



**HAL**  
open science

# Etude structurale et fonctionnelle des systèmes de modification t A des ARNt dans les trois domaines du vivant.

Sophia Missouri

► **To cite this version:**

Sophia Missouri. Etude structurale et fonctionnelle des systèmes de modification t A des ARNt dans les trois domaines du vivant.. Biologie structurale [q-bio.BM]. Université Paris-Saclay, 2020. Français. NNT : 2020UPASQ034 . tel-03594896

**HAL Id: tel-03594896**

**<https://theses.hal.science/tel-03594896>**

Submitted on 3 Mar 2022

**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.

# Etude structurale et fonctionnelle des systèmes de modification t<sup>6</sup>A des ARNt dans les trois domaines du vivant.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 569, Innovation Thérapeutique : du  
Fondamental à l'Appliqué (ITFA).

Spécialité de doctorat : Biochimie et Biologie structurale.

Unité de recherche : Université Paris-Saclay, CEA, CNRS,  
Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

Référent : Faculté de Pharmacie.

Thèse présentée et soutenue à ORSAY, le 15 Décembre 2020, par

**Sophia MISSOURY**

## Composition du Jury

**Daniel GAUTHERET**

Professeur, Université Paris-Saclay, CNRS  
UMR 9198

Président

**Carine TISNE**

Directrice de Recherche, CNRS UMR 8261  
Université de Paris

Rapporteur & Examinatrice

**Wim VERSEES**

Professeur, Vrije Universiteit Brussel

Rapporteur & Examineur

**Emmanuelle SCHMITT**

Directrice de Recherche, CNRS UMR 7654

Examinatrice

**Domenico LIBRI**

Directeur de Recherche, CNRS, UMR 7592  
Université de Paris

Examineur

**Herman VAN TILBEURGH**

Professeur, Université Paris-Saclay, CNRS  
UMR 9198

Directeur de thèse

**Bruno COLLINET**

Maître de conférences, Sorbonne  
Université, CNRS UMR 7590

Co-Encadrant de thèse



## Remerciements

Je remercie l'ensemble des membres de l'équipe Fonction et Architecture des Assemblages Macromoléculaires. Dominique Liger, je te remercie pour tes conseils et ta gentillesse. Inès Li de la Sierra-Gallay et Stéphane Plancqueel, merci pour votre aide avec la cristallographie, Noureddine Lazar pour tes conseils précieux concernant la purification de protéines. Sophie Cheruel, merci, pour ta bonne humeur et tes excellents conseils en purification et cristallogénèse mais aussi pour ton soutien, je te dois énormément. Raoudha Dammak, merci pour ton aide. Sylvie Nessler, je te remercie aussi pour tes conseils en bio-informatique. Karine Blondeau et Stéphanie Marsin merci pour votre gentillesse. Hélène Walbott merci pour avoir partagé tes précieuses connaissances concernant le monde des ARNt et pour ta gentillesse. Les nuits au synchrotron sont de bons souvenirs. Je remercie profondément Dominique Durand : pour ta gentillesse, tes heures consacrées à me former au SAXS, mais aussi pour ton écoute et tes précieux conseils. Merci également à Patrice Vachette, pour vos conseils. Je remercie également Catherine Vénien-Bryan et Charline Fagnen pour leur aide en microscopie électronique et également Stéphane Bressanelli que j'ai beaucoup sollicité.

Je remercie aussi Marion Malhéo et Sylvie Auxilien, ce fut un plaisir de travailler à vos côtés. Et je ne peux pas oublier de remercier les étudiants de passage et ceux qui sont restés plus longtemps avec qui j'ai passé de merveilleux moments de rigolade : Louisa Celma pour nos heures passées ensemble ! Jordhan Thuillier pour nos voyages au synchrotron de Grenoble. Et enfin Adrien Fleck et Claire Cargemel, "les affreux" !

Un immense merci aussi à Tamara Basta, Marie-Claire Daugeron et Violette Da Cunha. Vous m'avez toujours très bien accueilli, quelle belle aventure fut la caractérisation de Pcc2, je suis très contente d'avoir vécu une partie de cette aventure à vos côtés.

Aussi je tiens à remercier le service informatique pour avoir supporté toutes mes demandes !

Je souhaiterais remercier chaleureusement mes directeurs de thèse, Herman van Tilbeurgh et Bruno Collinet. Merci pour tous vos enseignements qui m'ont permis de me construire un esprit critique que je continuerais à affiner. Merci aussi pour votre confiance qui n'a jamais failli malgré mes doutes lors de ces dernières années. Je serais toujours reconnaissante pour tout ce que vous m'avez appris.

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. Travailler avec vous a été un vrai plaisir et

m'a permis de grandir et de m'améliorer. Je vous serai toujours reconnaissante pour cela et je n'oublierai jamais ces dernières années et ce qu'elles m'ont apprises.

Enfin, je remercie également ma famille, mes amis, mes parents et mes sœurs pour leur soutien infailible durant toutes ces années. A mon père qui m'a appris la persévérance et à ne jamais abandonner...

# Table des matières

<b>Remerciements</b> .....	<b>3</b>
<b>Introduction</b> .....	<b>11</b>
1. Généralités sur la traduction. ....	11
2. L'acide ribonucléique de transfert : la molécule au centre de la traduction.....	16
2.1. Le repliement des ARNt .....	16
2.1.1. Structure primaire. ....	16
2.1.2. Structure secondaire .....	17
2.1.3. Structure tertiaire.....	18
2.1.4. L'ARNt mitochondrial .....	20
2.2. Fonctions des ARNt. ....	21
2.2.1. La traduction .....	22
2.2.2. Fonctions alternatives des ARNt. ....	24
2.3. Synthèse des ARNt.....	25
2.3.1. Régulation de la synthèse des ARNt.....	25
2.3.2. Maturation des extrémités .....	26
2.3.2.1. Maturation de l'extrémité 5'.....	26
2.3.2.2. Maturation de l'extrémité 3' .....	26
2.3.2.3. Epissage de l'ARNt .....	27
2.4. Modifications des ARNt.....	28
2.4.1. Modifications post-transcriptionnelles des ARNt .....	28
2.4.2. Fréquence et distribution des modifications post-transcriptionnelles des ARNt.....	30
2.5. Effet des modifications .....	31
2.5.1. Effet des modifications sur la structure tridimensionnelle des ARNt. ....	32
2.5.1. Effet des modifications sur le bras AC. ....	33
2.5.1.1. Le cas de la base 34 ou " Wobble " base.....	34
2.5.1.2. Le cas de la base 37 ou "dangling" base :.....	37
2.5.2. Défauts de modifications reliées à des maladies génétiques.....	39
3. La modification t <sup>6</sup> A .....	41
3.1. La découverte d'un groupement supplémentaire sur la base 37 .....	41

3.2.	Impacts structuraux de la modification t <sup>6</sup> A.....	42
3.3.	Variations autour de la modification t <sup>6</sup> A. ....	44
3.3.1.	Mécanisme de la modification i <sup>6</sup> A, un variant du nucléotide t <sup>6</sup> A. ....	45
4.	Les enzymes de modification t <sup>6</sup> A. ....	47
4.1.	La génomique comparative. ....	47
4.2.	La modification t <sup>6</sup> A est une réaction nécessitant deux étapes catalytiques effectuées par les protéines des familles YrdC/Sua5 et Kae1/TsaD/Qri7. ....	50
4.3.	Structure des enzymes t <sup>6</sup> A synthases. ....	52
4.3.1.	Structure tridimensionnelle des protéines composants la famille YrdC/Sua5 ou TsaC <sub>1</sub> /TsaC <sub>2</sub> .....	52
4.3.2.	La structure tridimensionnelle des membres de la famille Kae1/TsaD/Qri7. ....	54
4.4.	Similitude avec les enzymes de carbamoylation HypF et TobZ.....	56
4.5.	La modification t <sup>6</sup> A chez les procaryotes. ....	58
4.5.1.	La biosynthèse de la modification t <sup>6</sup> A par les protéines TsaC, TsaB, TsaD et TsaE.....	58
4.5.1.1.	TsaB est le paralogue tronqué de TsaD.....	59
4.5.1.2.	TsaE, une ATPase comme troisième partenaire. ....	61
4.5.1.3.	Le complexe TsaBDE chez les procaryotes. ....	63
4.6.	La modification t <sup>6</sup> A chez les archées et les eucaryotes.....	64
4.6.1.	Découverte du complexe KEOPS.....	64
4.6.2.	Les partenaires de Kae1 dans le complexe KEOPS. ....	65
4.6.2.1.	La protéine Bud32, une kinase atypique.....	65
4.6.2.2.	Cgi121.....	67
4.6.2.3.	La sous-unité Pcc1.....	68
4.6.3.	Les différentes sous-unités Kae1, Bud32, Cgi121 et Pcc1 s'agencent de façon linéaire. ....	69
4.6.3.1.	Une cinquième sous-unité compose le complexe KEOPS chez les eucaryotes. ....	72
4.6.4.	La structure quaternaire du complexe KEOPS chez les eucaryotes et les archées. ....	73
4.6.5.	Fonction des sous-unités du complexe KEOPS chez les eucaryotes et les archées.....	74
4.7.	Un système minimaliste dans les mitochondries. ....	75
4.7.1.	Qri7, l'orthologue mitochondrial de Kae1/TsaD.....	75
4.7.2.	La formation d'un homodimère Qri7 assure la biosynthèse t <sup>6</sup> A sur les ARNt mitochondriaux. ....	75
4.8.	Les t <sup>6</sup> A synthases à travers les trois domaines du vivant.....	77

4.9. Le défaut de la modification t <sup>6</sup> A affectent différemment les espèces. ....	78
4.10. Effet du défaut t <sup>6</sup> A sur les organismes pluricellulaires. ....	80
5. Problématique et objectifs de la thèse.....	81
<b>Chapitre I</b> .....	<b>85</b>
I. Etude du rôle de TsaE dans la modification t <sup>6</sup> A chez les procaryotes.....	85
I.1. Présentation de l'article. ....	85
I.2 Discussion des résultats obtenus par le groupe de Manal Swarjo (Luthra <i>et al.</i> , 2019).....	117
I.3. Conclusion de l'article I. ....	119
<b>Chapitre II</b> .....	<b>121</b>
I. Etude du site actif de TsaD chez les procaryotes. ....	121
II.1. Présentation de l'article II. ....	121
II.2. Représentations schématiques du site actif de TsaBD et du domaine Kae1 de TobZ en présence de différents substrats. ....	169
II.3. Conclusion de l'article. ....	173
<b>Chapitre III</b> .....	<b>175</b>
III. Effets de YrdC et de GON7 sur la modification t <sup>6</sup> A chez l'Homme. ....	175
III.1. Présentation de l'article « Defects in t <sup>6</sup> A tRNA modification due to GON7 et YRDC mutations leads to Galloway-Mowat ». ....	175
III.2. Conclusion de l'article. ....	223
III.3. Résultats supplémentaires. ....	225
III.3.1. Evolution des interfaces GON7/LAGE3 et Gon7/Pcc1. ....	225
III.3.2. Modélisation de la structure 3D du complexe entier KEOPS humain par diffusion des rayons X aux petits angles.....	231
III.3.2.1. Matériels et Méthodes. ....	231
III.3.2.2. Résultats. ....	233
III.3.3. Etude structurale du complexe KEOPS humain avec son substrat : l'ARNt <sup>lle</sup> <sub>AUU</sub> . ....	237
III.3.3.1. Matériels et Méthodes. ....	237
III.3.3.2. Résultats. ....	239
III.3.3.3. Etudes préliminaires du complexe KEOPS humain en présence de l'ARNt <sup>lle</sup> par microscopie électronique en transmission. ....	243
III.3.3.3.1. Matériels et Méthodes. ....	248
III.3.3.3.2. Résultats. ....	252



<b>Chapitre IV</b> .....	<b>261</b>
<b>IV. Identification de la cinquième sous-unité du complexe KEOPS chez les archées.</b> .....	<b>261</b>
<b>IV.1. Présentation de l'étude.</b> .....	<b>261</b>
<b>IV.2. Conclusion.</b> .....	<b>299</b>
<b>CONCLUSION GENERALE ET PERSPECTIVES.</b> .....	<b>301</b>
Rappels sur la modification t <sup>6</sup> A. ....	301
Les t <sup>6</sup> A synthases au sein des procaryotes.....	302
Les t <sup>6</sup> A synthases au sein des eucaryotes et des archées. ....	310
Adaptation des enzymes t <sup>6</sup> A synthases selon les règnes du vivant. ....	314
<b>ANNEXES</b> .....	<b>319</b>
Annexe 1 : Expression, production et purification du complexe TsabD d' <i>Escherichia coli</i> .....	319
Annexe 2 : Comparaison des séquences de la sous-unité catalytique. ....	322
Annexe 3 : Séquences des gènes codant les protéines du complexe KEOPS humain. ....	327
Annexe 4 : séquence du pUC18DNA <sup>ARNt</sup> .....	327
Annexe 5 : Expression, production et purification des complexes archées.....	328
<b>Liste des abréviations</b> .....	<b>333</b>
<b>BIBLIOGRAPHIE</b> .....	<b>335</b>



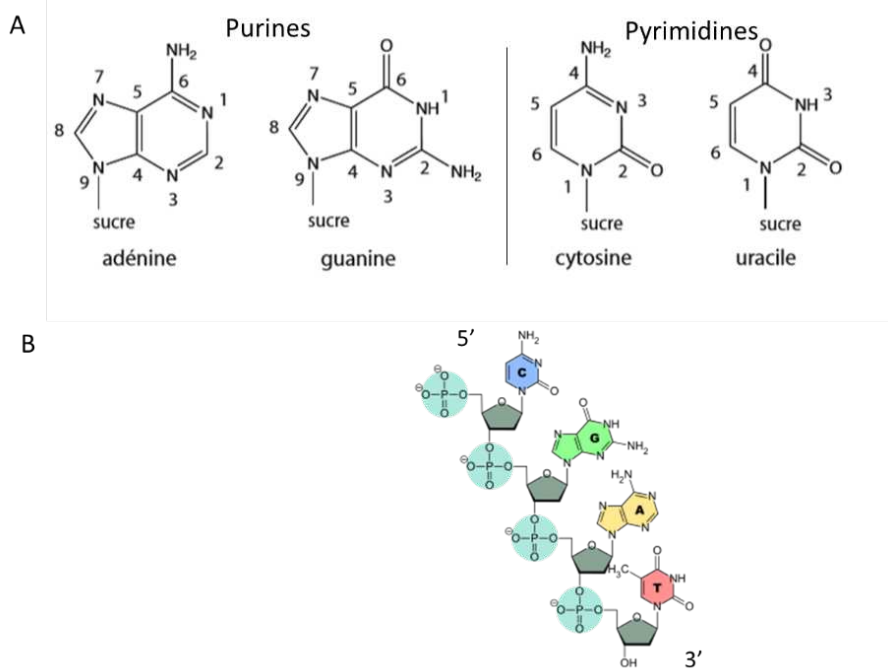


Figure 1 A : Formule chimique des bases composants les acides désoxyribonucléiques. B : Représentation de la structure primaire d'un brin d'ADN composé d'une cytosine (C), d'une guanine (G), d'une adénine (A) et d'une thymine (T).

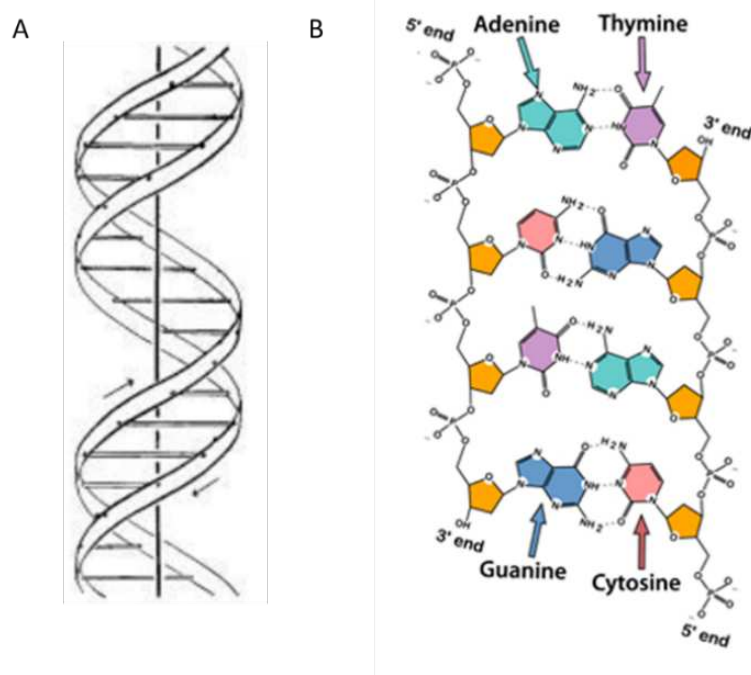


Figure 2 A : L'ADN selon Watson et Crick, les deux rubans schématisent les deux chaînes composées par les phosphates et les sucres. Les paires de bases sont représentées par les traits horizontaux. Le squelette (les phosphates et les sucres) et les paires de bases provoquent l'agencement en double hélice de l'ADN. B : Complémentarité canonique des paires de bases entre adénine et thymine et entre guanosine et cytosine formant deux et trois liaisons hydrogènes respectivement. D'après Watson et Crick, 1953.

# Introduction

## 1. Généralités sur la traduction.

Dans le début des années 1940, Avery et ses collaborateurs déclarent que l'hérédité de la cellule mère à la cellule fille est transportée par un "agent transformant" (Avery *et al.*, 1944). Cette molécule de composition inconnue est pressentie comme porteuse de l'information génétique et appelée acide désoxyribonucléique (ADN). Avec l'émergence de la chromatographie, Chargaff et ses collègues analysent la composition de cette molécule et concluent qu'elle est composée de nucléotides ; les quatre différentes bases nucléotidiques (deux bases purines : adénosine (A) et guanosine (G) ainsi que deux bases pyrimidines : thymine (T) et cytosine (C)) (figure 1 A) sont liées au désoxyribose, lui-même lié à un groupement phosphate (Chargaff, 1950 ; Magasanik et Chargaff, 1951) (figure 1 B). Malgré la connaissance de la composition de l'ADN, la communauté scientifique reste sceptique concernant sa fonction au sein de la cellule : en effet les protéines restent à l'époque considérées comme porteuse de l'information génétique (Kahn, 2003). Ce sont les travaux de Hershey et ses collègues, principalement, qui ont convaincu la communauté scientifique : l'ADN seul d'un phage infectant une bactérie, pénètre la paroi bactérienne et la bactérie hérite alors de l'information génétique du phage (Hershey, 1956).

