed into scintillation vials and soaked with 3 mL of Emulsifier Safe scintillation cocktail (Zinsser analytic). Radioactivity was recorded as average counts per minute (CPM) for 5 min, with Packard Liquid Scintillation Analyzer.

The ATP hydrolysis reactions (conversion into AMP and PPi) were carried out using *Pa*-Sua5 (1,5 μM) and, when indicated, with *Pa*-KEOPS (2 μM) and *Pa*-tRNA^{Lys}(UUU) (5μM) in the same buffer as for the t⁶A assay, supplemented with [α³²-P] ATP (3 Ci/μmol) (Perkins Elmer). The reaction was performed at 50 °C and stopped on ice. Radioactive nucleotides were separated by thin layer chromatography using 0.4 M KH₂PO₄- pH 3.4 as mobile phase. 1 μL of the mix was spotted on a 10 x 10 cm PEI-Cellulose plates (Merck), pre-run in distilled water. Plates were dried and radioactivity was revealed by phosphorimaging using the scanner Typhoon Trio (GE Healthcare). Quantification was carried out with ImageQuantTL software.

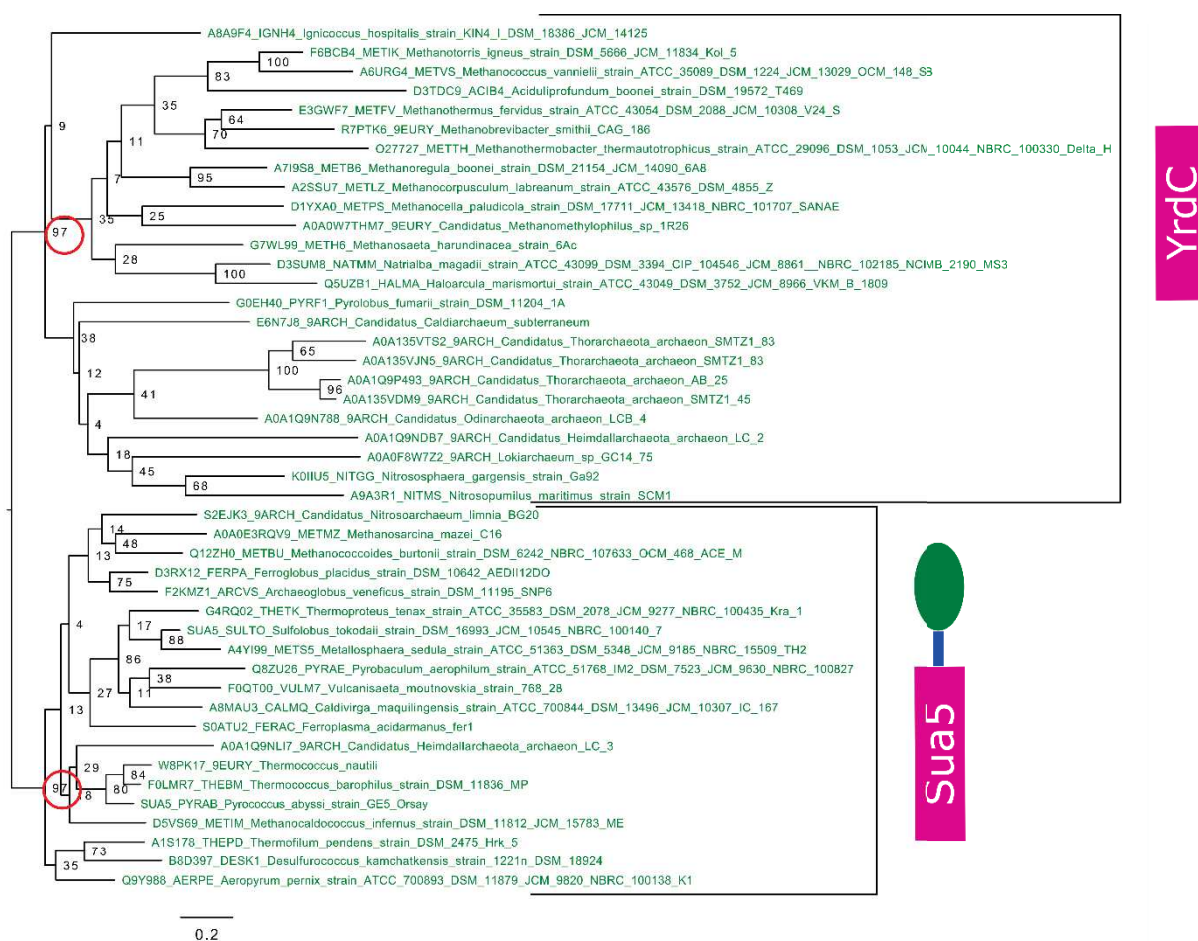


Figure 51. ML phylogeny tree of the YrdC/Sua5 proteins from Archaea

Alignment was done with the sequence for YrdC and YrdC-like domains. The archaeal YrdC variants (up) and the Sua5 variants (down) are segregated with high bootstrap values (indicated by a red circle). The scale-bar represents the average number of substitutions per site. The tree is based on the 176 positions remaining after alignment and trimming. Values at nodes represent support calculated by nonparametric bootstrap (out of 100).

Results and discussion

Phylogenetic analysis of Sua5/YrdC family

For phylogenetic analyses of Sua5/YrdC proteins, a species set was chosen that has been previously vetted for its ability to produce a robust phylogenetic signal using universal proteins (Da Cunha et al., 2017). Sua5/YrdC sequences encoded by the species in this set were obtained from the Uniprot database. These sequences were aligned and used to generate phylogenetic trees. Despite repeated attempts to simplify the dataset by taxonomic domain and/or by protein variant (Sua5 or YrdC), the phylogenetic signal obtained in all cases was very weak (as determined by tree topology, and bootstrap values <60) (Figure 51).

It is notable that Sua5 sequences appear to cluster together in these trees, in contrast to the position of these species in classical phylogenetic analyses. This is likely due to the presence of highly conserved residues in the inter-domain interface which are only found in this variant. In order to analyze the phylogeny of the entire Sua5/YrdC dataset, it is necessary to use residues which can provide a phylogenetic signal across most of the sequences. This therefore excludes the SUA5 domain and linker, and limits analysis to the YrdC-like domain, comprising only 200 amino acids.

Following gap removal and trimming with BMGE, the phylogenetically relevant positions were reduced to 150 amino acids – far too few to provide a robust phylogenetic signal for an evolutionarily ancient protein. Comprising more conserved residues, the Sua5 subset of sequences does result in a stronger phylogeny, however the resulting trees are still weakly supported, and of course do not contain any information on YrdC-encoding species. It is therefore our conclusion that Sua5/YrdC is not an appropriate marker for deep-phylogenetic analyses, due to its small size, and variation in domain composition.

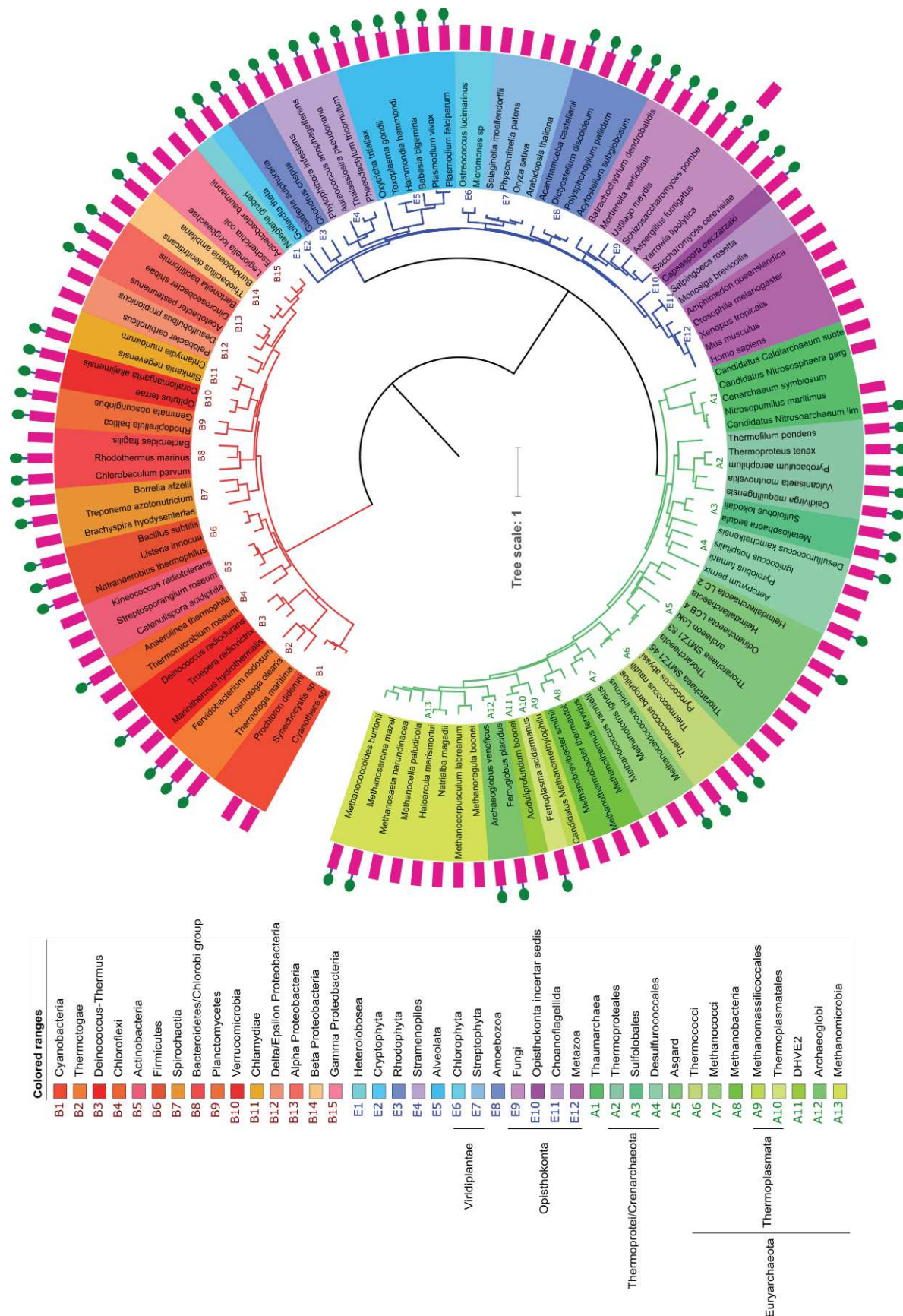


Figure 52. Distribution of Sua5 and YrdC across the three domains of life

The distribution of YrdC and Sua5 proteins was mapped on the universal phylogeny based on RNA polymerase (Da Cunha et al., 2017) using iTOL (Letunic and Bork, 2016). The presence of YrdC is depicted by a magenta ellipse while Sua5 depiction has an additional green rectangle, representing the SUA5 domain.

Distribution of Sua5 and YrdC proteins across the universal tree of life

To obtain a first insight into the distribution of Sua5 and YrdC among different taxonomic groups, we mapped their position on the universal RNA polymerase tree (Da Cunha et al., 2017) (Figure 52). For each taxonomic group in this tree we extracted all available Sua5/YrdC sequences (from Uniprot) and determined the percent of species using an Sua5 variant (Table 7). This analysis confirmed that the only organisms missing Sua5/YrdC are symbionts with greatly reduced genomes, such as the bacterium *Prochloron didemni* and the archaeon *Nanoarchaeum equitans*, or organisms whose genomes are incomplete.

Data further show that all Eukarya carry a Sua5 variant except Metazoa (noted E12 group in Figure 52) and the Choanoflagellida (E11 group in Figure 52) that both belong to Opisthokonta and the Viridiplantae (E6 and E7 group in Figure 52 respectively). A manual inspection of the sequences confirms that all current genomes of Viridiplantae species encode YrdC, except in the Chlorophyceae, a class in the Chlorophyta phylum (green algae) (E6 group in Figure 52) which carry Sua5 proteins.

BLASTP search showed that the Chlorophyceae sequences are most similar to Sua5 sequences found in Fungi (E9 group in Figure 52), the neighboring branch of Viridiplantae in the RNA polymerase tree, e.g. *Spizellomyces punctatus* (64% similarity, E-value 6×10^{-122}), *Parasitella parasitica* (65,3% similarity, E-value $8,9 \times 10^{-122}$) and *Rhizopus microsporus* (65,4% similarity, E-value $1,8 \times 10^{-120}$). To test the origin of the YrdC variant from Viridiplantae, we used BLASTP approaches which showed that the sequences within Viridiplantae are more similar to those of eukaryotes than to the one from Cyanobacteria (data not shown) suggesting it was probably not acquired during chloroplast endosymbiosis.

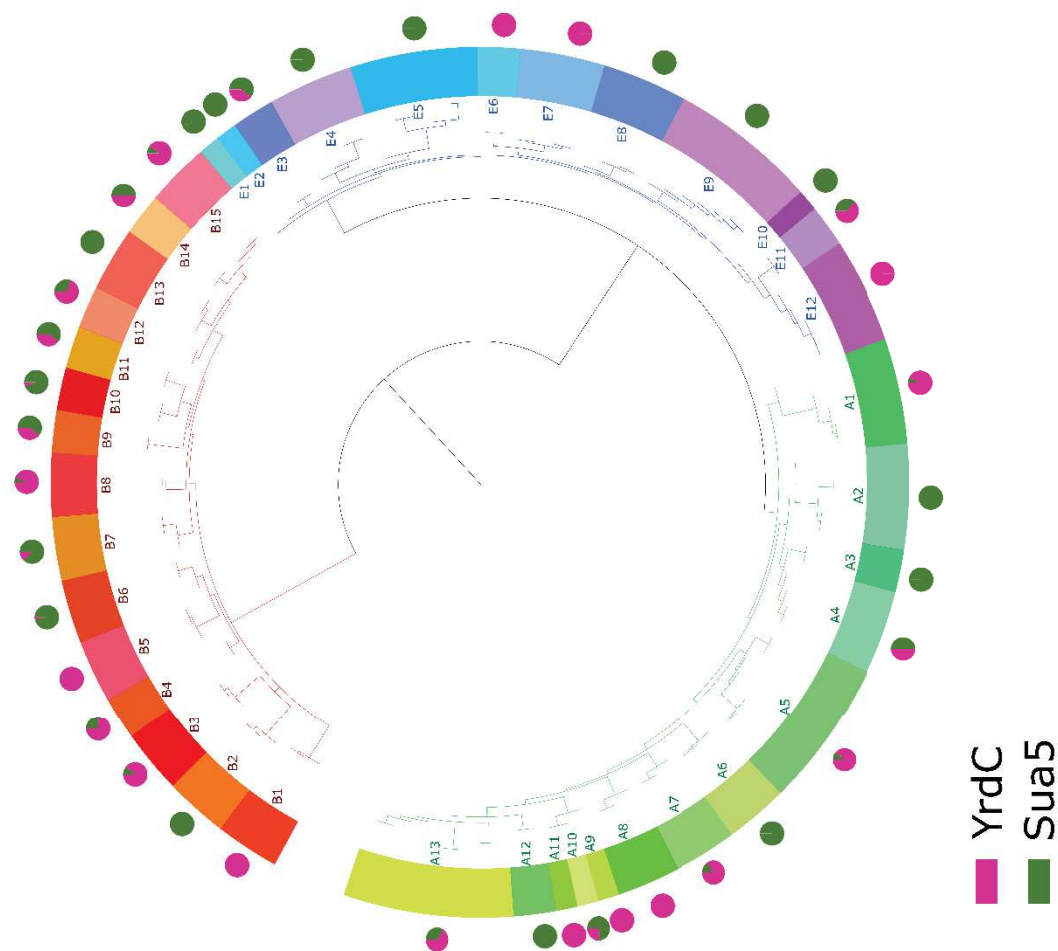
B

Group	Sub-group	Nb of species	% Sua5-users
B1 Cyanobacteria		198	0
B2 Thermotogae		40	100
B3 Deinococcus-Thermus		35	11
	Anaerolineae	18	22
	Arcticatenaceae	2	0
	Chloroflexia	8	75
	Dehalococcidia	9	0
	Korarchaeum	3	100
	Thermomicrobia	3	0
	Unclassified	58	22
	Total:	101	26
	Acidimicrobia	9	44
	Actinobacteria	1528	1
	Coriobacteria	47	0
	Rubrobacter	1	0
	Solirubrobacterales	3	0
	Unclassified	33	6
	Total:	1621	1.4
	Bacilli	962	100
	Clostridia	741	99
	environmental samples	39	15
	Erysipelotrichaceae	38	13.15
	Limnochorda pilosa	1	100
	Negativicutes	79	100
	Tissierella	41	100
	unclassified Firmicutes sensu strict	28	96
	Total:	1929	96.5
	Brachyspira	8	0
	Leptospiraceae	29	100
	Spirochaetales	54	7.5
	Total:	91	87
	Bacteroidetes	916	3.3
	Chlorobi	18	100
	Ignatiibacteriia	21	38
	Total:	955	6
	Planctomycetes	70	61
	Verrucomicrobia	51	94
	Chlamydiae	28	60
	Delta Proteobacteria	146	37
	Epsilon Proteobacteria	31	0
	Total:	177	30
	B13 Alpha Proteobacteria	ND	100
	B14 Beta Proteobacteria	ND	50
	B15 Gamma Proteobacteria	ND	10
	E1 Heterobiosae	1	100
	E2 Cryptophyta	1	100
	E3 Rhodophyta	3	67
	E4 Stramenopiles	24	100
	E5 Alveolata	49	100
	E6 Chlorophyta	4	100
	E7 Streptophyta	7	0
	E8 Archaezoa	82	0
	E9 Fungi	8	100
	E10 Opisthokonta Incertae sedis	518	100
	E11 Choanoflagellida	2	100
	E12 Metazoa	2	0
	A1 Thaumarchaea	64	6.25
	A2 Thermoproterales	27	100
	A3 Sulfolobales	11	100
	A4 Desulfurococcales	12	67
	A5 Asgard	4	0
	A6 Thermococci	9	11
	A7 Methanococci	30	100
	A8 Methanobacteria	13	15
	A9 Methanomassiliococcales	33	0
	A10 Thermoplasmatales	12	0
	A11 DHWE2	23	74
	A12 Archaeoglobi	2	100
	A13 Methanomicrobia	8	100
	Methanosarcinaceae	34	100
	Others	65	0

A

Table 7. Distribution of Sua5-users across the three domains of life

(A) For all the groups depicted in the Figure 52, the total number of species and the percent of them using Sua5 are indicated. For practical reasons, the detailed analysis of Alpha, Beta and Gamma Proteobacteria was done by counting the total Sua5 genes identified in Uniprot without limiting the analysis to one sequence per species. (B) For each taxonomic group (according to Taxonomy NCBI database), the pie chart corresponding to the percent of Sua5 (in green) and YrdC (in pink) are mapped onto the tree.



The only representative members of the classes Heterolobosea and Cryptophyta, respectively *Naegleria gruberi* and *Guillardia theta*, both have Sua5 (E1 and E2 group in [Figure 52](#) respectively). All current Stramenopiles, Alveolata and Amoebozoa organisms also have Sua5 (E4, E5, and E8 group in [Figure 52](#), respectively). Among the Rhodophyta (E3 group in [Figure 52](#)), *Galdieria sulphuria* and *Chondrus crispus* have Sua5 while *Cyanidioschyzon merolae* (not shown) has a YrdC protein (Uniprot: M1VLY8) whose closest homolog are Viridiplantae such as *Klebsormidium nitens* (55% similarity, E-value 3.10^{-46}).

All the current species belonging to the Fungi phylum encode a Sua5 variant. Curiously, all the leotiomyceta fungi such as the fungus *Aspergillus fumigatus* have a second gene encoding a YrdC protein with a complete KxR/SxN motif (Uniprot: A0A0J5PIE0). This protein is most similar to homologs within Actinobacteria, such as *Streptomyces sp. PRh5* (68% similarity, E-value 6×10^{-70}), and Proteobacteria, such as *Candidimonas nitroreducens* (66% similarity, E-value 8×10^{-64}) suggesting a horizontal gene transfer (HGT) between a bacterium and an ancestral leotiomyceta fungus. Altogether, the data show that major groups of eukaryotes carry either Sua5 or YrdC variant indicating that the genes encoding these proteins, were vertically inherited from the common ancestor of each group. The few exceptions to this rule likely correspond to gene acquisitions via horizontal transfer.

In contrast, the distribution in prokaryotes is much patchier with more groups using one variant or the other. This patchy distribution is globally observed in Archaea, for example, in the Euryarchaeota phylum, the current genomes in the classes Archaeoglobi, Thermococci (A12 and A6 group in [Figure 52](#) respectively) and Hadesarchaea (not shown in [Figure 52](#)) encode exclusively Sua5. In the Thermoplasmata class (A10 group in [Figure 52](#)), most organisms also encode Sua5 except for some species of Thermoplasmatales such as *Thermoplasmatales archaeon SG8-52-1* (not shown in [Figure 52](#), Uniprot: A0A151ECH4), recently identified by metagenomics analyses (Zaremba-Niedzwiedzka et al., 2017).

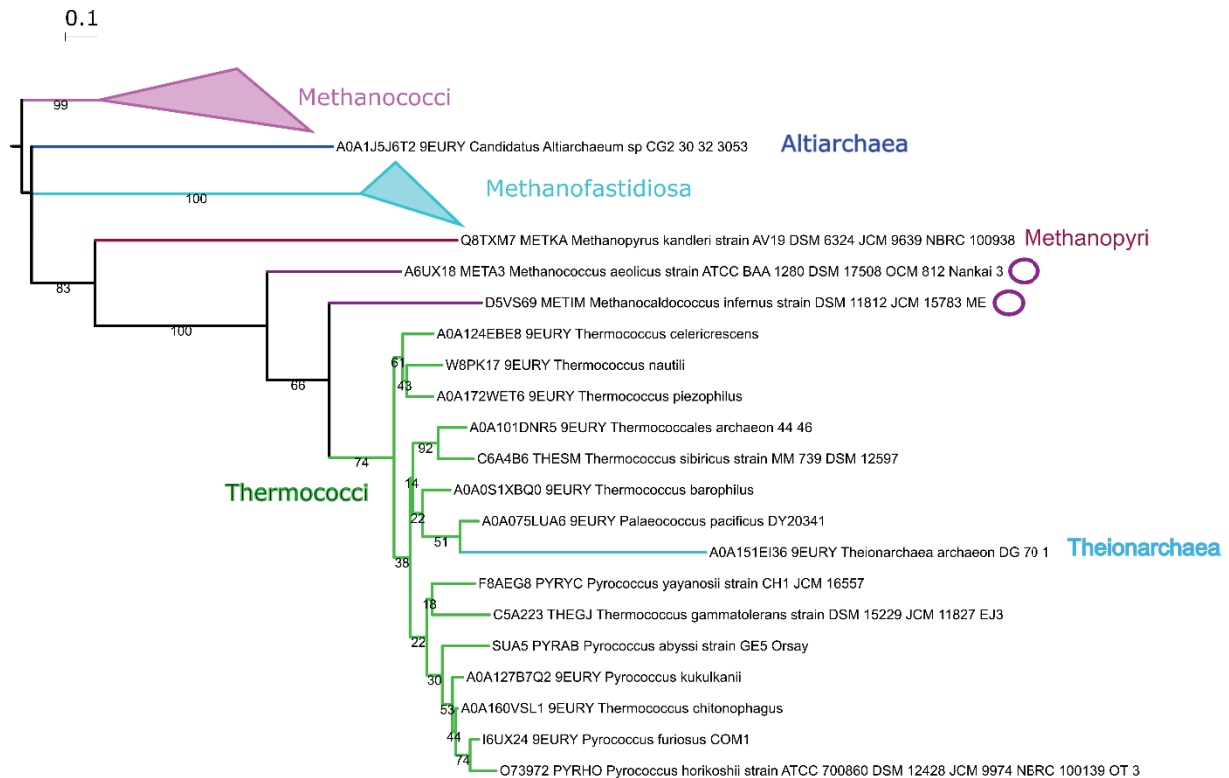


Figure 53: ML phylogeny tree of the YrdC/Sua5 proteins from Methanococcales and Thermococcales and their close relatives

The tree is based on the 194 positions remaining after alignment and trimming. In this tree, branches corresponding to Methanococcales are indicated in pink, the Altiarchaea is indicated in dark-blue, Methanopyri is indicated in dark-red. The members of the Super class Acherontia (Adam et al., 2017) that contains the Methanofastidiosa, the Thermococci, and the Theionarchaea are indicated in light-blue, green and blue respectively. The Methanococcales sequences corresponding to putative ancient HGT transfer are indicated by purple circle, the Sua5 found in the genome of the Theioarchaea could correspond to recent HGT or contamination. The scale-bar represents the average number of substitutions per site. Values at nodes represent support calculated by nonparametric bootstrap (out of 100).

The Aciduliprofundales (DHVE-2) (A10 group in [Figure 52](#)) all encode a YrdC protein. Similarly, all members of the Methanomicrobia class (A13 group in [Figure 52](#)) encode YrdC except the Methanosarcinaceae family (represented by *Methanococcoides burtonii* and *Methanosarcina mazei*). Furthermore, all Methanococcales (A7 group in [Figure 52](#)) encode YrdC as well with the exception of *Methanocaldococcus infernus* (Uniprot: D5VS69) and *Methanococcus aeolicus* (Uniprot: A6UX18). The sequences of these two Sua5 proteins are most similar to homologs from Thermococcales ([Figure 53](#)) and the bacterial candidatus phylum Kryptonia.

The Sua5/YrdC family could not be identified in 8 out of the 11 genomes available for organisms classified within the Asgard group (A5 group in [Figure 52](#)), namely Thorarchaeota, Odinarchaeota, Lokiarchaeota and Heimdellarchaeota, most likely because their genomes are incomplete. Two of them encode a YrdC variant while *Heimdallarchaeota LC-3* encodes for Sua5 (Uniprot: A0A1Q9NLI7). The latter protein is most similar to those found in Thermococci (65% sequence similarity and E-value 2×10^{-105} with *Pyrococcus kukulkanii*). The Crenarchaeota/Thermoprotei phylum mostly uses Sua5. The YrdC-encoding species are found exclusively in the genera *Staphylothermus* and *Ignicoccus* from the Desulfurococcaceae family (A4) as well as in the Pyrodictiaceae family (*Pyrolobus fumarii*). In the Thaumarchaeota phylum (A1), members of the Sua5/YrdC family could be identified in 64 species and only 4 (like *Candidatus Nitrosoarchaeum limnia*, Uniprot: S2EJK3) encode a Sua5 protein with sequence similarity to the Thermococci version. This suggests that Thaumarchaeota are mostly YrdC-users.

In Bacteria, Cyanobacteria and Thermotogae encode YrdC and Sua5 respectively (B1 and B2 group in [Figure 52](#) respectively). Most members of the Deinococcus-Thermus group (B3 group in [Figure 52](#)) have YrdC except *Truepera radiovictrix* (Uniprot: D7CVC7) and the bacteria from the *Meiothermus* genus that encode a Sua5. Either variant can be found in the Chloroflexi (B4 group in [Figure 52](#)) but out of the hundred genomes encoding Sua5/YrdC in this phylum more than half are not yet classified, making it impossible to draw a clear idea of the phylogenetic context of Sua5/YrdC distribution in Chloroflexi.

The Actinobacteria phylum is mostly YrdC specific (B5 group in [Figure 52](#)), the only exceptions are found in incomplete genomes yet to be classified, whereas Firmicutes species (B6 group in [Figure 52](#)) are mostly all Sua5 users. In the Spirochaetia phylum only the Brachyspora class use YrdC (B7 group in [Figure 52](#)), with almost all other spirochaetes encoding Sua5. In the Bacteroidetes/Chlorobi group (B8 group in [Figure 52](#)), Bacteroidia, Flavobacteriia and Sphingobacteriia as well as most of unclassified Bacteroidetes use a YrdC while the few sequenced Rhodothermaceae and Saprospirales use Sua5, and intriguingly, Cytophagales encode both variants. The Planctomycetes and Verrucomicrobia mostly contains Sua5 (B9 and B10 group in [Figure 52](#) respectively). In Chlamydiae (B10 group in [Figure 52](#)), YrdC is found only in some Chlamydia species. Delta/Epsilon Proteobacteria (B12 group in [Figure 52](#)) are all YrdC-users except several unclassified Delta Proteobacteria identified by metagenomics. Alpha Proteobacteria are Sua5-users while Gamma Proteobacteria are mostly YrdC-users (B13 and B15 group in [Figure 52](#) respectively). Beta Proteobacteria are Sua5- or YrdC-users (B14 group in [Figure 52](#)). The number and quality of the genomes available for Proteobacteria in addition to the lack of consensus phylogenies for this phylum make the Sua5/YrdC distribution analysis difficult in these organisms. For a strain by strain distribution of YrdC and Sua5 in about 9200 bacterial genomes, refer to <http://tinyurl.com/t6A-bacteria> (Thiaville et al., 2015a).

The distribution analysis of Sua5/YrdC shows that, in most cases, groups of phylogenetically related organisms encode specifically one of the two variants. Several exceptions to this rule exist suggesting that *sua5* and *yrdC* genes can be exchanged between groups of organisms via horizontal gene transfer. Notably, in all domains of life we find major taxonomic groups that carry Sua5 and those that carry YrdC proteins indicating a complex evolutionary history for this family of proteins.

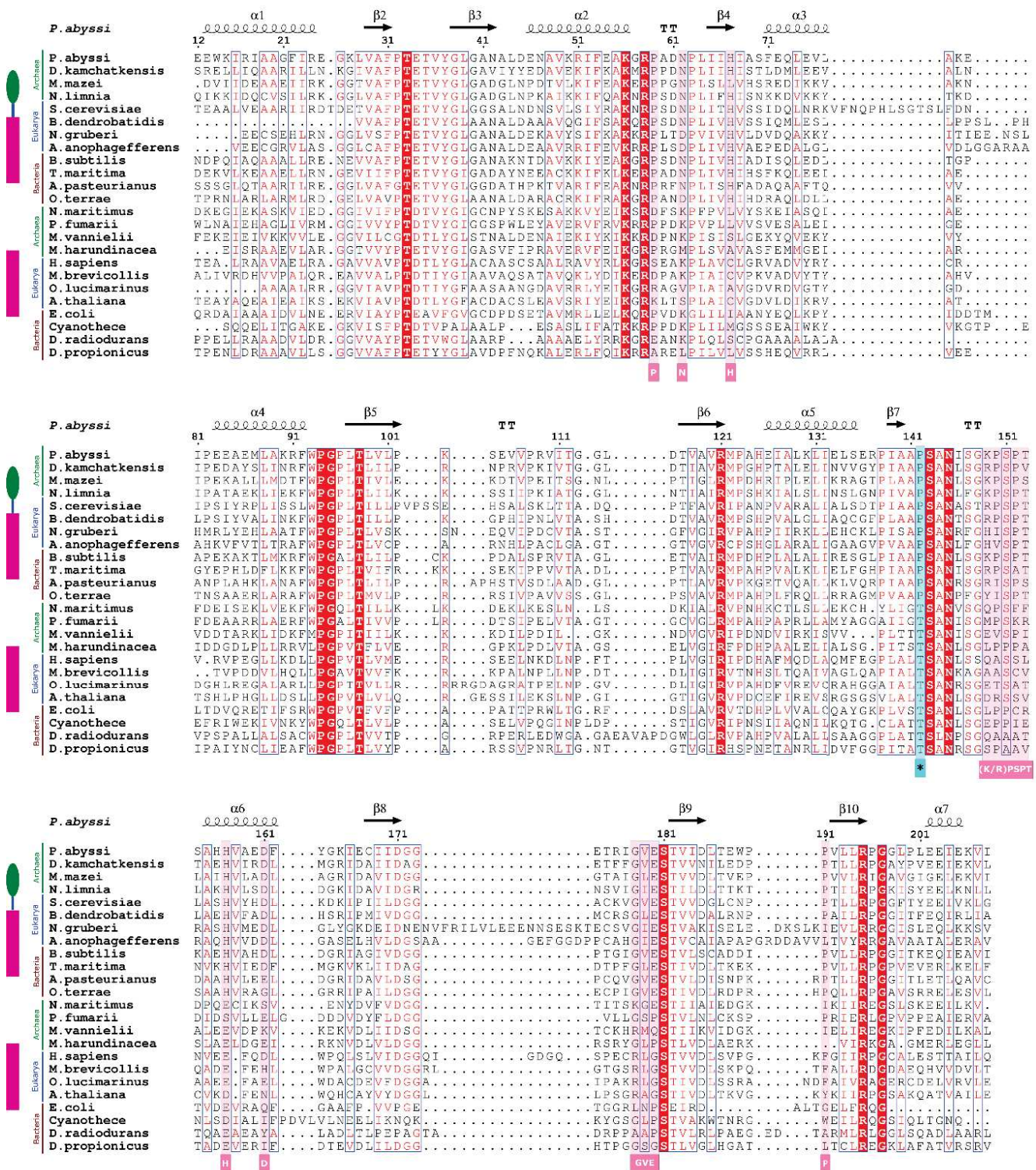


Figure 54. Alignment of the catalytic domain of Sua5 and YrdC

The sequences of the YrdC proteins (lower part) and the catalytic domains of the Sua5 members (upper part) were aligned with MAFFT. For clarity, only four Sua5 sequences and four YrdC sequences per domain were used for this figure. Depiction was done with ESPript3 (Robert and Gouet, 2014). Identical positions are indicated by white text and highlighted in red while similar positions are indicated by red text and framed by a blue rectangle. The corresponding secondary structure from Sua5 of *P. abyssi* are indicated at the top. The residues conserved only in Sua5 are highlighted in pink and the consensus sequence is indicated below. The signature residue is highlighted in cyan and is indicated by an asterisk.

Sequence alignments reveal differences between Sua5 and YrdC catalytic domains

In order to identify sequence patterns specific of either variant, the Sua5 and YrdC sequences used to compute the universal RNA polymerase tree were aligned (Figure 54). About 90 positions are similar in all YrdC domain sequences, in line with a strong conservation of structure and function across the tree of life for this domain. Interestingly, the motifs (Lys/Arg)¹⁴⁹-Pro¹⁵⁰-Ser¹⁵¹-Pro¹⁵²-Thr¹⁵³ (numbered by *Pa*-Sua5) and Gly¹⁷⁸-Hyd¹⁷⁹-Glu¹⁸⁰ (where Hyd corresponds to the hydrophobic residue Val, Ile or Leu) are conserved only in the catalytic domain of Sua5 proteins. The structure of Sua5 proteins (codes PDB: 4E1B and 3AJE) shows that these motifs are found in the inter-domain interface, only found in Sua5 proteins, and interact respectively with Glu¹⁸⁰, His¹⁵⁷ and Arg³⁰¹ for the former and Arg³²⁸ and Thr³⁵ for the latter (Figure 55). Of note, Arg³⁰¹ and Arg³²⁸ were shown to be important for *Pa*-Sua5 activity, likely through their role in stabilizing the domain-domain interaction (Pichard-Kostuch et al., Article I). In addition, the Gly¹⁷⁸-Hyd¹⁷⁹-Glu¹⁸⁰ motif is adjacent to the universally conserved residue Ser¹⁸¹ that is essential for the binding of the substrate L-threonine (Harris et al., 2015; Kuratani et al., 2011; Wan et al., 2013).

Out of the 14 positions conserved only in Sua5, only two are found in the proximity of ligands, Asn⁶² and His⁶⁷. The latter is found in the deep part of the catalytic cavity and forms an H-bond (2.8 Å) with the hydroxyl function of L-threonine in the structures of *St*-Sua5 (Code PDB: 3AJE) (Kuratani et al., 2011) and *Pa*-Sua5 (Pichard-Kostuch et al., Article I). The corresponding position in YrdC is a conserved cysteine in eukaryotes while it is usually a hydrophobic residue in prokaryotes suggesting that these residues have different role in the YrdC proteins. Asn⁶² forms an H-bond with PPi in the structure of *Pa*-Sua5 and with the 3'-OH function of the ribosyl moiety of TC-AMP and AMPPNP in the structures of Sua5 from *S. tokodaii*. This position is often occupied by a lysine in YrdC, whose side chain is longer and charged and may thus have a different role in the active site.

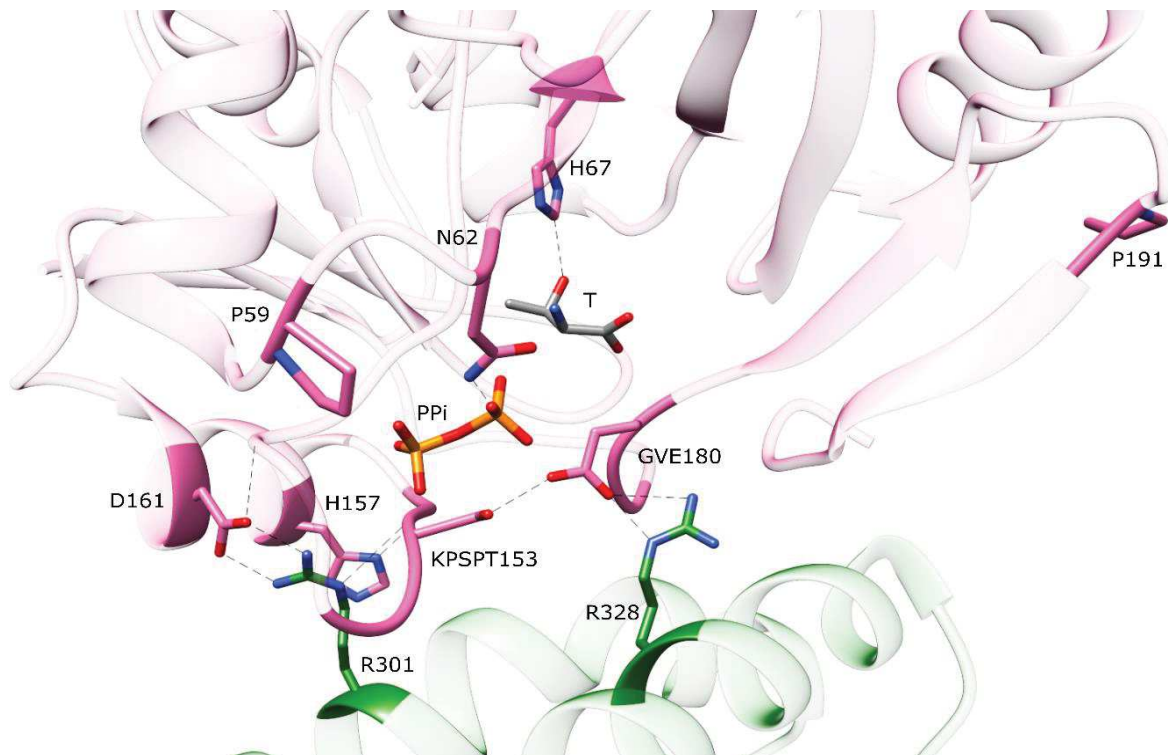


Figure 55. The Sua5-specific residues

Amino acid positions conserved exclusively in Sua5 sequences are represented in the structure of Sua5 from *P. abyssi* bound to PPi and L-threonine. Domain colors are as in [Figure 50](#). Except for the Sua5 specific residues, the main chain is in transparent. Their side chains as well as those for Arg³⁰¹ and Arg³²⁸ from the SUA5 domain are shown. H-bonds are indicated by dashed black lines. The Sua5 motifs are annotated KPSPT153 and GVE180 respectively. For these two motifs, only the side chain of Ser¹⁵¹ and Glu¹⁸⁰ are shown.

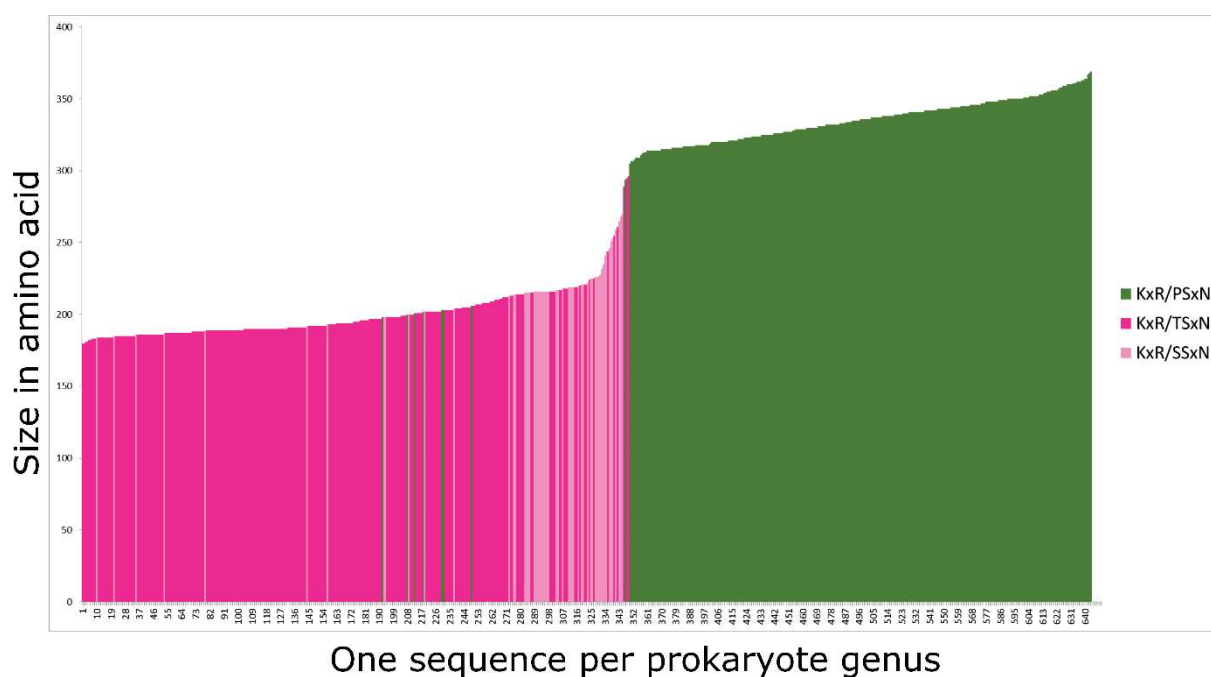


Figure 56. Nature of the signature residue as a function of the size of Sua5/YrdC

YrdC and Sua5 sequences were retrieved from all prokaryote complete genomes with BLASTN on the Complete Genome Data Bank using separately *Pa*-Sua5 and *Ec*-YrdC as query. One UniRef 90 sequence per genus was kept (644 in total) and used to analyze the correlation between the nature of the signature residue and the protein variant. Sequences presenting a signature Proline are in green while these presenting a threonine/serine counterpart are in pink.

Unlike Sua5, the YrdC variant has no variant specific highly conserved motifs. Yet a single peculiar residue could be identified. This highly conserved residue is a Proline in Sua5 (Pro¹⁴³ in *Pa*-Sua5) while it is a Threonine (Thr¹³⁸ in *Ec*-YrdC) in YrdC or, less often and mostly in bacteria, a serine. To check the relevance of this observation, we extracted 26.036 sequences of complete genomes of prokaryotes and used one sequence per genus to map the nature of the signature residue as a function of the size of the protein (Figure 56). This analysis confirms that Sua5 variants all have a Pro¹⁴³ which is in contrast to threonine or serine in the same position in YrdC proteins. The same correlation is observed in eukaryotes (data not shown).

Some exceptions to this correlation do exist, with YrdC containing an equivalent of Pro¹⁴³ found in the bacterial genera *Chlamydia*, *Brachyspira*, *Lactococcus*, *Erysipelothrix* as well as the entire *Aquificae* phylum. No coding sequence corresponding to the SUA5 domain could be found 3' downstream of the *YrdC* gene (where the SUA5 domain would usually be encoded) in these organisms. In addition, these proteins lack the Sua5 specific motifs important for the interdomain interaction (data not shown). These observations suggest that these proteins are genuine exceptions (i.e. YrdC proteins carrying Pro¹⁴³) and do not result from an error in annotation. Taken together the data indicate that, although exceptions exist, Pro¹⁴³ and Thr¹³⁸ are signature residues specific for the YrdC-domain found in Sua5 proteins or YrdC proteins, respectively.

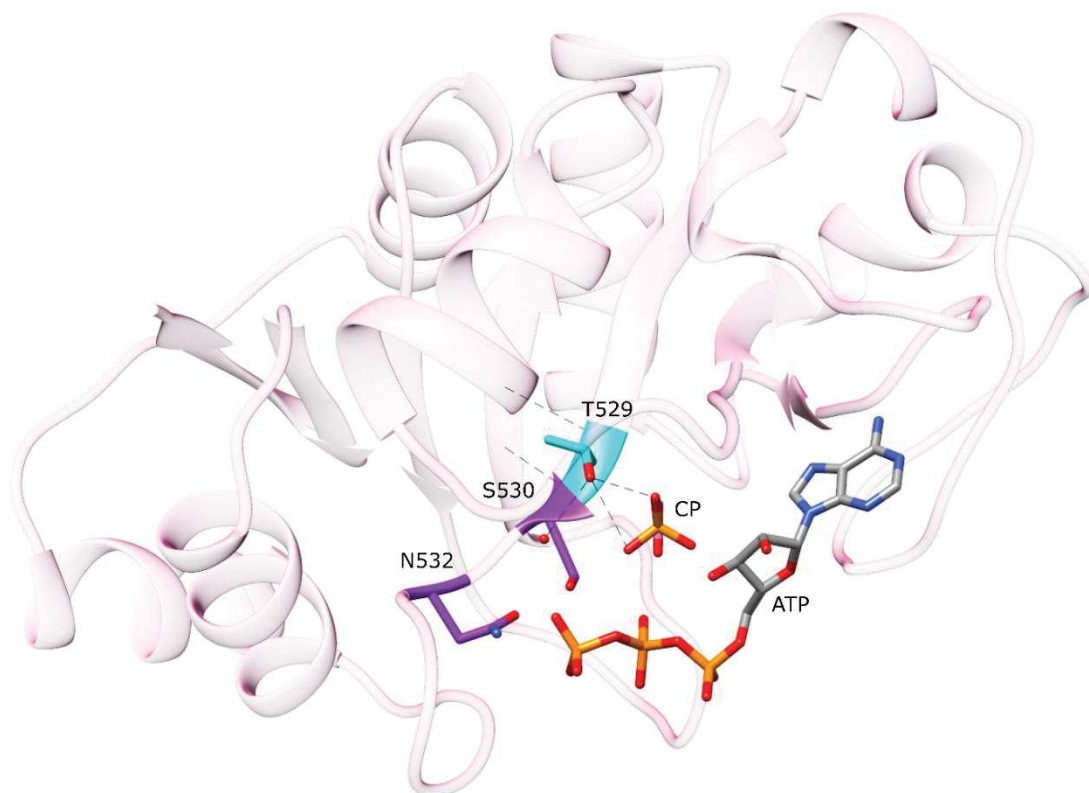


Figure 57. Position of the signature threonine residue in TobZ

The structure of YrdC-like domain of TobZ is shown [residues 372-570], bound to ATP and carbamoyl-phosphate (CP), in the same orientation as Sua5 and YrdC in [Figures 50 and 61](#). Residues Ser⁵³⁰ and Asn⁵³² from the KxR/SxN tetrad are shown in purple. The signature residue Thr⁵²⁹ is shown in cyan. H-bonds are shown by dashed lines.

The signature residue is in the active site

The signature residue is positioned in the $\beta 8$ strand in the active site of *Pa*-Sua5, adjacent to the conserved Ser¹⁴³ of the K⁵⁶xR⁵⁸/S¹⁴³xN¹⁴⁵ motif. The crystallographic structures of *Ec*-YrdC (PDB: 1HRU) (Teplova et al., 2000) and of YrdC from *Streptococcus mutans* (PDB: 3L7V) (Fu et al., 2010) indicate that the signature residues Thr¹³⁸ and Ser¹⁵³ are oriented towards the active site. Unfortunately, there is currently no structure of YrdC in complex with a ligand. However, orthologous domains of YrdC are found in the carbamoylases TobZ and HypF where they catalyze the formation of carbamoyl-adenylate from carbamoyl-phosphate and ATP, thus these proteins can be used to give some indications of the function of the signature residues in true YrdC proteins.

The structure of TobZ from *Streptoalloteichus tenebrarius* bound to both substrates has been resolved (3VEZ) (Parthier et al., 2012). The YrdC-like domain of TobZ [residues 372-570] and *Ec*-YrdC structures share a similar topology. Notably the side-chain of Thr¹³⁸ and Thr⁵²⁹, its equivalent in TobZ, have a similar orientation. In this position, the hydroxyl function of Thr⁵²⁹ is able to form H-bonds with the phospho-moiety of the carbamoyl-phosphate (2.75 Å) (Figure 57). This phospho-moiety occupies the same position as the phospho-moiety of TC-AMP and as the β phosphate of AMPPNP in the structures of *St*-Sua5 (PDB: 4E1B and 3AJE) (Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012). In the structure of HypF from *E. coli* (PDB: 3TTC) (Petkun et al., 2011), the signature residue Thr³²¹ interacts with the β phosphate of the bound ADP.

It is likely then that the presence of a hydroxyl residue at this position indicates a role for this group in the binding of ATP/TC-AMP in YrdC proteins. However, the structures of *Ec*-YrdC resolved by RMN (PDB: 2MX1) (Harris et al., 2015) show a different orientation for the side chain of Thr¹³⁸, away from the active site. In addition, no major chemical shift for this residue was detected by H¹-N¹⁵-HSQC in presence of ATP (Harris et al., 2015).

The conservation of the signature threonine/serine and its strategic position in the active site suggests it is important for the function of YrdC proteins. However, in Sua5 proteins, this residue is a proline, whose side-chain is unable to form H-bonds. Prolines are known to break β strand secondary structure in proteins. Indeed the presence of a proline in Sua5 is correlated with a shorter $\beta 8$ strand (as compared with the corresponding $\beta 6$ strand in *Ec*-YrdC) which is likely to weaken the β -sheet formed by $\beta 3$, $\beta 4$ and $\beta 8$ and thus to influence the flexibility of this region. This property may explain the conservation of this proline in the Sua5 proteins.

The His-Tyr linker motif is conserved among Sua5 proteins and was recently shown to be important for the activity of *Pa*-Sua5, likely by being involved in the binding of PPi/ATP (Pichard-Kostuch et al., Article I). Indeed the conserved histidine forms H-bonds with the corresponding β phospho-moieties in AMPPNP and in PPi in the Sua5 structures. Interestingly, we noticed that the presence of a proline as a signature residue was correlated with the presence of a conserved His-Tyr motif. It is therefore tempting to suggest that simultaneous presence of a signature threonine/serine and a linker histidine, both of which may bind PPi/ATP, could have a negative impact on the protein efficiency. Consequently, this combination is counter selected, resulting in the two variant-specific motifs observed.

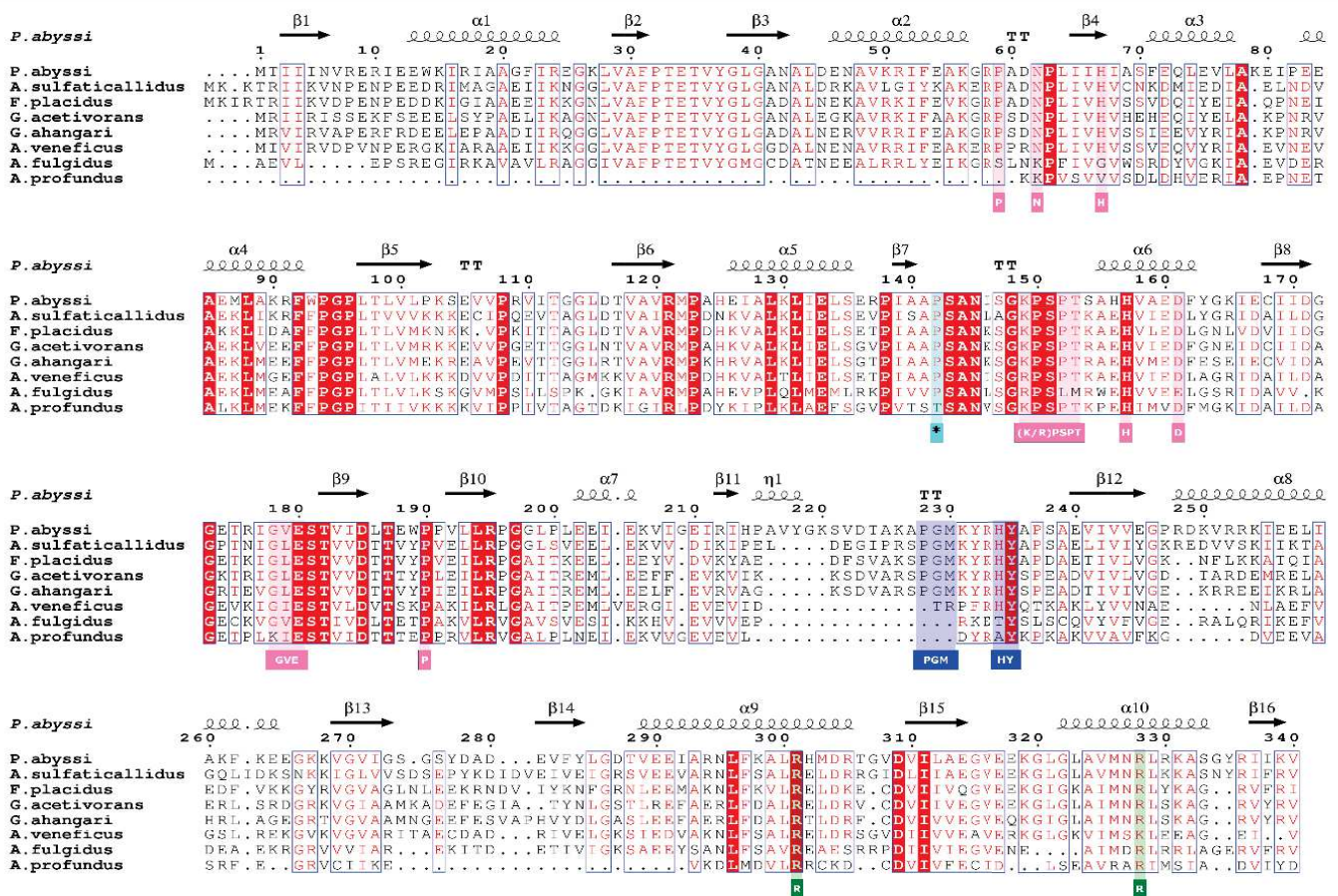


Figure 58. Alignment of Sua5 sequences from the Archaeoglobi

The complete Sua5 sequences are depicted as in [Figure 55](#). The Sua5-specific residues of the YrdC-like domain are highlighted in pink, the conserved motifs of the linker Pro-Gly-Met and His-Tyr are highlighted in blue and the conserved Arginines of the SUA5 domain are highlighted in green. Consensus sequence for these residues are indicated below the alignment.

An atypical Sua5 in Archaeoglobus profundus

The mapping of the signature residue in prokaryotes highlighted the existence in the hyperthermophilic archaeon *Archaeoglobus profundus* of an anomalous Sua5 protein that has a threonine as signature residue. The Archaeoglobi class is currently composed of 7 species, all encoding Sua5 suggesting vertical inheritance of *sua5* gene from their last common ancestor (Figure 58). Four of these sequences have all the fourteen Sua5 specific conserved residues. In contrast, in the Sua5 protein sequence of *Archaeoglobus veneficus*, *Archaeoglobus fulgidus* and *A. profundus* residues which are functionally important in the linker and in the inter-domain interface are missing.

These sequences also contain additional insertions and deletions in the linker and SUA5 domain sequences, especially for the *A. profundus* Sua5 protein (*Ap-Sua5*). Such sequences variations could lead to conformation changes of the SUA5 domain, as suggested by the structure prediction of *Ap-Sua5* (Figure 59). This structural model shows that while the catalytic domain is similar to *Pa-Sua5* structure, the SUA5 domain and the linker are significantly shortened. The linker is likely unable to adopt the closed conformation observed in *Pa-Sua5*. Furthermore, the SUA5 domain contains several atypical features: it is missing several secondary structures including the helix α_{11} ; the helix α_{10} is shortened; and, notably, the conserved arginines important for the stabilization of the interdomain interaction (Arg³⁰¹ and Arg³²⁸ in *Pa-Sua5*) are not interacting with the catalytic domain anymore in the predicted model suggesting the inter-domain interaction in *Ap-Sua5* is not conserved. Another explanation is that it is an artefact of the model.

Additionally, in the Sua5 protein sequence of *A. fulgidus* and *A. profundus* the conserved histidine of the linker is missing. Notably, in the *A. profundus* Sua5 sequence the signature proline residue (indicated by a blue star in Figure 58) is replaced by a threonine (Thr¹⁴¹) which is found typically in YrdC proteins.

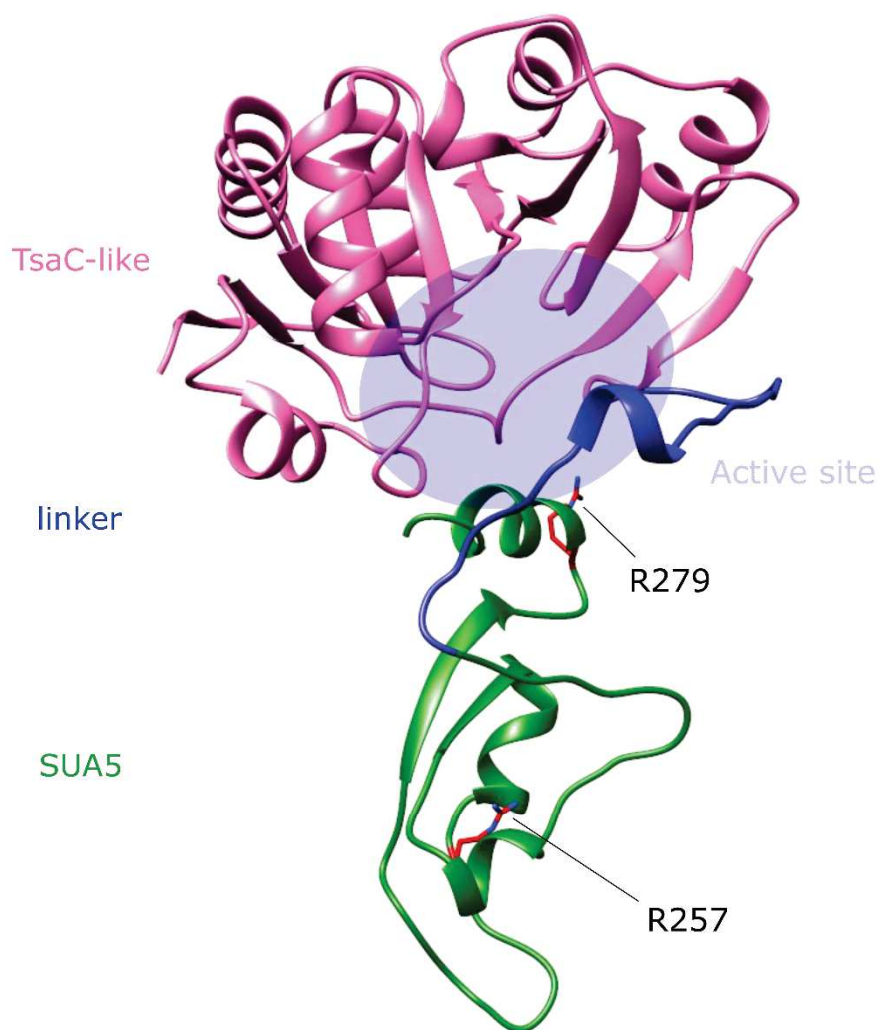


Figure 59. Structure prediction of *Ap-Sua5*^{wt}

The structure of *Ap-Sua5*^{wt} was predicted using the intensive mode of the Phyre2 software and default settings (Kelley et al., 2015) with 289 residues (100%) modelled at >90% accuracy. The catalytic domain (in pink) is highly similar to that of *Pa-Sua5*. In contrast, the linker (in blue) is significantly shorter and the SUA5 domain has lost several of its secondary structures as a result of multiple indels in its sequence. The helix α 11 is missing and the α 10 is shortened. Few conserved residues that are still conserved in α 10 interact with the catalytic domain. Notably, the two arginines (Arg257 and Arg279) found at the inter-domain interface are not interacting with the catalytic domain in the model, despite being conserved in the sequence.

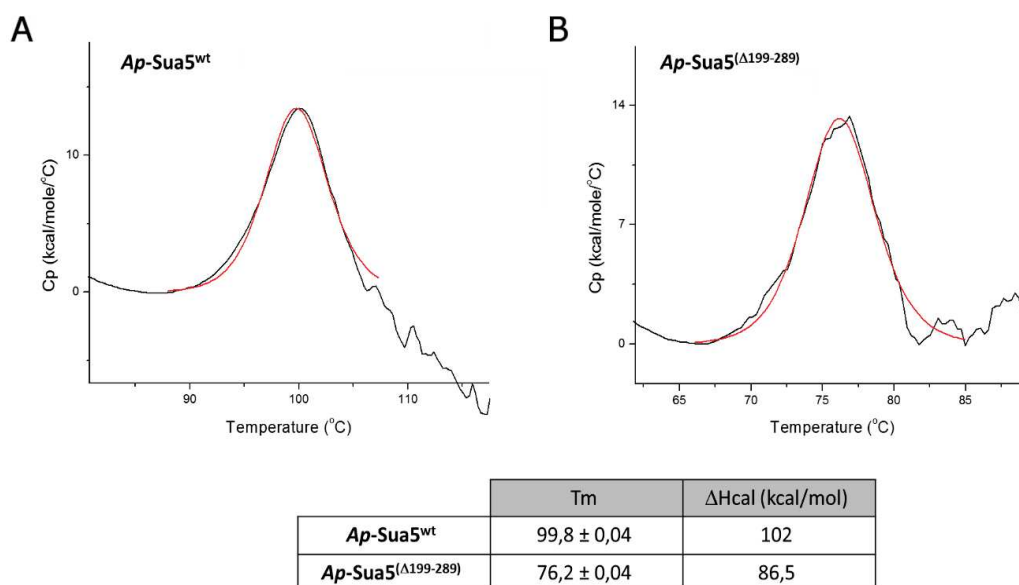


Figure 60. Differential scanning calorimetry analysis of *Ap-Sua5* proteins

(A) *Ap-Sua5*^{wt} and **(B)** *Ap-Sua5*^(Δ199-289) were analyzed at respectively 24 and 18 μM in 50 mM K-HEPES pH 8, 35 mM KCl. The scan was done from 20 °C to 120 °C by incrementing 1 °C every minute.

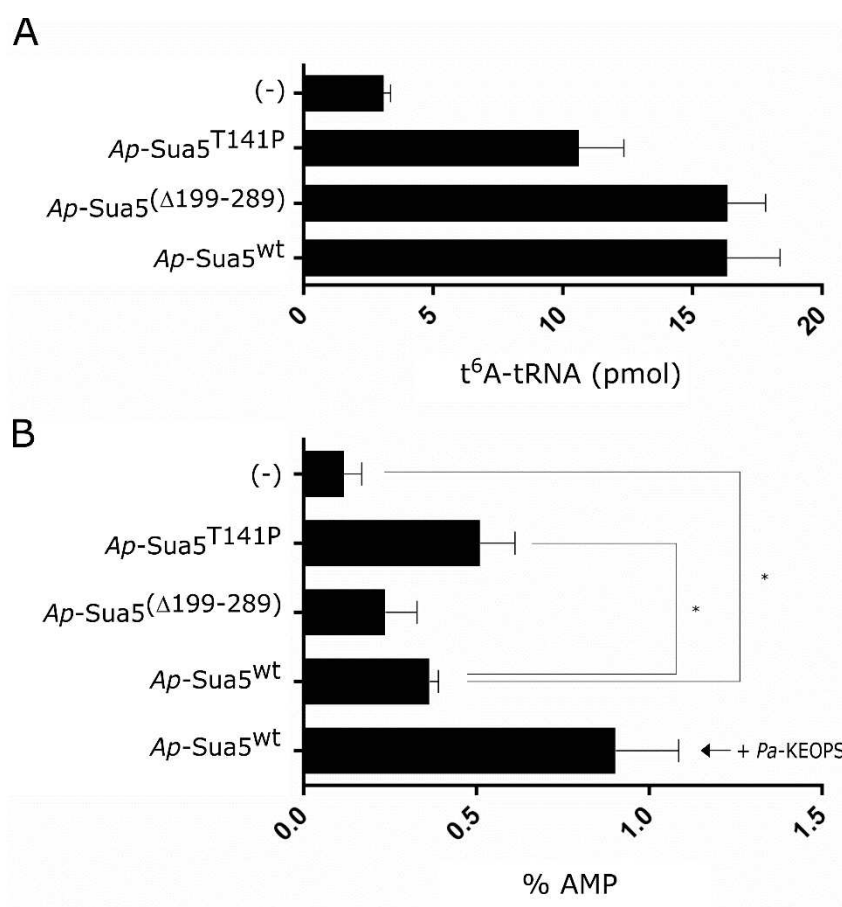


Figure 61. *In vitro* ATPase and t⁶A activities of *Ap-Sua5* proteins

(A) The t⁶A synthesis activity using 2 μM *Ap-Sua5* proteins and KEOPS from *P. abyssi* was determined after 1 h at 50 °C.

(B) The ATPase activity, i.e the formation of AMP from ATP, using 1.5 μM *Ap-Sua5* proteins as determined after 1 h at 50 °C. The data were analyzed using a t test. Significant differences are indicated by * where *p* value < 0.05.

In vitro activity of Sua5 from *A. profundus*

The degenerate sequence of SUA5 domain and presence of threonine signature residue in the active site of the *Ap*-Sua5 protein suggested that SUA5 domain lost its function in this protein. To test this hypothesis we constructed the truncated mutant *Ap*-Sua5(Δ 199-289) corresponding to the YrdC-like domain of the protein. *Ap*-Sua5(Δ 199-289) was predominantly soluble, contrary to the equivalent mutant of *Pa*-Sua5 we previously constructed (Pichard-Kostuch et al, Article I). The DSC analysis produced single relatively narrow and symmetric peaks for both the wild-type and truncated mutant typical for a cooperative unfolding process of compact globular proteins (Figure 60). *Ap*-Sua5^{wt} and *Ap*-Sua5(Δ 199-289) display a T_m of respectively 99.8 °C and 76.2 °C. The significant difference in T_m indicates that the protein is stabilized by the presence of the SUA5 domain.

The t⁶A-tRNA synthesis activity *in vitro* of *Ap*-Sua5^{wt} was tested and compared to the activity obtained with both the *Ap*-Sua5(Δ 199-289) and the point mutant of the signature residue *Ap*-Sua5^{T141P} in presence of the KEOPS complex from *P. abyssi* (*Pa*-KEOPS) (Figure 61A). The wild-type protein and the truncated mutant both produce a similar amount of t⁶A-tRNA. This suggests that the catalytic domain of *Ap*-Sua5 alone is sufficient for the synthesis of the intermediate TC-AMP. The *Ap*-Sua5^{T141P} mutant produced 60% of t⁶A as compared to the wild type protein. This shows that this residue is important but not essential for *Ap*-Sua5 activity.

In addition, we quantified the ATPase activity of the *Ap*-Sua5 proteins (Figure 61B). No significant amount of AMP compared to the negative control in absence of enzyme could be detected for the truncated mutant *Ap*-Sua5(Δ 199-289) while *Ap*-Sua5^{T141P} produced significantly more AMP than the wild-type protein. Interestingly, when compared with *Pa*-Sua5 *in vitro* activities under similar conditions (Pichard-Kostuch et al, Article I), the amounts of t⁶A-tRNA detected for both are similar whereas nearly tenfold less AMP is detected in the ATPase *in vitro* assay for full-length *Ap*-Sua5. Also, as observed for *Pa*-Sua5^{wt}, the production of AMP increases in the presence of *Pa*-KEOPS. These observations are currently difficult to interpret and their significance remains unknown.

ARCHAEA	Nb species	Sua5-users
(HYPER)THERMOPHILIC		
Crenarchaeota unclassified	10	1
Thermoprotei	54	46
Archeoglobi	8	8
Thermococci	30	30
Thermoplasmatales	23	17
DHVE2	2	2
Methanothermus (Methanobacteria)	1	0
Methanopyrus (not shown)	1	0
Korarchaeum (not shown)	1	0
Hadesarchaea (not shown)	2	2
% Sua5 = 80,3		
NON-HYPERTHERMOPHILIC		
Thaumarchaea	64	4
Asgard	9	1
Methanococci	13	2
Methanobacteria (except Methanothermus)	32	0
Methanomassiliicoccales	12	0
Methanomicrobia	99	34
Bathyarchaeota (not shown)	12	5
% Sua5 = 19,1		

Table 8. Distribution of Sua5-users in the (hyper)thermophilic and non-hyperthermophilic archaea

The total number of species per group and the number of species using Sua5 are indicated for hyperthermophilic and thermophilic organisms (upper table) and for the non-thermophilic organisms (lower table).

An integrative evolution model for the Sua5/YrdC family

Despite catalyzing the same reaction, i.e. the synthesis of TC-AMP, and sharing very similar catalytic domains, Sua5 and YrdC are easily distinguishable based on the presence of a supplementary C-terminal domain in the former. Conserved residues in the linker and SUA5 domain together with the structure-function analysis of *Pa*-Sua5 (Pichard-Kostuch et al., Article I) strongly suggest that the YrdC-like domain of Sua5 variants alone is not sufficient for efficient TC-AMP synthesis. Yet, many organisms in the three domains of life rely upon the standalone YrdC variant which lacks a linker and SUA5 domain. How did two variants come to exist and what evolutionary forces determine their distribution across the tree of life remain open questions.

Mapping the Sua5/YrdC protein on the RNA polymerase-based universal phylogenic tree showed that this distribution is not completely random. Indeed, there appears to be specificity for one or the other variant at various taxonomic levels (such as phylum, class or order). The distribution is *a priori* simple in Eukarya as all the Sua5 encoding species are clustered in two taxonomically distinct groups and all the YrdC encoding species in two others. Parsimony analysis, based on eukaryotic phylogeny such as in (Eme et al., 2014), suggests that the LECA (Last Eukaryotic Common ancestor) coded for the two-domain Sua5 variant. On the other hand, the distribution of Sua5/YrdC in prokaryotes is more difficult to interpret as many taxonomic groups using either variants are distributed across the two domains.

To date, no clear link between the presence of YrdC or Sua5 and a specific lifestyle has been determined (Thiaville et al., 2015a). We noticed that in Archaea, 80% of (hyper)thermophilic species use Sua5 while only 20% of non-thermophiles encode this variant (Table 8). It could be suggested that the extra domains of Sua5 variants provides an advantage by stabilizing the protein at high temperature. However, some hyperthermophilic archaea, such as some species of *Methanothermus* (in the Methanobacteria class), *Methanopyrus* and *Korarchaeum* (not shown), indeed use YrdC. Similarly in Bacteria, the thermophilic Aquificae phylum encode YrdC while Thermodesulfobacteria appear to show a mixed distribution of variants. In conclusion, despite clear prevalence of Sua5, this protein does not appear to be a marker of the thermophilic lifestyle.

Horizontal gene transfers (HGT) appear to be responsible for the presence of both variants in several organisms, mostly in prokaryotes but also in some eukaryotes such as in the fungi leotiomyceta. Still, we could observe that most organisms encode only one variant. This is in accordance with (Thiaville et al., 2015a) who found that only 9% of the 9200 prokaryotes strains had both forms. This value is based on analyses at the strain level, with some species being therefore more highly represented than others, indicating that the actual percent of organisms having both variants may be even smaller. This suggests that HGT is not frequent enough to explain the patchy distribution of both Sua5 and YrdC variants across the universal tree and that another mechanism is likely in action. Incidentally, even if HGT may in some extent account for the present repartition of Sua5 and YrdC variants in living organisms it does not explain how the two orthologous variants Sua5 and YrdC came to existence.