Avec l'apparition de la technique de diffraction des rayons X et notamment de la cristallographie, la détermination de la structure de l'ADN permet de comprendre la fonction de cette molécule. Dans les années 1950, Rosalind Franklin enregistre des clichés de diffraction de l'ADN dont le pattern de diffraction forme une croix (Cramer, 2020). A partir de ces travaux et des suggestions de leur compétiteur, Linus Pauling qui suggère que l'ADN forme une hélice, Watson et Crick publient en 1953 le premier article décrivant la structure des acides nucléiques (figure 2 A) (Watson et Crick, 1953). Ils proposent que l'ADN s'agence en double hélice, les phosphates sont localisés à l'extérieur de la double hélice. L'oxygène en 3' du désoxyribose est relié avec le phosphate en 5' du nucléotide suivant formant une liaison phosphodiester (figure 2 B). Les chaînes de désoxyriboses sont perpendiculaires aux bases. Les bases nucléotidiques sont, quant à elles, orientées à l'intérieur de la double hélice, perpendiculaires à son axe et appariées par des liaisons hydrogènes. Les paires de bases formées sont restreintes entre adénine (purine) et thymine (pyrimidine) liées par deux liaisons

hydrogènes, ou entre guanosine (purine) et cytosine (pyrimidine) liées par trois liaisons hydrogènes, formant ainsi un polymère double brins (Watson et Crick, 1953) (figure 2 B).

La succession de ces bases compose la séquence de l'ADN. La complémentarité des bases confère deux copies de séquences identiques d'une molécule d'ADN après la réplication. Le génome est transmis aux cellules filles issues d'une division et conservé d'une génération à l'autre, l'information génétique portée par l'ADN est ainsi héréditaire. Le génome comporte des gènes, contenant l'information nécessaire pour produire les protéines, acteur de la fonction au sein des cellules.

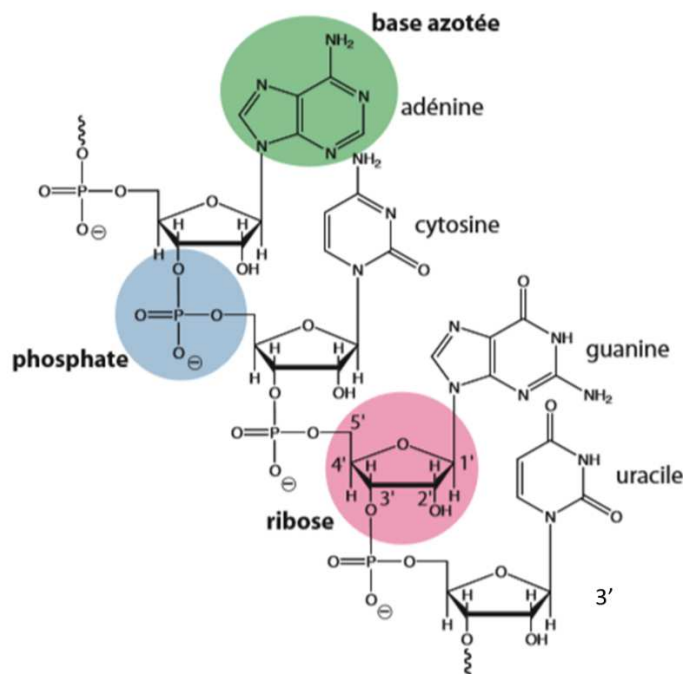


Figure 3 Représentation schématique de la structure primaire des ARN. Les trois éléments constitutifs des nucléotides sont représentés sur fond de couleur. Le groupement phosphate (en bleu), le sucre ribose (en rose) avec la numérotation des carbones du cycle (de 1' à 5'), et une base azotée (en vert) reliée par une liaison N-glycosidique au C1 du ribose.

Crick, en 1958, stipule que l'ADN est transcrit en une molécule d'ARN messager (ARNm) simple brin, qui sert de matrice pour la synthèse de la protéine (Crick, 1958). Les acides ribonucléiques (ARN) sont composés de 4 types de bases nucléotidiques, adénosine (A), guanosine (G), cytosine (C) et uridine (U), cette dernière est une thymine non méthylée. L'agencement de l'ARNm simple brin est similaire à un brin d'ADN ; la base nucléotidique est liée au ribose, lui-même lié à un groupement phosphate. La présence d'un groupement

hydroxyle (OH) à la position 2' du ribose confère à l'ARN des propriétés chimiques et structurales différentes de celles de l'ADN (chez qui ce groupement est absent) (figure 3). Les expériences de Hoagland et ses collaborateurs mettent en évidence que l'acide aminé leucine, marqué au carbone 14, peut se lier à un ARN soluble en présence d'adénosine triphosphate (Littlefield *et al.*, 1955; Hoagland *et al.*, 1958).

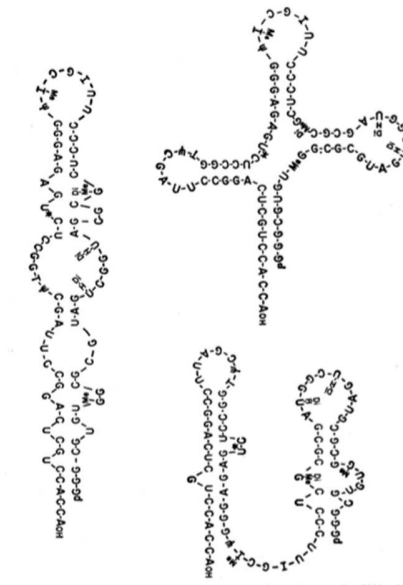


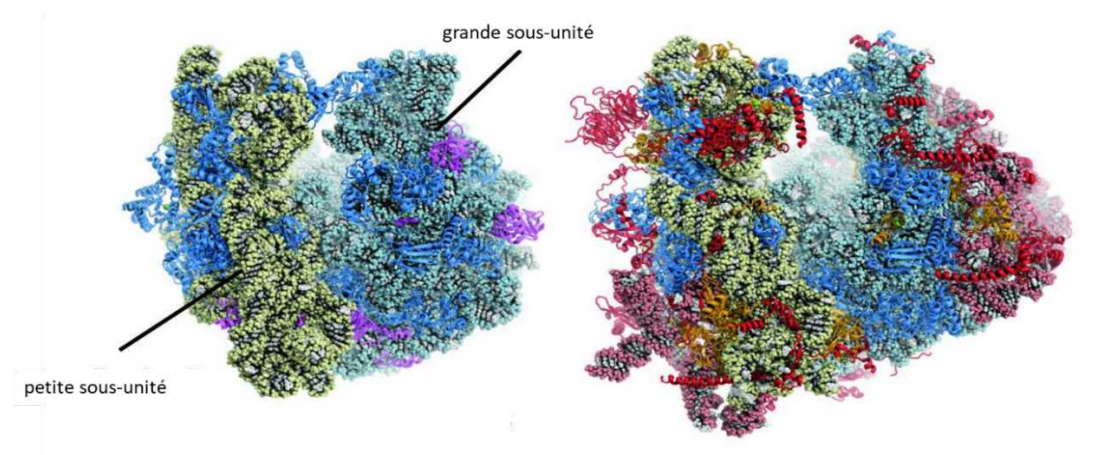
Figure 4 Représentation schématique de la structure secondaire de trois formes de l'ARNt<sup>Ala</sup> d'après Holley *et al.*, 1965.

Également, les travaux de Robert W. Holley et ses collègues mettent en évidence que certaines régions de l'ARN forment des liaisons hydrogènes, similaires à l'appariement des bases au sein de l'ADN (Holley *et al.*, 1965) (figure 4). Ces interactions impliquent un repliement secondaire de l'ARN. Ces derniers travaux prouvent que l'ARNm linéaire n'est pas le seul type d'ARN qui existe dans la cellule.

Afin de traduire l'ARNm en protéines, composées d'acides aminés, une molécule adaptatrice doit contrôler cette étape ; c'est la théorie de l'adaptateur ("the adaptor hypothesis") (Crick, 1958; Crick, 1966). Très vite, ces molécules adaptatrices sont nommées les acides ribonucléiques de transfert (ARNt). Les ARNt sont des petites molécules d'ARN faisant 76 à 98 nucléotides. La première structure cristallographique de l'ARNt<sup>Phe</sup> de levure, résolue en 1977 par Alexander Rich et Sung Hou Kim, révèle un repliement tridimensionnel inattendu en forme de L (Rich et Kim, 1977). Les trois nucléotides localisés sur la boucle anticodon (AC), en position 34, 35 et 36, forment l'anticodon qui s'apparie de façon complémentaire aux triplets de l'ARNm, permettant la traduction de l'information génétique

(séquence nucléotidique) en une protéine (séquence d'acide aminé). 21 acides aminés différents (dont la sélénocystéine et la pyrrolysine) composent les protéines dans les trois domaines du vivant. Ces acides aminés sont codés par les 63 triplets de nucléotides sur l'ARNm, complémentaires de l'anticodon porté par la boucle AC des ARNt venant s'y appairer au sein du ribosome lors de la traduction (Nirenberg et Leder, 1964). Théoriquement, 63 ARNt différents devrait décoder les 63 codons. Cependant, à titre d'exemple, le séquençage du génome bactérien a révélé que le nombre de gènes codant pour les ARNt varie entre 33 (*Mycoplasma genitalium*) et 88 (*Bacillus subtilis*), avec une moyenne comprise vers 44 ARNt. Comment expliquer que seulement environ 44 ARNt puissent décoder les 61 codons ? Crick propose la théorie du " wobble base " : l'appariement " bancal " de type non canonique entre la base 34 de l'AC et le troisième nucléotide du codon, permet de palier la disparité entre le nombre de codons et de séquences d'ARNt (Crick, 1966). Ainsi, un même ARNt peut reconnaître plusieurs codons synonymes.

Avec l'émergence du séquençage, de nombreux nucléotides modifiés sont identifiés dans les séquences des ARNt. Ces modifications consistent en l'ajout de groupement chimiques, catalysées par des enzymes. Ainsi, la séquence de l'ARNt<sup>Ala</sup> de *Saccharomyces cerevisiae* présente un nucléotide différent, l'inosine (Holley *et al.*, 1965). A ce jour, plus de 91 nucléotides modifiés sur les ARNt ont été identifiés (Boccaletto *et al.*, 2018). Ces nucléotides modifiés participent au repliement et à la stabilité des ARNt, mais également à la fidélité de la traduction (*cf.* § 2.4. page 28).



**Figure 5** Structures du ribosome de l'eucaryote *Saccharomyces cerevisiae* (à droite) et de la bactérie *Thermus thermophilus* (à gauche). Les ribosomes eucaryotes (4.3 MDa) sont plus grands et possèdent généralement plus de protéines ribosomales que les ribosomes procaryotes (2.6 MDa). En jaune et vert sont représentés les ARN ribosomiaux, en bleu, les protéines ribosomales (protéines r) universellement conservées, en violet les protéines r procaryotiques et enfin en rouge sont les protéines r eucaryotiques. D'après Ben-Shem *et al.*, 2011.

Parmi les différents types d'ARN présent dans la cellule, certains ARN interagissent avec des protéines et forment des particules atypiques observées pour la première fois par Albert Claude en 1938 dans des cellules d'embryons de poulet (Claude, 1938). Dans les années 1950, Paul Zamecnik et ses collaborateurs utilisent des acides aminés radioactifs et découvrent que ces particules " fabriquent " les protéines (Zamecnik et Keller, 1954). Ces particules sont retrouvées au sein de tous les domaines du vivant, composées de deux sous-unités : une petite sous-unité de coefficient de sédimentation 30S ou 40S (chez les procaryotes et les eucaryotes respectivement) et une grande sous-unité de coefficient de sédimentation 50S ou 60S (chez les procaryotes et les eucaryotes respectivement) (figure 5). Cette particule est nommée ribosome, car majoritairement composée d'acides ribonucléiques (ARNr) mais aussi de protéines dites ribosomales. En 1964, James Watson, propose que l'ARN ribosomal établit des contacts directs avec les ARN de transferts et les ARN messagers. Le ribosome formerait le lieu de la traduction (Watson, 1964). En 1966, cette hypothèse est validée par Peter Moore (Moore et Asano, 1966). Dans les années 2000, la structure tridimensionnelle complète du ribosome est résolue, révélant que l'enzyme responsable de la fabrication des protéines, n'est pas une protéine, mais une molécule d'ARN (Ben-Shem *et al.*, 2011; Yoshizawa, 2003; Yusupov *et al.*, 2001).

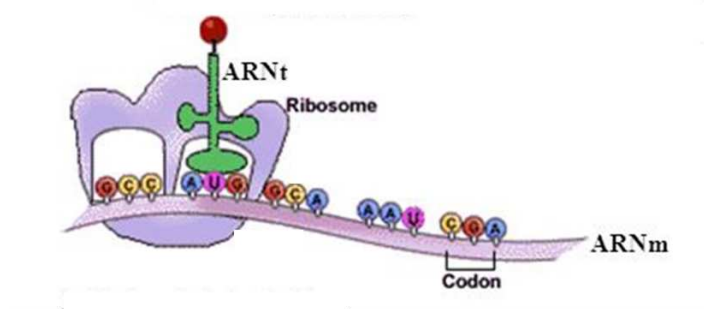


Figure 6 Représentation schématique de l'interaction entre le ribosome, l'ARNm et l'ARNt lors de la traduction. Le ribosome, composé d'ARNr et de protéines ribosomales, est la machinerie centrale de synthèse du polypeptide, où a lieu le décodage de l'information génétique, portée par l'ARN messager sous forme de codons et où s'effectue la formation de la liaison peptidique. Les ARN de transfert assurent ce décodage. Chaque ARNt porte un acide aminé spécifique et s'apparie avec l'ARNm au sein du ribosome selon les règles du code génétique qui assurent la correspondance entre codon et acide aminé.

L'ARNm, l'ARNt et le ribosome font un ménage à trois pour assurer le décodage de l'information génétique (figure 6). Concernant la traduction, bien qu'il existe des variations mécanistiques entre les archées, les procaryotes et les eucaryotes, certains mécanismes



moléculaires de ce processus restent parmi les plus conservés au sein du vivant (Ganoza *et al.*, 2002). La fonction de l'ARNt comme molécule adaptatrice fut une découverte très importante dans le domaine de la biologie et dans la compréhension des mécanismes de la traduction.

## **2. L'acide ribonucléique de transfert : la molécule au centre de la traduction.**

### **2.1. Le repliement des ARNt**

#### **2.1.1. Structure primaire.**

Les ARNt sont des molécules d'acides ribonucléiques faisant 76 à 98 nucléotides. Après leur transcription, les ARNt sont composés d'une succession de quatre types de nucléotides canoniques (A, U, C et G), mais également de pseudouridine ( $\Psi$ ) et de ribothymidine (rT) des modifications de l'uridine et de la thymidine respectivement (Hoagland *et al.*, 1958). Nous avons vu que les ARNt sont composés de trois éléments : un groupement phosphate lié à un ribose, lui-même relié à une base azotée. La première molécule d'ARNt entière a été séquencée en 1965 (Holley *et al.*, 1965). Après de nombreuses années d'études sur les ARNt, un système conventionnel a été adopté pour numéroter les nucléotides dans la séquence de l'ARNt : du numéro 1 pour celui situé à l'extrémité 5' jusqu'au numéro 76 pour celui situé à l'extrémité 3'. Les trois nucléotides de l'anticodon portent ainsi les numéros 34, 35 et 36 (figure 3). L'alignement des séquences des ARNt a permis de mettre en lumière qu'ils contiennent des nucléotides invariants et semi-variants à des positions relativement identiques (figure 7). Les nucléotides invariants sont représentés par les lettres classiques : A, U, C, G, T et  $\Psi$ . Les lettres, représentant les nucléotides semi-variants, R, Y et H évoquent, respectivement, les purines, pyrimidines et les purines hautement modifiés (Kahn, 1989). La majorité de ces nucléotides conservés sont impliqués dans les interactions intramoléculaires, qui stabilisent le repliement de l'ARNt.

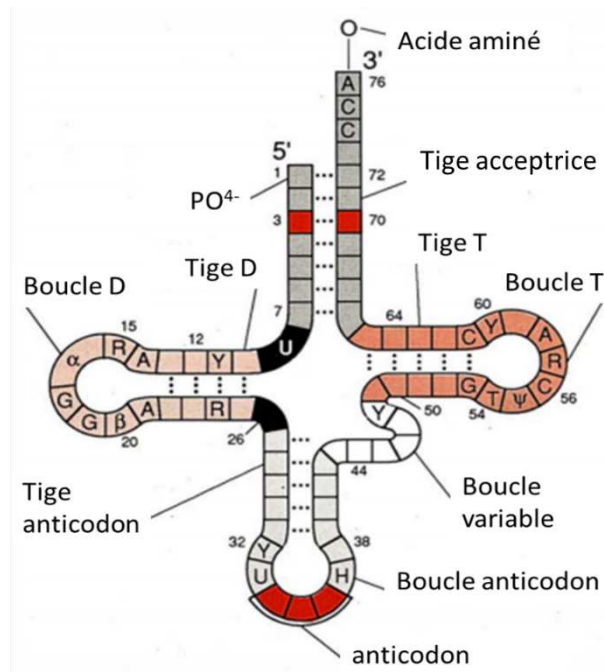


Figure 7 Représentation et numérotation des nucléotides dans la structure secondaire de l'ARNt. Les cinq régions des ARNt sont nommées sur la Figure. Les positions nommées A, G, U, C, T ou ψ représentent des nucléotides invariants toujours présents et les positions nommées R, Y ou H représentent les nucléotides semi-invariants. Les symboles α et β dans la boucle D représentent les positions variables (positions 16 et 20) occupés très souvent par une dihydrouridine. D'après Kahn, 1989.

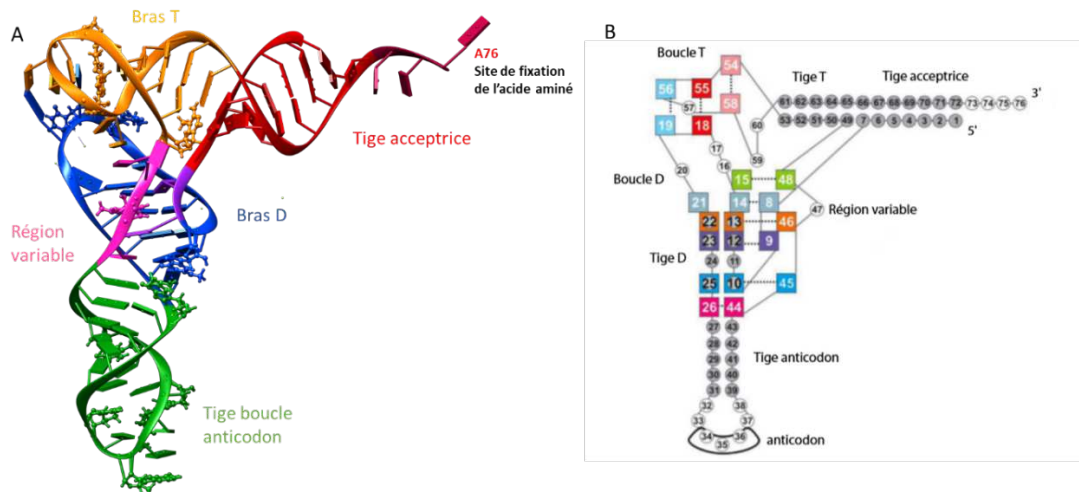
### 2.1.2. Structure secondaire

Les ARN ont la capacité de se replier sur eux-mêmes, d'adopter des structures secondaires stabilisées par l'appariement des bases complémentaires de type Watson-Crick (A-U, C-G) et parfois G-U. Les ARNt, quant à eux, adoptent une structure secondaire très stable et conservée en forme de feuille de trèfle (Holley *et al.*, 1965). Cinq régions sont ainsi formées par cette structure secondaire et sont communément appelées comme suit : la tige-boucle D, ou bras D, dénommée ainsi car elle contient souvent une ou plusieurs dihydrouridines (D) ; la tige-boucle anticodon (AC) ou bras AC ; la région variable ; la tige-boucle T ou bras T, dénommée ainsi car elle contient souvent une ribothymidine (rT) en position 54 ; et la tige acceptrice. Cette dernière tige ne possède pas de boucle car elle contient les extrémités 5' et 3' de la molécule. A l'extrémité 3', les ARNt présentent une séquence ubiquitaire, CCA, ajoutée par une transférase de nucléotides, nommée l'ARNt nucléotidyl-transférase ou CCAse, permettant l'attachement de l'acide aminé correspondant par l'amino-acyl transférase. La tige acceptrice est formée par 7 paires de bases et 4 nucléotides, incluant la séquence CCA. Le bras D est composé de 15 à 18 nucléotides, avec 3 à 4 paires de bases et 7 à 11 nucléotides localisés dans la tige et dans la boucle, respectivement. La région variable, comme son nom

l'indique, présente un nombre de nucléotides variables ([figure 7](#)), allant de 4 à 5 nucléotides mais pouvant s'allonger de 13 à 21 nucléotides pour les ARNt<sup>Ser</sup>, ARNt<sup>Leu</sup> et ARNt<sup>Tyr</sup> des bactéries (Rich et RajBhandary, 1976). La taille variable de cette région permet de classer les ARNt selon deux classes : la classe I est composée des ARNt possédant 4 à 5 nucléotides dans le bras variable ; et la classe II, où le bras variable possède 12 à 21 nucléotides (Jühling *et al.*, 2009; Chan et Lowe, 2016). La tige AC contient 5 paires de bases, tandis que la tige T en contient 5. Les boucles AC et T ont une longueur constante de 7 nucléotides. La conservation de la taille de la boucle AC lui confère 3 nucléotides invariablement numérotés 34, 35, 36, qui forment l'anticodon.

### 2.1.3. Structure tertiaire

La première structure tridimensionnelle d'un ARNt résolue est celle de l'ARNt<sup>Phe</sup> de *Saccharomyces cerevisiae* (Rich et Kim, 1977). Les extrémités de cette structure sont constituées de la tige acceptrice et de la boucle anticodon. Les structures cristallographiques à haute résolution ont permis d'observer de nombreuses interactions stabilisant le repliement tertiaire ([figure 8 A](#)). La tige T s'empile sur la tige acceptrice formant la première branche du L tandis que les tiges D et anticodon forment la seconde branche perpendiculaire à la première. La boucle T, quant à elle, s'empile sur la boucle D, formant le coude de la forme L. Cet empilement est stabilisé par des liaisons hydrogènes entre les bases ou entre les bases et les groupements du squelette sucres-phosphates ([figure 8 B](#)). Le bras D réalise un grand nombre d'interaction, d'une part avec la boucle T et d'autre part avec la tige acceptrice. Les appariements entre les nucléotides G<sub>15</sub> - C<sub>48</sub>, G<sub>18</sub> - U<sub>55</sub>, U<sub>18</sub> - C<sub>56</sub>, A<sub>26</sub> - G<sub>44</sub> et G<sub>19</sub> - C<sub>56</sub> maintiennent la structure tridimensionnelle. Il existe également des appariements entre trois nucléotides : U<sub>8</sub> - A<sub>14</sub> - A<sub>21</sub>, G<sub>10</sub> - C<sub>25</sub> - G<sub>44</sub>, A<sub>9</sub> - U<sub>12</sub> - A<sub>23</sub> et C<sub>13</sub> - G<sub>22</sub> - G<sub>46</sub>. Des appariements de type Hoogsteen inverse U<sub>8</sub> - A<sub>14</sub> et rT<sub>54</sub> - A<sub>58</sub> conservés participent aussi à la stabilisation des boucles D et T (Quigley et Rich, 1976).



**Figure 8** Structure cristallographique de l'ARNt<sup>Phe</sup> de *Saccharomyces cerevisiae* (code PDB 1EHZ). A : Les cinq régions des éléments de structure secondaire de l'ARNt<sup>Phe</sup> sont représentées par des couleurs différentes. B : Interaction tertiaire entre les bases nucléotidiques conférant la structure tridimensionnelle en forme de L des ARNt.

La structure en U-turn est un motif caractéristique des ARN et correspond à un changement de direction du squelette de la chaîne nucléotidique pouvant être causé par la présence d'une uridine. Dans les ARNt, la boucle AC possède un motif U-turn canonique induit par la présence d'une uridine invariante en position 33 (U<sub>33</sub>), en 5' de l'anticodon (Agris *et al.*, 2007; Gustilo *et al.*, 2008; Sundaram *et al.*, 2000). Cette uridine, s'empilant sur les bases 32 et 35, est essentielle pour le maintien et la conformation du U-turn de la boucle AC (Ashraf *et al.*, 1999). Cette conformation est cruciale pour la reconnaissance codon-anticodon et la translocation (Phelps *et al.*, 2002) (figure 9), assurant une distance de 75 Å entre le domaine AC et l'extrémité de la tige acceptrice qui porte l'acide aminé (Ramakrishnan, 2002). Cette distance est optimale pour permettre à l'ARNt de lire l'ARNm, via l'appariement codon - anticodon, tout en délivrant l'acide aminé au ribosome, via l'extrémité 3' lors de la traduction (Ramakrishnan, 2002). Ainsi, le repliement canonique en forme de L des ARNt est très finement lié à sa fonction.

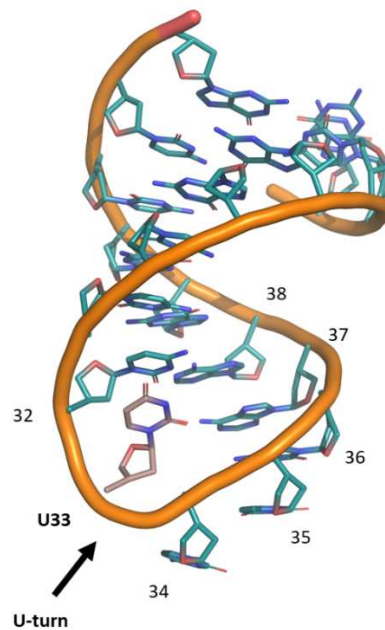


Figure 9 Structure tridimensionnelle de l'anticodon, les bases de 34 à 36 correspondent aux bases de l'AC. Un motif " U-turn " est engendré par la présence de l'uridine 33. La structure de la boucle AC provient de l'ARNt<sup>Glu</sup> de *Thermus thermophilis* (code PDB 2CVO).

#### 2.1.4. L'ARNt mitochondrial

Les mitochondries contiennent un génome réduit et linéaire, sont capables de leur propre synthèse protéique. Le séquençage du génome de la mitochondrie a permis de mettre en évidence une origine bactérienne aux mitochondries contribuant à l'élaboration du modèle endosymbiotique. Les génomes mitochondriaux codent pour deux classes majeures de gènes conservés. La première classe est composée des gènes nécessaires à la production d'énergie, telle que les sous-unités des complexes de la chaîne respiratoire. La seconde, quant à elle, est composée des facteurs requis pour la traduction, telle que les ARNt. La traduction des protéines codées dans le génome mitochondrial nécessite au moins 20 à 22 ARNt différents, selon le code génétique et les règles du "wobble". Lorsque le génome mitochondrial n'encode pas un ARNt spécifique, celui-ci est importé du cytoplasme de la cellule (Salinas-Giegé *et al.*, 2015). De nombreux génomes mitochondriaux expriment des ARNt tronqués. Par exemple, l'ARNt<sup>Ser</sup><sub>GCU</sub> mitochondrial chez les mammifères ne possèdent pas la tige D (Pütz *et al.*, 2007; Schimmel, 2018) (figure 10 A). Malgré ces changements de la structure secondaire, des études structurales montrent que le repliement tridimensionnel en forme de L est préservé, conservant la distance minimale de 75 Å entre l'extrémité CCA, localisée en 3', et l'anticodon

qui est requise pour la traduction au sein du ribosome (figure 10 B) (Suzuki *et al.*, 2011) montrant que la structure tridimensionnelle de l'ARNt a été conservée au cours de l'évolution.

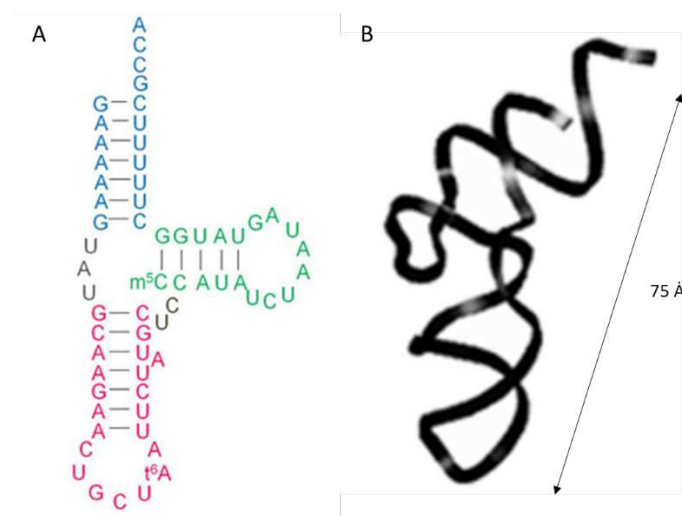


Figure 10 Structure secondaire (A) et tridimensionnelle (B) de l'ARNt<sup>Ser</sup><sub>AGY</sub> mitochondrial de vache (*Bos taurus*) montrant la conservation de la structure tertiaire de l'ARNt<sup>Ser</sup><sub>AGY</sub> malgré l'absence du bras D, d'après Watanabi *et al.*, 2014.

## 2.2. Fonctions des ARNt.

La structure tridimensionnelle en forme de L implique que les ARNt au sein des organismes interagissent avec des partenaires semblables, tels que le ribosome et les facteurs de traductions. Les variations de séquences, quant à elles, sont impliquées dans des associations spécifiques avec d'autres partenaires, tels que les aminoacyl - ARNt synthétases (aaRS) qui diffèrent selon l'acide aminé correspondant au codon de l'ARNm (Pang *et al.*, 2014). Les aaRS activent l'acide aminé par hydrolyse de l'ATP, formant un complexe aminoacyl - AMP puis catalysent le transfert de l'acide aminé activé sur la fonction hydroxyle 2' ou 3' de l'adénosine A<sub>76</sub> de l'ARNt substrat en libérant une molécule d'adénosine monophosphate (AMP), selon la classe de l'aaRS (figure 11) (Alford et Hecht, 1978; Cusack *et al.*, 1991; Eriani *et al.*, 1995). Ces enzymes doivent être très spécifiques vis-à-vis de l'acide aminé et de l'ARNt sur lequel l'acide aminé doit être chargé ; on considère qu'elles sont très fiables puisqu'elles produisent moins d'une erreur pour 10 000 catalyses (Yarus, 1979; Kurland, 1992).

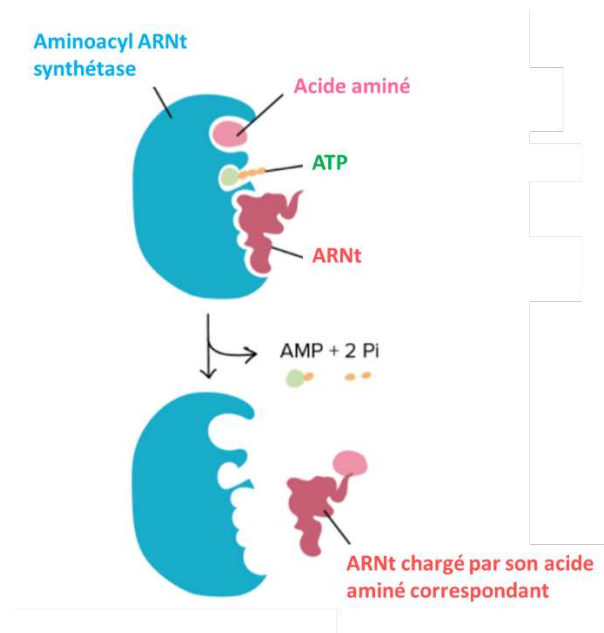
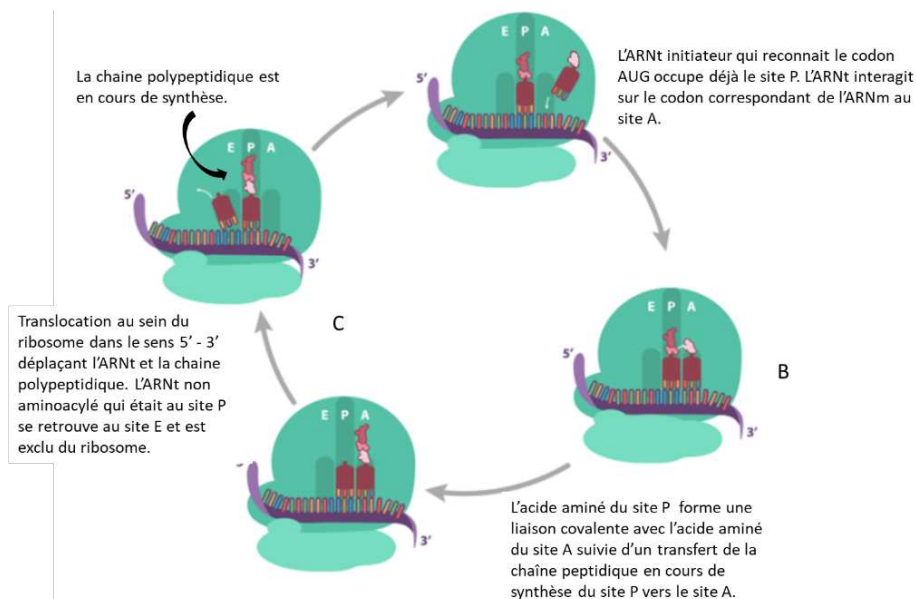


Figure 11 Représentation schématique du chargement de l'acide aminé sur son ARNt correspondant en deux étapes. L'aaRS reconnaît l'ARNt non aminoacylé. En présence de l'acide aminé et d'une molécule d'ATP, l'enzyme hydrolyse l'ATP et forme un aminoacyl-adénylate. Puis, l'aaRS par une réaction d'estérification, transfère l'acide aminé sur la position 2' ou 3'OH du ribose de l'adénosine situé à l'extrémité 3' de l'ARNt selon la classe de l'aaRS.

### 2.2.1. La traduction

Chez les eucaryotes et les procaryotes, les ribosomes sont composés de deux sous-unités formées par les ARN ribosomiaux (ARNr) et les protéines ribosomiales (figure 5). La petite sous-unité présente trois sites différents requis pour la synthèse de la chaîne polypeptidique ; les sites A (aminoacyl), P (peptidyl) et E (exit). Pendant la traduction, l'ARNm est lu dans le sens 5' vers 3' par le ribosome. L'anticodon de l'ARNt (position 34, 35 et 36) s'apparie, via des liaisons hydrogènes, sur trois nucléotides du codon de l'ARNm ; la base 34 s'apparie à la troisième base du codon, la base 35 à la seconde et enfin la base 36 à la première. La première étape de la traduction est l'initiation : un ARNt<sup>Met</sup> initiateur décode le codon AUG de l'ARNm et se positionne à la petite sous-unité ribosomale au niveau du site P (figure 12 A). Suite à cette interaction, la grande sous-unité ribosomale se lie à la petite pour former le ribosome, sous l'action des facteurs d'initiation de la traduction spécifiques aux procaryotes et eucaryotes, et ainsi la synthèse de la chaîne polypeptidique débute. Les ARNt élongateurs, quant à eux, vont décoder les codons suivants de l'ARNm et se lient au site A du ribosome (figure 12 A). Lorsque l'ARNt aminoacylé est fixé au site A, le polypeptide en formation fixé sur l'ARNt qui le précède au site P est transféré vers l'ARNt élongateur positionné au site A, via une réaction de transpeptidation (figure 12 B). Lorsque la liaison peptidique est formée, le

ribosome se transloque sur le codon suivant. L'ARNt déacylé au site P est transféré au site E et relargué du ribosome. L'ARNt liant le polypeptide se trouve au site P, libérant le site A pour que le prochain ARNt puisse s'y fixer (Reece *et al.*, 2011) (figure 12 C). Le processus se répète jusqu'à la lecture d'un codon STOP (UAA, UGA, UAG). Lorsqu'un codon STOP est reconnu, des facteurs de relargage sont mobilisés et la chaîne polypeptidique est libérée de l'ARNt par hydrolyse de la liaison peptidyl-ARNt au niveau du site P (Nakamura *et al.*, 1996).