We can still consider three main hypothetical scenarios consistent with the results presented: (i) the LUCA encoded both Sua5 and YrdC, and one was lost in some organisms, (ii) the LUCA encoded YrdC, and Sua5 emerged by acquisition of SUA5 domain in several organisms (iii) the LUCA encoded Sua5, and YrdC formed through domain loss in several organisms. The first scenario appears the most unlikely, as almost all extant species have only one homolog of the Sua5/YrdC family, thus suggesting that the LUCA likely also had only one. In addition, two homologous proteins encoded by the same genome and having similar cellular function tend to diverge over time (Gabaldón and Koonin, 2013). This is the observed situation for YciO, a paralog of YrdC found in some organisms, that has lost conserved residues of the KxR/SxN tetrad and is not involved in t⁶A-tRNA synthesis (El Yacoubi et al., 2009; Jia et al., 2002).

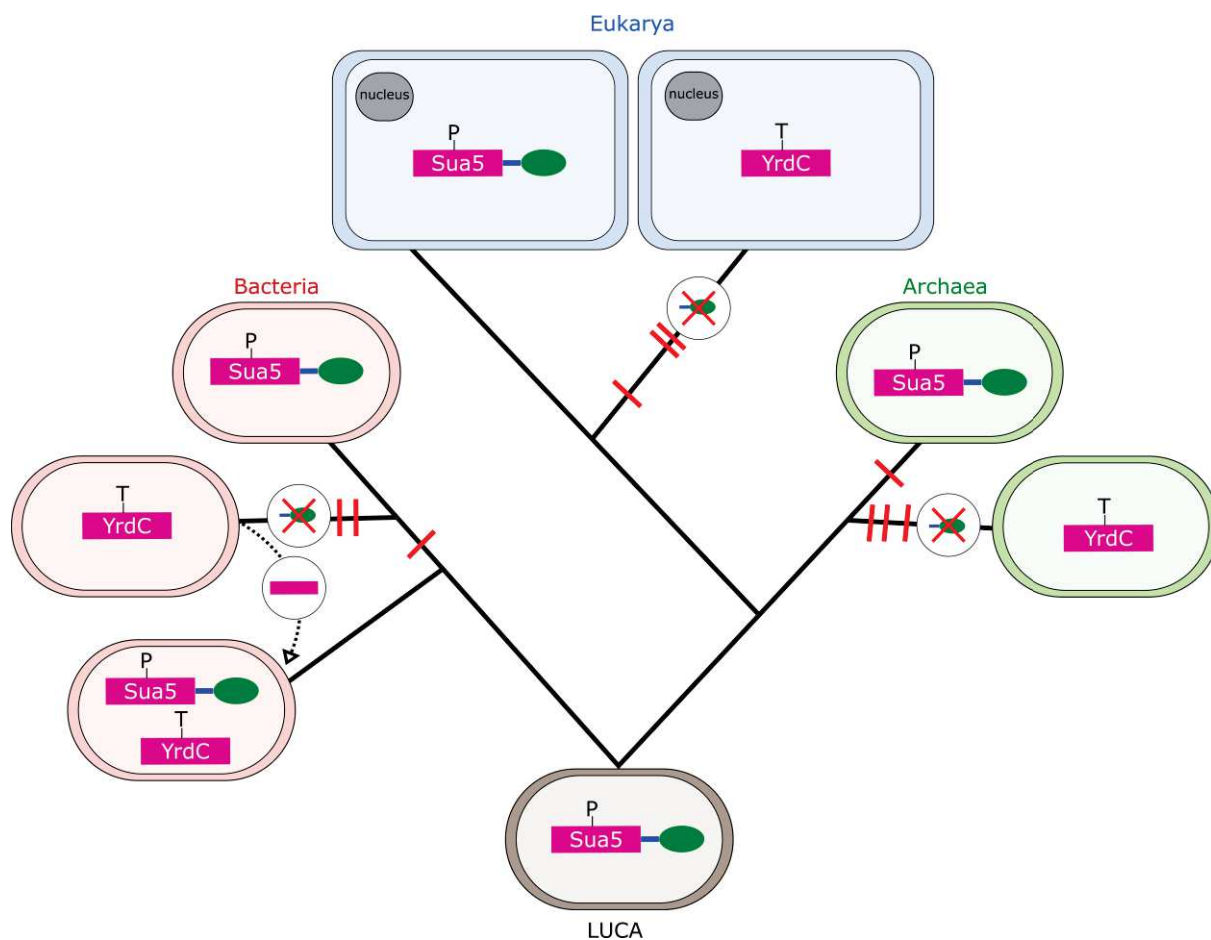


Figure 62. Evolutionary model of the Sua5/YrdC family

In our model the LUCA encoded for a Sua5 protein whose activity depended on the presence of the linker and SUA5 domain and that bore the proline signature residue in its active site. In several lineages throughout the three domains of life, the accumulation of key mutations made the YrdC-like domain autonomous for t⁶A synthesis. The inter-domain interface, no longer being required, started to diverge. Eventually, the appearance of a stop codon induced the expression of the single domain YrdC variant, characterized by the presence of a hydroxyl amino acid as signature residue. This shortened variant was then spread by HGT, especially between prokaryotes, resulting in the presence of both variants in some species. The LUCA is in grey, Archaea in green, Bacteria in red and Eukarya in blue. Point mutations involved in the shortening process are indicated by red line. Their position on the tree is arbitrary. An HGT between two bacteria is indicated by a dotted arrow.

The hypothesis that the LUCA had only one variant, together with the modern existence of two variants that fulfill the same role in cells, implies that either YrdC was derived from Sua5 by shortening or Sua5 from YrdC by lengthening. The latter case would require either that the additional C-terminal extremity of Sua5 was gained by domain fusion or that it emerged *de novo*. Domain fusion is a frequent method for new proteins to arise (Bornberg-Bauer et al., 2010), with evolution famously being described as a tinkerer rather than an inventor (Jacob, 1977). However, the absence of obvious sequence homologs in current databases which may act as a source of the linker and SUA5 domain (data not shown) is in disfavor of the domain fusion hypothesis. It should be noted that the Rossmann fold (adopted by the SUA5 domain) is an ancient and widely distributed structure (Ma et al., 2008) and therefore we cannot exclude a reassignment of an existing domain into a SUA5 domain.

The presence of highly conserved motifs at the interface between the two domains supports the idea that all current Sua5 proteins originated from one common ancestral protein. This implies that the Sua5 variant emerged only once during evolution. In the lengthening scenario, multiple HGT are required to explain the current observed patchy distribution of Sua5 across the tree of life, especially in prokaryotes.

These analyses then favor only one evolutionary hypothesis for the distribution of the Sua5/YrdC family. In this scenario, YrdC emerged from Sua5 through domain loss (Figure 62). One could envisage two pathways by which this could occur: 1. The occurrence of a stop codon causes the shortening of the Sua5 variant, and subsequent mutations occur to produce the modern YrdC protein. However, fragmentation of a domain/protein may disrupt its stability and/or activity (Weiner et al., 2006). This was observed for the truncated mutants of *Pa*-Sua5 (Pichard-Kostuch et al., Article I). Combined with the fact that Sua5/YrdC activity is essential in most organisms, such mutants would be rapidly eliminated by natural selection. 2. Key mutations in the YrdC-like domain, in particular in the active site such as at the signature residue, would limit its dependence on the linker and SUA5 domain presence. The pressure to maintain the integrity of the linker and inter-domain interface would consequently be less strong. Erosion of this region could therefore occur, as observed in the Archaeoglobi class, to yield *in fine* YrdC protein.

The case of *Ap*-Sua5 is one example of such a process since this protein no longer requires the presence of the SUA5 domain to be active and stable. The data suggest that the key (but not only) mutation required for this process is the substitution of signature proline by threonine. Our model fits with the observed distribution of YrdC and Sua5, as it is plausible that domain loss may have occurred several times across the tree of life.

Domain loss is a frequent mechanism for protein evolution in the three domains of life and may even be a helpful tool to decipher the phylogeny of species by analyzing patterns of protein domain gain and loss across the tree of life (Nasir et al., 2014). The shortest known YrdC proteins end in an arginine residue (Arg¹⁹⁵ in *Pa*-Sua5) (Figure 56) that is highly conserved among the whole Sua5/YrdC family. In the structures of *St*- and *Pa*-Sua5, this conserved Arg¹⁹⁵ is found in between the L-threonine and the adenylyl moiety of TC-AMP/AMPPNP binding sites. Of note, the formation of an adenylate derivate from ATP and a carboxylate group, such as in TC-AMP, is well-known to use an arginine, a lysine or a histidine to polarize the α phosphate of ATP to enhance its electrophilicity and favor the condensation reaction with the carboxylate (Schmelz and Naismith, 2009). The highly conserved Arg¹⁹⁵ is therefore likely a determining factor in the limit of shortening possible in these proteins.

Incidentally, the tendency of thermophilic organisms to use Sua5 variant may find a possible explanation through our domain loss model. Intermediate forms between Sua5 and YrdC may cause instability of the protein (Bloom et al., 2004). Indeed, our results show that removal of the remaining SUA5 domain from *Ap*-Sua5 significantly decreased the melting temperature of the protein to one below the optimum growth temperature of the source organism. The high temperature lifestyle may thus be a supplementary constraint for the domain loss process in hyperthermophilic organisms (though clearly not making it impossible).

It is also worth noting that *Homo sapiens* encode YrdC while some parasites, like *Plasmodium yoelii*, encode Sua5. According to our model, Sua5 would be a poor target for Sua5-specific antibiotic development, as shortening toward an YrdC form is possible and may lead to resistance.

Whether there is a constrained order in mutations during the Sua5-YrdC remains to be elucidated. Also, our model is based upon the assumption that the SUA5 domain has no proper function for the protein outside for TC-AMP synthesis. Still, we do not exclude other roles for the SUA5 domain which have yet to be identified. Finally, this study is reflective of the current genomic sequence databases and it may need to be updated in the future, especially with the expanding numbers of sequenced genomes. To this end, metagenomic data non-culture based sequencing methods are a gold mine of information. However, the quantity and quality of some of these data make the precise analysis of the Sua5/YrdC distribution among some groups already impossible. For instance, we were able to identify several partial Sua5/YrdC proteins as well as sequence contaminations during our analysis. If no proper classification and validation are done, similar approaches, for Sua5/YrdC as well as for other enzymes, may become uninterpretable in the future. This may be a particular obstacle for phylogenies quality.

Conclusion

The Sua5/YrdC is an atypical fundamental enzyme family found in all organisms. Despite the significant amount of data obtained in the last ten years, many issues remain unclear e.g. their mechanism for TC-AMP synthesis, the functional differences between both variants and the evolutionary pressures behind their emergence, persistence and distribution. We propose a preliminary model in which the LUCA possessed a Sua5 variant and that shortening towards YrdC in several organisms, followed by HGT explain the existence of both variants and their current patchy distribution. It may just be that in the Sua5/YrdC case, to quote Leonard da Vinci, “simplicity is the ultimate sophistication”.

Acknowledgments

This work has benefited from the platform and expertise of the Macromolecular interactions measurements Platform of I2BC

ARTICLE III

(Projet en cours)

A complete pathway for the universal t⁶A tRNA modification in pandoraviruses

Adeline Pichard-Kostuch, Paul Villain, Lorenzo Carré, Herman van Tilbeurgh, Patrick Forterre, Tamara Basta

Introduction

The enzymatic modification of the canonical A, C, U and G nucleosides is a key RNA maturation process. It involves the addition of more or less complex chemical groups on the base or the ribose part of the nucleosides. The diverse chemical and physical properties provided by these modifications are needed for the cellular functions of RNA (Agris, 2015). More than one hundred different modified nucleosides have been identified among RNA molecules and the greatest number and variety of modifications are found in transfer RNA (tRNA) (Cantara et al., 2011; Machnicka et al., 2013). These modifications, depending on their chemical nature and position in the tRNA, stabilize the tertiary structure of tRNA, modulate its decoding properties and/or regulate the interaction with protein factors and the translational machinery (El Yacoubi et al., 2012; Väre et al., 2017). Several of these modifications are universally found in all three domains of life and were likely present in the last universal common ancestor (LUCA). The N⁶-threonylcarbamoyladenosine (t⁶A) is one of these primordial modified nucleosides. t⁶A is found in position 37, 3' adjacent to the anticodon, of tRNA decoding 5'ANN3' codon (where N = A, C, U or G). The only exception to this rule, not yet understood, is its absence in the initiation tRNA^{Met} in Archaea and Bacteria.

Modifications of the A₃₇ base have been associated with the stabilization of codon-anticodon interactions through base-stacking at the ribosomal decoding site. When bound to mRNA at the A site of the small ribosomal subunit, its planar structure allows t⁶A to stack with the N₃₈ in the anticodon loop and with A₁, the first base of the codon (Murphy et al., 2004). In addition, bulky modifications such as t⁶A prevent by steric hindrance the intraloop pairing between the A₃₇ and the conserved U₃₃ so that the anticodon loop can adopt its canonical U-turn structure (Sundaram et al., 2000).

This open conformation of the loop is required for ribosomal binding and correct pairing between anticodon and codon (Stuart et al., 2000). In some prokaryotes such as *Escherichia coli* (*E. coli*), t⁶A synthesis was shown to be strictly essential. In other prokaryotes, such as *Deinococcus radiodurans*, and in the eukaryote *Saccharomyces cerevisiae* (*S. cerevisiae*) (Thiaville et al., 2015b, 2016), its absence had a severe impact on the organisms growth despite not being *sensu stricto* essential. In *Drosophila melanogaster*, larvae with homozygous mutant alleles of Kae1 exhibit a smaller size compared with wild-type animals and die at larvae stages before pupariating (Rojas-Benitez et al., 2015). Recently, a defect in t⁶A biosynthesis was linked to severe neurological and renal disorders in human called Galloway-Mowat syndrome (Braun et al., 2017; Edvardson et al., 2017).

The maturation of tRNA into t⁶A-tRNA involves two enzymatic reactions governed by two universal enzyme families, Sua5/YrdC and Kae1/YgjD/Qri7 (El Yacoubi et al., 2009, 2011). First, Sua5/YrdC uses ATP, L-threonine and bicarbonate/carbon dioxide to produce a molecule of L-threonyl-carbamoyl-adenylate (TC-AMP) (Lauhon, 2012; Wan et al., 2013). Second, Kae1/YgjD/Qri7 catalyzes the transfer of the threonyl-carbamoyl moiety from the TC-AMP to the substrate tRNA to form t⁶A-tRNA. To be functional in bacteria, YgjD forms a complex called DEZ with two essential proteins, YeaZ and YjeE (Deutsch et al., 2012). In mitochondria, Qri7 works as a homodimer. Finally, in archaea and in the cytoplasm of eukaryotes, Kae1 needs to associate with accessory proteins Bud32, Cgi121 and Pcc1 to form a complex called KEOPS (Perrochia et al., 2013a, 2013b; Srinivasan et al., 2011). In Fungi and Metazoa, an additional fifth subunit was identified, named Gon7 and C14ORF42, respectively (Downey et al., 2006; Wan et al., 2017). The exact role of each subunit of the DEZ and KEOPS complexes as well as the catalytic mechanism behind the t⁶A synthesis are not fully understood. Kae1/YgjD/Qri7 family share a common structure and a conserved active site configuration. The dimers Kae1-Pcc1, YgjD-YeaZ and Qri7-Qri7 form almost identical interfaces suggesting that dimerization might play a key role for the activity of YgjD/Kae1/Qri7 (Mao et al., 2008; Wan et al., 2016; Zhang et al., 2015a).

The modalities behind the transfer of the TC-AMP molecule between the Sua5/YrdC enzymes and the DEZ/KEOPS complexes or Qri7 protein remain unclear. The separation of Qri7 and Sua5 from *S. cerevisiae* by a porous membrane did not suppress t⁶A-tRNA synthesis suggesting that TC-AMP is a diffusible molecule (Wan et al., 2013). However, TC-AMP is an unstable molecule that hydrolyzes spontaneously in solution to produce AMP and an equimolar mixture of L-threonine and its cyclic derivative 5-methyl-2-oxazolidinone-4-carboxylate as final products (Lauhon, 2012). Orthologous domains of Sua5/YrdC and Kae1/YgjD/Qri7 are found in TobZ and HypF, two enzymes that use carbamoyl-phosphate and ATP to form carbamoyl-AMP as an intermediate. Interestingly, the respective structures of TobZ (code PDB: 3VEZ) (Parthier et al., 2012) and HypF (code PDB: 3TTC) (Petkun et al., 2011) show that the organization of the Sua5/YrdC and Kae1/YgjD/Qri7 domains form a channel that protects the carbamoyl-AMP intermediate during the transfer.

Except in some symbionts with very reduced genomes, Sua5/YrdC and YgjD/Kae1/Qri7 are ubiquitously distributed among cellular organisms (Thiaville et al., 2015a). In several species of bacteria, fungi, plants and protists, t⁶A is further modified to the cyclic t⁶A (ct⁶A) via dehydration (Matuszewski et al., 2017; Miyauchi et al., 2013). This additional modification was shown to be involved, together with ms²U₃₄, in efficient decoding of tRNA^{Lys}(UUU) (Miyauchi et al., 2013). In *E. coli*, TcdA catalyzes the ATP-dependent dehydration of t⁶A to form ct⁶A. TcdA belong to the E1-like enzyme superfamily and was shown to be active as a dimer and to be able to bind tRNA (López-Esteva et al., 2015). An homolog of TcdA is found in the only complete genome of an acanthamoeba, *Acanthamoeba castellanii str. Neff*, suggesting it may also use ct⁶A (Clarke et al., 2013).

The pandoraviruses are part of the NucleoCytoplasmic Large DNA Viruses group (NCLDV) and are able to infect *acanthamoeba* species. They are visible using optical microscope as ovoid particles of 1 μm in length and 0.5 μm in diameter. So far three strains have been completely sequenced: *Pandoravirus salinus* (*P. salinus*) (2.5 Mb), *P. dulcis* (1.9 Mb) (Philippe et al., 2013) and *P. inopinatum* (2.2 Mb) (Scheid, 2016). Of note, the difference in size between *P. salinus* and *P. dulcis* is linked to the presence of four repeat-rich regions in *P. salinus* genome. Despite their atypical cellular-like dimensions and the large size of their genomes, the fact that the pandoraviruses lack operational translation apparatus and several core genes that encode essential components of DNA replication implies they still require host functions to replicate. Interestingly, *P. salinus* and *P. dulcis* encode a putative protein corresponding to a fusion between Sua5, YgjD and TcdA (Philippe et al., 2013). This prompted us to suggest that these viruses carry a compact and minimal kit for t⁶A synthesis.

To test this hypothesis, we initiated the characterization of the putative t⁶A synthetase from *P. salinus* (*Ps-t⁶A*). This work may help us to understand how the unstable TC-AMP intermediate is transferred between the two universal families Sua5/YrdC and Kae1/YgjD/Qri7. Incidentally, the study of this atypical enzyme may provide new insight in the pandoraviruses infection process as well as their evolution. Indeed, understanding whether the virus uses this tRNA modification, and how, has yet to be elucidated.

Materials and methods

Structure prediction

The structure of the t⁶A synthetase from *Pandoravirus salinus* was modeled using the program Phyre2 (Kelley et al., 2015). Intensive mode that combines multiple template modeling and simplified *ab initio* folding simulation was used with standard settings. After gathering homologous sequences by scanning the query against protein sequence database with HHblit, secondary structures are predicted using PSIPRED. Both the alignment and secondary structure prediction are combined into a query hidden Markov model. The model is scanned against a database of known structure. The top-scoring alignments from this search are used to construct crude backbone-only models. Finally, loops are modelled and side-chain are placed. Structure depiction and analysis were done using UCSF Chimera (Pettersen et al., 2004).

***In silico* analysis**

BLAST analysis were made using the NCBI server. Protein and nucleotide sequence alignments were made using standard settings of MAFFT (Yamada et al., 2016), manually refined using Seaview and depicted with ESPript (Robert and Gouet, 2014). Plot between nucleotide sequences of *P. salinus* and *P. dulcis* was rendered by MAFFT.

***In vitro* t⁶A activity assay**

The reaction was performed in reaction buffer composed of 50 mM K-HEPES pH 8, 35 mM KCl, 10 mM MgCl₂, 1 mM MnCl₂, 5 mM DTT, 1 mM ATP, 10 mM NaHCO₃⁻, supplemented with 182 μM L-[1-¹⁴C]-threonine (0,05 μCi, 55 Ci/mol) (American Radiolabeled Chemicals). After incubation for 30 min, macro-molecules were precipitated by addition of 1 mL of TCA 15 %, and incubated on ice for 1 hour. Precipitated material was applied to glass microfiber filters GF/F (Whatmann), pre-wet with TCA 15%, using a vacuum apparatus (Millipore). Assay tubes were rinsed with 1 mL of TCA 15%, and filters were washed three times with 1 mL of TCA 5% and three times with 1mL of Ethanol 95%. After extensive drying, filters were placed into scintillation vials and soaked with 3 mL of Emulsifier Safe scintillation cocktail (Zinsser analytic). Radioactivity was recorded as average counts per minute (CPM) during 5 min, with Packard Liquid Scintillation Analyzer, and converted into pmol of t⁶A-tRNA produced.

<i>Pandoravirus salinus</i>		Description	Taxonomy	Max score	Total score	Query cover	E value	Ident	Accession
t⁶A synthetase	YgjD [1-369]	mitochondrial Qri7 <i>Alligator sinensis</i>	Metazoa	216	216	96%	6,00E-63	39%	XP_006036829.1
		YgjD <i>Asticcacaulis</i> sp. AC460	a-proteobacteria	213	213	88%	2,00E-62	44%	WP_023453171.1
		YgjD <i>Sphingomonas</i> sp. PAMC 26605	a-proteobacteria	209	209	92%	2,00E-61	43%	WP_010184392.1
	Sua5 [370-724]	<i>Bacillus</i> sp. SA1-12	firmicutes	275	275	95%	3,00E-87	45%	WP_046588535.1
		<i>Anoxybacillus</i> sp. DT3-1	firmicutes	268	268	95%	2,00E-84	46%	WP_009360973.1
		<i>Fictibacillus</i> sp. FJAT-27399	firmicutes	264	264	95%	1,00E-82	44%	WP_062238491.1
	TcdA [725-979]	<i>Carboxydotherrmus hydrogenoformans</i>	firmicutes	135	135	74%	2,00E-35	40%	WP_011345089.1
		<i>Labilibacter aurantiacus</i>	CFB group bacteria	135	135	89%	2,00E-35	35%	WP_075603859.1
		<i>Fusobacterium</i> sp. oral taxon 203	fusobacteria	129	129	70%	3,00E-33	38%	WP_094242759.1
	bifunctional metallophosphatase/5' -nucleotidase	<i>Azospirillum lipoferum</i>	a-proteobacteria	236	236	98%	2,00E-68	33%	WP_014249418.1
		<i>Inquilinus limosus</i>	a-proteobacteria	224	224	94%	4,00E-64	33%	WP_026870673.1
		<i>Magnetospirillum</i> sp. 64-120	a-proteobacteria	218	218	96%	2,00E-61	33%	OJX81406.1
S-adenosyl-L-homocysteine hydrolase		<i>Bacillaceae bacterium G1</i>	firmicutes	321	321	90%	2,00E-102	41%	OJF17487.1
		<i>Caldicellulosiruptor obsidiansis</i>	firmicutes	303	303	90%	1,00E-95	40%	WP_013290131.1
		<i>Thermoanaerobacter cellulosyticus</i>	firmicutes	302	302	90%	5,00E-95	39%	WP_045165731.1

Table 9. Closest homologs of *Ps*-t⁶A and its neighboring genes

The analysis were done using BLAST with standard settings. Protein/domain used as query are indicated on the left. Three hits within the best ones are shown. The hits for the Cupin-5 genes are not indicated because their E-value is too low.

Results

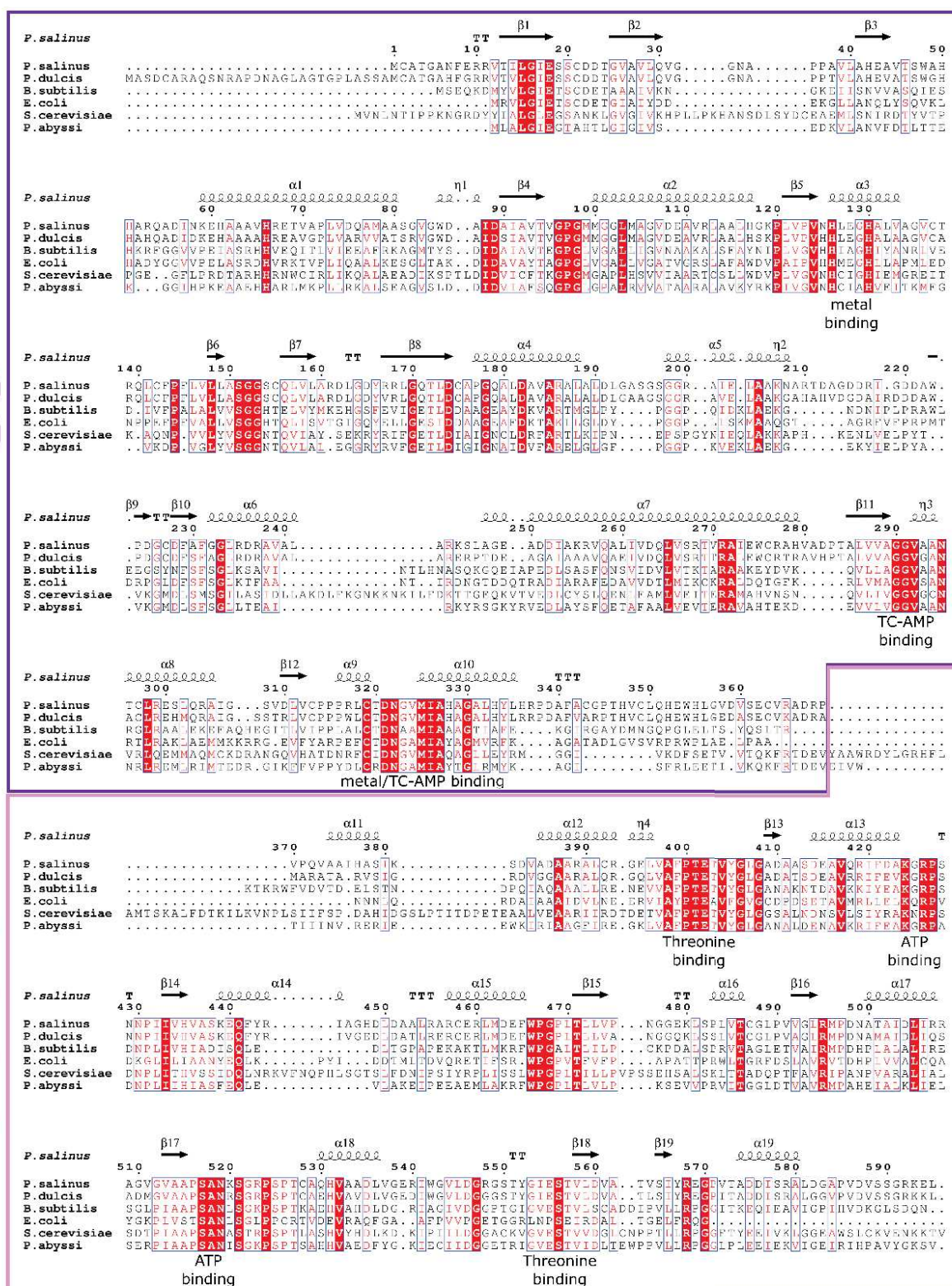
Sequence analysis of the t⁶A synthetase of pandoraviruses

The t⁶A synthetases from *P. salinus* and *P. dulcis* share 79.5 % of sequence identity. This indicates that they are close orthologs sharing a common, and likely recent, ancestor. Multiple sequence alignment with the Sua5/YrdC, Kae1/YgjD and TcdA orthologs from the bacteria *Bacillus subtilis* and *E. coli*, the eukaryote *S. cerevisiae* and the archaeon *Pyrococcus abyssi* reveal significant sequence conservation in the pandoraviruses t⁶A synthetases. Notably the residues known to be involved in ligand binding in Sua5/YrdC and Kae1/YgjD/Qri7 are conserved ([Figure 63](#)), suggesting these enzymes are able to synthesize t⁶A-tRNA. For instance, in Sua5, the KxR/SxN tetrad involved in ATP binding as well as the linker motif PGMxxxHY, shown to be important for TC-AMP synthesis, are present. In the YgjD domain, the HxxxH/D motif involved in metal binding as well as residues involved in nucleotide binding are also conserved. The overall good sequence conservation, 26,5 % identity (50 % similarity) with *Ec*-TcdA sequence, suggests the TcdA domain of pandoraviruses t⁶A synthetases may be active. Of note, the gene region encoding the TcdA domain starts with an AUG codon (Met⁷²⁵ in protein), suggesting that it may be translated separately.

The respective domains extremities of *Ps*-t⁶A were determined using multiple sequence alignments: YgjD [residues 1-369 in *Ps*-t⁶A], Sua5 [370-724], and TcdA [725-979]. In order to identify their probable origin, we made BLAST analysis using the sequence for separated domain as input ([Table 9](#)). Close orthologs of the YgjD domain are the YgjD proteins of alpha-proteobacteria and the Qri7 proteins of eukaryotes. This is not surprising since Qri7 is a mitochondrial ortholog related to YgjD from alpha-proteobacteria (Hecker et al., 2007). Another possibility is that the YgjD domain sequence of *Ps*-t⁶A diverged from YgjD to be functional without accessory proteins, thus becoming closer to the mitochondrial Qri7 that underwent the same process. Closest orthologs of Sua5 domain are found in the bacteria firmicutes which are phylogenetically distant from alpha-proteobacteria. Similarly, the first hits obtained for the TcdA domain correspond to proteins from fusobacteria, firmicutes or CFB group. This indicates different bacterial origins for the three domains composing the pandoraviruses t⁶A synthetases and raises the question of the mechanism leading to such assembly.

YgjD

Sua5



TcdA

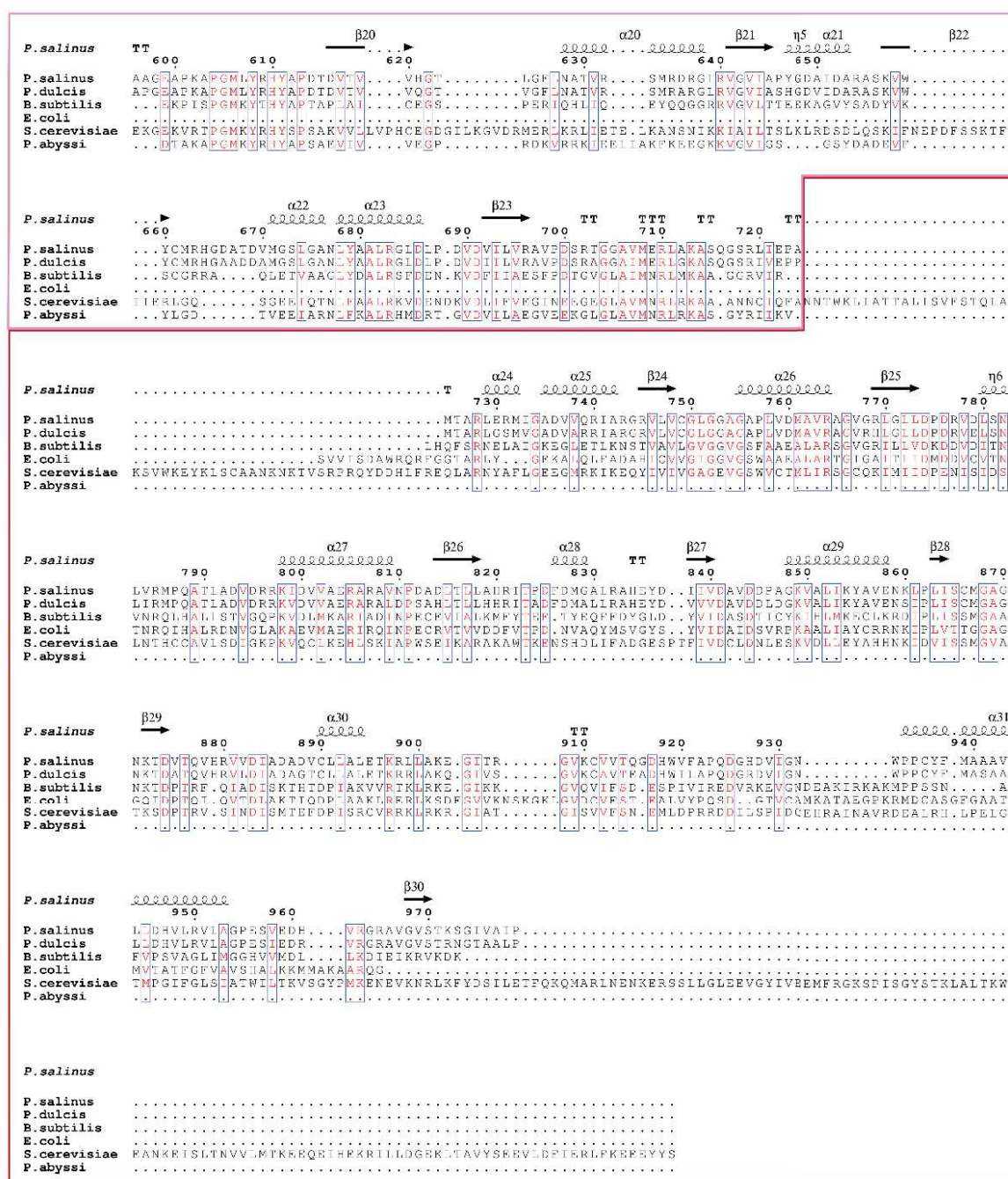


Figure 63. Sequence comparison of the pandoraviruses t⁶A synthetases with Sua5/YrdC, Kae1/YgjD and TcdA sequences from model organisms

The sequences were aligned using MAFFT (Yamada et al., 2016) and depicted using ESPript 3 (Robert and Gouet, 2014). The identical position are highlighted in red while similar positions are written in red. The sequence regions corresponding to YgjD, Sua5 and TcdA domains of pandoraviruses t⁶A synthetase are indicated. For several conserved motifs, their identified role is indicated below the sequence in black.