**Figure 12** Schéma des différentes étapes de traduction. A : Après la reconnaissance de l'ARNt initiateur au sein du site P du ribosome qui decode le codon AUG, un ARNt élongateur interagit avec le codon suivant au site A. B : Une réaction de transpeptidation est catalysée entre l'acide aminé du site P et l'acide aminé du site A. C : Lorsque la liaison covalente est formée, le ribosome se déplace sur l'ARNm et transloque l'ARNt et la chaîne polypeptidique naissante au site P. L'ARNt non aminoacylé restant au site P est transloqué au site E et expulsé du ribosome.

Outre la discrimination entre les ARNt des différents codons, celle entre les ARNt initiateurs et les ARNt élongateurs est également cruciale pour l'efficacité et fidélité de la traduction. Les ARNt sont donc des molécules adaptatrices essentielles assurant le décodage de l'information génétique. Plusieurs codons peuvent coder un acide aminé identique. Le code génétique est donc dit "dégénéré". Il existe donc des ARNt qui portent des séquences anticodon différentes et sont chargés d'un même acide aminé ; ce sont des ARNt iso-accepteurs. Chez certains eucaryotes, des ARNt possèdent la même séquence anticodon, mais une séquence nucléotidique différente sur le corps de l'ARNt : ce sont les ARNt isodécodateurs (Schimmel, 2018). A ce jour, le nombre d'ARNt différents s'élève de 30 à 46 selon les organismes, ce qui est insuffisant pour décoder la totalité des codons.



### 2.2.2. Fonctions alternatives des ARNt.

Ainsi, l'ARNt présente un rôle majeur dans la traduction. Malgré cette implication indispensable et universelle au sein des organismes, les ARNt peuvent être détournés de leur fonction majeure dans la traduction pour participer à des activités différentes au sein des êtres vivants (Chatterjee *et al.*, 2018) (figure 13).

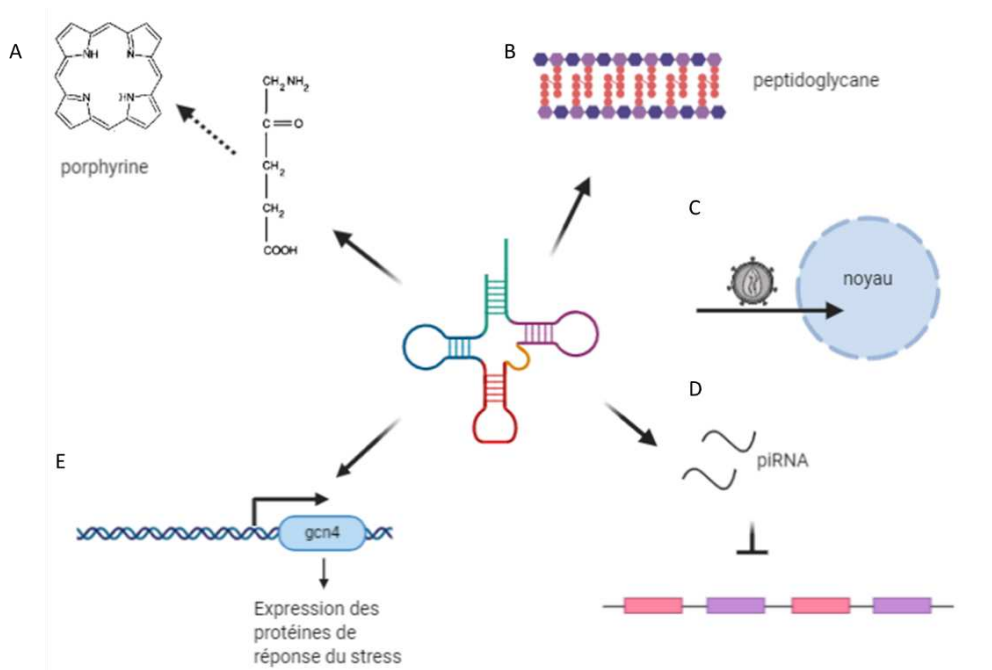


Figure 13 Fonctions alternatives des ARNt au sein de divers processus cellulaires. A : Le glutamate de l'ARNt<sup>Glu</sup> joue le rôle de co-facteur dans la synthèse de l'acide 5-aminolevulinique (ALA) qui lui-même est impliqué dans la formation de la porphyrine. B : Certains acides aminés des ARNt sont impliqués dans la synthèse du peptidoglycane chez les bactéries à gram +. C : Les ARNt peuvent servir d'amorce à la transcriptase inverse lors de la répllication de certains ARN viraux. D : Les piRNA empêchent l'expression de certains gènes transposons. E : Lors de lésions à l'ADN, des ARNt non matures s'accumulent dans le noyau et active la transcription du gène *gcn4*.

Par exemple, chez les archées, les plantes et certaines bactéries, le glutamate de l'ARNt<sup>Glu</sup> participe à la synthèse de l'acide 5-aminolevulinique qui assure la synthèse de la porphyrine (Jahn *et al.*, 1992) (figure 13 A). Également, les ARNt aminoacylés peuvent servir de sources d'acide aminés utilisés pour d'autres voies biochimiques, telles que la biosynthèse du peptidoglycane chez les bactéries à gram positive (Dare et Ibba, 2012; Shepherd et Ibba, 2013b, 2013a) (figure 13 B). Dans les cellules de levure traitées par des agents entraînant des dommages à l'ADN, l'accumulation rapide des pré-ARNt dans le noyau active la transcription de *gcn4* qui inhibe la phase G1 du cycle cellulaire (figure 13 E). De plus, il a été observé que l'activation de Gcn4 potentialise la synthèse générale des protéines, tout en induisant en

même temps l'expression des protéines nécessaires à l'adaptation au stress (Mohler *et al.*, 2017; Chatterjee *et al.*, 2018). Certains ARNt peuvent aussi servir de première amorce lors de l'initiation de la transcription inverse des rétrovirus et de certains rétrotransposons (Marquet *et al.*, 1995) (figure 13 C). En effet, le virus HIV1 (Human Immunodeficiency virus 1) utilise l'ARNt comme amorce de la transcriptase inverse des lentivirus lors de la synthèse du brin complémentaire de l'ARN viral permettant l'intégration de l'ADN du virus au noyau (Zaitseva *et al.*, 2006). Enfin, piRNA, pour PIWI interacting RNA, sont des petits ARN de 24 à 31 nucléotides, qui ont un rôle dans la régulation des gènes cellulaires, et dans l'extinction des transposons (Iwasaki *et al.*, 2015). Ces piRNA sont composés à partir des moitiés de séquence 5' des ARNt<sup>Asp</sup><sub>GUC</sub> et ARNt<sup>His</sup><sub>GUG</sub> (Honda *et al.*, 2013) (figure 13 D).

## 2.3. Synthèse des ARNt

### 2.3.1. Régulation de la synthèse des ARNt.

La synthèse des ARNt est un processus complexe qui se déroule en plusieurs étapes incluant la synthèse du transcrit initial et la maturation de l'ARNt précurseur, aboutissant au chargement de l'acide aminé à l'extrémité 3'. Le mécanisme de production est hautement performant. Entre deux divisions cellulaires, environ trois millions d'ARNt sont produits au sein de l'organisme modèle *Saccharomyces cerevisiae*, contre 60000 pour les ARNm (Waldron et Lacroute, 1975). Ces chiffres montrent que la production des ARNt contribue en grande partie aux dépenses énergétiques de la cellule (Boguta et Reese, 2018). Avant la reconnaissance par les aaRS ou par des protéines de la machinerie de traduction, les ARNt doivent subir de nombreuses étapes de maturation. Tout d'abord, le gène est transcrit en ARNt précurseur pourvu d'une séquence de nucléotides canoniques (A, U, C, G) et possédant 20 nucléotides de plus que dans la molécule mature ; ces séquences supplémentaires sont nommées extensions 5' et 3', en gris sur la figure 14 (Ramsay et Vannini, 2018).

### 2.3.2. Maturation des extrémités

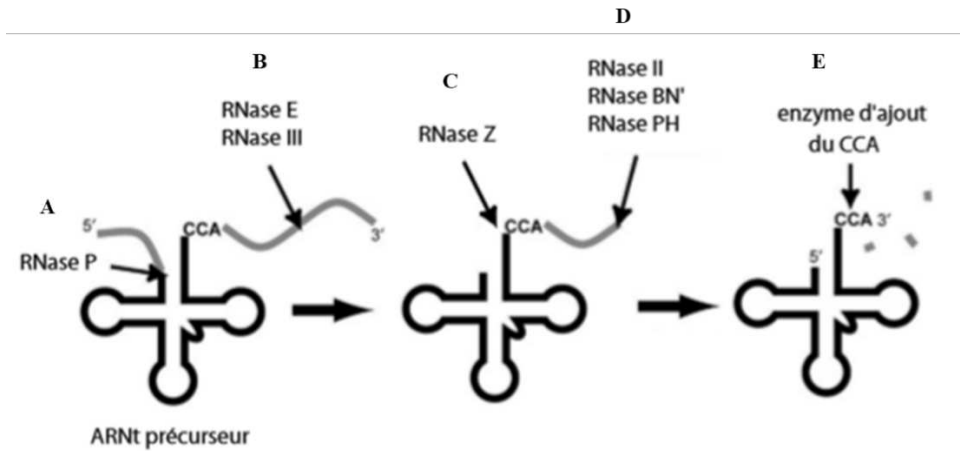


Figure 14 Etapes de maturation de l'ARNt précurseur. L'ARNt transcrit subit des étapes de maturation qui consistent en l'excision des extrémités 5' et 3'. A. L'endonucléase RNaseP clive l'extension 5'. D. L'extension 3' est clivée par des exonucléases (Rnase II, RnaseBN' et RNase PH) et B. par des endonucléases (dont la RNase Z). E. Une enzyme CCase d'ajout de la séquence CCA à l'extrémité 3' de l'ARNt finalise la maturation (dans certains cas) de l'ARNt.

#### 2.3.2.1. Maturation de l'extrémité 5'.

L'excision de l'extension à l'extrémité 5' des ARNt précurseurs est réalisée par la ribonucléase P (RNase P) (figure 14 A) (Walker et Engelke, 2006). Plusieurs sous-unités s'assemblent pour former la RNase P. Dimérique chez les procaryotes, elle est composée des sous unités : RnpA et RnpB. Plus élaborée chez les archées et les eucaryotes, elle est composée de plusieurs sous-unités différentes. A titre d'exemples, chez l'Homme, la RNase P est composée de 10 sous-unités (Rpp14, 20, 21, 25, 29, 30, 38, 40, hPop1, 5) et chez *Methanothermobacter thermoautotrophicus*, 4 sous-unités la composent (Mth11p, Mth687p, Mth688p et Mth1618p) (Hartmann et Hartmann, 2003).

#### 2.3.2.2. Maturation de l'extrémité 3'

Des endonucléases et des exonucléases s'allient pour former un ARNt mature et actif, possédant la séquence CCA à l'extrémité 3' (Hopper et Phizicky, 2003) (figure 14). Des exonucléases permettent de digérer l'extrémité 3' (figure 14 D). Aux côtés de ces exonucléases, la RNaseZ, une endonucléase conservée universellement, clive la liaison phosphodiester entre les nucléotides 73 et 74 (figure 14 C). Cette RNase appartient à la famille des métallob- $\beta$ -lactamases (de la Sierra-Gallay *et al.*, 2005). Chez certaines bactéries et archées, quelques ARNt possèdent déjà la séquence CCA aux positions 74, 75 et 76 (Vogel *et al.*, 2005; Hartmann *et al.*, 2009). Chez *Bacillus subtilis*, l'ARNt substrat possédant une séquence CCA à

la position 74 améliore l'activité catalytique de la RNaseZ. En effet la RNaseZ possède un  $K_M$  2,5 fois plus faible pour un ARNt précurseur possédant une séquence CCA que pour un ARNt précurseur possédant une séquence UAA (Pellegrini *et al.*, 2003; Redko *et al.*, 2007). Également, l'activité RNaseZ peut être inhibée par la présence de l'extension à l'extrémité 5' non encore clivée, ainsi le processus de la RNaseP précède l'activité endonucléase de la RNaseZ, ce qui ordonne la maturation de l'ARNt précurseur (Kunzmann *et al.*, 1998). Pour finir, une enzyme CCAse vient ajouter la séquence CCA à l'extrémité 3' (Augustin *et al.*, 2003).

### 2.3.2.3. Epissage de l'ARNt

Chez les eucaryotes et les archées, la maturation de certains précurseurs d'ARNt fait de plus intervenir l'épissage d'intron. L'intron est localisé majoritairement en 3' de l'anticodon, entre les nucléotides 37 et 38. Son excision fait intervenir une endonucléase, une ligase et une 2' phosphotransférase (Abelson *et al.*, 1998; Culver *et al.*, 1997; Hopper et Phizicky, 2003) (figure 15).

La maturation de l'ARNt, plus simple que celui de l'ARNm, utilise un nombre limité de protéines, chacune effectuant une réaction définie. L'ARNt mature possède une demi-vie longue (estimée à 50 heures dans les cellules des muscles de poulet) (Nwagwu et Nana, 1980), 3 jours dans les cellules de foie aviaire (Kanerva et Maenpaa, 1981) qui atteste de sa stabilité.

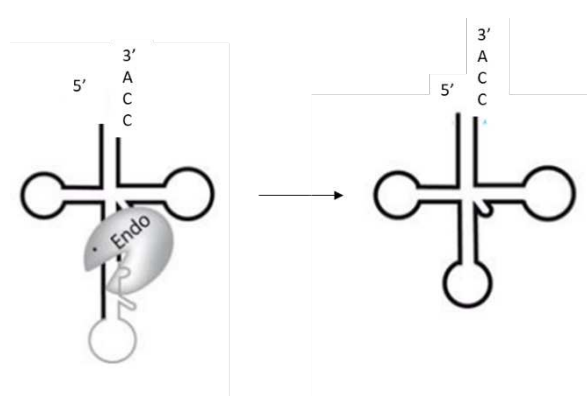


Figure 15 La réaction d'épissage des introns des pré-ARNt se réalise en trois étapes. Le pré-ARNt est clivé par une endonucléase. Les produits de réaction de l'endonucléase sont les deux moitiés épissées de l'ARNt et l'intron linéaire avec des extrémité 5'-OH et 3'-PO<sub>4</sub>- cyclique. Les deux moitiés d'ARNt sont ensuite utilisé comme substrat pour la réaction de ligation. L'enzyme ARNt-ligase catalyse la réaction de phospho-estérification entre les extrémités 3'-OH et 5'-PO<sub>4</sub>-. Enfin une 2'phosphotransférase catalyse le transfert du 2'-PO<sub>4</sub>-, présent au site 5' épissé, au NAD. Ces deux dernières étapes ne sont pas schématisées sur la figure ci-contre. D'après Hopper et Phizicky, 2003.

## 2.4. Modifications des ARNt

Les nucléotides composant l'ARNt sont le siège de nombreuses modifications chimiques post-transcriptionnelles, impliquées d'une part dans la stabilité de la structure tridimensionnelle et d'autre part dans l'efficacité et la fidélité de la traduction. À l'exception des nucléotides modifiés queuosine et archeosine qui sont introduites dans l'ARNt à la place d'une guanosine par un mécanisme de transglycosylation (Ishitani *et al.*, 2003) et l'addition spécifique d'une guanosine supplémentaire à l'extrémité 5' de l'ARNt<sup>His</sup> chez les levures par une guanylyltransférase (Nakamura *et al.*, 2018), la majorité des modifications sont réalisées sur des nucléotides déjà présents sur les ARNt. La nature de ces modifications varie et celles-ci peuvent être localisées sur les différents atomes des nucléotides (Boschi-Muller et Motorin, 2013).

### 2.4.1. Modifications post-transcriptionnelles des ARNt

Des études ont montré que les modifications chimiques des nucléotides d'ARNt peuvent se produire dans le noyau des organismes. En effet, des ARNt<sup>Tyr</sup> de la levure ont été injectés dans des noyaux d'ovocytes de Xenope, certaines modifications ont été identifiées sur ces transcrits précurseurs avant maturation (Melton *et al.*, 1980; Nishikura et De Robertis, 1981). Cependant la majorité des modifications sont effectuées après maturation des ARNt, dans le cytoplasme (Morin *et al.*, 1998), après l'excision des extrémités 5' et 3' et l'ajout de la séquence ubiquitaire CCA. Après ces étapes de maturation, l'ARNt est soumis à de nombreuses modifications chimiques post-transcriptionnelles. Le premier nucléotide non canonique, la pseudouridine  $\Psi$ , est identifié dans les années 1950. La  $\Psi$  consiste en une isomérisation de l'uridine. Le C<sup>6</sup> du noyau uracile est activé par une attaque nucléophile (réalisée par un aspartate conservé), conduisant à la rupture de la liaison glycosidique entre le ribose et le noyau. Le noyau subit alors une rotation et une nouvelle liaison entre le C<sup>5</sup> et C<sup>1</sup> du ribose se forme (Pan *et al.*, 2003).

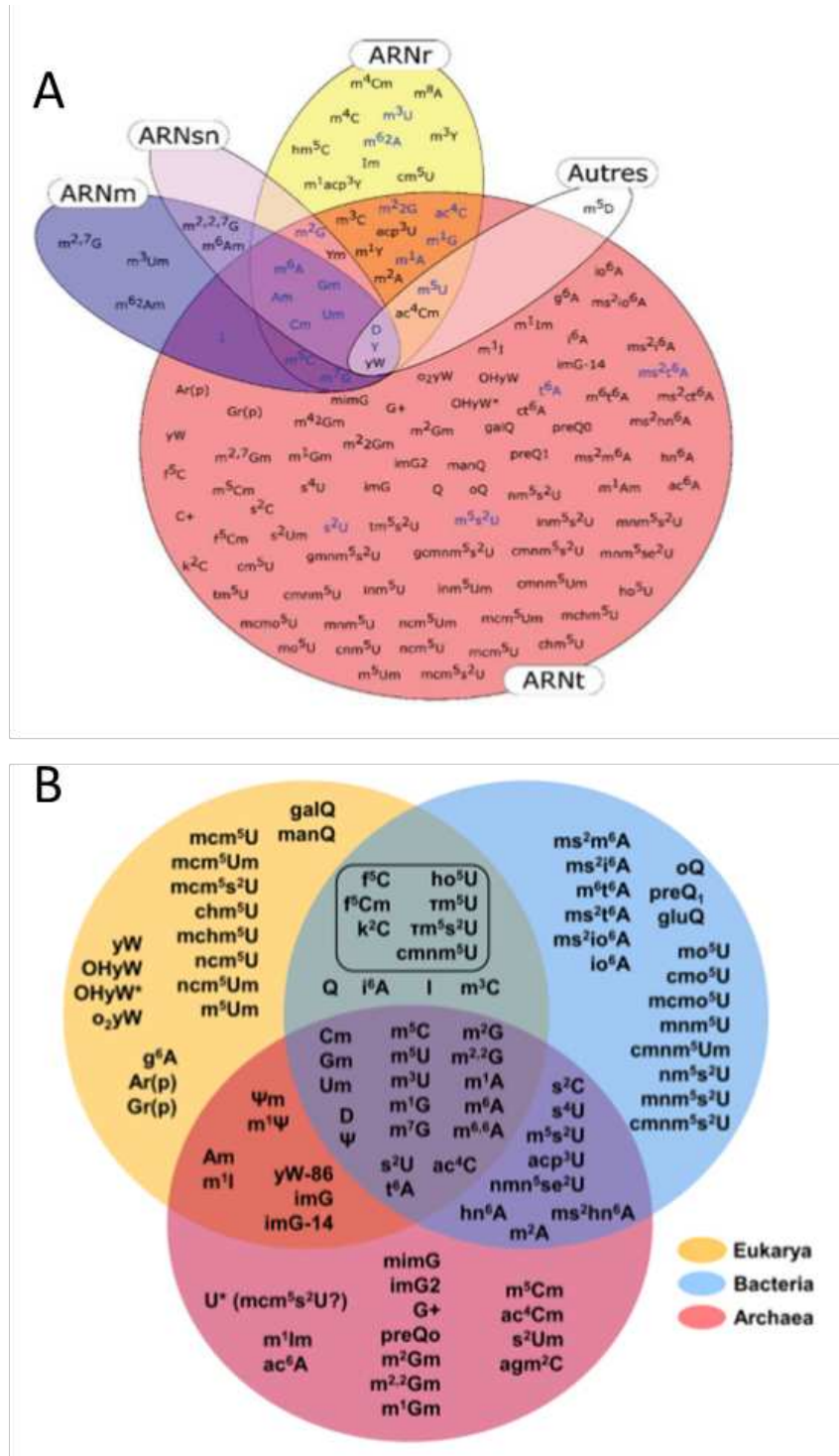


Figure 16 Distribution des nucléotides modifiés par type d'ARN (A), et sur les ARNt selon les domaines du vivant (B). Chaque couleur de fond correspond aux modifications chimiques d'un type d'ARN ou aux modifications des ARNt d'un domaine du vivant. B : Les nucléotides modifiés retrouvés dans les trois domaines du vivant sont à l'intersection des trois cercles. Les chiffres en exposant indiquent la position des groupements chimiques sur la base azotée du nucléotide. Adapté des bases de données MODOMICS (Machnicka *et al.*, 2013) et RNAMDB (Cantara *et al.*, 2011).

## 2.4.2. Fréquence et distribution des modifications post-transcriptionnelles des ARNt.

Les progrès des méthodes de séquençage ont permis de référencer plus de 172 différentes bases modifiées (Cantara *et al.*, 2011; Boccaletto *et al.*, 2018) avec une forte occurrence et diversité au sein des ARNt (Sokołowski *et al.*, 2018) (figure 16 A). Environ 91 modifications différentes sont enregistrées au sein des ARNt, avec quelques rares modifications retrouvées au sein de tous les êtres du vivant (Sokołowski *et al.*, 2018) (figure 16 B). Les ARNt appartiennent ainsi à la famille d'ARN la plus modifiée, avec en moyenne 17 % de nucléotides modifiés sur un ARNt (Jackman et Alfonzo, 2013). La nature de ces bases modifiées peut aller de simples modifications chimiques, telles que des méthylation, des nucléotides à des ajouts de groupements chimiques plus complexes. Les modifications sont localisées de façon très diverses sur l'ensemble de l'ARNt. Elles affectent n'importe quel type de base mais certaines positions sont plus fréquemment modifiées (figure 17).

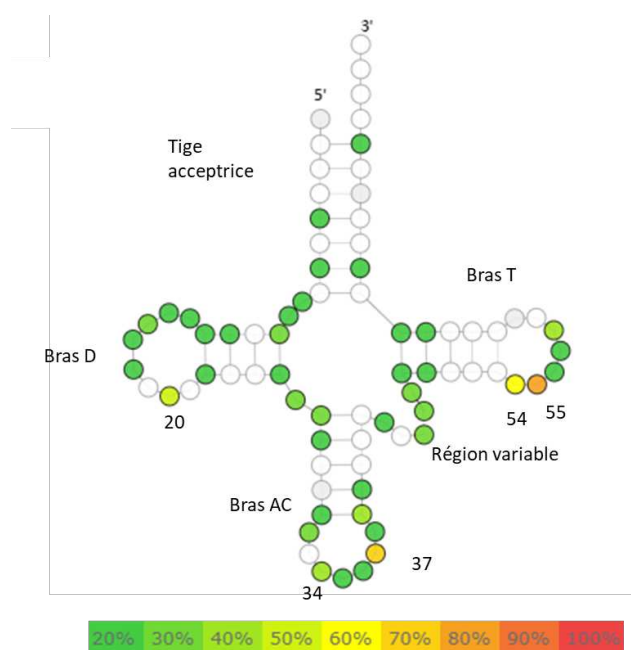


Figure 17 Représentation schématique de la proportion de modification des nucléotides de l'ARNt. Chaque nucléotide est coloré en fonction du pourcentage d'ARNt chez lesquels cette position est modifiée. Les nucléotides non colorés sont très faiblement modifiés. Les numéros (20, 34, 37, 54, 55) sont les positions les plus modifiées. Adapté de la base de données MODOMICS (<https://iimcb.genesilico.pl/modomics/>) (Machnicka *et al.*, 2013).

Les nucléotides majoritairement modifiés se situent dans les boucles T, D, dans la région variable (positions 46, 47 et 48) et dans la boucle anticodon (position 32, 34, 37 et 39). Il a été observé que les positions 32 et 37 de la boucle anticodon, 54, 55 et 58 de la boucle T et 48 de

la région variable sont des positions de nucléotides dits "invariants" (c'est à dire possédant toujours la même base) (cf. § 2.1.1. page 16) et sont souvent modifiées. A noter également qu'au niveau des positions 34 (position du "wobble") et 37 (celle qui suit la dernière base de l'anticodon) une grande diversité de modifications est retrouvée, alors que les bases 35 et 36 subissent très peu de modifications (Machnicka *et al.*, 2013). Ainsi, les nucléotides conservés à certaines positions précises sont souvent le lieu d'une forte variabilité de modifications post transcriptionnelles (cf. § 2.1.1. page 16) (Jackman et Alfonzo, 2013).

On peut distinguer deux catégories de modifications selon leur localisation sur l'ARNt. La première concerne les modifications au niveau du corps de l'ARNt. La seconde, quant à elle, concerne les modifications localisées au niveau du bras anticodon. Ces deux classes sont impliquées soit dans la stabilité de l'ARNt pour favoriser les interactions avec certaines enzymes ou soit dans l'appariement codon-anticodon pour réguler les mécanismes de traduction (Lorenz *et al.*, 2017).

Également, les modifications sur les ARNt sont classifiées selon leur degré de complexité. La première classe est celle des modifications chimiques dites simples car elle nécessite une seule réaction enzymatique. Il peut s'agir de méthylation, d'isomérisation, de thiolation ou de réduction. La seconde classe, quant à elle, regroupe les modifications dites complexes, et sollicitent plusieurs réactions enzymatiques, telles que l'isopentylation et la thréonylcarbamoylation (Sokołowski *et al.*, 2018). Il convient de noter que la plupart des modifications dans le corps de l'ARNt sont des modifications simples, tandis que la plupart des modifications complexes sont introduites dans la région de la boucle anticodon des ARNt (Grosjean et Westhof, 2016).

## **2.5. Effet des modifications**

Il est important de noter que les fonctions biologiques des ARN, et particulièrement les ARNt, au sein des cellules sont dépendantes de ces modifications.

Au sein de tous les ARNt, des modifications ont été identifiées, certaines sont universelles, suggérant leur conservation au cours de l'évolution et donc leur nécessité à la survie des organismes. Les modifications essentielles sont, de plus, très souvent localisées au niveau de la boucle AC. Les modifications en dehors de la boucle anticodon ne semblent pas, quant à elles, essentielles à la survie cellulaire (Han et Phizicky, 2018). La structure des ARNt



en forme de L, y compris pour les ARNt qui ne possèdent pas de bras D, a été conservée au cours de l'évolution, permettant d'une part, d'assurer la reconnaissance des codons par l'extrémité anticodon et, d'autre part, d'assurer la synthèse polypeptidique par l'extrémité 3' de l'ARNt au sein du ribosome.

La comparaison des structures cristallographiques de l'ARNt<sup>Phe</sup> de *Escherichia coli* non modifié (Byrne *et al.*, 2010) et de l'ARNt<sup>Phe</sup> de *Saccharomyces cerevisiae* modifié (Robertus *et al.*, 1974) révèle de légères différences concernant la structure tertiaire, sans toutefois altérer la structure globale en forme de L. L'ARNt non modifié possède donc, déjà avant maturation, sa forme canonique en L.

### 2.5.1. Effet des modifications sur la structure tridimensionnelle des ARNt.

Comme mentionné précédemment, les modifications en dehors du bras AC peuvent avoir un effet sur la structure tridimensionnelle des ARNt (Helm, 2006). Des exemples, non exhaustifs, seront décrits dans cette partie.

Il est admis que le maintien de la structure de l'ARNt est majoritairement assuré par les interactions tertiaires dans la partie coudée de la molécule, où les boucles D et T s'empilent (figure 8) (Zagryadskaya *et al.*, 2004). Il a été vu précédemment que certaines positions de l'ARNt sont souvent le lieu de modifications post-transcriptionnelles (cf. § 2.4.2. figure 17) telles que les positions 20, 54 et 55. De plus, les modifications m<sup>5</sup>U<sub>54</sub>, Ψ<sub>55</sub> et m<sup>1</sup>A<sub>58</sub> sont conservées au sein des trois domaines du vivant (figure 16 B). Il a été mis en évidence que l'addition de ces groupements méthyles contribue fortement à la stabilisation des interactions tertiaires qui maintiennent la structure tridimensionnelle de l'ARNt, en favorisant l'empilement des bases grâce au caractère hydrophobe de ces nucléotides modifiés (Agris *et al.*, 2017; Helm, 2006; Motorin et Helm, 2010). Aussi, la modification N<sup>1</sup>-méthylguanosine perturbe l'appariement canonique G-C en ajoutant un groupement méthyle sur l'azote N<sub>1</sub>. La présence de ce nucléotide modifié, m<sup>1</sup>G<sub>9</sub>, chez l'ARNt<sup>Lys</sup> mitochondrial stabilise sa structure tridimensionnelle en forme de L (Voigts-Hoffmann *et al.*, 2007). A l'inverse, une grande flexibilité est promue lorsque la dihydrouridine est présente sur les ARNt des archées psychrophiles et hyperthermophiles (Grosjean et Oshima, 2007; Ishida *et al.*, 2011). Cette modification semble empêcher certains appariements canoniques de type Watson-Crick et permettrait à l'ARNt de s'adapter aux environnements extrêmes. De même, la thiolation de la ribothymidine en position 54 (s<sup>2</sup>T<sub>54</sub>) maintient le sucre dans une conformation C3'-endo, à

cause de la répulsion stérique entre le soufre et le groupement 2'OH du ribose. Cette modification  $s^2T_{54}$  est spécifique des ARNt des organismes thermophiles et hyperthermophiles et engendre notamment une augmentation de la température de fusion ( $T_m$ ) de 3°C pour l'ARNt<sup>lle</sup> de *Thermus thermophilus* (Yokoyama, 1987). Ainsi, les modifications post-transcriptionnelles, localisées sur le corps des ARNt, peuvent stabiliser et rigidifier l'ensemble de la structure, mais aussi peuvent assouplir ce repliement afin de favoriser l'interaction avec des partenaires dans la cellule (Jackman et Alfonzo, 2013). Les modifications post-transcriptionnelles semblent maintenir une balance entre stabilité et flexibilité de l'ARNt, en fonction de l'environnement.

Il a été également observé que l'acquisition de nouvelles modifications peut dépendre de l'existence au préalable de nucléotides modifiés. Ce phénomène est connu, dans la littérature, sous le nom de "circuit des modifications" (Boguta et Reese, 2018; Han et Phizicky, 2018; Sokołowski *et al.*, 2018). Par exemple, chez *Thermus thermophilus*, un circuit de modifications est mis en place en réponse aux hautes températures. La modification  $s^2T_{54}$  est stimulée par la présence préalable de la modification  $m^1A_{58}$  (Yokoyama, 1987; Shigi *et al.*, 2006). Les modifications  $Gm_{18}$  et  $m^1G_{37}$  sont acquises après la modification  $m^7G_{46}$  en réponse aux hautes températures (Tomikawa *et al.*, 2010). Également, la modification 5-méthyl-2-thiouridine ( $m^5s^2$ ) aux positions  $U_{54}$  ou  $m^1A_{58}$  n'est pas observée si une pseudouridine en position 55 ( $\Psi_{55}$ ) est déjà présente en réponses aux basses températures (Ishida *et al.*, 2011).

Les circuits de modification ne sont pas spécifiques aux nucléotides localisés sur le corps de l'ARNt, ces événements ont été observés, également, sur la boucle AC, dont un autre exemple sera évoqué au § 2.5.1.2.

### **2.5.1. Effet des modifications sur le bras AC.**

Les modifications à proximité des trois bases de l'anticodon, au sein du bras AC, sont importantes pour le maintien de la structure tridimensionnelle de la boucle anticodon en stabilisant cette dernière et en prévenant les appariements potentiels de certaines bases adjacentes. La plupart des exemples bien documentés de modifications se produisent dans la région de la boucle anticodon et impliquent des modifications aux positions 34 et 37 qui sont très fréquemment modifiées et pour lesquelles les modifications les plus complexes sont retrouvées (Han et Phizicky, 2018; Barraud et Tisné, 2019). Les modifications des bases 34 et

37 influencent directement l'appariement de l'anticodon avec le codon de l'ARNm au sein du ribosome, modulant la traduction.

#### **2.5.1.1. Le cas de la base 34 ou " Wobble " base.**

La position 34 admet des appariements non Watson-Crick, permettant à un même ARNt de lire plusieurs codons dont la troisième base diverge : c'est la théorie du "wobble" énoncée par Francis Crick en 1966 (G peut s'apparier avec U). Cependant les propriétés physico-chimiques de la base 34 non modifiée ne sont pas suffisantes pour expliquer toutes les possibilités d'appariements codon-anticodon observées expérimentalement. En fait, le phénomène du "wobble" repose en grande partie sur la présence de nucléotides modifiés en position 34. En effet, 39 modifications sont identifiées sur la "wobble" base, devenant une des bases les plus modifiées des ARNt, en particulier quand cette dernière est une uridine U (Björk *et al.*, 1987; Yokoyama *et al.*, 1985). La base 34 est donc hyper-modifiée et ces modifications peuvent étendre ou restreindre l'appariement "wobble" (" extended " ou " restricted wobble ") (Agris *et al.*, 2017). L'extension signifie qu'en plus de l'appariement possible entre une uridine et une guanosine, s'ajoute la possibilité de l'appariement avec une cytosine. La restriction, par la présence d'une modification sur la base 34, limite l'appariement avec sa base complémentaire ou la change avec une autre base. Les modifications chimiques de la base 34 résolvent le problème causé par les codons synonymes, qui sont reconnus par différents ARNt iso-accepteurs. Voici des exemples non exhaustifs de modifications sur la base 34 qui sont récapitulés à la [figure 19](#).