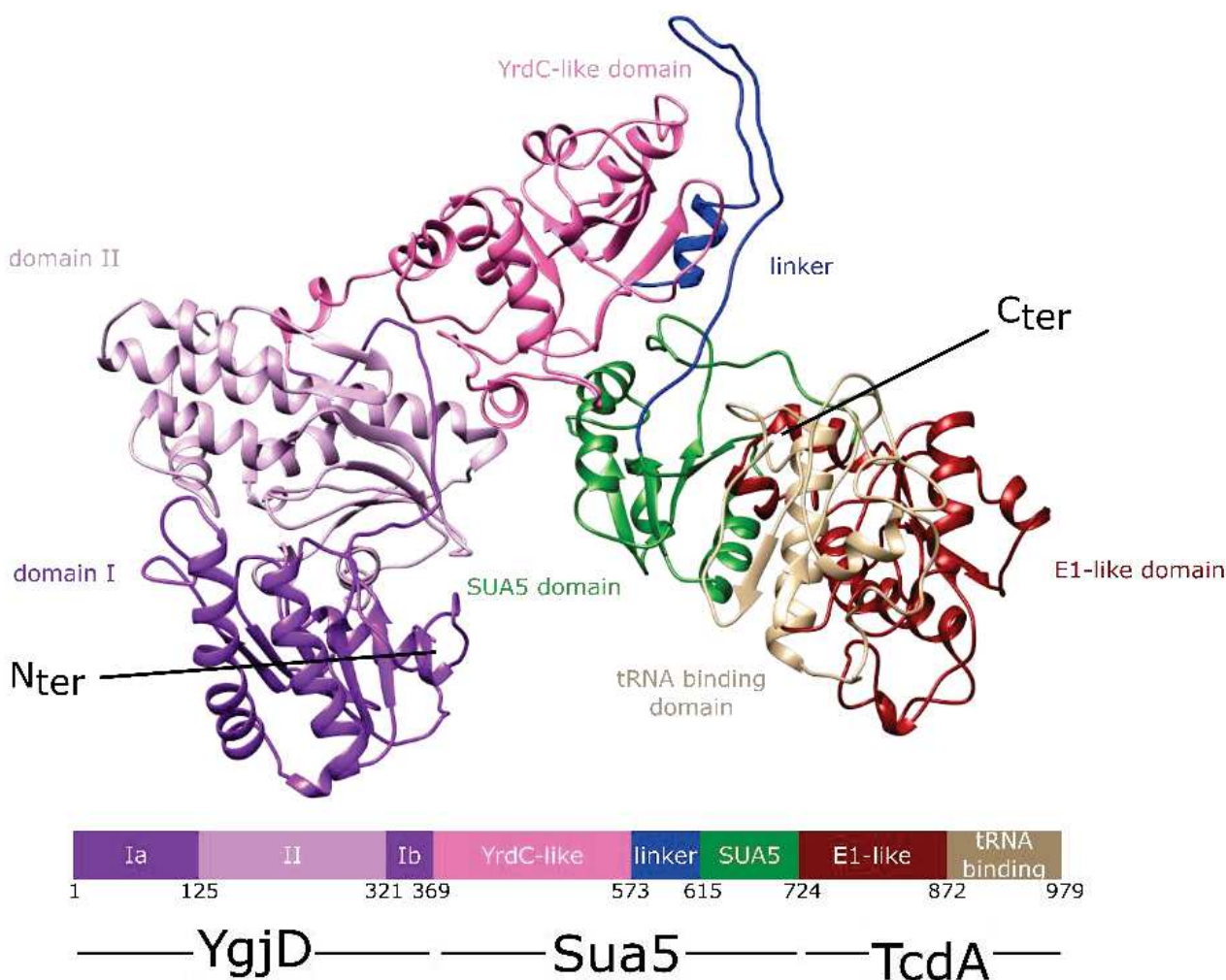


Figure 64. Structure prediction of the t⁶A synthetase from *P. salinus*

The structure of *Ps*-t⁶A was modeled with the program Phyre 2 (Kelley et al., 2015) using the intensive mode in standard settings. The structure was modelled over 920 residues (94%) with a > 90% accuracy. The domains of YgjD (I and II), Sua5 (YrdC-like, linker and SUA5) and TcdA (E1-like and tRNA binding domains) are depicted in different colors. Their corresponding position in the sequence is indicated below.

Structure prediction of *Ps-t⁶A synthetase*

The high sequence conservation of the three domains allowed us to make structure prediction of *Ps-t⁶A* using the software Phyre2 with good confidence (Figure 64). The YgjD domain of the *Ps-t⁶A* synthetase is similar to other known Kae1/YgjD/Qri7 structures with an RMSD of 1.011 Å between 271 pruned atom pairs (3.271 Å across all 328 pairs) when superposed with YgjD from *E. coli* (code PDB: 4WQ5) (Zhang et al., 2015a)). Similarly, the overall structure of Sua5 is close to the one of Sua5 from *S. tokodaii* (code PDB: 4E1B) (Agari et al., 2008; Parthier et al., 2012) with an RMSD of 0.618 Å between 248 pruned atom pairs (5.916 Å across all 328 pairs). However, the C-terminal helices α_{11} and η_3 in Sua5 from *S. tokodaii* (as described in (Kuratani et al., 2011)), is not modeled with good confidence in *Ps-t⁶A* synthetase and form a disordered region in the model.

The N-terminal domain of TcdA, belonging to the E1-like enzyme family and involved in the catalytic reaction, is highly similar the one of TcdA from *E. coli* (*Ec-TcdA*) (code PDB: 4D79) (RMSD between 124 pruned atom pairs is 0.873 Å - across all 144 pairs: 2.428) (López-Esteva et al., 2015). In contrast, the C-terminal domain of TcdA, involved in tRNA binding, is not superposing well (RMSD between 23 pruned atom pairs is 0.653 Å; across all 62 pairs: 11.123) despite an overall good confidence for this region in the model. This might be a genuine specificity of the TcdA domain in *Ps-t⁶A* synthetase or an artefact of the modelling method used.

TcdA is likely unable to dimerize in the *Ps-t⁶A* synthetase model since its last 20 residues as well as the SUA5 sub-domain are clashing with the expected position of the second TcdA monomer, as observed in the structure of *Ec-TcdA*. However, the low confidence in the model for the 20 last residues and for the linker between Sua5 and TcdA, suggesting it may be flexible, indicate that the TcdA domain is susceptible to adopt another conformation and/or position that is compatible with dimerization. The YgjD domain has its sub-domain II oriented towards Sua5. The sub-domain I remains accessible for dimerization with a partner, as required in all Kae1/YgjD/Qri7 orthologs. Interestingly, no homologs of the DEZ complex accessory proteins were found in pandoraviruses genomes. This hints that the pandoravirus t⁶A synthetase functions alone.

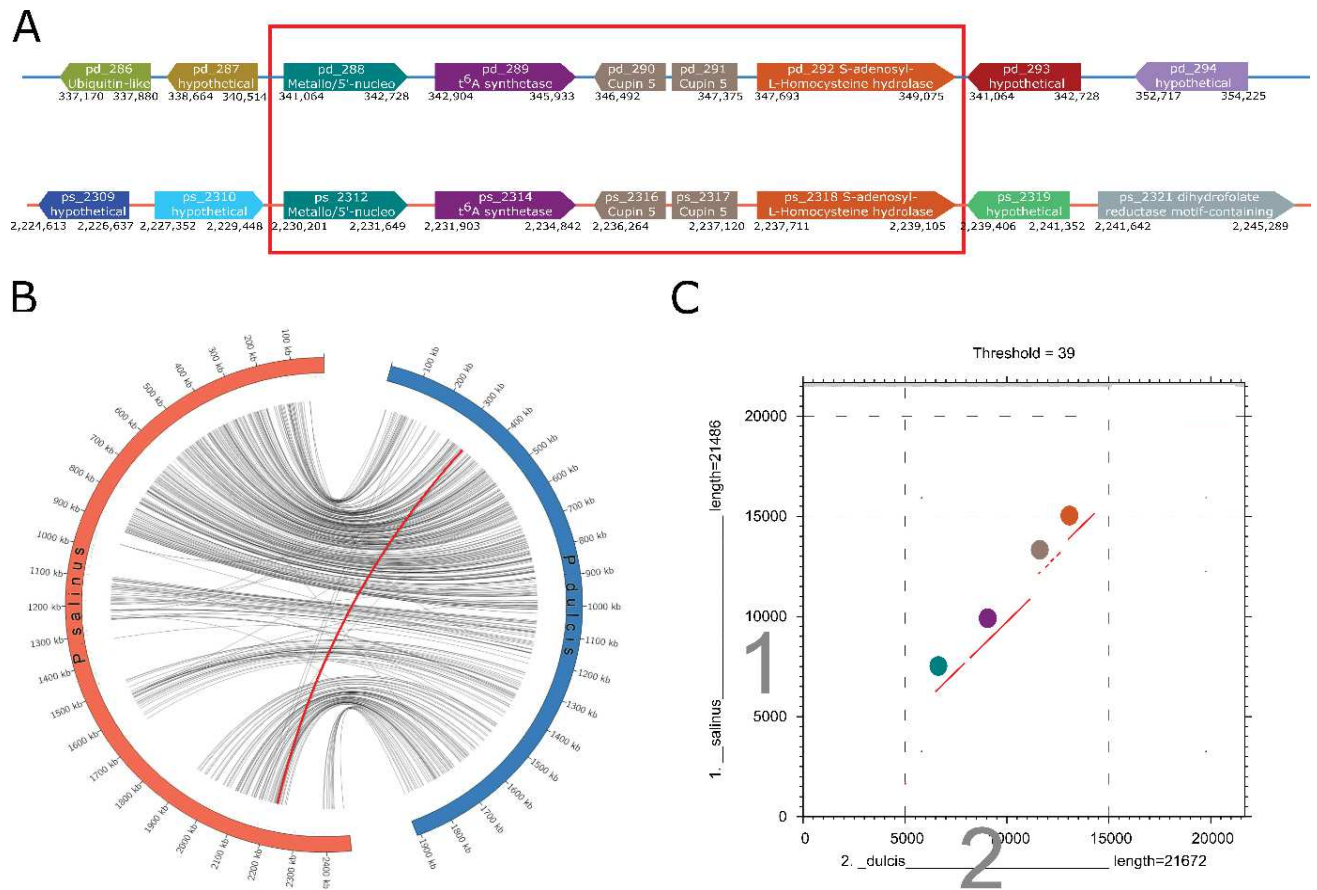


Figure 65. Genomic context of the pandoraviruses t⁶A synthetase genes

(A) Comparison of the insert region between *P. salinus* (down) and *P. dulcis* (up). The insert is framed by a red rectangle. Homologous genes have the same color. The name of the gene and the protein encoded are written in white. Nucleotide positions are indicated below the genes. **(B)** Colinearity of the *P. salinus* (left) and *P. dulcis* (right) genomes. The four repeat-rich regions of the *P. salinus* genome absent from *P. dulcis* are clearly visible, centered at positions 1000 kb, 1400 kb, 1750 kb and 2300kb. The red line indicates the positions of the insert on both genomes. Adapted from (Philippe et al., 2013) **(C)** Alignment of the regions [2,223,967-2,245,452] from *P. salinus* (noted 1) and [334,792-356,463] from *P. dulcis* (noted 2). The two regions were aligned using MAFFT and plotted one against the other: the positions conserved in both genomes are indicated in red. Each insert gene is indicated by a dot with the same color as in (A). The genes present in the insert are conserved between both pandoraviruses while intergenic regions and flanking regions of the insert are not.

Genomic context of the pandoraviruses *t*⁶A synthetase

The putative gene for *t*⁶A synthetase is found in a conserved 8,9 kb region of *P. salinus* and *P. dulcis*, positioned between nucleotides 2,230,201 and 2,239,105 (*P. salinus* genome coordinates) (Figure 65A). No significant similarity was found using BLASTn in *P. inopinatum* suggesting that this insert is completely absent in this strain. Notably, this region is positioned at two distinct locus within the *P. salinus* and *P. dulcis* genomes (Figure 65B) suggesting a genomic recombination in one strain or a horizontal gene transfer between two strains occurred. The insert region contains a putative gene in 5' encoding a protein annotated as a bifunctional metalloprotease/5'-nucleotidase composed of a N-terminal metalloprotease domain (Pfam: PF00149) and a C-terminal 5'-nucleotidase domain (Pfam: PF02872). Similarly to the YgjD domain gene, the closest homologs found by BLAST are uncharacterized proteins found in alpha-proteobacteria. More distant homologs of the 5'-nucleotidase domain are also found in eukaryotes, such as in *Homo sapiens* where it hydrolyzes extracellular nucleotides into membrane permeable nucleosides (Garavaglia et al., 2012). Still the role of the bifunctional metalloprotease/5'-nucleotidase protein in other organisms remains ambiguous.

Several putative genes are additionally found in 3' of the *t*⁶A gene: two putative genes belonging to the Cupin-5 superfamily (Pfam: PF06172), as well as a gene encoding a member of the S-adenosyl-L-homocysteine hydrolase superfamily (Pfam: PF00670), responsible for the reversible hydration of S-adenosyl-L-homocysteine into adenosine and homocysteine. The latter has close homologs among firmicutes bacteria. In contrast, the Cupin-5 genes have no significant homologs in database (first hit for Cupin gene ps_2317: Cupin gene from *Inquilinus limosus* with a cover 37 % and E-value 2,5) despite this superfamily being present in many organisms in the three domains of life. This suggest they have diverged significantly in pandoraviruses. In the end, the lack of knowledge about the respective function of these genes makes it impossible to evaluate the relevance behind their presence next to a *t*⁶A synthesis pathway in the insert. Interestingly, plot of the insert region between *P. salinus* and *P. dulcis* show clearly that the gene sequences are highly conserved as compared with the intergenic region, except in the diverging Cupin-5 region (Figure 65C). That indicates that there is a selective pressure to conserve the gene suggesting that it may have a function for the virus.

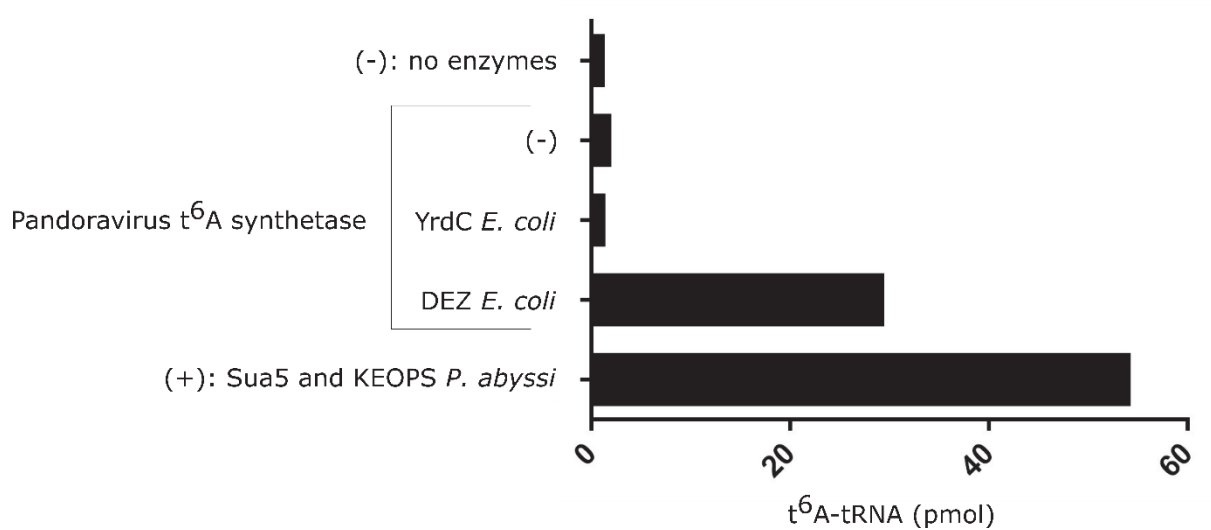


Figure 66. *In vitro* activity of the pandoravirus t⁶A synthetase

The activity of *Ps*-t⁶A synthetase (6,6 μM) at 25 °C for 30 min was tested in absence or in presence of YrdC or DEZ complex from *E. coli*. A positive control in presence of Sua5 (2μM) and KEOPS (2 μM) from *P. abyssi* was done (at 55 °C for 30 min).

In vitro activity of *P. salinus* t⁶A synthetase

To test whether this enzyme is able to synthesize t⁶A-tRNA, we expressed and purified *Ps*-t⁶A wild-type. The protein is mostly expressed in an insoluble form. In addition, several truncated products were identified suggesting it is sensible to proteolysis. Still, enough protein could be purified and used for a preliminary *in vitro* assay (Figure 66). No activity was observed for the enzyme alone in the conditions established previously (Perrochia et al., 2013). However, formation of t⁶A-tRNA was detected when the DEZ complex from *E. coli* was added to the reaction. In contrast, addition of YrdC from *E. coli* did not lead to production of t⁶A-tRNA. The fact that no technical nor biological replica were made have to be kept in mind when interpreting these data. However, these first results suggest that the Sua5 domain of *Ps*-t⁶A is active and produces TC-AMP. On the other hand no activity was observed for the YgjD domain. One explanation is that the conditions used for the *in vitro* assay are not adequate or that the YgjD domain requires the presence of yet unidentified partner proteins.

Discussion and perspectives

Viruses depend on the host protein expression machinery and have consequently evolved various strategies to optimize tRNA usage e.g. by encoding their own tRNA or tRNA-like structures. Similarly to other NCLDV viruses, such as mimiviruses and phycodnaviridae, pandoravirus encode indeed for several tRNA (Dreher, 2010): tRNA^{Pro, Met and Trp} for *salinus* and tRNA^{Pro} for *dulcis* and *inopinatum*. Another frequent strategy is adaption to the host codon usage. However similarity of the codon usage between virus and host results in competition for tRNAs for protein synthesis. Interestingly, modifications on tRNAs strongly impact their functionality in viruses. For instance, in *E. coli*, a decrease of thiolation of tRNA^{Lys(UUU)} led to decreased infection of the lambda phage (Maynard et al., 2012). Similarly, the modifications mcm⁵s²U₃₄ and ms²t⁶A₃₇ in the anticodon loop of the human tRNA^{Lys(UUU)} are important for the reverse transcription mechanism of HIV-1 (Spears et al., 2014).

Unfortunately there is currently no knowledge about the strategy used by pandoraviruses to translate their own proteins during infection. The case of the pandoraviruses t⁶A synthetase is therefore quite interesting. Indeed it is so far the only case reported of a complete synthesis pathway of a universal modified nucleoside in a virus. This is all the more intriguing that t⁶A is a complex modification, involving two covalent bonds and several enzymes in its synthesis. The first results of the *in vitro* activity assay are encouraging but further attempts to optimize purification and reaction conditions have to be done to draw clear conclusions. Additionally, we recently initiated *in vivo* complementation assay in *S. cerevisiae*. The capacity of the whole protein or the truncated mutant, corresponding to the YgjD and Sua5 domains, to complement the yeast strains Δsua5, Δkae1, Δbud32 and Δpcc1 will be tested.

Finally, if the stability of the protein and the purification yield can be improved, crystallogenesis assays will be done in order to resolve the structure of this unique t⁶A synthetase. These data should allow us to identify the contact surfaces between the domains, and notably those between Sua5 domain and YgjD domain. This data could help understand how the TC-AMP intermediate is transferred between Sua5/YrdC and Kae1/YgjD/Qri7 proteins.

Whether it is active or not in the cell has yet to be determined but the sequence conservation of this gene and its presence in two strains of pandoraviruses make it likely to be functional. Nevertheless, its absence in the genome of the third pandoravirus, *P. inopinatum*, indicates that it is not strictly required for this group of viruses. The presence of a t⁶A synthetase may be an evolutive advantage for the giant virus. Indeed defect of t⁶A was linked to pleiotropic phenotypes including major translation errors (Thiaville et al., 2016). Hypothetically, the pandoravirus t⁶A synthetase may be used to favor and/or optimize viral protein synthesis during its replication cycle in *acanthamoeba*. Another possibility is that the viral t⁶A synthetase may be used as lure to hijack the host modification system. These questions may be answered in the future by analyzing the function of the t⁶A synthetase in the pandoravirus-*acanthamoeba* *in vivo* system. For instance, RT-PCR combined with Western Blot, using *Ps*-t⁶A specific antibodies, would allow us to clearly determine whether the gene is expressed, when and how much during the viral cycle.

The origin of pandoraviruses, and NCLDV in general, remains elusive. Despite the size of their genome, pandoraviruses are highly divergent and share only 17 of the 49 conserved genes found among NCLDV (Yutin et al., 2009). Based on phylogenies of several of these markers, such as RNA polymerase subunits 1 and 2, it was proposed that NCLDV represent a monophyletic phylum, subsequently renamed Megavirales, that composes a genuine fourth domain of life (Boyer et al., 2010; Colson et al., 2011; Sharma et al., 2015). Current NCLDV would have thus inherited universal markers from a cellular common ancestor. However this view is challenged by others groups that argue that horizontal gene transfer between ancestors of current NCLDV and proto-eukaryotic host cells may explain the presence of similar markers in both (Filée, 2014; Forterre and Gaïa, 2016; Williams et al., 2011). It was suggested that modern giant viruses (with genomes > 500 kb) originated several times independently by genome expansion of smaller ancestral NCLDV and do not form a fourth domain of life (Yutin et al., 2014). None of the conflicting hypothesis can be currently ruled out but it is generally accepted that the NCDLV behave as pick pockets i.e. acquire new genes from the host or from other organisms (Filée et al., 2008; Koonin and Yutin, 2010). This is further corroborated by the analysis of the origin of *Ps*-t⁶A.

The proximity of the pandoravirus t⁶A synthetase with bacterial orthologs suggests it was acquired from bacteria which were probably sharing the cytoplasmic environment with the virus. Indeed, *acanthamoeba* are known to host intracellularly replicating endocytobionts (Balczun and Scheid, 2017). This feature is a health threat as it may protect the endoparasitic microbes against disinfectants and antibiotics, especially when forming cysts. The concentration of organisms in a close environment e.i. the *acanthamoeba* cytoplasm is likely a favorable context for horizontal transfer gene. Interestingly, the source of the different domains of the t⁶A synthetase differ since the YgjD domain is close to alpha-proteobacteria orthologs while the Sua5 is close to firmicutes orthologs.

The gene fusion may have occurred in an ancestral pandoravirus. In this case, the giant virus prove itself to be able to remodel efficiently current existing pathway of cellular organisms. Another possibility is that the gene fusion occurred in another organism yet unidentified and that the whole gene was acquired by transfer in the pandoraviruses. In both cases, whether the ancestor that acquired this insert is also common to *P. inopinatum*, i.e. this strain has lost the t⁶A synthetase, or not is hard to tell. The sequencing of new pandoraviruses genomes is needed to evaluate the repartition and the importance of the t⁶A synthetase in this lineage. Incidentally, the study of this enzyme may led to interesting discoveries about the unique infection process of pandoraviruses.

Conclusion

The t⁶A synthetase of pandoraviruses possesses all the residues known to be important for t⁶A-tRNA formation. Its gene is present in a 9 kb insert found so far only in *P. salinus* and *P. dulcis*. Its functionality as well as its role for the virus remain to be elucidated.

Conclusion et perspectives

Conclusion et perspectives du travail de thèse

Notre objectif premier était de comprendre le rôle du linker et du domaine SUA5 chez les enzymes Sua5. Nous avons pu montrer par l'étude structure-fonction de Sua5 de *P. abyssi* que cette région C-terminale est importante pour la synthèse de TC-AMP (Article I). L'ensemble des données suggère que le linker est une *gating loop* capable de passer d'un état désordonné à un état ordonné, bloquant alors le site actif. Aucune fonction enzymatique précise n'est attribuée au domaine SUA5 ce qui laisse penser que son rôle est principalement de limiter la flexibilité du linker. L'étude plus poussée des mutants de *Pa-Sua5*, en particulier l'obtention des structures cristallographiques, permettra d'en apprendre d'avantages sur le rôle respectif de chacun des motifs mutés ainsi que sur les mécanismes de la réaction.

Le variant YrdC est quant à lui actif en absence de linker. L'étude RMN d'YrdC de *E. coli* suggère que les 20 résidus de l'extrémité C-terminal sont flexibles et puissent donc agir eux-aussi comme une *gating loop* (Harris et al., 2015). Cependant notre analyse des séquences de Sua5/YrdC procaryotes (Article II) montre qu'un nombre considérable de protéines YrdC ne possèdent pas une extrémité C-terminal complète ce qui suggère qu'elles utilisent un autre mécanisme réactionnel. Le mystère reste donc entier.

A la suite de l'étude structure-fonction de *Pa-Sua5*, nous nous sommes intéressés aux différences fonctionnelles existant entre les sites actifs d'YrdC et Sua5, susceptible d'expliquer l'autonomie du premier vis-à-vis des linker et domaine SUA5. Grâce à l'analyse des séquences, nous avons pu identifier plusieurs motifs et résidus conservés spécifiquement chez les variants Sua5 (Article II). Nous avons mis en évidence l'existence d'un résidu signature particulièrement intéressant au sein du site actif qui est une proline chez Sua5 et une thréonine/sérine chez YrdC. A la suite de ces découvertes, nous avons identifié une Sua5 atypique au sein d'*A. profundus* dont le domaine catalytique possède plusieurs caractéristiques du variant YrdC, dont le résidu signature. L'étude structure-fonction de cette protéine nous permettra de déterminer le rôle de ce résidu et, plus généralement, les paramètres déterminants les différences de fonctionnalité de chaque variant.

La famille Sua5/YrdC étant universelle, nous avons également tenté de comprendre son histoire évolutive (Article II). La faiblesse du signal phylogénétique au sein de cette famille n'a malheureusement pas permis d'obtenir une phylogénie robuste. Cependant, l'ensemble des données biochimiques, structurales et bio-informatiques sur la famille Sua5/YrdC, nous a permis de proposer un modèle évolutif susceptible d'expliquer l'existence de deux formes ainsi que la répartition de ces deux formes sur l'arbre du vivant. Dans ce modèle, LUCA possédait une protéine Sua5 qui aurait divergé dans certaines lignées jusqu'à perdre son extrémité C-terminale pour devenir une protéine YrdC active. Ce modèle, bien qu'étant le plus cohérent et parcimonieux avec les données actuelles, reste spéculatif.

L'importante similarité de séquence entre les domaines catalytiques d'YrdC et Sua5 suggère que l'absence du linker et du domaine SUA5 est compensée « seulement » par quelques mutations clés. C'est assez surprenant au vu du nombre de résidus conservés présents sur le linker et du rôle structural majeur, à savoir de *gating loop*, qui lui est attribué. Encore une fois l'étude approfondie d'Ap-Sua5, qui est *a priori* un état de transition entre une Sua5 et une YrdC, permettra de tester notre modèle et d'identifier les mécanismes derrière le processus de raccourcissement.

Un processus de simplification semble également être à l'origine de l'existence de la t⁶A synthétase des pandoravirus. La synthèse de t⁶A-ARNt est un processus cellulaire fondamental. Les effets délétères sur la traduction résultant d'une absence de t⁶A chez les cellules eucaryotes et procaryotes sont bien identifiés. Le rôle potentiel de cette modification pour des virus n'a cependant jamais été étudié ou mis en évidence. C'est pourquoi l'identification d'un système de t⁶A synthétase complet chez les pandoravirus est une découverte déroutante et inédite (Article III). L'étude de cette protéine *in vitro* puis *in vivo* nous permettra d'en apprendre d'avantages sur le mécanisme de synthèse de t⁶A d'une part et sur son rôle pour les virus géants d'autre part.

Projets à venir

De grandes avancées ont été faites ces dernières années dans la compréhension du mécanisme réactionnel et du rôle de la modification t⁶A. Cependant, de nombreuses questions intéressantes restent ouvertes telles que : Comment les t⁶A synthétases reconnaissent l'ARNt substrat ? Quelle est la fonction des protéines accessoires ? Pourquoi le système bactérien est différent de celui présent chez les archées/eucaryotes ? Comment les mutations dans le KEOPS humain provoquent des troubles touchant le rein et le cerveau ? Quel est le rôle du cinquième élément au sein du KEOPS ? Ce cinquième élément est-il présent ailleurs que chez les champignons et les métazoaires ? Les deux laboratoires qui m'ont accueillie lors de ma thèse poursuivront l'étude des t⁶A synthétases à la suite de mon départ en tentant de répondre à certaines de ces questions.

L'équipe BCA se focalisera sur l'étude *in vivo* de la thréonylcarbamoylation des ARNt chez les archées. Il semble important et pertinent d'initier ces travaux car une telle étude n'a jamais été menée chez ces organismes. Cela permettra donc d'établir un premier état des lieux pour ce domaine du vivant et de le comparer avec ce qui est connu chez les eucaryotes et les bactéries. De plus, ces données peuvent se révéler pertinentes pour l'étude du système homologue eucaryote. L'équipe se servira essentiellement d'approches génétiques et biochimiques avec comme organisme modèle l'archée hyperthermophile *Thermococcus kodakaerensis* (*T. kodakarensis*). En effet il existe maintenant une série d'outils génétiques pour les *Thermococcales* (Farkas et al., 2013).

Dans un premier temps, ils ont l'intention d'étudier l'impact de l'altération de la synthèse de t⁶A chez *T. kodakaerensis*, ceci en procédant à des altérations systématiques des différents gènes impliqués dans la synthèse de t⁶A chez cet organisme. Ces souches recombinantes seront soumises à une série d'analyses. La quantité de t⁶A produite par ces mutants sera quantifiée par HPLC. L'effet des mutations sur la quantité d'ARNt dans les cellules sera également mesuré par RT-PCR et Northern Blot. Enfin, un profil de polysomes sera établi afin d'évaluer l'effet global des mutations sur l'activité traductionnelle.

Pour le domaine des eucaryotes et des archées, la question demeure si oui ou non les protéines accessoires du KEOPS assurent d'autres fonctions dans la cellule indépendamment de la synthèse de la modification t⁶A. Pour répondre à cette question, l'équipe BCA a l'intention de construire une souche de *T. kodakarensis* modifiée pour exprimer en *trans* le complexe DEZ de la bactérie hyperthermophile *Thermotoga maritima*. Cette stratégie permet d'étudier les conséquences des délétions et/ou modifications des gènes codant sous-unités du complexe KEOPS dans d'autres voies métaboliques que celle de la t⁶A, puisque tout défaut de synthèse de t⁶A du KEOPS sera compensée par l'activité de DEZ.

Dans le but d'identifier des éventuels nouveaux partenaires protéiques de KEOPS chez les archées des expériences de co-purification par affinité à partir des lysats de *T. kodakarensis*. Ils espèrent également découvrir le cinquième élément du KEOPS chez les archées, cette sous-unité qui n'est pas détectable *in silico* par comparaison de séquences. En effet, la découverte, très récemment, chez l'homme d'un analogue de Gon7 rend plus crédible l'hypothèse de son existence également chez les archées.

De son côté, l'équipe FAAM s'intéresse d'une part à l'étude du complexe DEZ bactérien et d'autre part à l'étude du complexe KEOPS des archées/eucaryotes. Le caractère universel de t⁶A ainsi que son essentialité pour de nombreuses bactéries font du complexe DEZ une cible potentiellement intéressante pour le développement d'antibiotiques, en particulier l'enzyme YjeE qui n'a pas d'homologue chez les eucaryotes. Une première stratégie est de cibler l'interface entre les trois sous-unités du complexe DEZ en combinant des approches structurale, chimique et bioinformatique. Pour y arriver, l'équipe FAAM réalise actuellement des tests de cristallogénèse avec des complexes DEZ provenant de diverses bactéries hyperthermophiles. Une fois la structure complète de DEZ obtenue, une approche *in silico* de *screening* permettra d'identifier des petites molécules susceptibles d'interférer avec la formation du complexe et donc d'inhiber la synthèse de t⁶A.

Une seconde stratégie consiste à cibler le site actif d'YgjD en utilisant des molécules analogues de l'intermédiaire TC-AMP, qui sont susceptibles d'agir comme des inhibiteurs compétitifs. La similarité des structures d'YgjD et de Kae1 représente un obstacle potentiel à cette approche. Cependant, l'identification d'analogues du TC-AMP liant Kae1/YgjD/Qri7, même si elle n'aboutit pas des inhibiteurs spécifiques de YgjD, s'avérera être des outils utiles pour l'étude du mécanisme moléculaire de cette famille.

Récemment, plusieurs mutations au sein des sous-unités du KEOPS humain ont été identifiées dans une trentaine de famille dont les enfants présentent un GAMOS, qui se caractérise par une microcéphalie, un retard du développement et un syndrome néphrotique cortico-résistant. L'équipe FAAM a participé au travail de caractérisation de ces mutations qui a permis de montrer que ces altérations du KEOPS humain entraînaient une baisse d'activité de synthèse de t⁶A, susceptible d'être responsable du GAMOS (Braun et al., 2017), et travaille désormais en collaboration avec l'équipe de Corinne Antignac (Institut Imagine –Paris), spécialiste des aspects moléculaire des maladies du rein.

La caractérisation du mécanisme de synthèse de t⁶A par le complexe KEOPS, en particulier via l'identification du rôle précis de chaque sous-unité ainsi que des modalités d'interaction avec les ARNt substrats par des approches structurales et biochimiques est primordiale pour mieux comprendre *in fine* les mécanismes moléculaire et cellulaire liant le complexe KEOPS humain et le GAMOS. L'étude enzymologique *in vitro* du complexe KEOPS humain et de ses versions mutantes combinée à une caractérisation de leur activité *in vivo* chez la levure ainsi que chez des lignées cellulaires humaines et chez la souris permettront d'évaluer l'impact de chacune de ces mutations sur l'activité.

Enfin, l'équipe FAAM s'intéresse également à la sous-unité C14. Une approche RMN leur a déjà permis de montrer que, comme Gon7, C14 est désordonnée en solution. Les complexes C14-Pcc1 ou C14-Pcc1-Kae1 humains ont été purifiés avec succès et des essais de cristallogénèse sont actuellement en cours. La structure cristallographique de C14 lié à Pcc1 permettra entre autres de déterminer s'il s'agit bien d'un homologue de Gon7. Elle permettra également d'introduire des mutations à l'interface afin de tester leur impact sur l'activité du complexe et donc de comprendre le rôle de C14 dans le contexte cellulaire.

Annexes

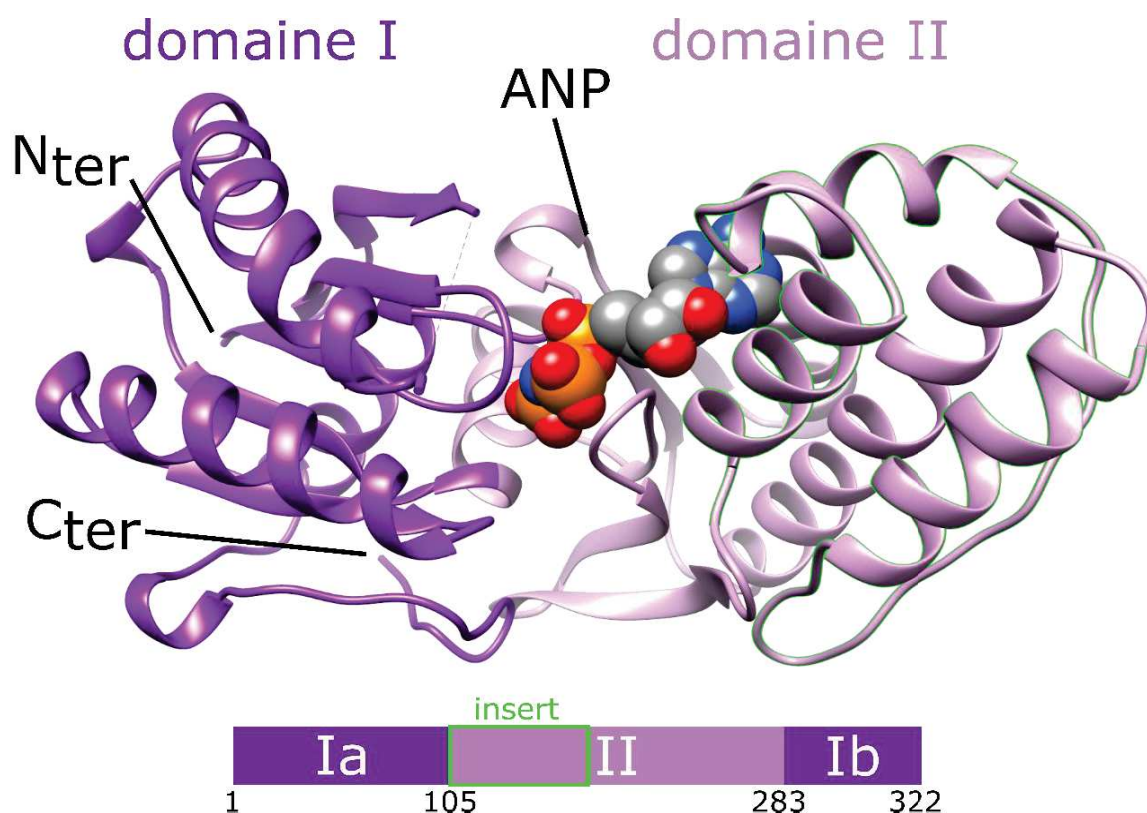


Figure 67. Structure de Kae1 de *M. jannashii*

La structure de Kae1 a été résolue en complexe avec un analogue non hydrolysable de l'ATP, le *phosphoaminophosphonic acid-adenylate ester* (ANP) (code PDB : 2VWB) (Hecker et al., 2008). Le domaine I est en violet foncé et le domaine II en violet clair. La position des domaines dans la séquence de Kae1 est indiquée en dessous de la structure. L'atome de métal est positionné en arrière des phosphates de l'ANP et n'est donc pas visible sur cette image. Les trois hélices correspondant à l'insert spécifique de Kae1/YgjD/Qri7 sont entourées d'un trait vert.