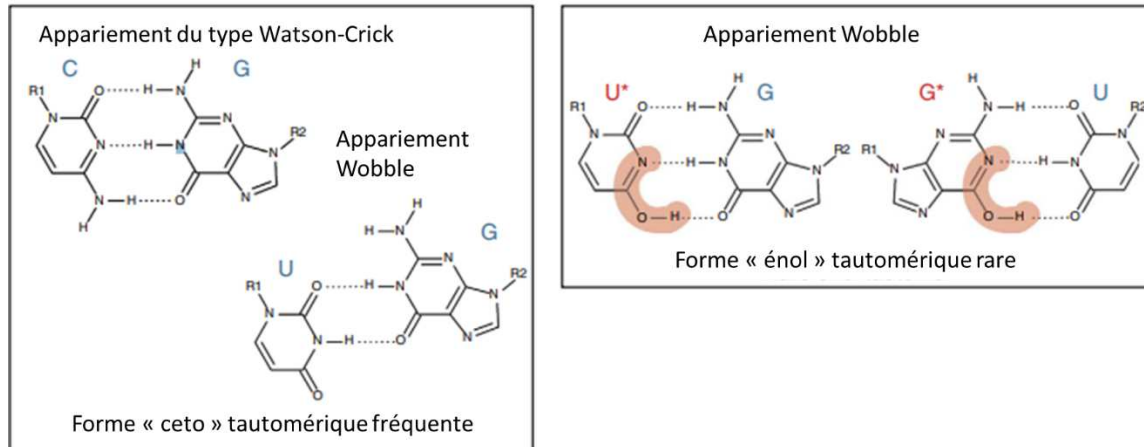
modifications	Position 34 anticodon	Appariement sans modification avec nucléotide du codon	Appariement avec modification avec nucléotide du codon	Espèces
Q (queuosine)	G	C	U	Procaryotes
mQ (mannosyl-queuosine)	G	C	U	Eucaryotes
I (Inosine)	A	U	A, G, U	Procaryotes
mcm <sup>5</sup> s <sup>2</sup>	U	A, G	A	Eucaryotes
mn <sup>5</sup> s <sup>2</sup>	U	A, G	A, G	Procaryotes
cm <sup>5</sup>	U	A, G	A, G, C, U	Procaryotes
mo <sup>5</sup> (methoxy)	U	A, G	A, G, C, U	Procaryotes
k <sup>2</sup> (lysidine)	C	G	A	Procaryotes
agm <sup>2</sup> (agmatidine)	C	G	A	Archées
ac <sup>4</sup> (acetyl)	C	G	G	Procaryotes

Figure 19 Exemples tirés de la littérature de modifications de la base 34 et effet sur l'appariement anticodon-codon.

La présence d'une queuosine, chez les bactéries, ou d'une mannosyl-queuosine chez les eucaryotes, à la place d'une guanosine en position 34 favorise l'appariement avec une uridine plutôt qu'avec une cytosine (Harada et Nishimura, 1972; Meier *et al.*, 1985). La modification mcm<sup>5</sup>s<sup>2</sup> sur la base U<sub>34</sub> des ARNt<sup>Lys,Gln,Glu</sup> chez les eucaryotes est la modification la plus caractérisée dans la littérature, elle consiste en une double méthylation par la méthyltransférase Trm9-Trm112 et une thiolation par URM1 sur les positions C5 et C2 respectivement de l'uridine (Agris *et al.*, 1973; Yokoyama *et al.*, 1985; Lustig *et al.*, 1993 ; Glatt *et al.*, 2012). Le nucléotide modifié mcm<sup>5</sup>s<sup>2</sup>U s'apparie de façon stricte à l'adénosine et empêche l'appariement U-G, prévenant une mauvaise lecture des codons AUG (Met). La lysidine sur la base C<sub>34</sub> (k<sup>2</sup>C<sub>34</sub>) reconnaît une adénosine au lieu d'une guanosine, et facilite le décodage du codon AUA (Ile) par l'ARNt<sup>Ile</sup><sub>CAU</sub> au sein des bactéries en prévenant la lecture du codon AUG (Met). Tandis que la modification N<sup>4</sup>-acétylcytidine (ac<sup>4</sup>C) sur les ARNt<sup>Met</sup><sub>CAU</sub> prévient la lecture du codon AUA et favorise l'appariement avec le codon AUG (Harada et Nishimura, 1974 ; Muramatsu *et al.*, 1988b ; Weber *et al.*, 1990 ; Matsugi *et al.*, 1996 ; Grosjean et Björk, 2004). L'équivalent de la modification k<sup>2</sup>C au sein des archées est l'agmatidine agm<sup>2</sup>C.

Au sein des bactéries, la modification mn<sup>5</sup>s<sup>2</sup> sur la base U<sub>34</sub> des ARNt<sup>Lys</sup><sub>UUU</sub> (ARNt iso-accepteur unique pour la moitié des procaryotes) permet d'étendre le décodage du codon AAA au codon synonyme AAG (Rozov *et al.*, 2016). La paire wobble G – U a été caractérisée,

le nucléotide modifié  $mnm^5s^2U34$  forme deux liaisons hydrogènes identiques à celle de l'appariement wobble, tandis que sous sa forme protonée, une liaison hydrogène supplémentaire peut se former entre l'atome  $N^3$  de l'uracile avec le groupement amine de la guanosine formant un état tautomères rares (Rozov *et al.*, 2016) (figure 20).



**Figure 20** Différents types d'appariement de la base G avec les bases nucléotidiques. L'appariement Watson-Crick (à gauche) admet seulement des liaisons hydrogènes entre les paires de base G-C, trois liaisons hydrogènes stabilisent cet appariement. L'appariement "wobble" permet de former la paire de base G-U, deux liaisons hydrogènes le stabilise. A droite un réarrangement des bases (via une forme énole tautomérique rare) permet de stabiliser par trois liaisons hydrogènes la paire de base G-U.

La désamination de l'adénosine 34 en inosine  $I_{34}$  permet son appariement avec les bases A, U mais aussi G. Chez les procaryotes, très peu de gènes codent les ARNt<sub>CNN</sub>. La modification  $I_{34}$  permet de combler le défaut de ces ARNt décodant les codons NNG (Kanaya *et al.*, 1999). La modification  $cmo^5U$  à la position 34 des ARNt<sup>Val</sup><sub>UAC</sub> d'*Escherichia coli* permet de décoder efficacement les quatre codons valines GUU/GUC/GUA/GUG. Récemment, un nouveau nucléotide modifié a été découvert : la methoxyuridine  $mo^5U_{34}$  qui, au sein des procaryotes sur les ARNt<sup>Ala</sup>, permet l'appariement de l' $U_{34}$  avec les bases A, C, G et U (Ryu *et al.*, 2018). Dans les mitochondries, l'ARNt<sup>Met</sup><sub>CAU</sub> peut être modifié par un groupement 5-formyle sur la cytidine 34 permettant à cet ARNt modifié de reconnaître et décoder le codon usuel AUG mais également le codon AUA, alors que ce dernier est normalement assigné à l'isoleucine dans l'usage codon canonique (Agris *et al.*, 2017; Haag *et al.*, 2016; Jia et Higgs, 2008). Ces exemples mettent en évidence que certaines modifications permettent à un seul ARNt iso-accepteur de s'apparier à tous les codons synonymes. Au final, il a été mis en évidence que les modifications de la base 34 induisent des changements physico-chimiques affectant l'appariement de l'ARNt.

En majorité selon les exemples documentés, les modifications de la base 34 engendrant une restriction de la paire "wobble" impliquent généralement la fidélité de la traduction lors de la reconnaissance du codon initiateur AUG (Met), alors que les modifications engendrant une liberté d'appariement semblent combler un défaut d'ARNt au sein de l'espèce.

### 2.5.1.2. Le cas de la base 37 ou "dangling" base :

Tout comme la position 34, le nucléotide 37 de l'ARNt est la cible de nombreuses modifications aux effets divers. La position 37, en 3' de l'anticodon, est occupée par une purine (majoritairement une adénine), qui est fortement modifiée (Westhof et Auffinger, 2001). La purine 37 change de conformation lors de la liaison à l'ARNt au sein du site A du ribosome. Pour cette raison elle est appelée la "dangling base" (la base se balançant) (Grosjean *et al.*, 1976). En complexe avec l'ARNm au sein du ribosome, elle se positionne au-dessus du couple formé par la troisième base de l'anticodon, en position 36, et la première base du codon de l'ARNm (Agris *et al.*, 2007; Väre *et al.*, 2017). La boucle AC non modifiée en position 37 a tendance à adopter une conformation fermée en raison de liaisons hydrogènes intra-boucle entre la base conservée U<sub>33</sub> et la base purique à la position 37 (Stuart *et al.*, 2000) (figure 21 A).

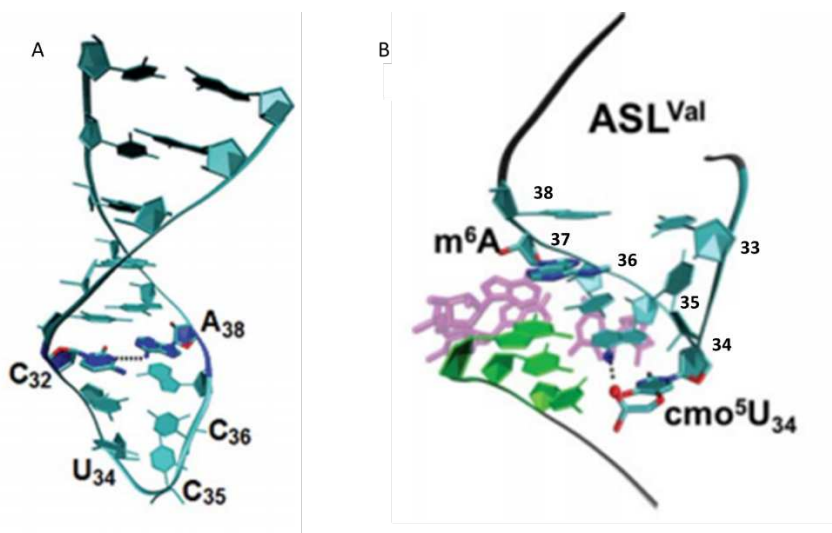


Figure 21 Repliement de la boucle anticodon stabilisé par des appariements entre les bases en présence (B) et en absence (A) de modifications sur la base 37. D'après Stuart *et al.*, 2000

Les modifications de la position 37 aident à ouvrir la conformation de la boucle AC, prévenant l'appariement Watson-Crick avec des nucléotides voisins (figure 21 B) et contribuent également à la formation de la structure canonique de demi-tour (U-turn)

importante pour l'appariement codon-anticodon lors de la traduction. Voici, ci-après, quelques exemples de modifications de la base 37.

La modification 2-methylthio-N6-threonylcarbamoylation  $ms^2t^6A_{37}$  sur l'adénosine 37 de l'ARNt<sup>Lys</sup> chez l'Homme prévient l'appariement entre  $U_{33}$  et  $A_{37}$ , stabilisant la forme de la boucle anticodon fonctionnelle pour la traduction (Vendeix *et al.*, 2012).

La méthylation de la guanine 37 ( $m^1G_{37}$ ) est une modification conservée à travers les 3 domaines du vivant et est présente dans les ARNt décodant les codons CNN et UNN, prévenant les erreurs de décalage du cadre de lecture du codon (Goto-Ito *et al.*, 2017).

La modification  $t^6A_{37}$ , qui consiste en l'ajout d'un groupement thréonyl-carbamoyle joue un rôle similaire à la modification  $ms^2t^6A_{37}$  mais dans les ARNt reconnaissant les codons ANN (Stuart *et al.*, 2000).

Une modification de la base 37 ajoute souvent un groupement hydrophobe (Björk *et al.*, 1987), ce qui renforce l'empilement de la base 37 avec la base 38 dans la boucle AC (Witts *et al.*, 2013). Ceci a pour conséquence de maintenir l'appariement entre la base 36 de l'anticodon et la première base du codon (Auffinger et Westhof, 2001) mais aussi de favoriser le repliement de la boucle en conformation ouverte " u-turn " plus adaptée à l'appariement sur l'ARNm au sein du ribosome. Les modifications de la "dangling" base jouent donc un rôle important dans la structure de la boucle anticodon, ce qui confère un impact capital lors de la lecture des codons de l'ARNm et donc l'efficacité et la fidélité de la traduction.

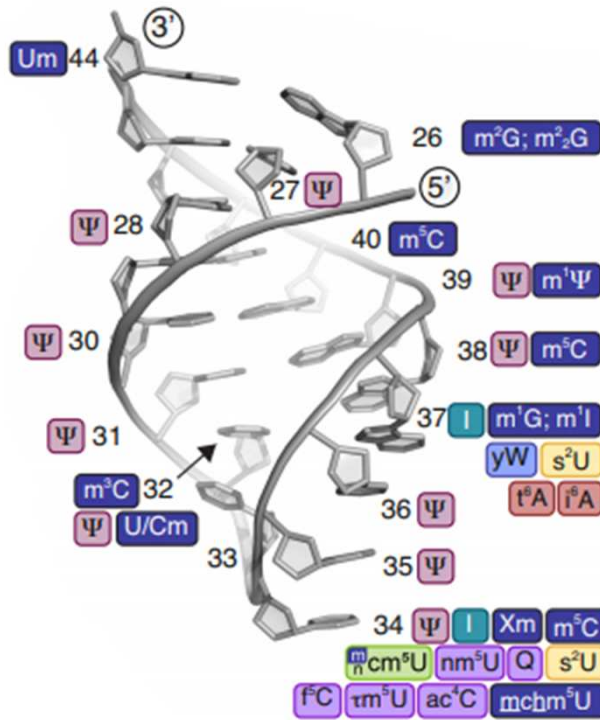


Figure 22 Localisation des nucléotides modifiés sur la boucle AC de l'ARNt. D'après Krutyhołowa *et al.*, 2019.

Beaucoup de modifications se trouvent généralement aux positions 34 et 37 des ARNt (figure 22). De plus, les modifications retrouvées sur les nucléotides de la boucle AC sont souvent dépendantes les unes des autres. Parmi les exemples récents de ces circuits de modifications, la formation de wybutosine (yW) sur  $m^1G_{37}$  de ARNt<sup>Phe</sup>, chez *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* et *Homo sapiens*, est grandement stimulée par la présence du 2'-O-methylribose des positions C<sub>32</sub> (Cm) et G<sub>34</sub> (Gm) (Guy *et al.*, 2012; Guy et Phizicky, 2015; Guy *et al.*, 2015); la 3-méthycytidine ( $m^3C$ ) en position 32 est fortement stimulée par la formation antérieure  $i^6A_{37}$  chez *Schizosaccharomyces pombe* et *Saccharomyces cerevisiae* (Arimbasseri *et al.*, 2016), ou par  $t^6A_{37}$  (Han et Phizicky, 2018) ; alors que la modification 5-méthycytidine ( $m^5C$ ) à la position C<sub>38</sub> chez *Schizosaccharomyces pombe* est stimulée par la formation antérieure que la queuosine (Q) au résidu 34 (Boschi-Muller et Motorin, 2013). Pour finir, chez les procaryotes, la modification  $k^2C$  semble possible si la modification  $t^6A$  est déjà présente (Agris, 2004; Grosjean et Björk, 2004; Sonawane et Sambhare, 2015).

### 2.5.2. Défauts de modifications reliées à des maladies génétiques.

L'importance de certaines modifications post-transcriptionnelles des ARNt s'illustre également dans certains exemples de maladies génétiques.



Il a été vu précédemment que l'acquisition de ces nucléotides modifiées est maintenant reconnue pour stabiliser et réguler le repliement tridimensionnel, d'une part, et affiner la reconnaissance et la lecture des 61 codons du code génétique, afin de synthétiser de façon fidèle les protéines, d'autre part. Selon leur complexité, ces modifications sont synthétisées par une ou plusieurs enzymes et peuvent être influencées par des modifications antérieures à d'autres positions sur l'ARNt (circuit des modifications). L'étude des enzymes de modifications a mis en évidence un lien direct entre le phénotype des organismes (voire des pathologies chez l'Homme) et les modifications de l'ARNt. En voici quelques exemples.

Les méthyl-transférases FtsJ (chez l'Homme) et Trm7 (chez *Saccharomyces cerevisiae*) induisent la méthylation de la base 32 et 34 des ARNt<sup>Leu</sup>, ARNt<sup>Phe</sup>, ARNt<sup>Trp</sup>. Des mutations dans les gènes codant FtsJ/Trm7 ont été reliées à des retards mentaux chez l'Homme et à des défauts sévères de croissance chez la levure (Townes et Begley, 2012)

Le gène *NOP2/SUN* code une méthyltransférase qui catalyse la méthylation sur la position 34 des ARNt<sup>Leu</sup> mais également aux positions 48, 49 et 50 d'autres ARNt (Brzezicha *et al.*, 2006). Des mutations dans ce gène induisent chez l'Homme un syndrome autosomique récessif entraînant des difficultés intellectuelles.

D'autres gènes codant pour les enzymes de modifications sont reliés à l'apparition de cancers : la surexpression de TRM12, qui catalyse la formation de la wybutosine sur la position 37 des ARNt<sup>Phe</sup>, est observée dans les cellules cancéreuses du sein (Frye *et al.*, 2016).

Pour finir, un défaut de la modification ms<sup>2</sup>t<sup>6</sup>A sur la base 37 des ARNt<sup>Lys</sup> induit une synthèse anormale de la proinsuline. Ce dernier exemple relie les modifications aux pathologies de diabète de type II (Wei et Tomizawa, 2012).

Ainsi, les modifications des ARNt présentent un rôle central dans le développement cellulaire des organismes (Torres *et al.*, 2014). Il semblerait que les défauts de modifications soient corrélés à l'altération de l'efficacité et de la fidélité de la traduction qui induit une mauvaise synthèse ou un mauvais repliement des protéines synthétisées (Sokołowski *et al.*, 2018).

### 3. La modification t<sup>6</sup>A

#### 3.1. La découverte d'un groupement supplémentaire sur la base 37

A partir des années 1960, la caractérisation de groupements chimiques sur les ARN s'est accentuée fortement. Le premier groupement thréonine sur une adénosine des ARNt d'*Escherichia coli*, de *Saccharomyces cerevisiae* et des tissus mammifères (foie de veau), est isolé par Chheda en 1969 (Chheda *et al.*, 1969). Il a été montré depuis que la caractérisation de ce dérivé d'adénosine par Résonance Magnétique Nucléaire notamment, résulte en l'ajout d'un groupement thréonyle qui est lié par sa fonction amine via un groupement carbamoyle, à l'azote en position 6 (N<sup>6</sup>) de l'adénosine 37 (A<sub>37</sub>) (figure 23). (Schweizer *et al.*, 1969 ; Takemura *et al.*, 1969).

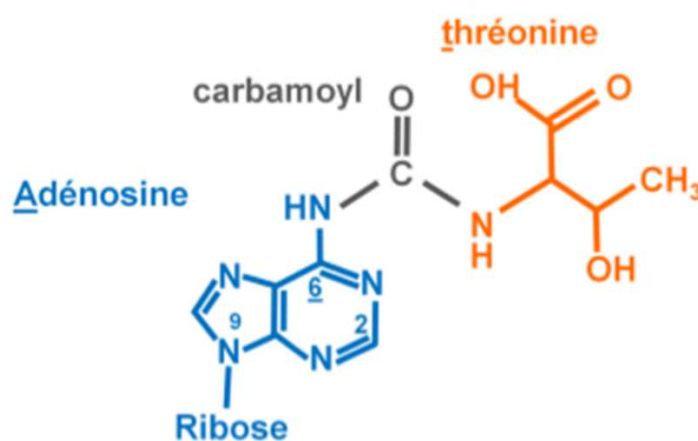


Figure 23 Structure chimique de la modification t<sup>6</sup>A sur l'adénosine 37. Une réaction de thréonylcarbamoylation est catalysée sur la position N<sup>6</sup> de l'adénine. En orange est représenté le groupement thréonine et en gris le carbonyle.

Après la découverte du monde des archées, le nucléotide modifié est retrouvé également sur l'ARNt<sup>Met</sup> de *Thermoplasma acidophilus* (Kilpatrick et Walker, 1981) puis sur certains ARNt décodant les codons ANN au sein de tous les archées (Edmonds *et al.*, 1991), généralisant cette modification aux eucaryotes, procaryotes et archées. La conservation de cette modification est donc l'objet de nombreuses études. L'évolution des techniques de séquençage et la mise en place des banques de données (Sprinzl et Vassilenko, 2005; Sprinzl *et al.*, 1998) permettent de compléter la répartition des modifications sur les ARNt au sein des domaines du vivant. Le nucléotide modifié t<sup>6</sup>A est alors retrouvé au sein de toutes les espèces, y compris les mitochondries et chez les organismes aux génomes réduits, tels que les

Mycoplasmes sur certains ARNt décodant exclusivement les codons ANN (Grosjean *et al.*, 1995). Ce caractère universel indique l'importance de cette modification pour la vie des organismes.

Afin d'identifier la modification t<sup>6</sup>A, des substrats ARN radiomarqués sont injectés dans l'ovocyte de xénope. Les ARNt radiomarqués sont ensuite purifiés avant d'être digérés en nucléosides et leur contenu analysé. Cette approche permet de suivre la cinétique des réactions de synthèse de t<sup>6</sup>A sans connaître les enzymes qui en sont responsables (Grosjean *et al.*, 1987). Différents ARNt de *Saccharomyces cerevisiae* ont été injectés. Au final, la structure en forme de L de l'ARNt et la maturation des extrémités 5' et 3 sont nécessaires à la thréonylcarbamoylation de l'adénosine A37 (Morin *et al.*, 1998). Ces études suggèrent que, pour que cette modification ait lieu, l'ARNt n'accepte que de légères perturbations de sa structure tridimensionnelle et que la machine de biosynthèse t<sup>6</sup>A est sensible aux étapes de maturation de l'ARNt. Également, une uridine en position 36 est absolument nécessaire pour la formation de la modification t<sup>6</sup>A, et une adénosine en position 38 la facilite (Morin *et al.*, 1998), suggérant que cette modification soit dépendante de la séquence des bases nucléotidiques qui l'encadrent.

### **3.2. Impacts structuraux de la modification t<sup>6</sup>A**

La structure de l'ARNt<sup>Phe</sup> de *Saccharomyces cerevisiae* possédant le nucléotide modifié t<sup>6</sup>A montre que la base 37 modifiée ne peut plus former de liaisons hydrogènes du type Watson-Crick avec d'autres bases de la boucle AC (Parthasarathy *et al.*, 1978; 1974). L'ajout du groupement thréonyl-carbamoyle confère donc des propriétés physico-chimiques particulières à l'ARNt ([figure 24](#)).

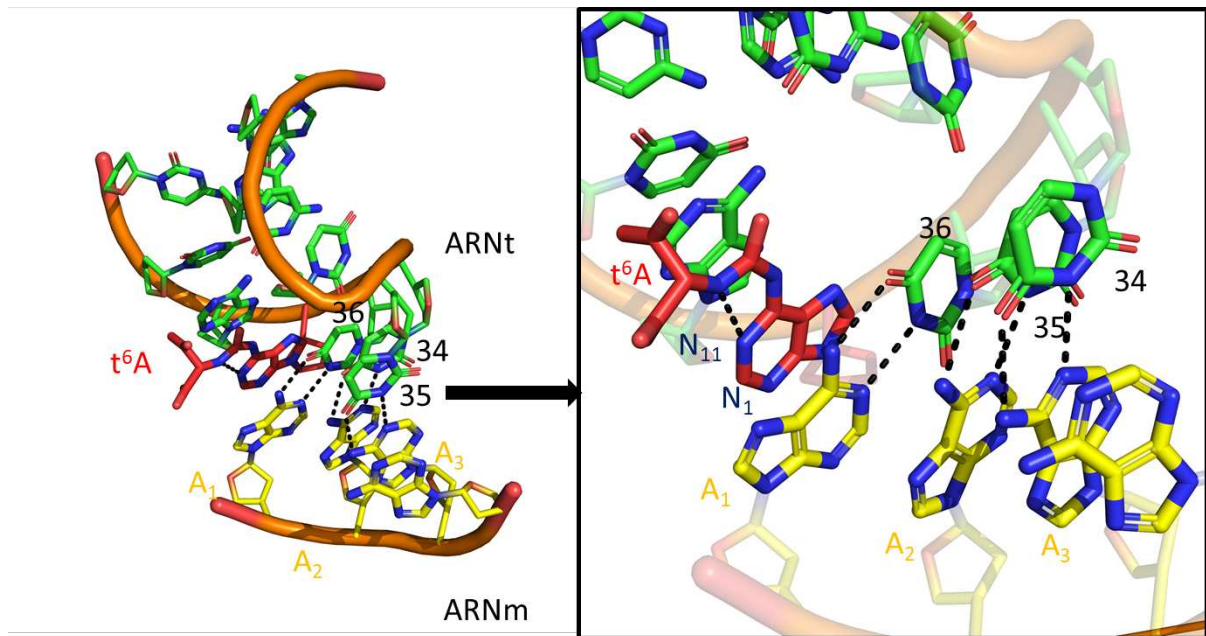


Figure 24 Structure de l'ARNt<sup>Lys</sup><sub>UUU</sub> (dont les nucléotides sont en verts) apparié à l'ARNm AAAA (dont les nucléotides sont en jaune) au sein du ribosome (code PDB 1XMQ), l'encadré à droite est un zoom sur l'appariement codon-anticodon. Le t<sup>6</sup>A colorié en rouge est planaire à l'anneau de l'adénine et son orientation à l'extérieur de la boucle AC est stabilisé par une liaison hydrogène entre les atomes N<sup>11</sup> et N<sup>1</sup>. Cet encombrement stérique empêche une interaction potentielle entre l'atome N<sup>1</sup> de l'adénosine 37 modifiée et l'atome N<sup>6</sup> de l'uridine 33.

En effet, la base modifiée s'oriente vers l'extérieur de la boucle AC (Parthasarathy *et al.*, 1974). La structure de l'ARNt<sup>Lys</sup><sub>UUU</sub> (code PDB 1XMQ) au sein de la sous-unité 30S du ribosome montre qu'une liaison hydrogène entre les atomes N<sup>11</sup> et N<sup>1</sup> se forme et empêche les deux azotes N<sup>6</sup> et N<sup>1</sup> de l'adénine de s'apparier avec d'autres nucléotides ; ceci prévient donc l'appariement intra-boucle entre l'adénosine 37 et l'uridine 33 (figure 24 A) (Stuart *et al.*, 2000; Yarian *et al.*, 2002; Agris *et al.*, 2007). L'incapacité de former des liaisons hydrogènes stabilise la boucle AC en conformation ouverte, assurant la formation du motif " U-turn " canonique, favorable pour la liaison au codon de l'ARNm (Lescrinier *et al.*, 2006). Cette conformation favorise également l'effet d'empilement de l'adénine modifiée avec l'adénine 38 et la première base du codon A<sub>1</sub> (Sonawane et Sambhare, 2015). L'encombrement stérique apporté par le groupement hydrophobe thréonyl-carbamoyle protège l'appariement A<sub>1</sub>-U<sub>36</sub> énergétiquement faible de l'hydrolyse (car seulement deux liaisons hydrogènes stabilisent cette paire de base) (Freier et Tinoco, 1975; Högenauer *et al.*, 1972; Weissenbach et Grosjean, 1981). t<sup>6</sup>A induit un réarrangement de la boucle anticodon et contribue à un meilleur empilement de l'anticodon avec le codon (Murphy *et al.*, 2004). De plus, il a été reporté que l'ARNt<sup>Lys</sup><sub>UUU</sub> non modifié peut reconnaître non seulement les codons lysines AAA et AAG, mais également les codons AAC et

AAU, ce qui engendrait la mauvaise lecture de ces codons asparagines. La combinaison des modifications de la base 37 et de la base wobble ( $t^6A_{37} + s^2U_{34}$  ou  $mnm^5U_{34}$ ) sur l'ARNt<sup>Lys</sup><sub>UUU</sub> permet la reconnaissance spécifique des codons AAA et AAG et empêche celle des codons AAC et AAU. De façon intéressante, il est à noter aussi que si l'ARNt<sup>Lys</sup><sub>UUU</sub> présente seulement une des deux modifications citées précédemment, seul le codon AAA est reconnu (Yarian *et al.*, 2002; Murphy *et al.*, 2004) (figure 25). Ceci est encore un exemple de coopérativité entre deux modifications : la modification  $t^6A$  "communique" avec la base 34 pour affiner la reconnaissance des codons.

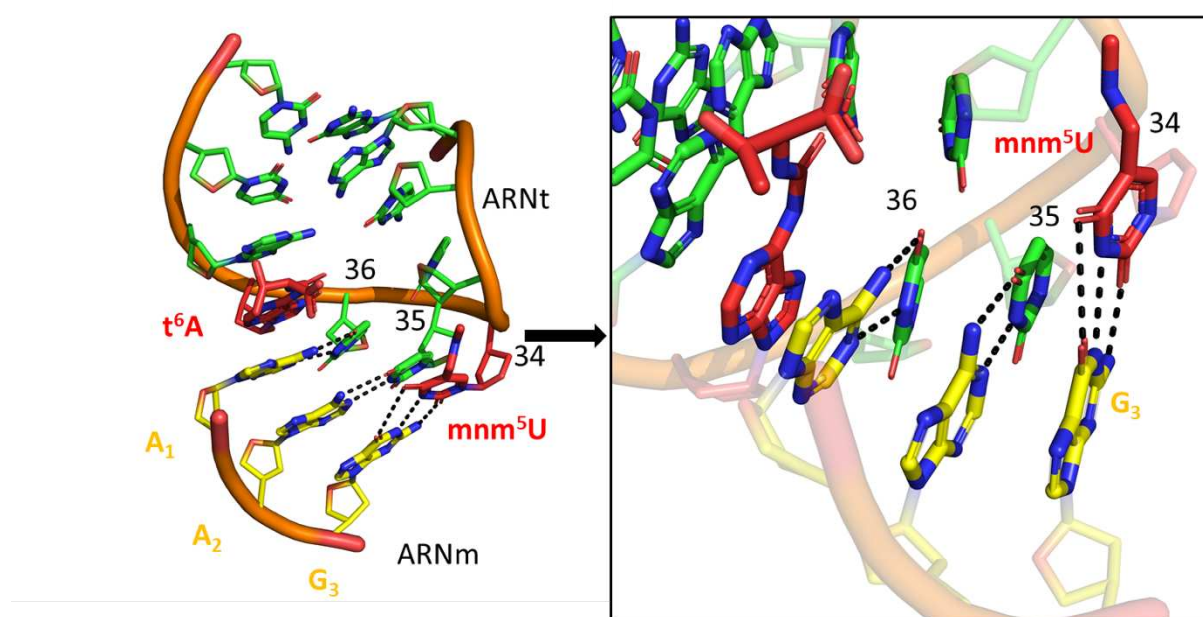


Figure 25 Effet de l'association du  $t^6A_{37}$  et  $mnm^5U_{34}$ . L'ARNt<sup>Lys</sup><sub>UUU</sub> modifié reconnaît (représenté si dessus) le codon AAG de l'ARNm. Le nucléotide  $mnm^5U_{34}$  modifié forme 3 liaisons hydrogènes avec G3. D'après Murphy *et al.*, 2004.

### 3.3. Variations autour de la modification $t^6A$ .

Il existe également des variants de la modification  $t^6A$  (figure 26), tel que l'ajout d'un groupement glycine ( $g^6A$ ) qui est retrouvé sur quelques ARNt au sein des levures (Schweizer *et al.*, 1969). Un groupement isopentenyl sur l'amine N<sup>6</sup> de l'adénosine 37 ( $i^6A$ ) est présent chez les eucaryotes et les procaryotes sur les ARNt<sup>Ser</sup> (Kline *et al.*, 1969). Un dernier variant est le  $hn^6A_{37}$ , une hydroxynorvaline, trouvé chez certaines bactéries et archées thermophiles (Reddy *et al.*, 1992).

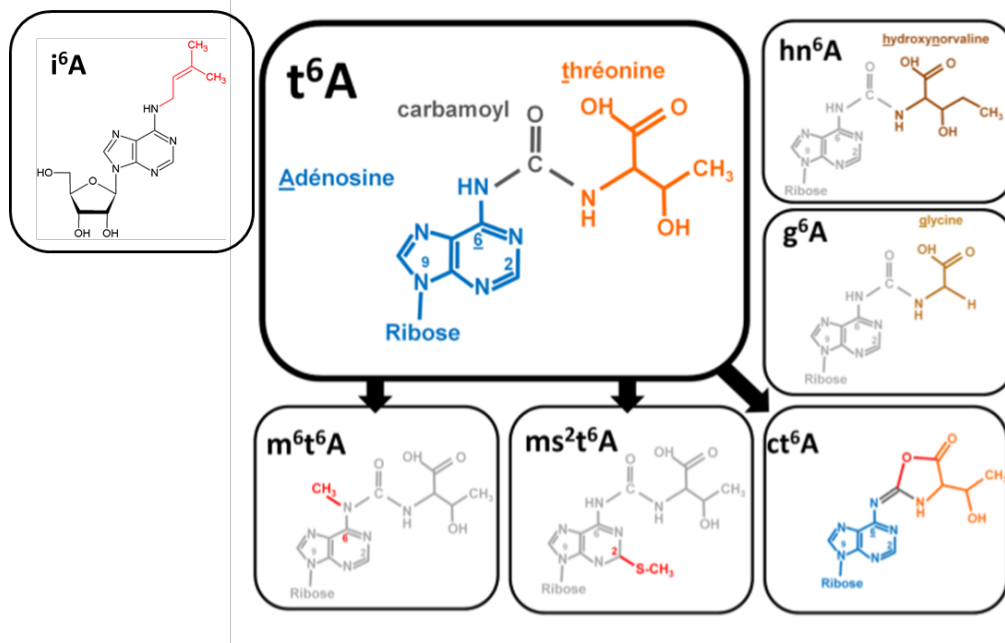


Figure 26 Représentation chimique du nucléotide modifié  $t^6A$ , de ses dérivées et de ses variants. Les flèches indiquent les nucléotides modifiés qui découlent du  $t^6A$ , c'est-à-dire qu'une seconde catalyse se réalise sur la modification  $t^6A$  ( $m^6t^6A$ ,  $ms^2t^6A$ ,  $ct^6A$ ), l'absence de flèches indiquent les nucléotides modifiés qui varient du nucléotides  $t^6A$  ( $hn^6A$ ,  $g^6A$ ,  $i^6A$ ).

Enfin, de nombreuses modifications découlent de la modification  $t^6A$  (figure 26). MtaB et SAM, deux enzymes de modification, catalysent la méthylation et la thiolation de l'adénosine 37 sur le carbone 2, présentant au préalable la modification  $t^6A$  chez les procaryotes et les eucaryotes, formant ainsi le nucléotide  $ms^2t^6A$  (Arragain *et al.*, 2010). Chez la levure, les plantes et certaines bactéries, une enzyme conservée CdsL (appartenant au cluster COG1179) est responsable de la cyclisation du  $t^6A$ . Ainsi, le nucléotide  $ct^6A$  est un ester cyclique avec un anneau d'oxazolone formé par déshydratation du carboxylate de la thréonine qui peut former alors une liaison avec le carbonyle. Cette modification peut être hydrolysée et reconvertie en  $t^6A$  (Miyachi *et al.*, 2013). Enfin, sur certains ARNt, le nucléotide modifié  $t^6A$  peut subir une méthylation, formant ainsi, le  $mt^6A$  (Parthasarathy *et al.*, 1978; Sprinzl *et al.*, 1998) (figure 26).

### 3.3.1. Mécanisme de la modification $i^6A$ , un variant du nucléotide $t^6A$ .

Parmi ces modifications, les mécanismes des réactions sont plus ou moins élucidés. La caractérisation du mécanisme catalytique aboutissant à l'isopentenylation ( $i^6A$ ) est la plus achevée. En effet,  $i^6A$  est une des premières modifications découvertes sur les ARNt<sup>Ser</sup><sub>GCA</sub> au sein des eucaryotes et des procaryotes (Bartz *et al.*, 1970; Kline *et al.*, 1969). Les eucaryotes utilisent une enzyme nommée Mod5, tandis que les procaryotes utilisent l'enzyme MiaA pour

catalyser l'isopentenylation (Dihanich *et al.*, 1987 ; Najarian *et al.*, 1987 ; Golvko *et al.*, 2000). Ces isopentyl transférases interagissent avec la boucle AC et la déforment. Même si la structure en forme de L de l'ARNt est essentielle à l'interaction enzyme-substrat, ce sont surtout des purines et des pyrimidines retrouvées aux positions 29-30 et 40-41, respectivement, qui déterminent la reconnaissance de l'ARNt substrat par l'enzyme (Motorin *et al.*, 1997). Les bases U<sub>33</sub>, G<sub>34</sub>, C<sub>35</sub> et A<sub>37</sub> exposées à l'extérieur de la boucle AC rendent l'azote N<sup>6</sup> de l'A<sub>37</sub> accessible au site catalytique des enzymes. L'azote N<sup>6</sup>, déprotoné par l'aspartate 46 de Mod5, attaque le substrat  $\Delta^2$ -isopentenylpyrophosphate (DMAPP), libérant un groupement pyrophosphate. Cet intermédiaire est synthétisé à partir de l'isopranoïde et d'un noyau stérol (Pierrel *et al.*, 2002; Zhou et Huang, 2008) (figure 27).