Annexe 1 : La famille Kae1/YgjD/Qri7

a. Propriétés structurales

Les protéines Kae1/YgjD/Qri7 font en moyenne 350 acide aminés et présentent toutes un repliement typique de la superfamille ASKHA (*acetate/sugar kinase, Hsp70, actin*) (Hecker et al., 2007; Nichols et al., 2013; Wan et al., 2013). Leur structure présente deux domaines semblables, notés I et II (Figure 67). Une insertion relativement importante dans le domaine II est spécifique à cette famille et n'existe pas chez les autres membres de la superfamille ASKHA. La jonction entre les domaines I et II forme un site de fixation pour un nucléotide (Hecker et al., 2008). Les protéines Kae1 de *M. jannaschii* et de *P. abyssi* peuvent y lier une molécule d'ATP (Hecker et al., 2007, 2008). Une activité ATPase *in vitro* faible a même été détectée pour Kae1 de *P. abyssi* et de *S. cerevisiae* (Hecker et al., 2007; Mao et al., 2008). Cependant l'absence d'un second site, suggère qu'il s'agit en réalité du site de liaison du TC-AMP. Ceci est corrélé avec le fait que la mutagenèse de ce site chez Kae1 de *P. abyssi* rend la protéine incapable de synthétiser de la t⁶A (Perrochia et al., 2013a). Kae1/YgjD/Qri7 n'a donc vraisemblablement pas un rôle d'ATPase dans la synthèse de t⁶A.

Kae1/YgjD/Qri7 a aussi la particularité d'être la seule famille de la superfamille ASKHA à utiliser un ion métal, zinc ou fer, au sein de son site actif. Chez Kae1, un atome de fer est présent, à proximité des phosphates du nucléotide. Le fer est coordonné par quatre résidus conservés : deux histidines, une tyrosine et un aspartate (Hecker et al., 2007). Chez YgjD et Qri7, seules les histidines sont conservées et elles lient un atome de zinc chez YgjD et un ion métallique non identifié chez Qri7 (Nichols et al., 2013; Tominaga et al., 2012; Wan et al., 2013). La mutation du site de liaison du fer de Kae1 de *S. cerevisiae* provoque un phénotype similaire à celui de la souche $\Delta kae1$ *in vivo* (Kisseleva-Romanova et al., 2006; Mao et al., 2008; Srinivasan et al., 2011). De même, l'altération du motif de liaison du métal chez la version mitochondriale Qri7 de *S. cerevisiae* abolit l'activité t⁶A synthétase de cette enzyme (Wan et al., 2013). La liaison de l'ion métal joue donc un rôle crucial pour l'activité de Kae1/YgjD/Qri7. Une différence entre les trois orthologues se situe au niveau de la distance séparant l'ion métal de l'atome d'oxygène du phosphate α du TC-AMP au sein du site actif, allant de 2.34 Å pour YgjD à 5.81 Å pour Qri7, ce qui est susceptible d'influencer leur activité enzymatique respective (Tominaga et al., 2012).

b. Le complexe DEZ chez les bactéries

Les partenaires d'YgjD

YeaZ (COG1214) fut le premier partenaire d'YgjD identifié lors d'une étude à grande échelle des interactions entre les protéines essentielles d'*E. coli* (Butland et al., 2005). Avec seulement 230 acides aminés, YeaZ est un paralogue tronqué d'YgjD. La structure d'YeaZ a été résolue chez plusieurs espèces de bactéries dont *E. coli* (Jeudy et al., 2005) et *Thermotoga maritima* (Xu et al., 2010). Comme YgjD, YeaZ présente un repliement caractéristique de la famille ASHKA (Nichols et al., 2006). Cependant son domaine II ne possède ni site de liaison au nucléotide ni site de liaison au métal. YeaZ forme des dimères seul en solution mais s'apparie préférentiellement avec YgjD de manière stable lorsque ce dernier est présent (Butland et al., 2005; Nichols et al., 2006, 2013; Zhang et al., 2015a). La même surface est impliquée dans les deux cas de figure. L'hétérodimère YgjD-YeaZ est capable de lier un nucléotide à l'interface des deux monomères mais est incapable de synthétiser de la t⁶A en absence d'YjeE ou d'ATP (Zhang et al., 2015a).

YjeE (COG0802) fait en moyenne 150 acides aminés et possède une activité ATPase faible *in vitro* (Allali-Hassani et al., 2004; Freiberg et al., 2001; Handford et al., 2009). Cette activité est fortement stimulée en présence d'YgjD-YeaZ, même en absence des substrats de la t⁶A (Allali-Hassani et al., 2004; Deutsch et al., 2012; Zhang et al., 2015a). La structure cristallographique d'YjeE de *Haemophilus influenza* révèle une similarité avec les petites Guanosine triphosphate hydrolases (GTPases), connue pour fonctionner comme des interrupteur moléculaire (*molecular switch*) utilisant l'hydrolyse du GTP pour passer d'un état structural à un autre (Edelheit et al., 2009; Teplyakov et al., 2002).

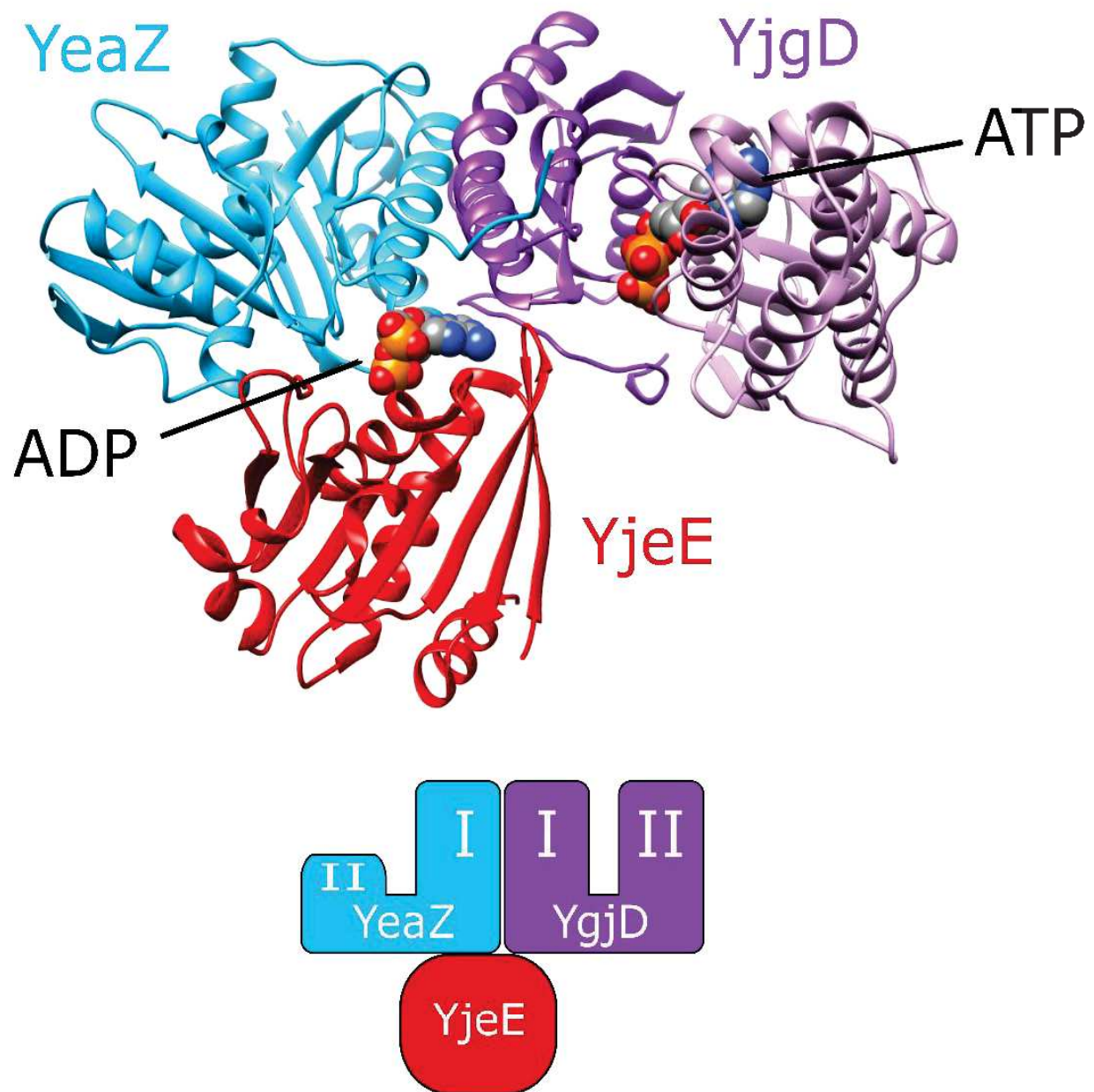


Figure 68. Organisation du complexe DEZ des bactéries

La position d'YjeE de *H. influenzae* en complexe avec l'ADP (code PDB : 1HTW) a été modélisée par rapport au dimère YeaZ-YjgD d'*E. coli* en complexe avec l'ADP et l'ATP (code PDB : 4YDU). YjgD est en violet (les domaines sont colorés comme dans la figure 23). YeaZ est en bleu et YjeE en rouge. Le modèle de (Zhang et al., 2015a) suggère que l'association d'YjeE au dimère est dépendant de la liaison en amont d'une molécule d'ATP (au niveau du site occupé par l'ADP dans la figure). Le complexe DEZ est représenté schématiquement en dessous de la structure.

Structure quaternaire du complexe DEZ

Il n'existe aucune structure du complexe entier. La position d'YjeE sur le dimère YeaZ-YgjD-ADP d'*E. coli* a donc été modélisée et corrélée par analyse SAXS (Figure 68) (Zhang et al., 2015a). YjeE se lie majoritairement sur YeaZ. Le dimère YgjD-YeaZ et YjeE sont susceptibles de lier simultanément une même molécule d'ATP. Des mutations inhibant la dimérisation d'YgjD-YeaZ abolissent l'interaction avec YjeE. L'interaction entre YeaZ-YgjD et YjeE est strictement dépendante de la présence d'ATP chez *E. coli* (Nichols et al., 2013; Zhang et al., 2015a). Dans le scénario proposé par Zhang et al, la liaison d'un ATP sur YjeE entraîne un changement de conformation de ce dernier, ce qui lui permet alors de se lier à YgjD-YeaZ afin de former un complexe DEZ-ATP actif. Par analogie avec les petites GTPases, YjeE pourrait être un *molecular switch* utilisant l'hydrolyse de l'ATP. Cependant, la mutation de Trp¹⁰⁹ d'YjeE d'*E. coli*, qui abolit l'activité ATPase mais n'affecte pas la liaison de l'ATP, n'entraîne pas une baisse d'activité de synthèse de t⁶A. Ceci suggère que la liaison elle-même d'YjeE au dimère YgjD-YeaZ, et non son activité ATPase, est le véritable critère déterminant pour l'activité du complexe DEZ (Zhang et al., 2015a).

De manière contradictoire, les orthologues YgjD, YeaZ et YjeE de *B. subtilis* peuvent synthétiser la t⁶A *in vitro* sans ajout d'ATP dans le mélange réactionnel ce qui suggère que leur mode de fonctionnement est différent de celui observé chez *E. coli* (Lauhon, 2012). L'apport d'YgjD de *B. subtilis* n'est d'ailleurs pas capable de compléter l'absence d'YgjD chez *E. coli*. En revanche, l'apport simultané d'YgjD et de son paralogue YeaZ restaure le phénotype normal (El Yacoubi et al., 2011). De manière cohérente avec ces observations, l'interaction *in vitro* entre YgjD et YeaZ d'une même espèce est plus forte qu'une interaction entre YgjD et YeaZ issues de deux espèces différentes (Rajagopala et al., 2010). Ceci indique que des spécificités propres à une ou plusieurs bactérie(s) ont été acquises au sein du complexe DEZ au cours de l'évolution (El Yacoubi et al., 2011).

Le mécanisme précis du fonctionnement du complexe DEZ et la fonction de ses sous-unités restent donc à confirmer. Curieusement l'orthologue mitochondrial, Qri7 ne complémente pas l'absence d'YgjD chez *E. coli* (El Yacoubi et al., 2011) alors qu'il restaure la voie de biosynthèse de t⁶A chez la souche $\Delta kae1$ de levure (Wan et al., 2013). Ce résultat fait penser que Qri7 n'est pas fonctionnelle dans l'environnement cellulaire d'*E. coli* ou alors que la protéine YgjD, seule ou en complexe, est susceptible d'avoir un autre rôle que celui déjà connu dans la voie de biosynthèse de t⁶A chez les bactéries.

En terme de distribution, plus de 90% des bactéries possèdent des homologues pour les trois sous-unités du complexe DEZ (Thiaville et al., 2015a). Les partenaires YeaZ et YjeE sont essentielles chez la plupart bactéries et n'ont aucun homologue chez les eucaryotes (Thiaville et al., 2015b). Les protéines partenaires de DEZ sont donc des cibles antibactériennes potentielles dans le futur. Cependant certaines bactéries n'ont pas YeaZ (Thiaville et al., 2015a) et d'autres n'ont ni YeaZ ni YjeE (Grosjean et al., 2014). Ce dernier cas de figure est le cas de certaines Mycoplasmes qui ne seraient donc pas sensible à des inhibiteurs spécifiques d'YeaZ ou YjeE.

c. Le complexe KEOPS chez les eucaryotes/archées

Chez les archées et les eucaryotes, la protéine Kae1 s'associe avec Bud32, Cgi121 et Pcc1 pour former le complexe KEOPS. Chez les champignons un cinquième élément nommé Gon7 fait également partie du complexe (Downey et al., 2006). Récemment un cinquième élément, nommé C14ORF142 (ou C14), a également été identifié chez les métazoaires (Wan et al., 2017). A l'exception de Gon7 et C14, Bud32, Cgi121 et Pcc1 sont conservées chez les eucaryotes et les archées, mais absentes chez les bactéries. Chez la levure la délétion individuelle des gènes codant pour Kae1, Bud32, Pcc1 et Gon7, mais pas de celui codant pour le Cgi121, abolit l'activité de synthèse de t⁶A du complexe *in vivo* (El Yacoubi et al., 2011; Srinivasan et al., 2011; Wan et al., 2013)(Wan et al., 2013)(Wan et al., 2013).

Les partenaires de Kae1

Bud32 est composé en moyenne de 260 acides aminés. *In vitro*, la protéine Bud32 seule possède une activité ATPase ainsi qu'une faible activité kinase. Cette dernière est abolie lorsque Bud32 est au sein du complexe KEOPS de *P. abyssi* (Hecker et al., 2008; Perrochia et al., 2013a). Des études de mutagenèse ont montré que l'activité ATPase de Bud32 est essentielle pour l'activité de synthèse de t⁶A de KEOPS de *P. abyssi* et de *S. cerevisiae*, respectivement *in vitro* et *in vivo* (Mao et al., 2008; Perrochia et al., 2013a).

Au niveau structural, Bud32 est composé des deux lobes distincts typiques des kinases. Le lobe N-terminal correspond au domaine de fixation du nucléoside et le lobe C-terminal est responsable de la fixation de la protéine substrat pour l'activité kinase (Facchin et al., 2002; Hecker et al., 2008; Mao et al., 2008). Plus particulièrement, Bud32 appartient à la famille des kinases dite atypiques : son domaine C-terminal a une taille réduite et la boucle d'activation caractéristique des kinases, impliquée dans la reconnaissance du substrat, est absente (LaRonde, 2014; LaRonde-LeBlanc and Wlodawer, 2005).

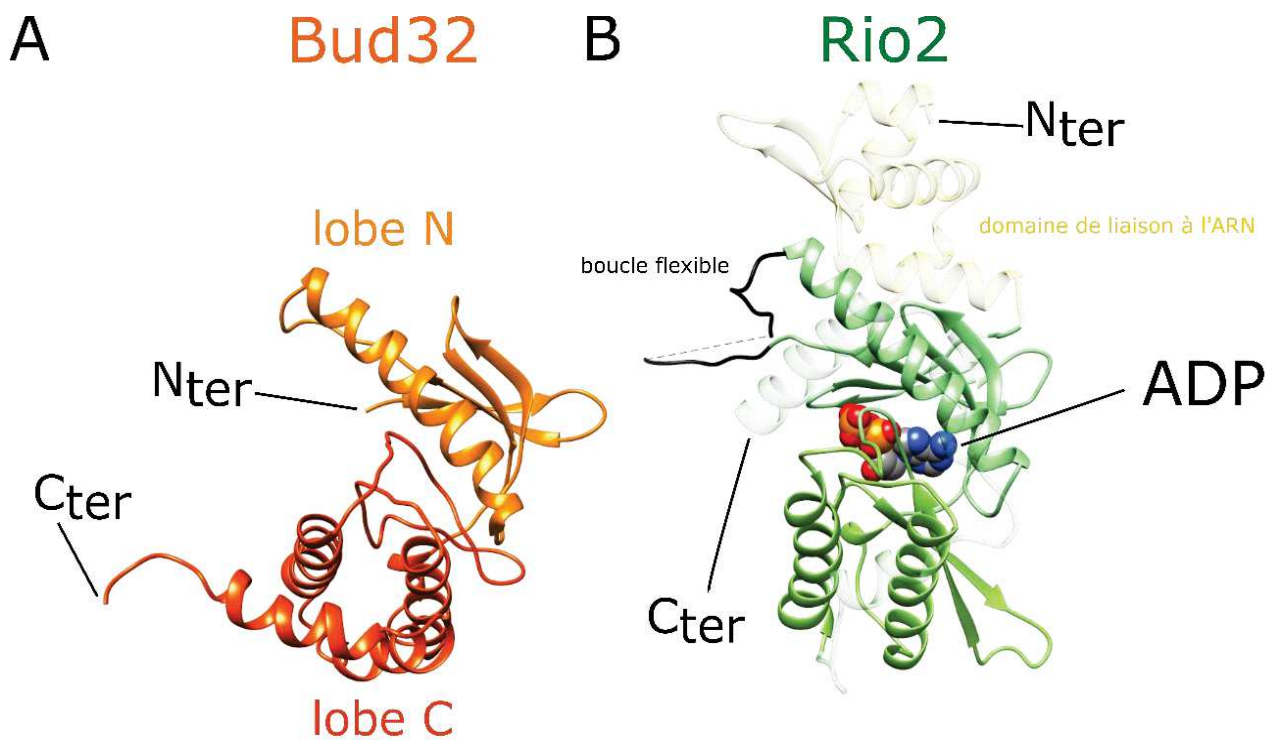


Figure 69. Comparaison des kinases atypiques Bud32 et Rio2

(A) Structure de Bud32 de *M. jannashii* (code PDB : 3ENH). Le lobe N et le lobe C ont des teintes différentes. Le site actif de Bud32 est situé entre ces deux lobes. L'extrémité C-terminale de Bud32 est flexible et est susceptible d'interagir avec Kae1 au sein du dimère Kae1-Bud32. **(B)** La structure de Rio2 de *Chaetomium thermophilum*, dans la même orientation que Bud32 (code PDB : 4GYI) (Ferreira-Cerca et al., 2012). Seule la région commune avec Bud32, le domaine Rio, est montré en opaque. Rio2 possède une boucle flexible spécifique (en noir), qui sert à interagir avec la petite sous-unité lors de la biogénèse du ribosome.

La protéine RIO-kinase Rio2 est également une kinase atypique. Elle a une structure proche de celle de Bud32 (Figure 69) et possède, comme cette dernière, une activité ATPase (Ferreira-Cerca et al., 2012). Rio2 est connue pour être impliquée dans la maturation de la petite sous-unité du ribosome chez les eucaryotes. L'hydrolyse de l'ATP par Rio2 lui permet de se dissocier du ribosome. Le résidu catalytique crucial de Rio2 (Asp²²⁹ chez *Chaetomium thermophilum*) est conservé chez Bud32 (Asp¹⁴² chez *P. abyssi* et Asp¹⁸² chez *S. cerevisiae*). Par extrapolation, l'activité ATPase de Bud32 sert donc potentiellement à réguler la liaison et la dissociation de l'ARNt substrat sur le complexe KEOPS pendant la synthèse de t⁶A (Perrochia et al., 2013a; Zhang et al., 2015b).

Cgi121 est composé en moyenne de 190 acides aminés (Mao et al., 2008)(Mao et al., 2008). La souche de levure Δ *cgi121* présente un phénotype de croissance beaucoup moins atteint que celui des souches Δ *pcc1*, Δ *bud32* et Δ *kae1* (Wan et al., 2013). De plus, l'absence de Cgi121 n'abolit pas la synthèse de t⁶A *in vitro* par le complexe KEOPS de *P. abyssi* (Perrochia et al., 2013a). Cgi121 n'est donc pas essentielle pour la synthèse de t⁶A. Cependant sa présence améliore l'efficacité de la réaction *in vitro* pour KEOPS de *P. abyssi* et stimule également l'activité d'autophosphorylation de Bud32 (Perrochia et al., 2013a). Cgi121 apparaît donc comme un régulateur positif de KEOPS, et plus précisément de Bud32.

Pcc1, avec une taille moyenne de 85 acides aminés, est plus petite que les autres sous-unités du KEOPS (Kisseleva-Romanova et al., 2006). L'analyse par ultracentrifugation analytique et la structure cristallographique de Pcc1 issue de *P. abyssi* montre qu'elle forme un homodimère en solution (Mao et al., 2008). Le dimère Pcc1-Pcc1 peut lier simultanément deux molécules de Kae1 pour former Kae1-Pcc1-Pcc1-Kae1. KEOPS peut donc dimériser via Pcc1 (Mao et al., 2008; Wan et al., 2016; Zhang et al., 2015b). Le complexe KEOPS de *P. abyssi* se purifie effectivement sous la forme d'un dimère actif (Perrochia et al., 2013a). Le dimère Pcc1-Kae1 est capable de lier les ARNt *in vitro* avec une meilleure affinité que la protéine Kae1 seule ou que le dimère Kae1-Bud32 (Perrochia et al., 2013a). Ainsi Pcc1 pourrait avoir une double fonction au sein du KEOPS en servant de plateforme pour la dimérisation du complexe d'une part et en participant à la liaison de l'ARNt d'autre part (Wan et al., 2016).

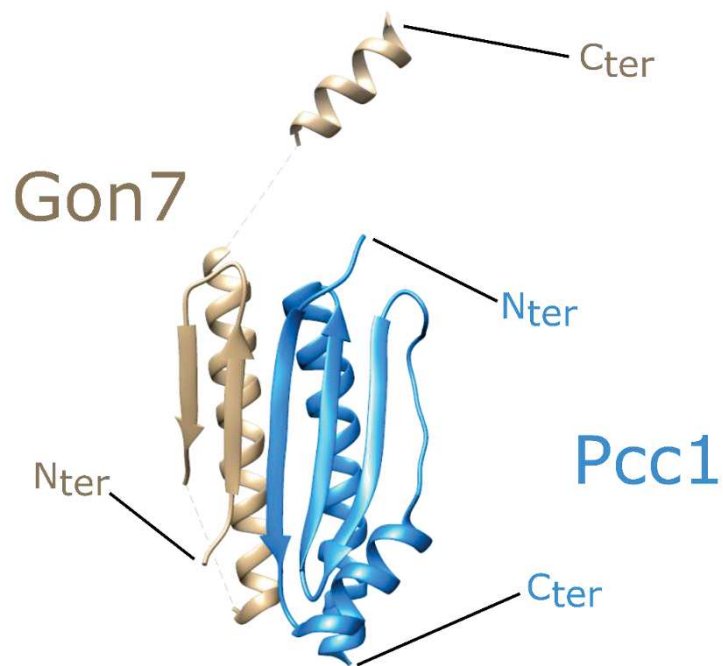


Figure 70. Structure du dimère Pcc1-Gon7

La structure du dimère Pcc1-Gon7 de *S. cerevisiae* (code PDB : 4WX8) (Zhang et al., 2015b) a permis de révéler que la protéine Gon7, désordonné en solution, se structure partiellement en se liant à Pcc1. Gon7 occupe la même surface d'interaction que celle de l'homodimère Pcc1. Cependant, contrairement à Pcc1, Gon7 n'est pas capable de lier Kae1.

Gon7 a été identifié chez *S. cerevisiae* comme étant un partenaire du complexe KEOPS par une approche de *pull-down* (Downey et al., 2006; Kisseleva-Romanova et al., 2006). Cette protéine d'environ 100 acides aminés est présente chez les champignons uniquement. Les souches $\Delta gon7$ et $\Delta kae1$ de levure ont un phénotype similaire suggérant que cette protéine est aussi essentielle pour l'activité de KEOPS. L'état désordonné de Gon7, prédit d'après sa séquence, a ensuite été confirmé par RMN (Zhang et al., 2015b). Gon7 est cependant capable de partiellement se structurer au contact de Pcc1 pour former des dimères Gon7-Pcc1 (Figure 70) (Zhang et al., 2015b).

Des essais de co-immunoprécipitation, réalisés dans les lignées cellulaires humaines K562 et HeLa53, avaient identifié une série d'interactants de KEOPS dont une protéine putative de fonction inconnue nommée C14 (Costessi et al., 2012). Très récemment, l'analyse protéomique plus précise de l'interactome du KEOPS humain au sein de la lignée HEK293 a permis de confirmer ces observations (Wan et al., 2017). Comme pour Gon7, l'interaction entre C14 et KEOPS se fait via Pcc1 et est stable, ce qui suggère qu'il est un membre à part entière du complexe. C14 serait l'équivalent chez les métazoaires de Gon7, avec lequel il partage un certain nombre de caractéristiques biophysiques et biochimiques. Cependant, l'absence de conservation de séquence évidente entre Gon7 et C14 ainsi que le manque de données structurales sur C14 (prédite comme étant également une protéine désordonnée) ne permettent pas de savoir il s'agit de deux protéines analogues ou homologues. Les surfaces de dimérisation des dimères C14-Pcc1 et Pcc1-Pcc1 se chevauchent : la liaison de C14 à Pcc1 empêche celle de Pcc1. C14, tout comme Gon7, aurait donc évolué afin de réguler négativement la dimérisation de Pcc1, et donc de KEOPS chez certains organismes (Wan et al., 2017).

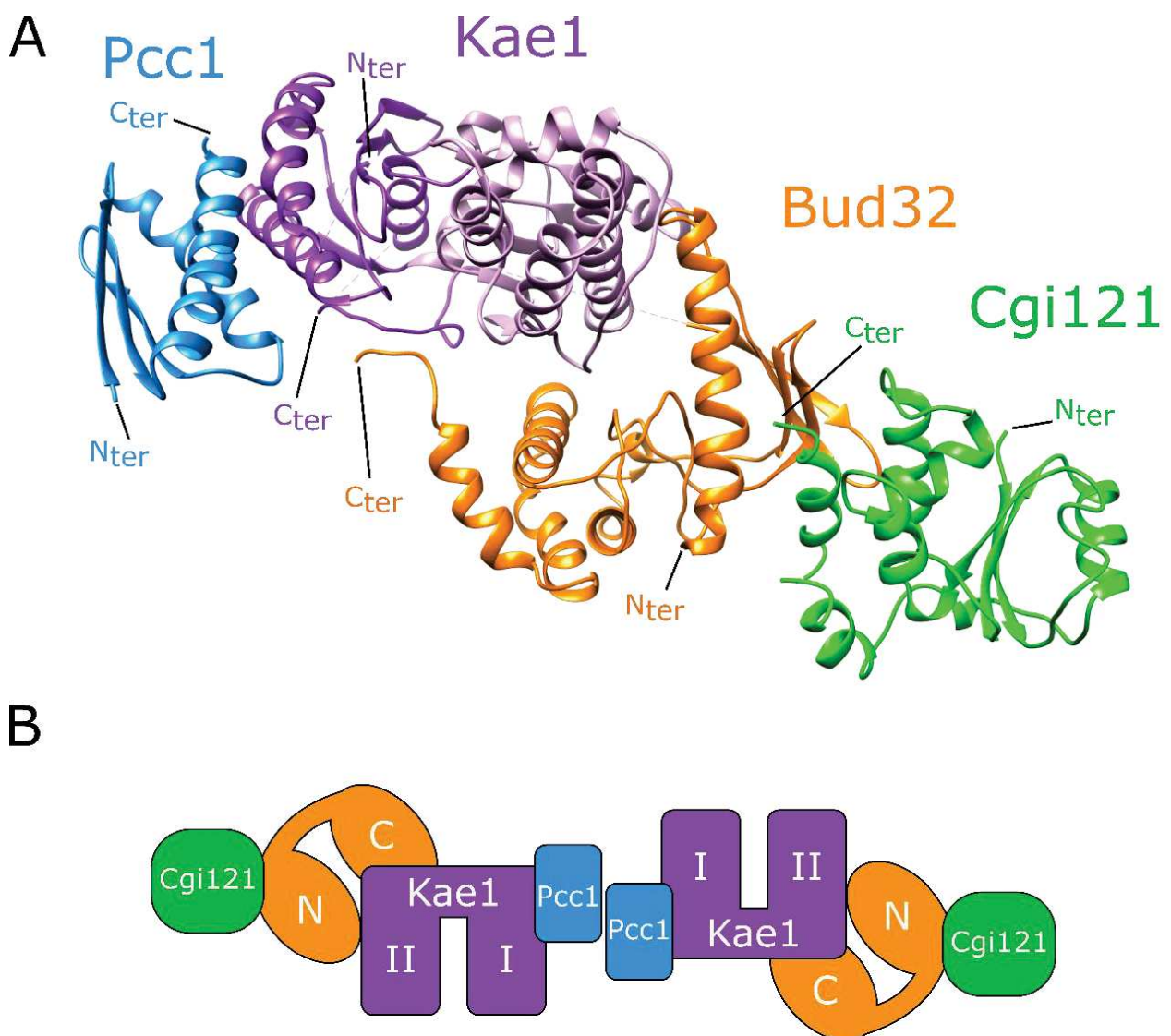


Figure 71. Organisation du complexe KEOPS des archées et des eucaryotes

(A) Modélisation du complexe KEOPS entier à partir des structures disponibles pour les sous-complexes Pcc1-Kae1 de *Thermoplasma acidophilum* (code PDB : 3ENO) et Kae1-Bud32-Cgi121 de *M. jannashii* (code PDB : 3ENH) (Mao et al., 2008). **(B)** Représentation schématique du dimère KEOPS-KEOPS.

Structure quaternaire de KEOPS

Aucune structure complète de KEOPS n'est résolue à ce jour. Cependant les structure des sous-complexes issus d'archées ainsi que des études de SAXS de celui de *S. cerevisiae*, entre autres, attribuent à KEOPS une architecture linéaire de type Cgi121-Bud32-Kae1-Pcc1 c'est à dire où Cgi121 se lie à Bud32, qui à son tour se lie à Kae1, qui à son tour lie Pcc1 ([Figure 71](#)) (Mao et al., 2008; Zhang et al., 2015b). Chez les champignons et les mammifères, le cinquième élément, Gon7 ou C14, vient à son tour se lier à Pcc1.

Cgi121 interagit avec le lobe N-terminal de Bud32 (Mao et al., 2008; Zhang et al., 2015b). L'interface Cgi121-Bud32 est la moins conservée au sein de KEOPS. Cette interaction provoque des changements conformationnels locaux mais significatifs au sein de Bud32. Bien qu'elle la stimule, Cgi121 n'est cependant pas requise pour la synthèse de t⁶A. La liaison de Cgi121 permettrait principalement de stabiliser le complexe (Perrochia et al., 2013a).

L'interaction entre Kae1 et Bud32 constitue le cœur du complexe KEOPS (Hecker et al., 2008; Mao et al., 2008). Bud32 interagit via ses deux lobes avec principalement le domaine II de Kae1. L'extrémité C-terminale de Bud32, qui est conservée et chargée positivement, se positionne à l'entrée du site actif de Kae1 (Zhang et al., 2015b). Dans cette conformation Bud32 interfère avec la liaison des ligands. L'activité ATPase de Bud32 pourrait engendrer un changement de conformation se traduisant par un déplacement de l'extrémité C-terminale de Bud32 afin de rendre le site actif de Kae1 accessible (Perrochia et al., 2013a; Zhang et al., 2015b).

Pcc1 interagit avec le domaine I de Kae1 sans entraîner de réarrangements conformationnels de ce dernier (Hecker et al., 2007; Mao et al., 2008). Pcc1 joue le rôle de plateforme de dimérisation pour le complexe lui conférant alors une structure quaternaire de type Cgi121-Bud32-Kae1-Pcc1=Pcc1-Kae1-Bud32-Cgi121. Cette dimérisation est observée en solution pour des complexes issus de diverses archées (Perrochia et al., 2013a; Wan et al., 2016). Cependant chez les champignons, la présence de Gon7 est susceptible d'empêcher la dimérisation de Pcc1. Une analyse SEC-MALLS du KEOPS de *S. cerevisiae* confirme que le complexe ne forme pas de dimère en présence de Gon7 (Zhang et al., 2015b). Curieusement, l'intégrité de la surface de dimérisation de Pcc1 de KEOPS d'archées n'est pas requise pour l'activité *in vitro*, c'est-à-dire que le complexe est actif à l'état monomérique (Wan et al., 2016). A l'inverse, cette interface est requise chez la levure *in vivo* (Mao et al., 2008). Cette différence de comportement peut s'expliquer par les limites du modèle *in vitro*, qui ne reflète pas la complexité d'une cellule, ou par des différences fonctionnelles authentiques entre les KEOPS archée et levure.

Le KEOPS humain se purifie quant à lui sous forme d'un dimère, d'après des expériences de co-immunoprécipitation faites dans des cellules HEK293T (Costessi et al., 2012). Cependant la découverte récente de C14 chez les métazoaires vient bousculer ce que l'on pensait savoir du complexe humain. Comme Gon7, C14 se lie avec le Pcc1 humain et est donc susceptible d'empêcher la dimérisation du complexe KEOPS. En outre, l'activité de synthèse de t⁶A par le KEOPS humain est 4 fois plus importante en présence de C14 qu'en son absence (Wan et al., 2017). Au final, il reste beaucoup à apprendre sur les modalités de la dimérisation du complexe KEOPS et sur son rôle pour l'activité du complexe.

D'autres fonctions pour KEOPS ?

Il est difficile de discriminer le phénotype provoqué par l'absence de t⁶A du phénotype provoqué par la perte d'une autre activité encore inconnue de KEOPS. Le second étant masqué par les effets pléiotropes du premier. Une approche a été tentée avec l'étude de la souche $\Delta cgi121$ de *S. cerevisiae* (Peng et al., 2015). En effet, Cgi121 n'est pas essentiel *in vivo* pour la synthèse de t⁶A. Les effets provoqués par sa délétion ne sont donc pas provoqués par une absence de t⁶A. Les résultats de cette méthode semblent indiquer que Cgi121 est impliqué dans la formation de coupure simple brin au niveau des télomères, ce qui aurait indirectement un rôle positif sur la recombinaison homologue des télomères (corrélé à une augmentation de la durée de vie des cellules).

En outre, Kae1 de *P. abyssi* se lie spécifiquement à l'ADN simple brin ou double brin d'une manière coopérative et possède même une activité endonucléase spécifique des sites apuriniques (Hecker et al., 2007). Chez l'humain, le gène de l'orthologue Kae1, aussi nommé OSGEP (d'après l'annotation initiale *O-sialglycoendopeptidase protein*), est exprimé de manière ubiquitaire et est co-transcrit avec un autre gène de ménage codant pour une endonucléase impliquée dans la réparation de l'ADN (Seki et al., 2002)

Le complexe KEOPS humain, via le dimère Kae1-Pcc1, est capable de lier la protéine PRAME (*preferentially expressed antigen in melanoma*). PRAME est fréquemment surexprimée lors de l'oncogenèse bien qu'on ignore son rôle dans ce processus. PRAME est capable de se lier spécifiquement à la chromatine et est une sous-unité de la *Cullin2-based E3 ubiquitin ligase*, ce qui suggère un lien fonctionnel entre l'ubiquitination et les propriétés oncogénique de PRAME (Costessi et al., 2011). Lorsque PRAME s'associe à KEOPS, elle recrute la *Cullin2-based E3 ubiquitin ligase* sur ce dernier (Costessi et al., 2012). En complexe avec la chromatine, PRAME reste capable de lier KEOPS (Costessi et al., 2012). Les complexes *Cullin2-based E3 ubiquitin ligase* et KEOPS semblent donc potentiellement fonctionnellement liés.

Toujours chez l'humain, l'homologue de Bud32 est appelé PRPK pour *p53 regulating protein kinase*. En effet, PRPK phosphoryle et active la protéine p53, régulateur transcriptionnel central du cycle cellulaire chez *H. sapiens* (Abe et al., 2001). De même, au sein de cellules neuronales primaires immortalisées de rat, PRPK est capable de réguler négativement l'activité de Rab35, un facteur clé du système de transport de l'endosome, qui provoque une élongation des axones chez ces dites cellules (Villarroel-Campos et al., 2016).

Les études de délétions des sous-unités de KEOPS chez la levure ont montré que l'inactivation de KEOPS provoquait en plus d'une absence de t⁶A, une réduction de la taille des télomères (Downey et al., 2006). La transcomplémentation par l'orthologue mitochondrial Qri7 restaure le niveau de t⁶A mais pas la taille des télomères ce qui suggère que ce phénotype n'est pas due à une absence de t⁶A (Wan et al., 2013). Chez l'humain, en plus de l'identification de C14, l'étude de l'interactome de KEOPS ou de ses sous-complexes a révélé près de 152 interactants spécifiques. Ces résultats font écho à ceux d'autres études réalisées depuis quelques années et semblent indiquer que les sous-unités du KEOPS aient un ou des rôles supplémentaires pour la cellule, autre la synthèse de t⁶A (Costessi et al., 2012; Wan et al., 2017).

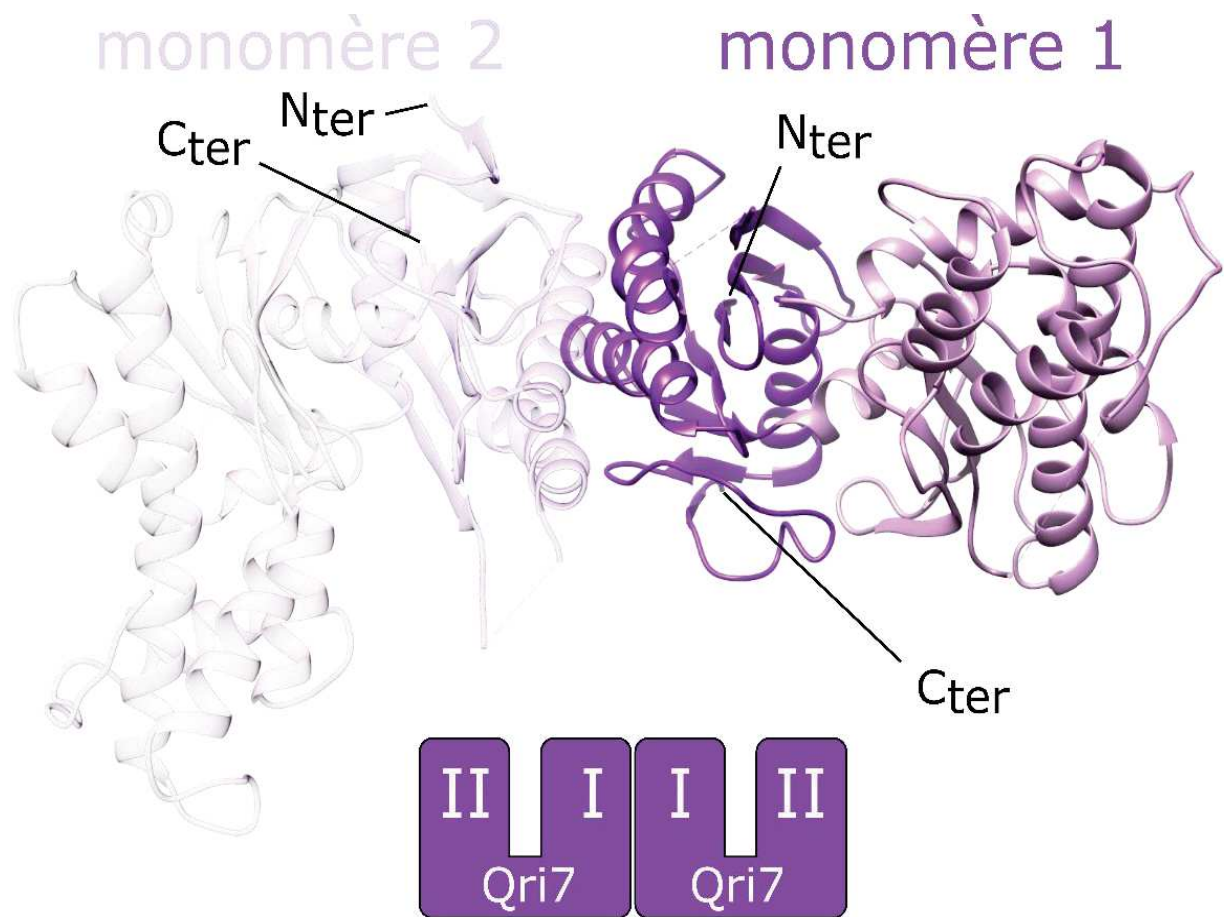


Figure 72. L'homodimère Qri7 des mitochondries

Structure de Qri7 de *S. cerevisiae* (code PDB : 4K25) (Wan et al., 2013). Seul le monomère 1 est montré en opaque. Le dimère est schématiquement représenté en dessous.

d. L'orthologue Qri7 chez les mitochondries

Les mitochondries ont des ribosomes spécialisés qui partagent certaines caractéristiques avec les ribosomes bactériens et cytoplasmiques (Amunts et al., 2014). Un dysfonctionnement de la traduction mitochondriale engendre des phénotypes sévères chez les cellules eucaryotes, tels que des maladies neurodégénératives chez l'humain (De Luca et al., 2006). La plupart des ARNt mitochondriaux sont synthétisés et maturés au sein des mitochondries elles-mêmes (Schneider, 2011). La t⁶A est l'un des nucléosides modifiés présents dans les ARNt mitochondriaux. Kae1 étant strictement cytoplasmique, la mitochondrie a recours à son orthologue Qri7 pour synthétiser la t⁶A. Qri7 est codée par le génome chromosomique et synthétisée dans le cytoplasme avant d'être importée dans la mitochondrie. Contrairement à Kae1 et YgjD, Qri7 ne nécessite aucun partenaire pour réaliser la synthèse de t⁶A et forme un homodimère en solution capable de lier le TC-AMP et l'ARNt (Figure 72) (Wan et al., 2013). Des mutations au niveau de l'interface du dimère engendrent une perte d'activité de Qri7 *in vitro* (Wan et al., 2013). Contrairement à KEOPS et DEZ, le système de t⁶A synthétase mitochondrial fonctionne sans avoir recours à une activité ATPase (Wan et al., 2013).

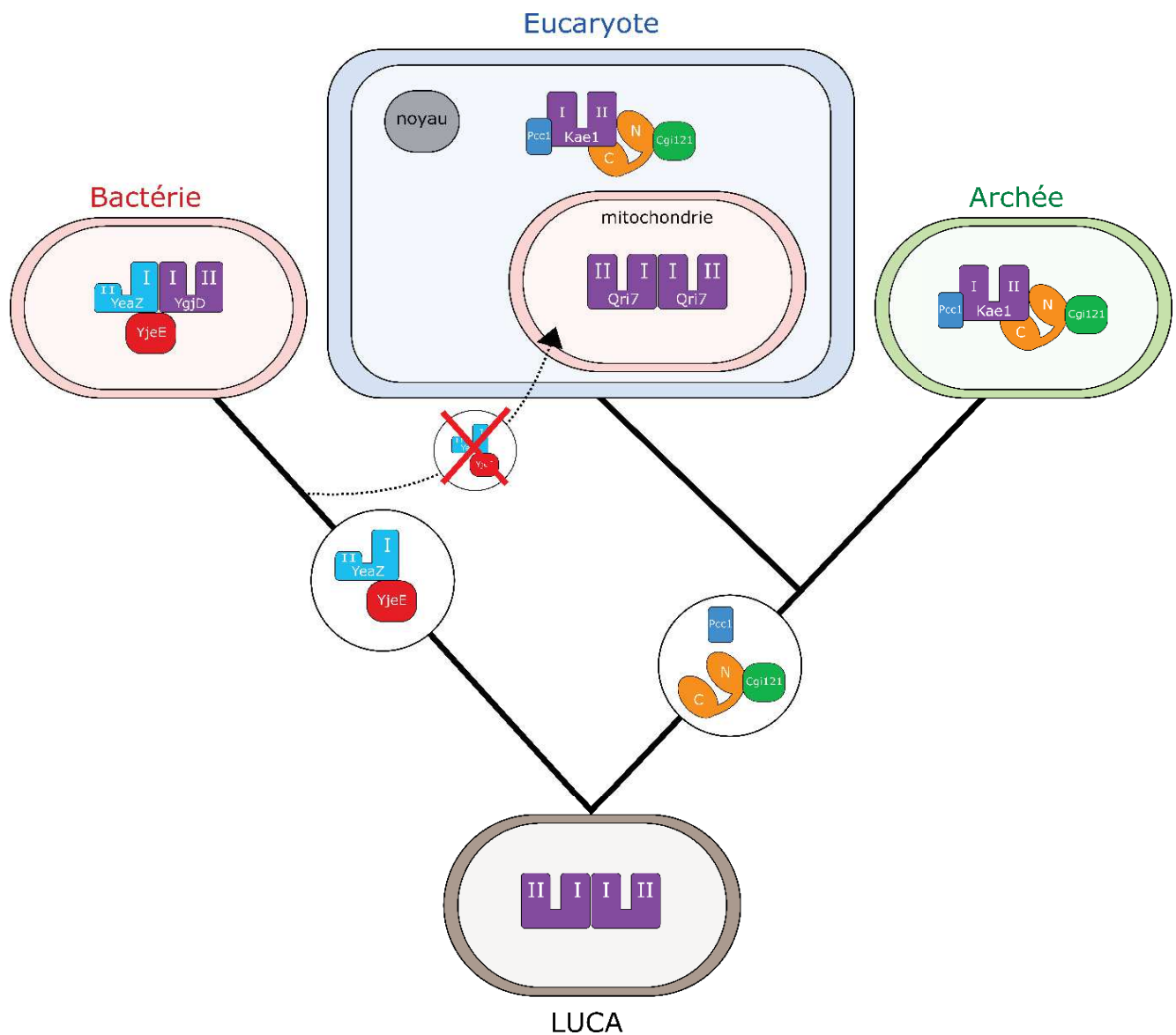


Figure 73. Scénario proposé pour l'évolution de la famille Kae1/YgjD/Qri7

Schéma récapitulatif de l'évolution théorique de la famille Kae1/YgjD/Qri7 depuis LUCA. L'évolution de la famille Kae1/YgjD/Qri7, telle que représentée dans cette figure, est discutée dans le texte principal.

e. Evolution de la famille Kae1/YgjD/Qri7

La t⁶A synthétase mitochondriale est la plus « simple » puisqu'elle est composée de Qri7 et de Sua5/YrdC (Thiaville et al., 2014; Wan et al., 2013). Les deux protéines universelles, sans aucun partenaire, sont donc suffisantes. Bien que le système mitochondrial ait été hérité du système bactérien, cette voie minimale supporte l'idée que LUCA réalisait la synthèse de la t⁶A avec un système similaire, composé uniquement des gènes ancestraux de Kae1/YgjD/Qri7 et de Sua5/YrdC. Des partenaires domaine-spécifique auraient alors été acquis au cours de l'évolution pour former les complexes KEOPS et DEZ (Figure 73). Ces protéines sont désormais essentielles et participent à la régulation de l'activité des complexes ainsi que probablement dans des processus cellulaires indépendant de t⁶A. Cependant la fonction précise de chaque sous-unité reste incomprise.

En particulier, les modalités de liaison de l'ARNt ne sont pas connues. Qri7 est capable de le lier sans avoir recours à des partenaires protéiques (Wan et al., 2013). En revanche, des études d'EMSA ont montré que, bien que Kae1 puisse lier l'ARNt seule, cette liaison est significativement stimulée en présence de Pcc1 (Perrochia et al., 2013a; Wan et al., 2013; Zhang et al., 2015a). Curieusement, aucune interaction avec DEZ n'a pu être détectée par cette même méthode, ce qui suppose que la liaison de l'ARNt sur ce complexe est plus faible et probablement transitoire (Zhang et al., 2015a). Dans les trois systèmes contemporains, la liaison d'un partenaire (Qri7, Pcc1 ou YeaZ) au domaine I de Kae1/YgjD/Qri7 est essentielle pour l'activité de synthèse de t⁶A et est donc potentiellement impliquée dans la liaison de l'ARNt substrat.

De manière intéressante, les interfaces de dimérisation des couples Pcc1-Kae1, Qri7-Qri7 et YgjD-YeaZ sont étrangement similaires entre elles (Wan et al., 2013). La conservation de cette interface est donc soumise à une pression évolutive forte, probablement liée à la fonction de cette famille, bien qu'on ignore encore son rôle exact (Wan et al., 2017). Par inférence, la forme ancestrale de LUCA fonctionnait donc probablement en tant qu'homodimère. D'autres observations également intéressantes sont que Qri7 forme un homodimère, YeaZ est un paralogue d'YgjD et enfin la structure de Pcc1 ressemble à celle du domaine I de Kae1. Ceci suggère qu'YeaZ et Pcc1 auraient évolué à partir de l'homodimère ancestral, suite à une duplication de gène, et auraient finalement divergés et acquis des fonctions différentes chez les organismes.

Annexe 2 : Chapitre : la synthèse de t⁶A chez les archées

Dans le cadre de la publication d'un livre dédié à la biologie des archées (en cours d'édition), nous avons écrit un chapitre visant à résumer les connaissances actuelles sur la voie de synthèse de la t⁶A, en se focalisant particulièrement sur le système Archée.

Function and biosynthesis of the universal tRNA modification N⁶-threonylcarbamoyl-adenosine

Adeline Pichard-Kostuch, Marie-Claire Daugeron, Patrick Forterre, Tamara Basta

Abstract:

Transfer RNAs (tRNAs) are essential components of the translation machinery that reads the genetic message and translates it into polypeptides. Apart from the four canonical nucleotides A, C, U and G, tRNAs contain a variety of modified nucleosides which are formed enzymatically during maturation process. Nucleosides 34 (wobble base) and 37 (dangling base) in the anticodon loop of tRNAs are frequently modified. These modifications were shown to be important for the efficiency and fidelity of translation. This chapter focuses on one of these modified nucleosides, N⁶-threonylcarbamoyl adenosine (t⁶A₃₇), which occurs exclusively at the position 37 of all tRNAs decoding ANN codons (N = A, C, U or G). Initial biochemical and structural studies established the function of this universal modification in facilitating the binding of tRNA on ribosome and preventing frameshifting in course of translation. Recently, the genes encoding the synthetic machinery for the biosynthesis of t⁶A₃₇ were discovered in the three domains of life providing the possibility to study the reaction mechanism and measure the effect of t⁶A₃₇ on cellular physiology. We describe in this chapter these experimental characterizations with the focus on the archaeal genes and proteins and the homologous eukaryotic system.

Annexe 3 : Table des oligonucléotides

Gene-mutation	Vector	Encoded protein	F or R	Sequence 5' -> 3'
ARTICLE I				
A) Directed mutagenesis of <i>Pa</i> -Sua5 :				
PAB1302-(H ₂₃₄ A/Y ₂₃₅ A)	pEt26b(+)	<i>Pa</i> -Sua5-(H ₂₃₄ A/Y ₂₃₅ A) ^{6H}	F	CCCGGGCATGAAGTACAGGGCCGCTGCTCCAAGTGCTGAAGTT
			R	AACTTCAGCACTTGGAGCAGCGGCCCTGTA
PAB1302-(P ₂₂₈ A/G ₂₂₉ A/M ₂₃₀ A)		<i>Pa</i> -Sua5-(P ₂₂₈ A/G ₂₂₉ A/M ₂₃₀ A) ^{6H}	F	CGATACCGCGAAGGCCGCGCCGCGAAGTACAGGCACTATG
			R	CATAGTGCCTGTACTTCGCGGCCGCGGCCTTCGCGGTATCG
PAB1302-(P ₂₃₇ A)		<i>Pa</i> -Sua5-(P ₂₃₇ A) ^{6H}	F	TATGCTGCAAGTGCTGAAGTTATAGTGGT
			R	CAGCACTTGCAGCATAGTGCCTGTA
PAB1302-(P ₂₃₇ S)		<i>Pa</i> -Sua5-(P ₂₃₇ S) ^{6H}	F	TATGCTTCAAGTGCTGAAGTTATAGTGGT
			R	CAGCACTTGAAGCATAGTGCCTGTA
PAB1302-(R ₃₀₁ A)		<i>Pa</i> -Sua5-(R ₃₀₁ A) ^{6H}	F	GCG TTG GCG CAC ATG GAT AGG ACT GGG
			R	ATC CAT GTG CGC CAA CGC TTT GAA TAA GTT CC
PAB1302-(R ₃₀₁ D)		<i>Pa</i> -Sua5-(R ₃₀₁ D) ^{6H}	F	GCG TTG GAC CAC ATG GAT AGG ACT GGG
			R	ATC CAT GTG GTC CAA CGC TTT GAA TAA GTT CC
PAB1302-(R ₃₂₈ A)		<i>Pa</i> -Sua5-(R ₃₂₈ A) ^{6H}	F	GTT ATG AAC GCG CTT AGG AAG GCC TCC GG
			R	CTT CCT AAG CGC GTT CAT AAC GGC CAA GCC
PAB1302-(R ₃₂₈ D)		<i>Pa</i> -Sua5-(R ₃₂₈ D) ^{6H}	F	GTT ATG AAC GAC CTT AGG AAG GCC TCC GG
			R	CTT CCT AAG GTC GTT CAT AAC GGC CAA GCC
PAB1302-(Δ240-340) ^{6H}		<i>Pa</i> -Sua5-(Δ240-340) ^{6H}	F-NdeI	GGAATTCATATGACTATCATAATAAACGTGAGGGAGAGAATC
			R-XhoI	CCCCCCCCCTCGAGTTAGTGATGGTGATGGTGATGAGCACTTGGAGCATAGTGCCTGTAC
PAB1302- ^{6H} (Δ240-340)		<i>Pa</i> -Sua5- ^{6H} (Δ240-340)	F-NdeI	GGAATTCATATGCATCACCATCACCATCACA
			R-XhoI	CCCCCCCCCTCGAGTTAAGCACTTGGAGCATAGTGCCTGTAC
PAB1302-(Δ1-239) ^{Strep}	<i>Pa</i> -Sua5-(Δ1-239) ^{Strep}	F-NdeI	GGAATTCATATGGAAGTTATAGTGGTCGAAGGGCCAAGGG	
		R-XhoI	CCCCCCCCCTCGAGTTATTTTCGAACTGCGGGTGGTCCAACTTTGATTATCCTATACCCGAGGCC	
B) Construction of YcpLac33-Sc-SUA5(wt)				
Sc-SUA5-(wt)-(entire gene)	YcpLac33	Sc-SUA5-(wt)	F-SalI	GCAGTCGACGAAGCGCAAGCAACAAAT
			R-SacI	CTCGAGCTCTACGCCGTTGGACTGTCT
C) Construction of pESC-LEU plasmids for complementation				
PAB1302-(full-lenght)	pESC-LEU	<i>Pa</i> -Sua5- ^{FLAG} (full-lenght)	F-BglII	TTTAGATCTCCATGACTATCATAATAAACTGAGG
			R-SacI	CCCGAGCTCAAACTTTGATTATCCTATACCCGGAG
SUA5-(wt)		Sc-SUA5- ^{FLAG} (wt)	F-BamHI	TTTGGATCCTTATGTACCTTGGACGACATTTTTTG
			R-PacI	CCCTTAATTAAACTGTATACAATTATTTGCAGCC
		Sc-SUA5-(wt) ^{Myc}	F-BamHI	TTTGGATCCTTATGTACCTTGGACGACATTTTTTG
			R-SalI	CCCGTCGACAACTGTATACAATTATTTGCAGCCG
		Sc-SUA5-(wt)	F-BamHI	TTTGGATCCTTATGTACCTTGGACGACATTTTTTG
			R-SalI	CCCGTCGACTTAAACTGTATACAATTATTTGCAG
SUA5-(Δ253-426)		Sc-SUA5- ^{FLAG} (Δ253-426)	F-BamHI	TTTGGATCCTTATGTACCTTGGACGACATTTTTTG
			R-PacI	CCCTTAATTA AAAACCGCCAGGTCGAGTAGCGTA

SUA5-(Δ297-426)		Sc-SUA5-FLAG(Δ297-426)	F-BamHI	TTTGGATCCTTATGTACCTTGGACGACATTTTTTG
			R-PacI	GGGTTAATTAAGCTGAAGGGGAATAGTGTCTATAC
SUA5-(Δ1-196)		Sc-SUA5-(Δ1-196) ^{Myc}	F-BamHI	CCCGGATCCATGAAAGTTGTTTTACTAGTTCCTCA
			R-Sall	CCCGTCGACAAACTGTATACAATTATTTGCAGCCG
SUA5-(Δ1-252)		Sc-SUA5-(Δ1-252) ^{Myc}	F-BamHI	CCCGGATCCATGACATATGAAGAAATTGTTAAATT
			R-Sal	CCCGTCGACAAACTGTATACAATTATTTGCAGCCG

D) Primers used for sequencing of pESC-LEU plasmids				
pESC-LEU-pGAL10	pESC-LEU		F	GGTGGTAATGCCATGTAATATG
			R	GGCAAGGTAGACAAGCCGACAAC
pESC-LEU-pGAL1		F	ATTTTCGGTTTGTATTACTTC	
		R	GTTCTTAATACTAACATAACT	
pESC-LEU-pGAL1-BIS		F	TGCAAAAACGCATAACCAC	
		R	GACCTAGACTTCAGGTTGTC	

ARTICLE II				
Directed mutagenesis of <i>Ap-Sua5</i> :				
ARCPR_RS00515- (Δ199-289)	pEt26b(+)	<i>Ap-Sua5</i> - ^{6H} (Δ199-289)	F-NdeI	CGATCATATGCATCATCATCATCATACCAAAATCTACAAGGTTGACC
R-XhoI			TGACTCGAGTTACAGCGCACCCACACGCAGAACAC	
ARCPR_RS00515- (T ₁₄₁ P)		<i>Ap-Sua5</i> -(T ₁₄₁ P)	F	CCGGTGACCAGCCCCAGCGCGAACGTTAG
			R	CGCGCTGGGGCTGGTCACCGGAACACCGC

Le gène codant pour le mutant *Pa-Sua5*-^{6H}(Δ200-340) a été directement synthétisé par Genscript.

Annexe 4 : Propriétés biochimiques des protéines recombinantes

Plasmid	Protein	Amino- acid	Molecular Mass kDa	1 g/L = μM	pI ^{theo}	Ext coeff	Abs
						M ⁻¹ .cm ⁻¹	(1g/l)
<i>Pyrococcus abyssi</i> (GE5)							
pET-26b(+)	<i>Pa</i> -Sua5-(wt) ^{6H}	346	38,23	26,1	6,38	28420	0,742
	<i>Pa</i> -Sua5-(H ₂₃₄ A/Y ₂₃₅ A) ^{6H}	346	38,11	26,2	6,34	26930	0,784
	<i>Pa</i> -Sua5 (P ₂₂₈ A/G ₂₂₉ A/M ₂₃₀ A) ^{6H}	346	38,2	26,2	6,38	28420	0,745
	<i>Pa</i> -Sua5-(P ₂₃₇ A) ^{6H}	346	38,24	26,15	6,38	28420	0,743
	<i>Pa</i> -Sua5 (P ₂₃₇ S) ^{6H}	346	38,26	26,14	6,38	28420	0,743
	<i>Pa</i> -Sua5-(R ₃₀₁ A/R ₃₂₈ A) ^{6H}	346	38,1	26,24	6,15	28420	0,746
	<i>Pa</i> -Sua5(R ₃₀₁ D/R ₃₂₈ D) ^{6H}	346	38,19	26,19	5,95	28420	0,744
	<i>Pa</i> -Sua5-(Δ240-340) ^{6H}	245	26,89	37,2	6.18	23950	0,887
	<i>Pa</i> -Sua5- ^{6H} (Δ240-340)	245	26,89	37,2	6.18	23950	0,887
	<i>Pa</i> -Sua5- ^{6H} (Δ200-340)	205	22,6	44,3	5,93	19480	0.862
	<i>Pa</i> -Sua5-(Δ1-239) ^{Strep}	113	12,72	78,7	7,95	9970	0,785
pEt-28a(+)	<i>Pa</i> -KEOPS- ^{Pcc1-6H} (wt)	798	90.47	11	8,74	63720	0,704
<i>Archaeoglobus profundus</i> (DSM 5631)							
pET-26b(+)	<i>Ap</i> -Sua5-(wt) ^{6H}	295	32,83	30	6,77	10680	0,325
	<i>Ap</i> -Sua5- ^{6H} (Δ199-289)	204	22,49	45	8,94	5960	0,265
	<i>Ap</i> -Sua5-(T ₁₄₁ P) ^{6H}	295	32,83	30	6,77	10680	0,325
<i>Pandoravirus salinus</i>							
pET-26b(+)	<i>Ps</i> -t ⁶ A-synthetase (wt) ^{6H}	985	103,82	9,6	6,03	77360	0,745

Liste des figures et de tables

Les figures et les tables présentées dans ce manuscrit sont résumées ci-dessous par ordre d'apparition.

Figure 1. Organisation universelle des ribosomes lors de la traduction	13
Figure 2. Les ribosomes Eucaryote et Bactérien.....	15
Figure 3. L'étape d'élongation lors de la synthèse protéique.....	19
Figure 4. Structures secondaire et tertiaire des ARNt.....	25
Figure 5. Système de numérotation des ARNt.....	27
Figure 6. Les aminoacyl-ARNt synthétases	31
Figure 7. Distribution des nucléosides modifiés des ARN	35
Figure 8. Exemples de modifications chimiques des nucléosides.....	37
Figure 9. Les modifications des bases 34 et 37	39
Figure 10. Représentation sphérique du code génétique.	43
Figure 11. Rôle structural de t ⁶ A au niveau de la boucle anticodon.....	45
Figure 12. Structure et rôle de t ⁶ A au niveau du duplex codon-anticodon.....	51
Figure 13. Essentialité de t ⁶ A au sein des trois domaines du vivant	59
Figure 14. Exemples de variants et dérivés de la t ⁶ A	63
Figure 15. Structures proposées du dérivé cyclique de la t ⁶ A.....	65
Figure 16. Structure cristallographique de TcdA d'E. coli	67
Figure 17. Phylogénie de Kae1/YgjD/Qri7 et YeaZ	77
Figure 18. Etude de la voie de biosynthèse de t ⁶ A in cellulo	89
Figure 19. La O-carbamoylase TobZ	93
Figure 20. Fonction cellulaire et structure de HypF	95
Figure 21. Apparence et génome de Pandoravirus salinus	97
Figure 22. Le paralogue YciO	99
Figure 23. Les enzymes Sua5 et YrdC	103
Figure 24. La protéine YrdC d'E. coli.....	105
Figure 25. Structure cristallographique de Sua5 de S. tokodaii.....	107
Figure 26. Coordination de la L-thréonine par Sua5/YrdC.....	109
Figure 27. Coordination de l'ATP par Sua5/YrdC.....	111
Figure 28. La dégradation du TC-AMP	115
Figure 29. Mécanisme proposé pour la synthèse de t ⁶ A-ARNt.....	119
Figure 30. Structure and biosynthesis pathway of t ⁶ A	141
Figure 31. Comparison of the Pa- and St-Sua5 structures.....	163
Figure 32. SEC-MALLS analysis of Pa-Sua5	164
Figure 33. The structure of Sua5 from Pyrococcus abyssi	165
Figure 34. Comparison of the Pa-Sua5 structures bound to PPi or water molecules	167
Figure 35. Linker acts as a lid closing the active site pocket of Sua5 proteins	169
Figure 36. Orientation of the His ²³⁴ -Tyr ²³⁵ motif side chains in the Sua5 structures.....	171
Figure 37. Kinetics of the ATP hydrolysis by Pa-Sua5 ^{wt}	172
Figure 38. Complementation of Δsua5 S. cerevisiae by truncated Sc-Sua5 proteins.....	173
Figure 39. Analysis of recombinant Pa-Sua5 proteins stability	175
Figure 40. Sequence conservation analysis of Pa-Sua5	178
Figure 41. Functional complementation of Δsua5 S. cerevisiae by Pa-Sua5 proteins	179

Figure 42. Validation of the in vivo complementation assays for Pa-Sua5 proteins	182
Figure 43. In vitro ATPase activity of Pa-Sua5 proteins	183
Figure 44. Nucleotide separation by TLC for in vitro ATPase assay	186
Figure 45. In vitro t ⁶ A synthesis activity of Pa-Sua5 proteins	187
Figure 46. In vitro ATPase activity of Pa-Sua5 in presence of Pa-KEOPS	187
Figure 47. Structure of Pa-Sua5 ^{H234A/Y235A}	189
Figure 48. Comparison of the YrdC domains of TobZ and Sua5	193
Figure 49. Structure and properties of the universal modified nucleoside t ⁶ A	203
Figure 50. The Sua5/YrdC universal enzyme family	207
Figure 51. ML phylogeny tree of the YrdC/Sua5 proteins from Archaea	217
Figure 52. Distribution of Sua5 and YrdC across the three domains of life	219
Figure 53: ML phylogeny tree of the YrdC/Sua5 proteins from Methanococcales and Thermococcales and their close relatives.....	223
Figure 54. Alignment of the catalytic domain of Sua5 and YrdC	227
Figure 55. The Sua5-specific residues	230
Figure 56. Nature of the signature residue as a function of the size of Sua5/YrdC	231
Figure 57. Position of the signature threonine residue in TobZ.....	233
Figure 58. Alignment of Sua5 sequences from the Archaeoglobi	237
Figure 59. Structure prediction of Ap-Sua5 ^{wt}	240
Figure 60. Differential scanning calorimetry analysis of Ap-Sua5 proteins	241
Figure 61. In vitro ATPase and t ⁶ A activities of Ap-Sua5 proteins.....	241
Figure 62. Evolutionary model of the Sua5/YrdC family	247
Figure 63. Sequence comparison of the pandoraviruses t ⁶ A synthetases with Sua5/YrdC, Kae1/YgjD and TcdA sequences from model organisms	268
Figure 64. Structure prediction of the t ⁶ A synthetase from <i>P. salinus</i>	269
Figure 65. Genomic context of the pandoraviruses t ⁶ A synthetase genes.....	271
Figure 66. In vitro activity of the pandoravirus t ⁶ A synthetase	273
Figure 67. Structure de Kae1 de <i>M. jannashii</i>	295
Figure 68. Organisation du complexe DEZ des bactéries.....	299
Figure 69. Comparaison des kinases atypiques Bud32 et Rio2.....	305
Figure 70. Structure du dimère Pcc1-Gon7	307
Figure 71. Organisation du complexe KEOPS des archées et des eucaryotes.....	309
Figure 72. L'homodimère Qri7 des mitochondries	317
Figure 73. Scénario proposé pour l'évolution de la famille Kae1/YgjD/Qri7	319

Table 1. Appariements possibles de la base 34 selon l'hypothèse du wobble initiale et sa version modifiée.....	41
Table 2. Les ARNt initiateurs sont modifiés en t ⁶ A-ARNt uniquement par la machinerie eucaryote.	61
Table 3. Résumé des nomenclatures existantes pour les t ⁶ A synthétases	87
Table 4. Résumé des affinités d'YrdC d'E. coli pour les acides nucléiques.....	121
Table 5. Les maladies humaines liées à des défauts de modifications des ARNt.....	123
Table 6. The X-ray data collection and structural model refinement statistics.....	153
Table 7. Distribution of Sua5-users across the three domains of life	221
Table 8. Distribution of Sua5-users in the (hyper)thermophilic and non-hyperthermophilic archaea.....	243
Table 9. Closest homologs of Ps-t ⁶ A and its neighboring genes	265

Liste des abréviations

Les abréviations utilisées dans le texte principal sont résumées ci-dessous par ordre alphabétique.

5FOA = Acide 5-fluoroorotic

A = Adénosine

A. profundus = *Archaeoglobales profundus*

A. thaliana = *Arabidopsis thaliana*

aa-AMP = Aminoacyl-AMP

aaRS = Aminoacyl-tRNA-synthetase

AC = Anticodon

ACP = Acylphosphatase

ADP = Adénosine diphosphate

agm²C = Agmatidine

AMP = Adénosine monophosphate

AMPPNP = Adénylyl-imidodiphosphate

ARN = Acide ribonucléique

ARNm = ARN messenger

ARNr = ARN ribosomique

ARNt = ARN de transfert

ATP = Adénosine triphosphate

ATPase = Adénosine triphosphate hydrolase

BCA = Biologie Cellulaire des Archées

C = Cytidine

C-AMP = carbamoyl-AMP

COG = Cluster of Orthologous Genes

CP = Carbamoyl-phosphate

ct⁶A = Cyclic-t⁶A

D = Dihydrouridine

D. melanogaster = *Drosophila melanogaster*

DEZ = YgjD-YeaZ-YjeE complexe

DHBP = 3,4-dihydroxy-2-butanone-4-phosphate

DLS = Dynamic light scattering

DSC = Differential Scanning Calorimetry

E. coli = *Escherichia coli*

EF = Elongation factor

FAAM = Fonction et Architecture des Assemblages Macromoléculaires

G = Guanosine

G+ = Archaeosine

GF = Gel filtration

GTPase = Guanosine triphosphate hydrolase

H. influenzae = *Haemophilus influenzae*

I = Inosine

i⁶A = N⁶-isopentenyl-adenosine

IF = Initiation factor

IleRS = Isoleucine RNAt synthetase

IPTG = Isopropyl β-D-1-thiogalactopyranoside

IRIP = Ischemia/reperfusion inducible protein

k²C = Lysidine

KEOPS = Kinase Endopeptidase and Other Proteins of small Size

LUCA = Last Universal Common Ancestor

M. jannaschii = *Methanocaldococcus jannaschii*

M. musculus = *Mus musculus*

m¹G = N¹-méthyl-guanosine

m⁶t⁶A = N⁶-méthyl-t⁶A

ms²ct⁶A = 2-méthylthio-cyclic-t⁶A

ms²t⁶A = 2-méthylthio-t⁶A

OSGEP = O-sialglycoendopeptidase protein

P. salinus = *Pandoravirus salinus*

PPi = Pyrophosphate inorganique

PRAME = Preferentially expressed antigen in melanoma

protéine r = protéine ribosomique

PRPK = p53 regulating protein kinase

PTC = Peptidyl transferase center

Q = Queuosine

RBBP10 = Retinoblastoma binding protein 10

RF = Release factor

RimN = Ribosome maturation N

RMN = Résonance magnétique nucléaire

RNase = Ribonucléase

RNP = Ribonucléoprotéique

RRF = Ribosome recycling factor

S = Svedberg

S. cerevisiae = *Saccharomyces cerevisiae*

S. pombe = *Schizosaccharomyces pombe*

S. tenebrarius = *Streptoalloteichus tenebrarius*

s²U = 2-méthylthiouridine

s²U = 2-thiouridine

SD = Shine-Dalgarno

SEC-MALLS = Size Exclusion Chromatography - Multi-Angle Laser Light Scattering

SU = Sous-unité

SUA = Suppressor of upstream AUG

T = Ribothymidine

t⁶A = N⁶-threonyl-carbamoyl adenosine

TCA = Acide trichloroacétique

TC-AMP = Thréonyl-carbamoyl-AMP

TcdA = Threonylcarbamoyladenose dehydratase A

Tcs = Threonyl-carbamoyl synthesis

TIR = Translation Initiation Region

T_m = Melting temperature

Tsa = Threonylcarbamoyl-6-adenosine

U = Uridine

UTR = Untranslated region

Y (ou Y) = Pseudouridine

yW = Wybotusine

Références bibliographiques

- Abe, Y., Matsumoto, S., Wei, S., Nezu, K., Miyoshi, A., Kito, K., Ueda, N., Shigemoto, K., Hitsumoto, Y., Nikawa, J., et al. (2001). Cloning and characterization of a p53-related protein kinase expressed in interleukin-2-activated cytotoxic T-cells, epithelial tumor cell lines, and the testes. *J. Biol. Chem.* 276, 44003–44011.
- Abelson, J., Trotta, C.R., and Li, H. (1998). tRNA splicing. *J. Biol. Chem.* 273, 12685–12688.
- Achenbach, J., and Nierhaus, K.H. (2015). The mechanics of ribosomal translocation. *Biochimie* 114, 80–89.
- Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.*
- Adamiak, R.W., and Górnicki, P. (1985). Hypermodified nucleosides of tRNA: synthesis, chemistry, and structural features of biological interest. *Prog. Nucleic Acid Res. Mol. Biol.* 32, 27–74.
- Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., Kuramitsu, S., and Shinkai, A. (2008). X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins* 70, 1108–1111.
- Agris, P.F. (1991). Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* 73, 1345–1349.
- Agris, P.F. (2008). Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.* 9, 629–635.
- Agris, P.F. (2015). The importance of being modified: an unrealized code to RNA structure and function. *RNA* 21, 552–554.
- Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* 366, 1–13.
- Ahel, I., Korencic, D., Ibba, M., and Söll, D. (2003). Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15422–15427.
- Allali-Hassani, A., Campbell, T.L., Ho, A., Schertzer, J.W., and Brown, E.D. (2004). Probing the active site of YjeE: a vital *Escherichia coli* protein of unknown function. *Biochem. J.* 384, 577–584.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.
- Amunts, A., Brown, A., Bai, X., Llácer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343, 1485–1489.

Arigoni, F., Talabot, F., Peitsch, M., Edgerton, M.D., Meldrum, E., Allet, E., Fish, R., Jamotte, T., Curchod, M.L., and Loferer, H. (1998). A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* *16*, 851–856.

Armache, J.-P., Anger, A.M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G.J., Beckmann, R., et al. (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Res.* *41*, 1284–1293.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* *44*, W344–350.

Ashraf, S.S., Ansari, G., Guenther, R., Sochacka, E., Malkiewicz, A., and Agris, P.F. (1999). The uridine in “U-turn”: contributions to tRNA-ribosomal binding. *RNA N. Y. N* *5*, 503–511.

Aström, S.U., and Byström, A.S. (1994). Rit1, a tRNA backbone-modifying enzyme that mediates initiator and elongator tRNA discrimination. *Cell* *79*, 535–546.

Baidya, N., and Uhlenbeck, O.C. (1995). The role of 2'-hydroxyl groups in an RNA-protein interaction. *Biochemistry (Mosc.)* *34*, 12363–12368.

Balczun, C., and Scheid, P.L. (2017). Free-Living Amoebae as Hosts for and Vectors of Intracellular Microorganisms with Public Health Significance. *Viruses* *9*.

Baldwin, R.L. (1953). Sedimentation coefficients of small molecules: methods of measurement based on the refractive-index gradient curve; the sedimentation coefficient of polyglucose A. *Biochem. J.* *55*, 644–648.

Baralle, F.E., and Brownlee, G.G. (1978). AUG is the only recognisable signal sequence in the 5' non-coding regions of eukaryotic mRNA. *Nature* *274*, 84–87.

Bartlett, J.M.S., Thomas, J., Ross, D.T., Seitz, R.S., Ring, B.Z., Beck, R.A., Pedersen, H.C., Munro, A., Kunkler, I.H., Campbell, F.M., et al. (2010). Mammostrat as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res. BCR* *12*, R47.

Basavappa, R., and Sigler, P.B. (1991). The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J.* *10*, 3105–3111.

Batey, R.T., and Kieft, J.S. (2007). Improved native affinity purification of RNA. *RNA N. Y. N* *13*, 1384–1389.

Bénas, P., Bec, G., Keith, G., Marquet, R., Ehresmann, C., Ehresmann, B., and Dumas, P. (2000). The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA N. Y. N* *6*, 1347–1355.

Benelli, D., and Londei, P. (2009). Begin at the beginning: evolution of translational initiation. *Res. Microbiol.* *160*, 493–501.

- Berg, J.M., Tymoczko, J.L., Stryer, L., Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). *Biochemistry* (W H Freeman).
- Betat, H., Long, Y., Jackman, J.E., and Mörl, M. (2014). From end to end: tRNA editing at 5'- and 3'-terminal positions. *Int. J. Mol. Sci.* *15*, 23975–23998.
- Bezerra, A.R., Guimarães, A.R., and Santos, M.A.S. (2015). Non-Standard Genetic Codes Define New Concepts for Protein Engineering. *Life Basel Switz.* *5*, 1610–1628.
- Biegert, A., and Söding, J. (2009). Sequence context-specific profiles for homology searching. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 3770–3775.
- Blaby, I.K., Phillips, G., Blaby-Haas, C.E., Gulig, K.S., El Yacoubi, B., and de Crécy-Lagard, V. (2010). Towards a systems approach in the genetic analysis of archaea: Accelerating mutant construction and phenotypic analysis in *Haloferax volcanii*. *Archaea Vanc. BC* *2010*, 426239.
- Blokesch, M., Paschos, A., Bauer, A., Reissmann, S., Drapal, N., and Böck, A. (2004). Analysis of the transcarbamoylation-dehydration reaction catalyzed by the hydrogenase maturation proteins HypF and HypE. *Eur. J. Biochem.* *271*, 3428–3436.
- Bloom, J.D., Wilke, C.O., Arnold, F.H., and Adami, C. (2004). Stability and the Evolvability of Function in a Model Protein. *Biophys. J.* *86*, 2758–2764.
- Bornberg-Bauer, E., Huylmans, A.-K., and Sikosek, T. (2010). How do new proteins arise? *Curr. Opin. Struct. Biol.* *20*, 390–396.
- Boyer, M., Madoui, M.-A., Gimenez, G., La Scola, B., and Raoult, D. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PloS One* *5*, e15530.
- Braun, D.A., Rao, J., Mollet, G., Schapiro, D., Daugeron, M.-C., Tan, W., Gribouval, O., Boyer, O., Revy, P., Jobst-Schwan, T., et al. (2017). Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat. Genet.* *49*, 1529–1538.
- Brenneis, M., Hering, O., Lange, C., and Soppa, J. (2007). Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet.* *3*, e229.
- Burggraf, S., Jannasch, H.W., Nicolaus, B., and Stetter, K.O. (1990). *Archaeoglobus profundus* sp. nov., Represents a New Species within the Sulfate-reducing Archaeobacteria. *Syst. Appl. Microbiol.* *13*, 24–28.
- Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* *433*, 531–537.
- Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A.P., Fabris, D., and Agris, P.F. (2011). The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* *39*, D195-201.

- Carmel, D., Dahlström, K.M., Holmström, M., Allahverdiyeva, Y., Battchikova, N., Aro, E.-M., Salminen, T.A., and Mulo, P. (2013). Structural model, physiology and regulation of Slr0006 in *Synechocystis* PCC 6803. *Arch. Microbiol.* *195*, 727–736.
- Carvalho, L.L., Salminen, T.A., and Dahlström, K.M. (2017). Slr0006-like proteins: A TsaC/TsaC2/YciO subfamily exclusive to cyanobacteria. *Mol. Phylogenet. Evol.* *109*, 1–10.
- Cedergren, R.J., Sankoff, D., LaRue, B., and Grosjean, H. (1981). The evolving tRNA molecule. *CRC Crit. Rev. Biochem.* *11*, 35–104.
- Chamberlain, J.R., Lee, Y., Lane, W.S., and Engelke, D.R. (1998). Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev.* *12*, 1678–1690.
- Chang, B., Halgamuge, S., and Tang, S.-L. (2006). Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene* *373*, 90–99.
- Chao, M.C., Pritchard, J.R., Zhang, Y.J., Rubin, E.J., Livny, J., Davis, B.M., and Waldor, M.K. (2013). High-resolution definition of the *Vibrio cholerae* essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Res.* *41*, 9033–9048.
- Chen, J., Ji, C., Gu, S., Zhao, E., Dai, J., Huang, L., Qian, J., Ying, K., Xie, Y., and Mao, Y. (2003). Isolation and identification of a novel cDNA that encodes human yrdC protein. *J. Hum. Genet.* *48*, 164–169.
- Chheda, G.B., Hall, R.H., Mozejko, J., Magrath, D.I., Schweizer, M.P., Stasiuk, L., and Taylor, P.R. (1969). Aminoacyl nucleosides. VI. Isolation and preliminary characterization of threonyladenine derivatives from transfer ribonucleic acid. *Biochemistry (Mosc.)* *8*, 3278–3282.
- Chheda, G.B., Hong, C.I., Piskorz, C.F., and Harmon, G.A. (1972). Biosynthesis of N-(purin-6-ylcarbamoyl)-L-threonine riboside. Incorporation of L-threonine in vivo into modified nucleoside of transfer ribonucleic acid. *Biochem. J.* *127*, 515–519.
- Christian, T., and Hou, Y.-M. (2007). Distinct determinants of tRNA recognition by the TrmD and Trm5 methyl transferases. *J. Mol. Biol.* *373*, 623–632.
- Clarke, M., Lohan, A.J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., Bertelli, C., Schilde, C., Kianianmomeni, A., Bürglin, T.R., et al. (2013). Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* *14*, R11.
- Cochella, L., Brunelle, J.L., and Green, R. (2007). Mutational analysis reveals two independent molecular requirements during transfer RNA selection on the ribosome. *Nat. Struct. Mol. Biol.* *14*, 30–36.
- Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Quérellou, J., Ripp, R., et al. (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* *47*, 1495–1512.

Colson, P., Gimenez, G., Boyer, M., Fournous, G., and Raoult, D. (2011). The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. *PloS One* 6, e18935.

Costessi, A., Mahrour, N., Tijchon, E., Stunnenberg, R., Stoel, M.A., Jansen, P.W., Sela, D., Martin-Brown, S., Washburn, M.P., Florens, L., et al. (2011). The tumour antigen PRAME is a subunit of a Cul2 ubiquitin ligase and associates with active NFY promoters. *EMBO J.* 30, 3786–3798.

Costessi, A., Mahrour, N., Sharma, V., Stunnenberg, R., Stoel, M.A., Tijchon, E., Conaway, J.W., Conaway, R.C., and Stunnenberg, H.G. (2012). The human EKC/KEOPS complex is recruited to Cullin2 ubiquitin ligases by the human tumour antigen PRAME. *PloS One* 7, e42822.

Crick, F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 19, 548–555.

Crick, F.H. (1968). The origin of the genetic code. *J. Mol. Biol.* 38, 367–379.

Crick, F.H., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227–1232.

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10, 210.

Cunningham, R.S., and Gray, M.W. (1974). Derivatives of N-(N-(9-beta-D-ribofuranosylpurin-6-yl)carbamoyl)threonine in phosphodiesterase hydrolysates of wheat embryo transfer ribonucleic acid. *Biochemistry (Mosc.)* 13, 543–553.

Cusack, S. (1995). Eleven down and nine to go. *Nat. Struct. Biol.* 2, 824–831.

Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å. *Nature* 347, 249–255.

Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13, e1006810.

Dalgliesh, C.E. (1953). The template theory and the role of transpeptidation in protein biosynthesis. *Nature* 171, 1027–1028.

Dalluge, J.J., Hamamoto, T., Horikoshi, K., Morita, R.Y., Stetter, K.O., and McCloskey, J.A. (1997). Posttranscriptional modification of tRNA in psychrophilic bacteria. *J. Bacteriol.* 179, 1918–1923.

Daugeron, M.-C., Lenstra, T.L., Frizzarin, M., El Yacoubi, B., Liu, X., Baudin-Baillieu, A., Lijnzaad, P., Decourty, L., Saveanu, C., Jacquier, A., et al. (2011). Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. *Nucleic Acids Res.* 39, 6148–6160.

- Davis, F.F., and Allen, F.W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* 227, 907–915.
- De Luca, C., Besagni, C., Frontali, L., Bolotin-Fukuhara, M., and Francisci, S. (2006). Mutations in yeast mt tRNAs: specific and general suppression by nuclear encoded tRNA interactors. *Gene* 377, 169–176.
- Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V., and Iwata-Reuyl, D. (2012). Biosynthesis of threonylcarbamoyl adenosine (t6A), a universal tRNA nucleoside. *J. Biol. Chem.* 287, 13666–13673.
- Dinman, J.D. (2009). The eukaryotic ribosome: current status and challenges. *J. Biol. Chem.* 284, 11761–11765.
- Downey, M., Houlsworth, R., Maringe, L., Rollie, A., Brehme, M., Galicia, S., Guillard, S., Partington, M., Zubko, M.K., Krogan, N.J., et al. (2006). A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell* 124, 1155–1168.
- Dreher, T.W. (2010). Viral tRNAs and tRNA-like structures. *Wiley Interdiscip. Rev. RNA* 1, 402–414.
- Dreyfus, M., and Heurgué-Hamard, V. (2011). Termination troubles in *Escherichia coli* K12. *Mol. Microbiol.* 79, 288–291.
- Dube, S.K., Marcker, K.A., Clark, B.F., and Cory, S. (1968). Nucleotide sequence of N-formyl-methionyl-transfer RNA. *Nature* 218, 232–233.
- Dumelin, C.E., Chen, Y., Leconte, A.M., Chen, Y.G., and Liu, D.R. (2012). Discovery and biological characterization of geranylated RNA in bacteria. *Nat. Chem. Biol.* 8, 913–919.
- Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* 9, 61.
- Edvardson, S., Prunetti, L., Arraf, A., Haas, D., Bacusmo, J.M., Hu, J.F., Ta-Shma, A., Dedon, P.C., de Crécy-Lagard, V., and Elpeleg, O. (2017). tRNA N6-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3 (OSGEP) mutation manifest by neurodegeneration and renal tubulopathy. *Eur. J. Hum. Genet. EJHG* 25, 545–551.
- El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J.R., Schimmel, P., Swairjo, M.A., and de Crécy-Lagard, V. (2009). The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyl-adenosine in tRNA. *Nucleic Acids Res.* 37, 2894–2909.
- El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., G Murzin, A., and de Crécy-Lagard, V. (2011). A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. *EMBO J.* 30, 882–893.
- El Yacoubi, B., Bailly, M., and de Crécy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* 46, 69–95.

- Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* 6.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347, 203–206.
- Facchin, S., Lopreiato, R., Stocchetto, S., Arrigoni, G., Cesaro, L., Marin, O., Carignani, G., and Pinna, L.A. (2002). Structure-function analysis of yeast piD261/Bud32, an atypical protein kinase essential for normal cell life. *Biochem. J.* 364, 457–463.
- Fahlman, R.P., Dale, T., and Uhlenbeck, O.C. (2004). Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol. Cell* 16, 799–805.
- Farkas, J.A., Picking, J.W., and Santangelo, T.J. (2013). Genetic techniques for the archaea. *Annu. Rev. Genet.* 47, 539–561.
- Ferreira-Cerca, S., Sagar, V., Schäfer, T., Diop, M., Wesseling, A.-M., Lu, H., Chai, E., Hurt, E., and LaRonde-LeBlanc, N. (2012). ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nat. Struct. Mol. Biol.* 19, 1316–1323.
- Filée, J. (2014). Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg? *Virology* 466–467, 53–59.
- Filée, J., Pouget, N., and Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *BMC Evol. Biol.* 8, 320.
- Files, J.G., Weber, K., Coulondre, C., and Miller, J.H. (1975). Identification of the UUG codon as a translational initiation codon in vivo. *J. Mol. Biol.* 95, 327–330.
- Forterre, P., and Gaïa, M. (2016). Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* 31, 44–49.
- Fox, G.E. (2010). Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* 2, a003483.
- Freiberg, C., Wieland, B., Spaltmann, F., Ehlert, K., Brötz, H., and Labischinski, H. (2001). Identification of novel essential *Escherichia coli* genes conserved among pathogenic bacteria. *J. Mol. Microbiol. Biotechnol.* 3, 483–489.
- Freier, S.M., and Tinoco, I. (1975). The binding of complementary oligoribonucleotides to yeast initiator Transfer RNA. *Biochemistry (Mosc.)* 14, 3310–3314.
- Frye, M., and Watt, F.M. (2006). The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. *Curr. Biol. CB* 16, 971–981.

- Fu, T.M., Liu, X., Li, L., and Su, X.D. (2010). The structure of the hypothetical protein smu.1377c from *Streptococcus mutans* suggests a role in tRNA modification. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* 66, 771–775.
- Gabaldón, T., and Koonin, E.V. (2013). Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14, 360–366.
- Galperin, M.Y., and Koonin, E.V. (2004). “Conserved hypothetical” proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–5463.
- Ganoza, M.C., Kiel, M.C., and Aoki, H. (2002). Evolutionary conservation of reactions in translation. *Microbiol. Mol. Biol. Rev. MMBR* 66, 460–485, table of contents.
- Garavaglia, S., Bruzzone, S., Cassani, C., Canella, L., Allegrone, G., Sturla, L., Mannino, E., Millo, E., De Flora, A., and Rizzi, M. (2012). The high-resolution crystal structure of periplasmic *Haemophilus influenzae* NAD nucleotidase reveals a novel enzymatic function of human CD73 related to NAD metabolism. *Biochem. J.* 441, 131–141.
- Gerdes, S., El Yacoubi, B., Bailly, M., Blaby, I.K., Blaby-Haas, C.E., Jeanguenin, L., Lara-Núñez, A., Pribat, A., Waller, J.C., Wilke, A., et al. (2011). Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC Genomics* 12, S2.
- Giegé, R., Sissler, M., and Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26, 5017–5035.
- Gietz, R.D., and Schiestl, R.H. (2007). Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 38–41.
- Goto-Ito, S., Ito, T., Ishii, R., Muto, Y., Bessho, Y., and Yokoyama, S. (2008). Crystal structure of archaeal tRNA(m¹G37)methyltransferase aTrm5. *Proteins* 72, 1274–1289.
- Goto-Ito, S., Ito, T., and Yokoyama, S. (2017). Trm5 and TrmD: Two Enzymes from Distinct Origins Catalyze the Identical tRNA Modification, m¹G37. *Biomolecules* 7.
- Green, R., and Noller, H.F. (1997). Ribosomes and translation. *Annu. Rev. Biochem.* 66, 679–716.
- Grill, S., Gualerzi, C.O., Londei, P., and Bläsi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J.* 19, 4101–4110.
- Grodberg, J., and Dunn, J.J. (1988). ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* 170, 1245–1253.
- Grosjean, H., and Chantrenne, H. (1980). On codon- anticodon interactions. *Mol. Biol. Biochem. Biophys.* 32, 347–367.
- Grosjean, H., and Westhof, E. (2016). An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* 44, 8020–8040.

- Grosjean, H., De Henau, S., Doi, T., Yamane, A., Ohtsuka, E., Ikehara, M., Beauchemin, N., Nicoghossian, K., and Cedergren, R. (1987). The in vivo stability, maturation and aminoacylation of anticodon-substituted *Escherichia coli* initiator methionine tRNAs. *Eur. J. Biochem.* *166*, 325–332.
- Grosjean, H., Sprinzl, M., and Steinberg, S. (1995). Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* *77*, 139–141.
- Grosjean, H., de Crécy-Lagard, V., and Marck, C. (2010). Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* *584*, 252–264.
- Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., Barré, A., Yoshizawa, S., Fourmy, D., de Crécy-Lagard, V., et al. (2014). Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. *PLoS Genet.* *10*, e1004363.
- Grosjean, H.J., de Henau, S., and Crothers, D.M. (1978). On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl. Acad. Sci. U. S. A.* *75*, 610–614.
- Gualerzi, C.O., and Pon, C.L. (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry (Mosc.)* *29*, 5881–5889.
- Hampsey, M., Na, J.G., Pinto, I., Ware, D.E., and Berroteran, R.W. (1991). Extragenic suppressors of a translation initiation defect in the *cyc1* gene of *Saccharomyces cerevisiae*. *Biochimie* *73*, 1445–1455.
- Handford, J.I., Ize, B., Buchanan, G., Butland, G.P., Greenblatt, J., Emili, A., and Palmer, T. (2009). Conserved network of proteins essential for bacterial viability. *J. Bacteriol.* *191*, 4732–4749.
- Harada, F., Matsubara, M., and Kato, N. (1984). Stable tRNA precursors in HeLa cells. *Nucleic Acids Res.* *12*, 9263–9269.
- Harris, K.A., Jones, V., Bilbille, Y., Swairjo, M.A., and Agris, P.F. (2011). YrdC exhibits properties expected of a subunit for a tRNA threonylcarbamoyl transferase. *RNA N. Y. N* *17*, 1678–1687.
- Harris, K.A., Shekhtman, A., and Agris, P.F. (2013). Specific RNA-protein interactions detected with saturation transfer difference NMR. *RNA Biol.* *10*, 1307–1311.
- Harris, K.A., Bobay, B.G., Sarachan, K.L., Sims, A.F., Bilbille, Y., Deutsch, C., Iwata-Reuyl, D., and Agris, P.F. (2015). NMR-based Structural Analysis of Threonylcarbamoyl-AMP Synthase and Its Substrate Interactions. *J. Biol. Chem.* *290*, 20032–20043.
- Haussuehl, K., Huesgen, P.F., Meier, M., Dessi, P., Glaser, E., Adamski, J., and Adamska, I. (2009). Eukaryotic GCP1 is a conserved mitochondrial protein required for progression of embryo development beyond the globular stage in *Arabidopsis thaliana*. *Biochem. J.* *423*, 333–341.

- Hecker, A., Leulliot, N., Gadelle, D., Graille, M., Justome, A., Dorlet, P., Brochier, C., Quevillon-Cheruel, S., Le Cam, E., van Tilbeurgh, H., et al. (2007). An archaeal orthologue of the universal protein Kae1 is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity in vitro. *Nucleic Acids Res.* *35*, 6042–6051.
- Hecker, A., Lopreiato, R., Graille, M., Collinet, B., Forterre, P., Libri, D., and van Tilbeurgh, H. (2008). Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex. *EMBO J.* *27*, 2340–2351.
- Helm, M., Giegé, R., and Florentz, C. (1999). A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNA^{Lys}. *Biochemistry (Mosc.)* *38*, 13338–13346.
- Higgs, P.G., and Lehman, N. (2015). The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* *16*, 7–17.
- Hinnebusch, A.G. (2017). Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* *42*, 589–611.
- Högenauer, G., Turnowsky, F., and Unger, F.M. (1972). Codon-anticodon interaction of methionine specific tRNAs. *Biochem. Biophys. Res. Commun.* *46*, 2100–2106.
- Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965). STRUCTURE OF A RIBONUCLEIC ACID. *Science* *147*, 1462–1465.
- Holm, L., and Laakso, L.M. (2016). Dali server update. *Nucleic Acids Res.* *44*, W351–355.
- Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* *38*, W545–549.
- Hou, Y.-M. (2010). CCA Addition to tRNA: Implications for tRNA Quality Control. *IUBMB Life* *62*, 251–260.
- Houssier, C., and Grosjean, H. (1985). Temperature jump relaxation studies on the interactions between transfer RNAs with complementary anticodons. The effect of modified bases adjacent to the anticodon triplet. *J. Biomol. Struct. Dyn.* *3*, 387–408.
- Ibba, M., and Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* *69*, 617–650.
- Ikeuchi, Y., Kimura, S., Numata, T., Nakamura, D., Yokogawa, T., Ogata, T., Wada, T., Suzuki, T., and Suzuki, T. (2010). Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat. Chem. Biol.* *6*, 277–282.
- Irwin, M.J., Nyborg, J., Reid, B.R., and Blow, D.M. (1976). The crystal structure of tyrosyl-transfer RNA synthetase at 2.7 Å resolution. *J. Mol. Biol.* *105*, 577–586.

- Ishikura, H., Yamada, Y., Murao, K., Saneyoshi, M., and Nishimura, S. (1969). The presence of N-[9-(beta-D-ribofuranosyl)purin-6-ylcarbamoyl]threonine in serine, methionine and lysine transfer RNA's from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 37, 990–995.
- Jackman, J.E., and Alfonzo, J.D. (2013). Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev. RNA* 4, 35–48.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127.
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166.
- Jakubowski, H., and Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* 56, 412–429.
- von Jan, M., Lapidus, A., Del Rio, T.G., Copeland, A., Tice, H., Cheng, J.-F., Lucas, S., Chen, F., Nolan, M., Goodwin, L., et al. (2010). Complete genome sequence of *Archaeoglobus profundus* type strain (AV18). *Stand. Genomic Sci.* 2, 327–346.
- Jeudy, S., Stelter, M., Coutard, B., Kahn, R., and Abergel, C. (2005). Preliminary crystallographic analysis of the *Escherichia coli* YeaZ protein using the anomalous signal of a gadolinium derivative. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* 61, 848–851.
- Jia, J., Lunin, V.V., Sauvé, V., Huang, L.-W., Matte, A., and Cygler, M. (2002). Crystal structure of the YciO protein from *Escherichia coli*. *Proteins* 49, 139–141.
- Jiang, W., Prokopenko, O., Wong, L., Inouye, M., and Mirochnitchenko, O. (2005). IRIP, a new ischemia/reperfusion-inducible protein that participates in the regulation of transporter activity. *Mol. Cell. Biol.* 25, 6496–6508.
- Jones, R.L., Jaskula, J.C., and Janssen, G.R. (1992). In vivo translational start site selection on leaderless mRNA transcribed from the *Streptomyces fradiae* aph gene. *J. Bacteriol.* 174, 4753–4760.
- Jørgensen, F., Adamski, F.M., Tate, W.P., and Kurland, C.G. (1993). Release factor-dependent false stops are infrequent in *Escherichia coli*. *J. Mol. Biol.* 230, 41–50.
- Jüdes, A., Bruch, A., Klassen, R., Helm, M., and Schaffrath, R. (2016). Sulfur transfer and activation by ubiquitin-like modifier system Uba4•Urm1 link protein urmylation and tRNA thiolation in yeast. *Microb. Cell Graz Austria* 3, 554–564.
- Jühling, F., Mörl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F., and Pütz, J. (2009). tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* 37, D159–162.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.

- Kaczanowska, M., and Rydén-Aulin, M. (2004). Temperature sensitivity caused by mutant release factor 1 is suppressed by mutations that affect 16S rRNA maturation. *J. Bacteriol.* *186*, 3046–3055.
- Kaczanowska, M., and Rydén-Aulin, M. (2005). The YrdC protein--a putative ribosome maturation factor. *Biochim. Biophys. Acta* *1727*, 87–96.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermini, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* *14*, 587–589.
- Kamp, H.D., Patimalla-Dipali, B., Lazinski, D.W., Wallace-Gadsden, F., and Camilli, A. (2013). Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathog.* *9*, e1003800.
- Kang, B.-I., Miyauchi, K., Matuszewski, M., D’Almeida, G.S., Rubio, M.A.T., Alfonzo, J.D., Inoue, K., Sakaguchi, Y., Suzuki, T., Sochacka, E., et al. (2017). Identification of 2-methylthio cyclic N6-threonylcarbamoyladenine (ms2ct6A) as a novel RNA modification at position 37 of tRNAs. *Nucleic Acids Res.* *45*, 2124–2136.
- Kapp, L.D., Kolitz, S.E., and Lorsch, J.R. (2006). Yeast initiator tRNA identity elements cooperate to influence multiple steps of translation initiation. *RNA N. Y. N* *12*, 751–764.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* *10*, 845–858.
- Kilpatrick, M.W., and Walker, R.T. (1981). The nucleotide sequence of the tRNA^{Met} from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Res.* *9*, 4387–4390.
- Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat. Biotechnol.* *28*, 617–623.
- Kim, S., Lee, H., and Park, S. (2015). The Structure of *Escherichia coli* TcdA (Also Known As CsdL) Reveals a Novel Topology and Provides Insight into the tRNA Binding Surface Required for N(6)-Threonylcarbamoyladenine Dehydratase Activity. *J. Mol. Biol.* *427*, 3074–3085.
- Kim, S.H., Quigley, G.J., Suddath, F.L., McPherson, A., Sneden, D., Kim, J.J., Weinzierl, J., and Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* *179*, 285–288.
- Kimura, S., Miyauchi, K., Ikeuchi, Y., Thiaville, P.C., de Crécy-Lagard, V., and Suzuki, T. (2014). Discovery of the β -barrel-type RNA methyltransferase responsible for N6-methylation of N6-threonylcarbamoyladenine in tRNAs. *Nucleic Acids Res.* *42*, 9350–9365.
- Kimura-Harada, F., Von Minden, D.L., McCloskey, J.A., and Nishimura, S. (1972). N-((9- -D-ribofuranosylpurin-6-yl)-N-methylcarbamoyl) threonine, a modified nucleoside

isolated from *Escherichia coli* threonine transfer ribonucleic acid. *Biochemistry (Mosc.)* 11, 3910–3915.

Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J.-C., Ilan, L., Hofmann, K., Namane, A., Mann, C., and Libri, D. (2006). Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J.* 25, 3576–3585.

Koonin, E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136.

Koonin, E.V., and Yutin, N. (2010). Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology* 53, 284–292.

Koski, R.A., and Clarkson, S.G. (1982). Synthesis and maturation of *Xenopus laevis* methionine tRNA gene transcripts in homologous cell-free extracts. *J. Biol. Chem.* 257, 4514–4521.

Kowalak, J.A., Dalluge, J.J., McCloskey, J.A., and Stetter, K.O. (1994). The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry (Mosc.)* 33, 7869–7876.

Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* 108, 229–241.

Kramer, E.B., and Farabaugh, P.J. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA N. Y. N* 13, 87–96.

Kramer, E.B., Vallabhaneni, H., Mayer, L.M., and Farabaugh, P.J. (2010). A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. *RNA N. Y. N* 16, 1797–1808.

Kuratani, M., Kasai, T., Akasaka, R., Higashijima, K., Terada, T., Kigawa, T., Shinkai, A., Bessho, Y., and Yokoyama, S. (2011). Crystal structure of *Sulfolobus tokodaii* Sua5 complexed with L-threonine and AMPPNP. *Proteins* 79, 2065–2075.

Kushnirov, V.V. (2000). Rapid and reliable protein extraction from yeast. *Yeast Chichester Engl.* 16, 857–860.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.

Lacasse, M.J., and Zamble, D.B. (2016). [NiFe]-Hydrogenase Maturation. *Biochemistry (Mosc.)* 55, 1689–1701.

Lahoud, G., Goto-Ito, S., Yoshida, K.-I., Ito, T., Yokoyama, S., and Hou, Y.-M. (2011). Differentiating analogous tRNA methyltransferases by fragments of the methyl donor. *RNA N. Y. N* 17, 1236–1246.

Lake, M.W., Wuebbens, M.M., Rajagopalan, K.V., and Schindelin, H. (2001). Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* 414, 325–329.

- LaRonde, N.A. (2014). The Ancient Microbial RIO Kinases. *J. Biol. Chem.* *289*, 9488–9492.
- LaRonde-LeBlanc, N., and Wlodawer, A. (2005). The RIO kinases: an atypical protein kinase family required for ribosome biogenesis and cell cycle progression. *Biochim. Biophys. Acta* *1754*, 14–24.
- Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* *51*, 2778–2786.
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* *32*, 11–16.
- Lauhon, C.T. (2012). Mechanism of N6-threonylcarbamoyladenonsine (t(6)A) biosynthesis: isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry (Mosc.)* *51*, 8950–8963.
- Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J.C., and Poch, O. (2001). Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* *11*, 981–993.
- Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* *30*, 5382–5390.
- Lee, C., Kramer, G., Graham, D.E., and Appling, D.R. (2007). Yeast mitochondrial initiator tRNA is methylated at guanosine 37 by the Trm5-encoded tRNA (guanine-N1-)-methyltransferase. *J. Biol. Chem.* *282*, 27744–27753.
- Lei, T., Yang, J., Zheng, L., Markowski, T., Witthuhn, B.A., and Ji, Y. (2012). The essentiality of staphylococcal Gcp is independent of its repression of branched-chain amino acids biosynthesis. *PloS One* *7*, e46836.
- Leimkühler, S., Wuebbens, M.M., and Rajagopalan, K.V. (2001). Characterization of *Escherichia coli* MoeB and its involvement in the activation of molybdopterin synthase for the biosynthesis of the molybdenum cofactor. *J. Biol. Chem.* *276*, 34695–34701.
- Lescrinier, E., Nauwelaerts, K., Zanier, K., Poesen, K., Sattler, M., and Herdewijn, P. (2006). The naturally occurring N6-threonyl adenine in anticodon loop of *Schizosaccharomyces pombe* tRNA_i causes formation of a unique U-turn motif. *Nucleic Acids Res.* *34*, 2878–2886.
- Lesnik, E.A., and Freier, S.M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry (Mosc.)* *34*, 10807–10815.
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* *44*, W242–W245.
- Li, L., and Wang, C.C. (2004). Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. *J. Biol. Chem.* *279*, 14656–14664.

- LI, Q., YANG, H., PENG, X., GUO, D., DONG, Z., POLLI, J.E., and SHU, Y. (2013). Ischemia/Reperfusion-Inducible Protein Modulates the Function of Organic Cation Transporter 1 and Multidrug and Toxin Extrusion1. *Mol. Pharm.* *10*, 2578–2587.
- Liger, D., Mora, L., Lazar, N., Figaro, S., Henri, J., Scrima, N., Buckingham, R.H., van Tilbeurgh, H., Heurgue-Hamard, V., and Graille, M. (2011). Mechanism of activation of methyltransferases involved in translation by the Trm112 “hub” protein. *Nucleic Acids Res.* *39*, 6249–6259.
- Lim, V.I., and Curran, J.F. (2001). Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA N. Y. N* *7*, 942–957.
- Lin, C.A., Ellis, S.R., and True, H.L. (2010). The Sua5 protein is essential for normal translational regulation in yeast. *Mol. Cell. Biol.* *30*, 354–363.
- Lin, C.-J., Smibert, P., Zhao, X., Hu, J.F., Ramroop, J., Kellner, S.M., Benton, M.A., Govind, S., Dedon, P.C., Sternglanz, R., et al. (2015). An extensive allelic series of *Drosophila* kae1 mutants reveals diverse and tissue-specific requirements for t6A biogenesis. *RNA N. Y. N* *21*, 2103–2118.
- Ling, J., O'Donoghue, P., and Söll, D. (2015). Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat. Rev. Microbiol.* *13*, 707–721.
- Liu, J., and Strâby, K.B. (2000). The human tRNA(m²(2)G(26))dimethyltransferase: functional expression and characterization of a cloned hTRM1 gene. *Nucleic Acids Res.* *28*, 3445–3451.
- Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* *29*, 185–200.
- López-Esteva, M., Ardá, A., Savko, M., Round, A., Shepard, W.E., Bruix, M., Coll, M., Fernández, F.J., Jiménez-Barbero, J., and Vega, M.C. (2015). The crystal structure and small-angle X-ray analysis of CsdL/TcdA reveal a new tRNA binding motif in the MoeB/E1 superfamily. *PloS One* *10*, e0118606.
- Lopreiato, R., Facchin, S., Sartori, G., Arrigoni, G., Casonato, S., Ruzzene, M., Pinna, L.A., and Carignani, G. (2004). Analysis of the interaction between piD261/Bud32, an evolutionarily conserved protein kinase of *Saccharomyces cerevisiae*, and the Grx4 glutaredoxin. *Biochem. J.* *377*, 395–405.
- Louw, R., Potgieter, H.C., and Vorster, W. (2005). Teratogenicity of 3-hydroxynorvaline in chicken and mouse embryos. *Amino Acids* *29*, 207–212.
- Ma, B.-G., Chen, L., Ji, H.-F., Chen, Z.-H., Yang, F.-R., Wang, L., Qu, G., Jiang, Y.-Y., Ji, C., and Zhang, H.-Y. (2008). Characters of very ancient proteins. *Biochem. Biophys. Res. Commun.* *366*, 607–611.
- Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013).

MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41**, D262-267.

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

Mandal, D., Köhrer, C., Su, D., Russell, S.P., Krivos, K., Castleberry, C.M., Blum, P., Limbach, P.A., Söll, D., and RajBhandary, U.L. (2010). Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2872-2877.

Mao, D.Y.L., Neculai, D., Downey, M., Orlicky, S., Haffani, Y.Z., Ceccarelli, D.F., Ho, J.S.L., Szilard, R.K., Zhang, W., Ho, C.S., et al. (2008). Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine. *Mol. Cell* **32**, 259-275.

Marintchev, A., and Wagner, G. (2004). Translation initiation: structures, mechanisms and evolution. *Q. Rev. Biophys.* **37**, 197-284.

Matuszewski, M., and Sochacka, E. (2014). Stability studies on the newly discovered cyclic form of tRNA N(6)-threonylcarbamoyladenine (ct(6)A). *Bioorg. Med. Chem. Lett.* **24**, 2703-2706.

Matuszewski, M., Wojciechowski, J., Miyauchi, K., Gdaniec, Z., Wolf, W.M., Suzuki, T., and Sochacka, E. (2017). A hydantoin isoform of cyclic N6-threonylcarbamoyladenine (ct6A) is present in tRNAs. *Nucleic Acids Res.* **45**, 2137-2149.

Maynard, N.D., Macklin, D.N., Kirkegaard, K., and Covert, M.W. (2012). Competing pathways control host resistance to virus via tRNA modification and programmed ribosomal frameshifting. *Mol. Syst. Biol.* **8**, 567.

McCloskey, J.A., Liu, X.H., Crain, P.F., Bruenger, E., Guymon, R., Hashizume, T., and Stetter, K.O. (2000). Posttranscriptional modification of transfer RNA in the submarine hyperthermophile *Pyrolobus fumarii*. *Nucleic Acids Symp. Ser.* **267-268**.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19**, 560-567.

Meng, F.-L., Chen, X.-F., Hu, Y., Tang, H.-B., Dang, W., and Zhou, J.-Q. (2010). Sua5p is required for telomere recombination in *Saccharomyces cerevisiae*. *Cell Res.* **20**, 495-498.

Miller, J.P., Hussain, Z., and Schweizer, M.P. (1976). The involvement of the anticodon adjacent modified nucleoside N-(9-(BETA-D-ribofuranosyl) purine-6-ylcarbamoyl)-threonine in the biological function of *E. coli* tRNA^{Ile}. *Nucleic Acids Res.* **3**, 1185-1201.

Minh, B.Q., Nguyen, M.A.T., and von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188-1195.

- Miyauchi, K., Kimura, S., and Suzuki, T. (2013). A cyclic form of N6-threonylcarbamoyladenine as a widely distributed tRNA hypermodification. *Nat. Chem. Biol.* 9, 105–111.
- Miyoshi, A., Kito, K., Aramoto, T., Abe, Y., Kobayashi, N., and Ueda, N. (2003). Identification of CGI-121, a novel PRPK (p53-related protein kinase)-binding protein. *Biochem. Biophys. Res. Commun.* 303, 399–405.
- Montero, M., Eydallin, G., Viale, A.M., Almagro, G., Muñoz, F.J., Rahimpour, M., Sesma, M.T., Baroja-Fernández, E., and Pozueta-Romero, J. (2009). Escherichia coli glycogen metabolism is controlled by the PhoP-PhoQ regulatory system at submillimolar environmental Mg²⁺ concentrations, and is highly interconnected with a wide variety of cellular processes. *Biochem. J.* 424, 129–141.
- Morin, A., Auxilien, S., Senger, B., Tewari, R., and Grosjean, H. (1998). Structural requirements for enzymatic formation of threonylcarbamoyladenine (t6A) in tRNA: an in vivo study with Xenopus laevis oocytes. *RNA N. Y. N* 4, 24–37.
- Morris, R.C., and Elliott, M.S. (2001). Queuosine modification of tRNA: a case for convergent evolution. *Mol. Genet. Metab.* 74, 147–159.
- Motorin, Y., and Helm, M. (2010). tRNA stabilization by modified nucleotides. *Biochemistry (Mosc.)* 49, 4934–4944.
- Msadek, T. (2009). Grasping at shadows: revealing the elusive nature of essential genes. *J. Bacteriol.* 191, 4701–4704.
- Mueller, E.G. (2006). Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nat. Chem. Biol.* 2, 185–194.
- Murphy, F.V., Ramakrishnan, V., Malkiewicz, A., and Agris, P.F. (2004). The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.* 11, 1186–1191.
- Myllykallio, H., Lopez, P., López-García, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288, 2212–2215.
- Na, J.G., Pinto, I., and Hampsey, M. (1992). Isolation and Characterization of Sua5, a Novel Gene Required for Normal Growth in Saccharomyces Cerevisiae. *Genetics* 131, 791–801.
- Naganuma, M., Sekine, S., Chong, Y.E., Guo, M., Yang, X.-L., Gamper, H., Hou, Y.-M., Schimmel, P., and Yokoyama, S. (2014). The selective tRNA aminoacylation mechanism based on a single G•U pair. *Nature* 510, 507–511.
- Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 478, 57–63.

- Nakagawa, S., Niimura, Y., and Gojobori, T. (2017). Comparative genomic analysis of translation initiation mechanisms for genes lacking the Shine-Dalgarno sequence in prokaryotes. *Nucleic Acids Res.* 45, 3922–3931.
- Nakamura, Y., Ito, K., and Isaksson, L.A. (1996). Emerging understanding of translation termination. *Cell* 87, 147–150.
- Naor, A., Thiaville, P.C., Altman-Price, N., Cohen-Or, I., Allers, T., de Crécy-Lagard, V., and Gophna, U. (2012). A Genetic Investigation of the KEOPS Complex in Halophilic Archaea. *PLoS ONE* 7.
- Nasir, A., Kim, K.M., and Caetano-Anollés, G. (2014). Global Patterns of Protein Domain Gain and Loss in Superkingdoms. *PLoS Comput. Biol.* 10.
- Nesin, M., Lupski, J.R., Svec, P., and Godson, G.N. (1987). Possible new genes as revealed by molecular analysis of a 5-kb *Escherichia coli* chromosomal region 5' to the rpsU-dnaG-rpoD macromolecular-synthesis operon. *Gene* 51, 149–161.
- Neveu, M., Kim, H.-J., and Benner, S.A. (2013). The “strong” RNA world hypothesis: fifty years old. *Astrobiology* 13, 391–403.
- Nichols, C.E., Johnson, C., Lockyer, M., Charles, I.G., Lamb, H.K., Hawkins, A.R., and Stammers, D.K. (2006). Structural characterization of *Salmonella typhimurium* YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog. *Proteins* 64, 111–123.
- Nichols, C.E., Lamb, H.K., Thompson, P., El Omari, K., Lockyer, M., Charles, I., Hawkins, A.R., and Stammers, D.K. (2013). Crystal structure of the dimer of two essential *Salmonella typhimurium* proteins, YgjD & YeaZ and calorimetric evidence for the formation of a ternary YgjD-YeaZ-YjeE complex. *Protein Sci. Publ. Protein Soc.* 22, 628–640.
- Nirenberg, M. (1965). Protein synthesis and the RNA code. *Harvey Lect.* 59, 155–185.
- Noll, H. (1966). Chain initiation and control of protein synthesis. *Science* 151, 1241–1245.
- Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet. TIG* 28, 574–581.
- Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., Giegé, R., and Yokoyama, S. (1994). Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. *J. Mol. Biol.* 236, 710–724.
- Nürenberg, E., and Tampé, R. (2013). Tying up loose ends: ribosome recycling in eukaryotes and archaea. *Trends Biochem. Sci.* 38, 64–74.
- Oberto, J., Breuil, N., Hecker, A., Farina, F., Brochier-Armanet, C., Culetto, E., and Forterre, P. (2009). Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance. *Nucleic Acids Res.* 37, 5343–5352.

- O'Donoghue, P., and Luthey-Schulten, Z. (2003). On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev. MMBR* 67, 550–573.
- Onodera, T., Satoh, K., Ohta, T., and Narumi, I. (2013). *Deinococcus radiodurans* YgjD and YeaZ are involved in the repair of DNA cross-links. *Extrem. Life Extreme Cond.* 17, 171–179.
- Orgel, L.E. (1968). Evolution of the genetic apparatus. *J. Mol. Biol.* 38, 381–393.
- Pallanck, L., and Schulman, L.H. (1991). Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine, and phenylalanine in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3872–3876.
- Pang, Y.L.J., Poruri, K., and Martinis, S.A. (2014). tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* 5, 461–480.
- Parthasarathy, R., Ohrt, J.M., and Chheda, G.B. (1974). Conformation and possible role of hypermodified nucleosides adjacent to 3'-end of anticodon in tRNA: N-(purin-6-ylcarbamoyl)-L-threonine riboside. *Biochem. Biophys. Res. Commun.* 60, 211–218.
- Parthasarathy, R., Ohrt, J.M., and Chheda, G.B. (1977). Modified nucleosides and conformation of anticodon loops: crystal structure of t6A and g6A. *Biochemistry (Mosc.)* 16, 4999–5008.
- Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Bräuer, U., Fandrich, U., Clausnitzer, D., Wehmeier, U.F., Böttcher, C., Scheel, D., et al. (2012). The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem. Int. Ed Engl.* 51, 4046–4052.
- Paschos, A., Bauer, A., Zimmermann, A., Zehelein, E., and Böck, A. (2002). HypF, a carbamoyl phosphate-converting enzyme involved in [NiFe] hydrogenase maturation. *J. Biol. Chem.* 277, 49945–49951.
- Peng, J., He, M.-H., Duan, Y.-M., Liu, Y.-T., and Zhou, J.-Q. (2015). Inhibition of telomere recombination by inactivation of KEOPS subunit Cgi121 promotes cell longevity. *PLoS Genet.* 11, e1005071.
- Perona, J.J., and Gruic-Sovulj, I. (2013). Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases. *SpringerLink* 1–41.
- Perrochia, L., Guetta, D., Hecker, A., Forterre, P., and Basta, T. (2013a). Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification. *Nucleic Acids Res.* 41, 9484–9499.
- Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., van Tilbeurgh, H., Forterre, P., and Basta, T. (2013b). In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.* 41, 1953–1964.
- Petkun, S., Shi, R., Li, Y., Asinas, A., Munger, C., Zhang, L., Waclawek, M., Soboh, B., Sawers, R.G., and Cygler, M. (2011). Structure of hydrogenase maturation protein HypF with reaction intermediates shows two active sites. *Struct. Lond. Engl.* 1993 19, 1773–1783.

Petrov, A.S., Gulen, B., Norris, A.M., Kovacs, N.A., Bernier, C.R., Lanier, K.A., Fox, G.E., Harvey, S.C., Wartell, R.M., Hud, N.V., et al. (2015). History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 15396–15401.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* *25*, 1605–1612.

Phelps, S.S., Jerinic, O., and Joseph, S. (2002). Universally conserved interactions between the ribosome and the anticodon stem-loop of A site tRNA important for translocation. *Mol. Cell* *10*, 799–807.

Phelps, S.S., Malkiewicz, A., Agris, P.F., and Joseph, S. (2004). Modified nucleotides in tRNA(Lys) and tRNA(Val) are important for translocation. *J. Mol. Biol.* *338*, 439–444.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* *341*, 281–286.

Phillips, G., Swairjo, M.A., Gaston, K.W., Bailly, M., Limbach, P.A., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2012). Diversity of archaeosine synthesis in crenarchaeota. *ACS Chem. Biol.* *7*, 300–305.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. *Genes Dev.* *24*, 1832–1860.

Pisareva, V.P., Skabkin, M.A., Hellen, C.U.T., Pestova, T.V., and Pisarev, A.V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J.* *30*, 1804–1817.

Powell, C.A., Nicholls, T.J., and Minczuk, M. (2015). Nuclear-encoded factors involved in post-transcriptional processing and modification of mitochondrial tRNAs in human disease. *Front. Genet.* *6*, 79.

Powers, D.M., and Peterkofsky, A. (1972a). Biosynthesis and specific labeling of N-(purin-6-ylcarbamoyl)threonine of Escherichiacoli transfer RNA. *Biochem. Biophys. Res. Commun.* *46*, 831–838.

Powers, D.M., and Peterkofsky, A. (1972b). The presence of N-(purin-6-ylcarbamoyl)threonine in transfer ribonucleic acid species whose codons begin with adenine. *J. Biol. Chem.* *247*, 6394–6401.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M., and Nagai, K. (1995). Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* *249*, 398–408.

Qian, Q., Curran, J.F., and Björk, G.R. (1998). The Methyl Group of the N6-Methyl-N6-Threonylcarbamoyladenosine in tRNA of Escherichia coli Modestly Improves the Efficiency of the tRNA. *J. Bacteriol.* *180*, 1808–1813.

Quigley, G.J., and Rich, A. (1976). Structural domains of transfer RNA molecules. *Science* 194, 796–806.

Rajagopala, S.V., Yamamoto, N., Zweifel, A.E., Nakamichi, T., Huang, H.-K., Mendez-Rios, J.D., Franca-Koh, J., Boorgula, M.P., Fujita, K., Suzuki, K., et al. (2010). The Escherichia coli K-12 ORFeome: a resource for comparative molecular microbiology. *BMC Genomics* 11, 470.

RajBhandary, U.L. (1994). Initiator transfer RNAs. *J. Bacteriol.* 176, 547–552.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557–572.

Ranea, J.A.G., Sillero, A., Thornton, J.M., and Orengo, C.A. (2006). Protein superfamily evolution and the last universal common ancestor (LUCA). *J. Mol. Evol.* 63, 513–525.

Rangarajan, E.S., Asinas, A., Proteau, A., Munger, C., Baardsnes, J., Iannuzzi, P., Matte, A., and Cygler, M. (2008). Structure of [NiFe] Hydrogenase Maturation Protein HypE from Escherichia coli and Its Interaction with HypF. *J. Bacteriol.* 190, 1447–1458.

Reddy, D.M., Crain, P.F., Edmonds, C.G., Gupta, R., Hashizume, T., Stetter, K.O., Widdel, F., and McCloskey, J.A. (1992). Structure determination of two new amino acid-containing derivatives of adenosine from tRNA of thermophilic bacteria and archaea. *Nucleic Acids Res.* 20, 5607–5615.

Rich, A., and RajBhandary, U.L. (1976). Transfer RNA: molecular structure, sequence, and properties. *Annu. Rev. Biochem.* 45, 805–860.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–324.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F., and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* 250, 546–551.

Rodriguez, V., Chen, Y., Elkahloun, A., Dutra, A., Pak, E., and Chandrasekharappa, S. (2007). Chromosome 8 BAC array comparative genomic hybridization and expression analysis identify amplification and overexpression of TRMT12 in breast cancer. *Genes. Chromosomes Cancer* 46, 694–707.

Rojas-Benitez, D., Thiaville, P.C., de Crécy-Lagard, V., and Glavic, A. (2015). The Levels of a Universally Conserved tRNA Modification Regulate Cell Growth. *J. Biol. Chem.* 290, 18699–18707.

Rould, M.A., Perona, J.J., Söll, D., and Steitz, T.A. (1989). Structure of E. coli glutamyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* 246, 1135–1142.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* 252, 1682–1689.

- Scheid, P. (2016). A strange endocytobiont revealed as largest virus. *Curr. Opin. Microbiol.* *31*, 58–62.
- Scheid, P., Zöller, L., Pressmar, S., Richard, G., and Michel, R. (2008). An extraordinary endocytobiont in *Acanthamoeba* sp. isolated from a patient with keratitis. *Parasitol. Res.* *102*, 945–950.
- Schmelz, S., and Naismith, J.H. (2009). Adenylate-forming enzymes. *Curr. Opin. Struct. Biol.* *19*, 666–671.
- Schneider, A. (2011). Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annu. Rev. Biochem.* *80*, 1033–1053.
- Schweizer, M.P., Chheda, G.B., Baczynskyj, L., and Hall, R.H. (1969). Aminoacyl nucleosides. VII. N-(Purin-6-ylcarbamoyl)threonine. A new component of transfer ribonucleic acid. *Biochemistry (Mosc.)* *8*, 3283–3289.
- Schweizer, M.P., McGrath, K., and Baczynskyj, L. (1970). The isolation and characterization of N-[9-(beta-D-ribofuranosyl)-purin-6-ylcarbamoyl]glycine from yeast transfer RNA. *Biochem. Biophys. Res. Commun.* *40*, 1046–1052.
- Seki, Y., Ikeda, S., Kiyohara, H., Ayabe, H., Seki, T., and Matsui, H. (2002). Sequencing analysis of a putative human O-sialoglycoprotein endopeptidase gene (OSGEP) and analysis of a bidirectional promoter between the OSGEP and APEX genes. *Gene* *285*, 101–108.
- Senger, B., Auxilien, S., Englisch, U., Cramer, F., and Fasiolo, F. (1997). The modified wobble base inosine in yeast tRNA^{Ile} is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. *Biochemistry (Mosc.)* *36*, 8269–8275.
- Sharma, V., Colson, P., Chabrol, O., Scheid, P., Pontarotti, P., and Raoult, D. (2015). Welcome to pandoraviruses at the “Fourth TRUC” club. *Front. Microbiol.* *6*, 423.
- Shine, J., and Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature* *254*, 34–38.
- Shomura, Y., and Higuchi, Y. (2012). Structural basis for the reaction mechanism of S-carbamoylation of HypE by HypF in the maturation of [NiFe]-hydrogenases. *J. Biol. Chem.* *287*, 28409–28419.
- Simonović, M., and Steitz, T.A. (2008). Peptidyl-CCA deacylation on the ribosome promoted by induced fit and the O3'-hydroxyl group of A76 of the unacylated A-site tRNA. *RNA N. Y.* *14*, 2372–2378.
- Sinclair, R.M., Ravantti, J.J., and Bamford, D.H. (2017). Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification. *J. Virol.* *91*.
- Slagter-Jäger, J.G., Puzis, L., Gutgsell, N.S., Belfort, M., and Jain, C. (2007). Functional defects in transfer RNAs lead to the accumulation of ribosomal RNA precursors. *RNA N. Y.* *13*, 597–605.

- Sloan, D.B., and Moran, N.A. (2012). Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* 29, 3781–3792.
- Smith, T.F., and Hartman, H. (2015). The evolution of Class II Aminoacyl-tRNA synthetases and the first code. *FEBS Lett.* 589, 3499–3507.
- Söll, D., Ohtsuka, E., Jones, D.S., Lohrmann, R., Hayatsu, H., Nishimura, S., and Khorana, H.G. (1965). Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 54, 1378–1385.
- Soma, A., Ikeuchi, Y., Kanemasa, S., Kobayashi, K., Ogasawara, N., Ote, T., Kato, J., Watanabe, K., Sekine, Y., and Suzuki, T. (2003). An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol. Cell* 12, 689–698.
- Sonawane, K.D., and Sambhare, S.B. (2015). The influence of hypermodified nucleosides lysidine and t(6)A to recognize the AUA codon instead of AUG: a molecular dynamics simulation study. *Integr. Biol. Quant. Biosci. Nano Macro* 7, 1387–1395.
- Spears, J.L., Xiao, X., Hall, C.K., and Agris, P.F. (2014). Amino acid signature enables proteins to recognize modified tRNA. *Biochemistry (Mosc.)* 53, 1125–1133.
- Spirek, M., Benko, Z., Carnecka, M., Rumpf, C., Cipak, L., Batova, M., Marova, I., Nam, M., Kim, D.-U., Park, H.-O., et al. (2010). *S. pombe* genome deletion project: an update. *Cell Cycle Georget. Tex* 9, 2399–2402.
- Sprinzel, M., and Cramer, F. (1975). Site of aminoacylation of tRNAs from *Escherichia coli* with respect to the 2'- or 3'-hydroxyl group of the terminal adenosine. *Proc. Natl. Acad. Sci. U. S. A.* 72, 3049–3053.
- Sprinzel, M., and Cramer, F. (1979). The -C-C-A end of tRNA and its role in protein biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* 22, 1–69.
- Sprinzel, M., and Vassilenko, K.S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 33, D139–140.
- Sprinzel, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R. (1989). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 17 *Suppl*, r1–172.
- Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E.V., Karzai, A.W., and Sternglanz, R. (2011). The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *EMBO J.* 30, 873–881.
- Srivastava, A., Gogoi, P., Deka, B., Goswami, S., and Kanaujia, S.P. (2016). In silico analysis of 5'-UTRs highlights the prevalence of Shine-Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively. *J. Theor. Biol.* 402, 54–61.
- Staehelin, T., Maglott, D.M., and Monro, R.E. (1969). On the catalytic center of peptidyl transfer: a part of the 50 S ribosome structure. *Cold Spring Harb. Symp. Quant. Biol.* 34, 39–48.

- Stewart, J.W., Sherman, F., Shipman, N.A., and Jackson, M. (1971). Identification and mutational relocation of the AUG codon initiating translation of iso-1-cytochrome c in yeast. *J. Biol. Chem.* **246**, 7429–7445.
- Stuart, J.W., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., and Agris, P.F. (2000). Functional anticodon architecture of human tRNA^{Lys3} includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t⁶A. *Biochemistry (Mosc.)* **39**, 13396–13404.
- Sundaram, M., Durant, P.C., and Davis, D.R. (2000). Hypermodified nucleosides in the anticodon of tRNA^{Lys} stabilize a canonical U-turn structure. *Biochemistry (Mosc.)* **39**, 12575–12584.
- Takaku, H., Minagawa, A., Takagi, M., and Nashimoto, M. (2003). A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. *Nucleic Acids Res.* **31**, 2272–2278.
- Takemura, S., Murakami, M., and Miyazaki, M. (1969). Nucleotide sequence of isoleucine transfer RNA from *Torulopsis utilis*. *J. Biochem. (Tokyo)* **65**, 489–491.
- Teplova, M., Tereshko, V., Sanishvili, R., Joachimiak, A., Bushueva, T., Anderson, W.F., and Egli, M. (2000). The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. *Protein Sci. Publ. Protein Soc.* **9**, 2557–2566.
- Teplyakov, A., Obmolova, G., Tordova, M., Thanki, N., Bonander, N., Eisenstein, E., Howard, A.J., and Gilliland, G.L. (2002). Crystal structure of the YjeE protein from *Haemophilus influenzae*: a putative ATPase involved in cell wall synthesis. *Proteins* **48**, 220–226.
- Thiaville, P.C., El Yacoubi, B., Perrochia, L., Hecker, A., Prigent, M., Thiaville, J.J., Forterre, P., Namy, O., Basta, T., and de Crécy-Lagard, V. (2014). Cross Kingdom Functional Conservation of the Core Universally Conserved Threonylcarbamoyladenosine tRNA Synthesis Enzymes. *Eukaryot. Cell* **13**, 1222–1231.
- Thiaville, P.C., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2015a). Diversity of the biosynthesis pathway for threonylcarbamoyladenosine (t⁶A), a universal modification of tRNA. *RNA Biol.* **11**, 1529–1539.
- Thiaville, P.C., El Yacoubi, B., Köhrer, C., Thiaville, J.J., Deutsch, C., Iwata-Reuyl, D., Bacusmo, J.M., Armengaud, J., Bessho, Y., Wetzel, C., et al. (2015b). Essentiality of threonylcarbamoyladenosine (t⁶A), a universal tRNA modification, in bacteria. *Mol. Microbiol.* **98**, 1199–1221.
- Thiaville, P.C., Legendre, R., Rojas-Benítez, D., Baudin-Baillieu, A., Hatin, I., Chalancon, G., Glavic, A., Namy, O., and de Crécy-Lagard, V. (2016). Global translational impacts of the loss of the tRNA modification t⁽⁶⁾A in yeast. *Microb. Cell Graz Austria* **3**, 29–45.
- Tocchini-Valentini, G.D., Fruscoloni, P., and Tocchini-Valentini, G.P. (2011). Evolution of introns in the archaeal world. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4782–4787.

- Tominaga, T., Watanabe, S., Matsumi, R., Atomi, H., Imanaka, T., and Miki, K. (2012). Structure of the [NiFe]-hydrogenase maturation protein HypF from *Thermococcus kodakarensis* KOD1. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* *68*, 1153–1157.
- Trotta, C.R., Miao, F., Arn, E.A., Stevens, S.W., Ho, C.K., Rauhut, R., and Abelson, J.N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* *89*, 849–858.
- Trotter, V., Vinella, D., Loiseau, L., Ollagnier de Choudens, S., Fontecave, M., and Barras, F. (2009). The CsdA cysteine desulphurase promotes Fe/S biogenesis by recruiting Suf components and participates to a new sulphur transfer pathway by recruiting CsdL (ex-YgdL), a ubiquitin-modifying-like protein. *Mol. Microbiol.* *74*, 1527–1542.
- Unni, S., Huang, Y., Hanson, R.M., Tobias, M., Krishnan, S., Li, W.W., Nielsen, J.E., and Baker, N.A. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* *32*, 1488–1491.
- Urbonavičius, J., Meškys, R., and Grosjean, H. (2014). Biosynthesis of wyosine derivatives in tRNA(Phe) of Archaea: role of a remarkable bifunctional tRNA(Phe):m1G/imG2 methyltransferase. *RNA N. Y. N* *20*, 747–753.
- Väre, V.Y.P., Eruysal, E.R., Narendran, A., Sarachan, K.L., and Agris, P.F. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* *7*.
- Vendeix, F.A.P., Dziergowska, A., Gustilo, E.M., Graham, W.D., Sproat, B., Malkiewicz, A., and Agris, P.F. (2008). Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry (Mosc.)* *47*, 6117–6129.
- Veyhl, M., Spangenberg, J., Püschel, B., Poppe, R., Dekel, C., Fritzsche, G., Haase, W., and Koepsell, H. (1993). Cloning of a membrane-associated protein which modifies activity and properties of the Na(+)-D-glucose cotransporter. *J. Biol. Chem.* *268*, 25041–25053.
- Veyhl, M., Wagner, C.A., Gorboulev, V., Schmitt, B.M., Lang, F., and Koepsell, H. (2003). Downregulation of the Na(+)- D-glucose cotransporter SGLT1 by protein RS1 (RSC1A1) is dependent on dynamin and protein kinase C. *J. Membr. Biol.* *196*, 71–81.
- Villarroel-Campos, D., Henríquez, D.R., Bodaleo, F.J., Oguchi, M.E., Bronfman, F.C., Fukuda, M., and Gonzalez-Billault, C. (2016). Rab35 Functions in Axon Elongation Are Regulated by P53-Related Protein Kinase in a Mechanism That Involves Rab35 Protein Degradation and the Microtubule-Associated Protein 1B. *J. Neurosci. Off. J. Soc. Neurosci.* *36*, 7298–7313.
- Wan, L.C.K., Mao, D.Y.L., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., Poda, G., Thevakumaran, N., Yuan, F., Szilard, R.K., et al. (2013). Reconstitution and characterization of eukaryotic N6-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Res.* *41*, 6332–6346.
- Wan, L.C.K., Pillon, M.C., Thevakumaran, N., Sun, Y., Chakrabartty, A., Guarné, A., Kurinov, I., Durocher, D., and Sicheri, F. (2016). Structural and functional characterization of

- KEOPS dimerization by Pcc1 and its role in t6A biosynthesis. *Nucleic Acids Res.* **44**, 6971–6980.
- Wan, L.C.K., Maisonneuve, P., Szilard, R.K., Lambert, J.-P., Ng, T.F., Manczyk, N., Huang, H., Laister, R., Caudy, A.A., Gingras, A.-C., et al. (2017). Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acids Res.* **45**, 805–817.
- Watanabe, K., and Yokobori, S.-I. (2011). tRNA Modification and Genetic Code Variations in Animal Mitochondria. *J. Nucleic Acids* **2011**, 623095.
- Watson, J.D. (1964). THE SYNTHESIS OF PROTEINS UPON RIBOSOMES. *Bull. Soc. Chim. Biol. (Paris)* **46**, 1399–1425.
- Wei, F.-Y., and Tomizawa, K. (2011). Functional loss of Cdkal1, a novel tRNA modification enzyme, causes the development of type 2 diabetes. *Endocr. J.* **58**, 819–825.
- Wei, F.-Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A., Matsui, H., Atta, M., Michiue, H., Fontecave, M., et al. (2011). Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. *J. Clin. Invest.* **121**, 3598–3608.
- Weiner, J., Beaussart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. *FEBS J.* **273**, 2037–2047.
- Weissenbach, J., and Grosjean, H. (1981). Effect of threonylcarbamoyl modification (t6A) in yeast tRNA Arg III on codon-anticodon and anticodon-anticodon interactions. A thermodynamic and kinetic evaluation. *Eur. J. Biochem.* **116**, 207–213.
- Williams, T.A., Embley, T.M., and Heinz, E. (2011). Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PloS One* **6**, e21080.
- Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242.
- Witts, R.N., Hopson, E.C., Koballa, D.E., Van Boening, T.A., Hopkins, N.H., Patterson, E.V., and Nagan, M.C. (2013). Backbone-base interactions critical to quantum stabilization of transfer RNA anticodon structure. *J. Phys. Chem. B* **117**, 7489–7497.
- Wohlgemuth, I., Pohl, C., and Rodnina, M.V. (2010). Optimization of speed and accuracy of decoding in translation. *EMBO J.* **29**, 3701–3709.
- Wohlgemuth, I., Pohl, C., Mittelstaet, J., Konevega, A.L., and Rodnina, M.V. (2011). Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **366**, 2979–2986.

- Xi, J., Ge, Y., Kinsland, C., McLafferty, F.W., and Begley, T.P. (2001). Biosynthesis of the thiazole moiety of thiamin in *Escherichia coli*: identification of an acyldisulfide-linked protein--protein conjugate that is functionally analogous to the ubiquitin/E1 complex. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8513–8518.
- Xu, Q., McMullan, D., Jaroszewski, L., Krishna, S.S., Elsliger, M.A., Yeh, A.P., Abdubek, P., Astakhova, T., Axelrod, H.L., Carlton, D., et al. (2010). Structure of an essential bacterial protein YeaZ (TM0874) from *Thermotoga maritima* at 2.5 Å resolution. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* 66, 1230–1236.
- Yamada, K.D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251.
- Yamamoto, Y., Hasegawa, J., and Ito, Y. (2012). Kinetic investigation on carbamate formation from the reaction of carbon dioxide with amino acids in homogeneous aqueous solution. *J. Phys. Org. Chem.* 25, 239–247.
- Yan, J., Lipka, A.E., Schmelz, E.A., Buckler, E.S., and Jander, G. (2015). Accumulation of 5-hydroxynorvaline in maize (*Zea mays*) leaves is induced by insect feeding and abiotic stress. *J. Exp. Bot.* 66, 593–602.
- Yarian, C., Townsend, H., Czystkowski, W., Sochacka, E., Malkiewicz, A.J., Guenther, R., Miskiewicz, A., and Agris, P.F. (2002). Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* 277, 16391–16395.
- Yutin, N., Wolf, Y.I., Raoult, D., and Koonin, E.V. (2009). Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6, 223.
- Yutin, N., Puigbò, P., Koonin, E.V., and Wolf, Y.I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PloS One* 7, e36972.
- Yutin, N., Wolf, Y.I., and Koonin, E.V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466–467, 38–52.
- Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358.
- Zavialov, A.V., Hauryliuk, V.V., and Ehrenberg, M. (2005). Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. *Mol. Cell* 18, 675–686.
- Zhang, W., Collinet, B., Perrochia, L., Durand, D., and van Tilbeurgh, H. (2015a). The ATP-mediated formation of the YgjD–YeaZ–YjeE complex is required for the biosynthesis of tRNA t₆A in *Escherichia coli*. *Nucleic Acids Res.* 43, 1804–1817.
- Zhang, W., Collinet, B., Graille, M., Daugeron, M.-C., Lazar, N., Libri, D., Durand, D., and van Tilbeurgh, H. (2015b). Crystal structures of the Gon7/Pcc1 and Bud32/Cgi121

complexes provide a model for the complete yeast KEOPS complex. *Nucleic Acids Res.* 43, 3358–3372.

Zhu, L., and Deutscher, M.P. (1987). tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. *EMBO J.* 6, 2473–2477.

van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725.

Titre : Structure, fonction et évolution de la famille universelle Sua5/YrdC impliquée dans la synthèse du nucléoside modifié t⁶A

Mots clés : modification ARNt, t⁶A, thréonycarbamoyl, protéine universelle, Sua5/YrdC, TC-AMP

Résumé : La t⁶A est universellement présente au sein des ARNt décodant les codons ANN et est essentielle pour la fidélité de traduction. Sa synthèse se déroule en deux étapes, dont la première implique la formation de l'intermédiaire de réaction thréonyl-carbamoyl-AMP (TC-AMP) par la famille Sua5/YrdC. Cette famille est retrouvée chez tous les organismes et était donc vraisemblablement présente chez le dernier ancêtre commun universel (LUCA). Elle est composée de deux variants distincts, YrdC et Sua5, qui partagent un domaine catalytique orthologue. A la différence du variant YrdC qui est composé d'un domaine unique, le variant Sua5 possède un domaine C-terminal additionnel nommé SUA5, de fonction inconnue. La plupart des espèces code pour un seul variant et les deux variants sont présents dans les trois domaines du vivant, Eucaryote, Archée et Bactérie.

Afin d'identifier le rôle du domaine SUA5 et du linker inter-domaine, nous avons étudié la protéine Sua5 de l'archée *Pyrococcus abyssi*. Nos résultats montrent que ces deux régions sont importantes pour l'activité de Sua5. Le linker est capable de contrôler le passage des ligands en changeant de conformation tandis que le domaine SUA5 agit comme une plateforme d'ancrage pour le linker. Afin de comprendre l'histoire évolutive de la famille Sua5/YrdC, nous avons ensuite étudié la distribution des variants et nous avons utilisé des approches *in silico* et *in vitro* afin de déterminer les différences fonctionnelles entre YrdC et Sua5. L'ensemble de ces données nous permet de proposer que LUCA possédait une protéine Sua5 et qu'YrdC serait apparu suite à une perte de domaine dans certaines lignées lors de l'évolution.

Title : Structure, function and evolution of the universal Sua5/YrdC family involved in the modified nucleoside t⁶A synthesis

Keywords : tRNA modification, t⁶A, threonycarbamoyl, universal protein, Sua5/YrdC, TC-AMP

Abstract: t⁶A is universally found in tRNAs that read ANN codons and is essential for translation fidelity. Its synthesis takes place in two stages, the first one involving the formation of the reaction intermediate threonyl-carbamoyl-AMP (TC-AMP) by the Sua5/YrdC family. The Sua5/YrdC family is found in all organisms and was thus presumably presents in the Last Universal Common Ancestor (LUCA). This family is composed of two distinct variants, YrdC and Sua5, which share an orthologous catalytic domain. While YrdC is a single domain protein, Sua5 has an additional C-terminal domain of unknown function named SUA5. Most species encode for either variant and both variants are found in the three domains of life, Eukarya, Archaea and Bacteria.

To discover the role of the SUA5 domain and the inter-domain linker, we studied the Sua5 protein from the archaeon *Pyrococcus abyssi*. We found that they are both important for the activity of Sua5. The linker is able to control the entry and exit of ligands by changing conformation while the SUA5 domain acts as an anchoring platform for the linker. To understand the evolutionary history of the Sua5/YrdC family, we then studied the distribution of Sua5 and YrdC across the tree of life and we used *in silico* and *in vitro* approaches to identify functional differences between YrdC and Sua5. Taken together, our work allows us to propose that the LUCA encoded a Sua5 protein and that YrdC emerged after domain loss in some lineages during evolution.