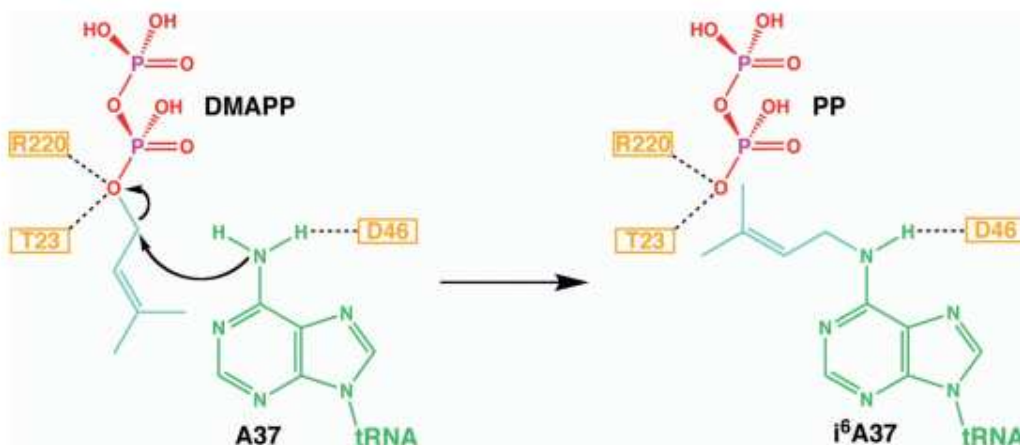


Figure 27 Synthèse de la modification N<sup>6</sup>-isopentényl sur l'adénosine 37 de l'ARNt<sup>Ser</sup><sub>UGA</sub> Mod5 chez les eucaryotes et MiaA chez les procaryotes en présence du DMAPP. D'après Zhou and Huang, 2008

La présence de la modification i<sup>6</sup>A sur l'ARNt<sup>Ser</sup><sub>UGA</sub> permet de reconnaître les codons UCU et UCC et en favorise donc la fidélité de lecture (Schweizer *et al.*, 2017). Les ARNt, présentant la modification i<sup>6</sup>A, peuvent être à leur tour substrat des enzymes aboutissant à la modification m<sup>2</sup>i<sup>6</sup>A (Geftter, 1969).

En raison de leur position en 3' de l'anticodon, les modifications des nucléotides 37 font l'objet d'études afin de comprendre non seulement leur rôle au niveau de l'ARNt mais aussi d'en élucider le mécanisme catalytique. Malgré les efforts des 40 dernières années, portant sur la compréhension des mécanismes enzymatiques des modifications retrouvées sur les ARNt, les connaissances concernant les mécanismes catalytiques des modifications restent très limitées. Leurs caractérisations complètes sont indispensables pour la recherche fondamentale et pour le développement de nouvelles molécules à visées thérapeutiques.

Nous verrons dans les paragraphes suivants que l'identification des enzymes t<sup>6</sup>A synthases fut l'objet de nombreuses années de recherche et que la caractérisation du mécanisme de la catalyse de la modification t<sup>6</sup>A est toujours en cours.

#### **4. Les enzymes de modification t<sup>6</sup>A.**

##### **4.1. La génomique comparative.**

La génomique comparative permet d'identifier des familles de gènes conservées au sein des espèces constituant les trois domaines du vivant. Les gènes conservés dans tous les organismes et de fonction inconnue sont une cible d'étude attractive, du fait de leur stricte conservation au cours de l'évolution (Koonin, 2005). Au début des années 2000, l'analyse des phénotypes des organismes à partir des constructions "knockout" (KO) de certains de leurs gènes a permis de mettre en évidence des familles de protéines essentielles à la survie de ces organismes et s'il s'avère que ces gènes sont aussi largement conservés dans les trois domaines du vivant alors le rôle crucial que jouent les protéines codées par ces gènes peut être confirmé. Parmi ces familles de protéines essentielles et conservées, les familles YrdC et Kae1 sont identifiées (Galperin, 2004; Koonin, 2005).

Les protéines appartenant à la famille YrdC forment un cluster de protéines orthologues : COG009 (figure 28) (Tatusov *et al.*, 2000). Dans certains organismes, la protéine YrdC est fusionnée à un domaine supplémentaire de 120 acides aminés. Cette version plus longue est nommée Sua5 car identifiée préalablement chez *Saccharomyces cerevisiae*. Sua5 de *Saccharomyces cerevisiae* fait partie de la large famille des protéines Sua (pour suppressor of upstream ATG) et son rôle est attribué à la régulation des défauts d'initiation de la traduction (Hampsey *et al.*, 1991; Na *et al.*, 1992). Chez les eucaryotes, la famille de protéine YrdC/Sua5 est présente à la fois dans le cytoplasme et dans la mitochondrie dont la version mitochondriale possède un peptide d'adressage spécifique issu d'une initiation alternative de la traduction (Thiaville *et al.*, 2014). Chez les procaryotes, une nouvelle nomenclature différencie la version à un domaine de celle à deux domaines : TsaC<sub>1</sub> et TsaC<sub>2</sub>, respectivement (Thiaville *et al.*, 2014). Il est important de noter que la répartition de la forme courte ou longue au sein des organismes n'est pas encore comprise.



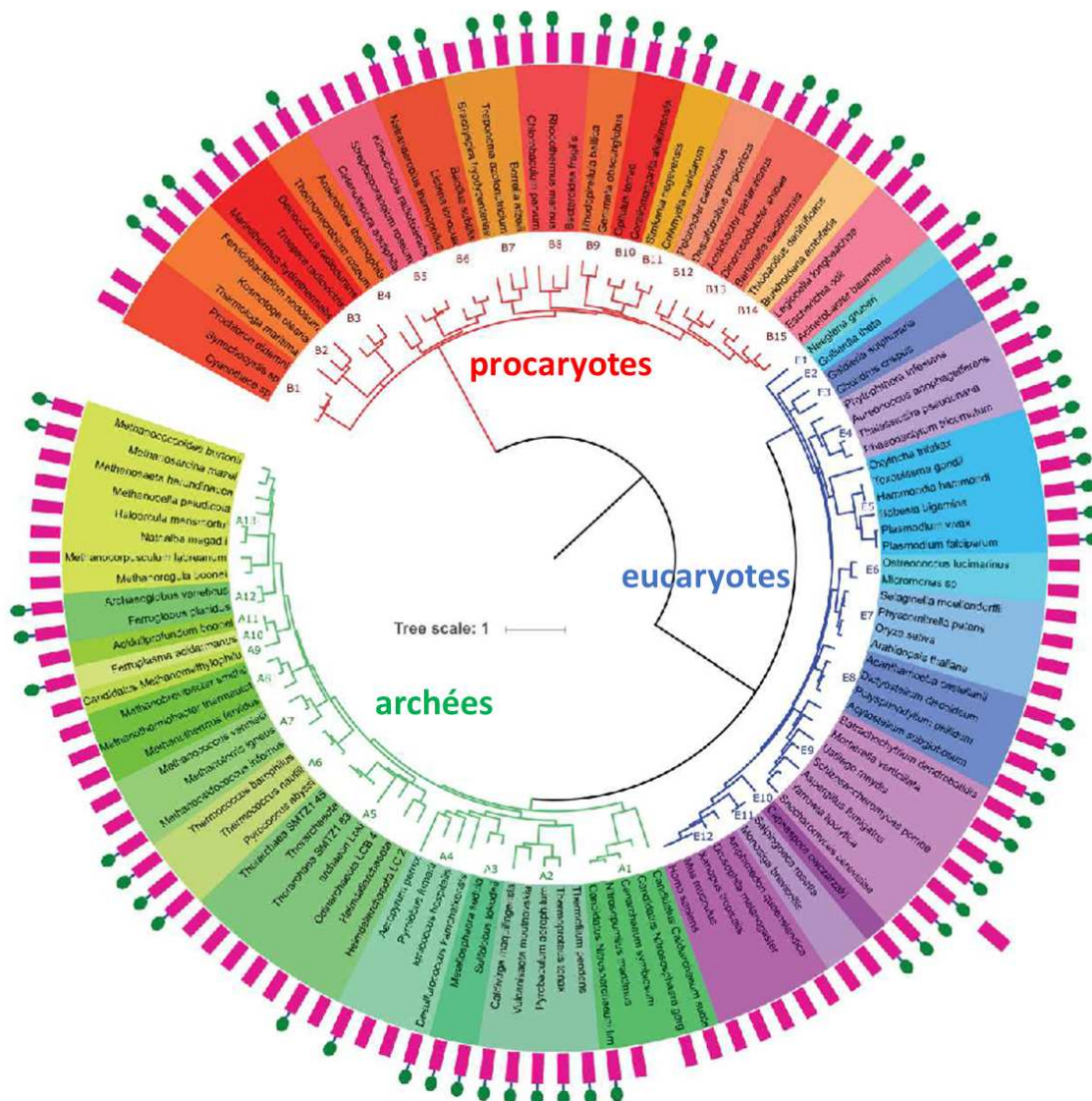


Figure 28 Distribution des protéines YrdC/Sua5 ou TsaC<sub>1</sub>/TsaC<sub>2</sub> selon les trois domaines du vivant. Les protéines YrdC ou TsaC<sub>1</sub> sont représentées par des rectangles roses. Les protéines possédant le domaine supplémentaire Sua5 ou TsaC<sub>2</sub> sont représentées par un cercle vert supplémentaire. Selon Adeline Pichard-Kostuch, thèse de doctorat, Université Paris-Saclay, 2017.

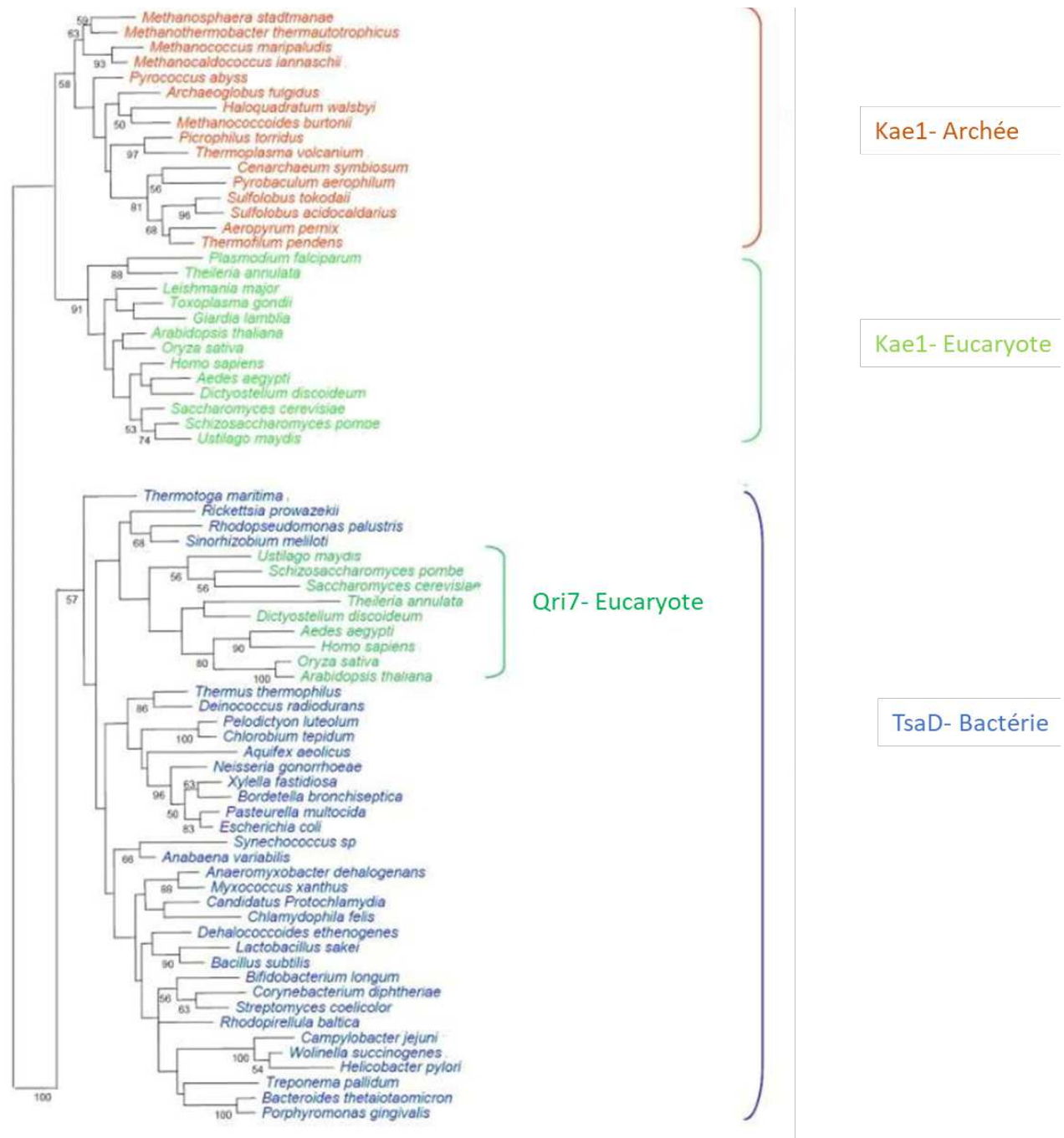


Figure 29 Phylogénie de la famille Kae1/TsaD/Qri7 COG0533. Un échantillon de 100 séquences issues d'espèces représentatives est montré. Les séquences d'archées sont en rouge, les séquences de bactéries en bleu et celles d'eucaryotes en vert. Les séquences des eucaryotes intercalées parmi les séquences bactériennes correspondent à celles de Qri7, l'orthologue mitochondrial. Selon Ludovic Perrochia, thèse de doctorat, Université Paris-Saclay, 2014.

Les protéines appartenant à la famille Kae1 sont conservées au sein des trois règnes du vivant et forment également un cluster de protéines orthologues : COG0533 (figure 29). Les organismes eucaryotes possèdent deux formes homologues de cette famille : Kae1 et Qri7 (Arigoni *et al.*, 1998). Qri7 est localisé au niveau des mitochondries (Huth *et al.*, 2003) et Kae1 au niveau du cytoplasme (Hecker *et al.*, 2007). Chez l'Homme, l'homologue cytoplasmique de Kae1 est nommé OSGEP (pour O-sialoglycoendopetidase protein) (Seki *et al.*, 2002). Contrairement aux eucaryotes, les organismes bactériens et archées possèdent une seule forme nommée TsaD.

#### **4.2. La modification t<sup>6</sup>A est une réaction nécessitant deux étapes catalytiques effectuées par les protéines des familles YrdC/Sua5 et Kae1/TsaD/Qri7.**

Les substrats indispensables à la synthèse de la modification t<sup>6</sup>A sont l'ATP, la L-thréonine, le bicarbonate et le magnésium (Körner et Söll, 1974; Elkins et Keller, 1974). Lors de la catalyse aboutissant à la modification t<sup>6</sup>A, un intermédiaire a été isolé. Des expériences d'HPLC ont mis en évidence que cet intermédiaire (de formule moléculaire C<sub>15</sub>H<sub>21</sub>NO<sub>6</sub>) était le L-thréonyl-carbamoyl-AMP (TC-AMP). L'existence de cet intermédiaire est une preuve que la modification t<sup>6</sup>A se déroule en deux étapes catalytiques prédites dès les années 1990 (Garcia et Goodenough-Lashua, 1998). Deux réactions distinctes dépendantes de l'ATP se produisent au cours de la synthèse de la modification t<sup>6</sup>A et sont effectuées par les protéines de la famille YrdC/Sua5 et Kae1/TsaD/Qri7 (El Yacoubi *et al.*, 2011; Lauhon, 2012; Perrochia *et al.*, 2013).

Au cours de ces dernières années, de nombreuses expériences ont permis d'affiner le rôle de ces protéines dans la voie de biosynthèse t<sup>6</sup>A.

Chez *Saccharomyces cerevisiae*, la fraction d'ARNt isolée à partir de souches mutantes  $\Delta$ Sua5 ne présente pas la modification t<sup>6</sup>A. Aussi le mutant  $\Delta$ Kae1 diminue drastiquement la production de la modification t<sup>6</sup>A alors que le double mutant  $\Delta$ Kae1 -  $\Delta$ Qri7 abolit totalement la synthèse de la modification t<sup>6</sup>A (Srinivasan *et al.*, 2011). Par ailleurs, dans des mutants  $\Delta$ Kae1 la délétion de la séquence d'adressage aux mitochondries de Qri7 permet de rétablir la production t<sup>6</sup>A (El Yacoubi *et al.*, 2011).

Dans les bactéries, il a été mis en évidence que les mutants  $\Delta$ TsaC et  $\Delta$ TsaD conduisent à l'abolition totale de la modification t<sup>6</sup>A sur les ARNt<sup>Thr</sup> et ARNt<sup>Lys</sup> (El Yacoubi *et al.*, 2009;

Deutsch *et al.*, 2012; Srinivasan *et al.*, 2011; Lei *et al.*, 2012). Chez *Escherichia coli*, TsaC seule ne conduit pas à la synthèse de la modification t<sup>6</sup>A des ARNt (Deutsch *et al.*, 2012).

Dans la même année, Charles T. Lauhon montre par spectrométrie de masse et résonance magnétique nucléaire que la famille YrdC/Sua5 n'est pas responsable du transfert du groupement thréonyl-carbamoylé sur les ARNt substrats mais de la formation de l'intermédiaire TC-AMP (Lauhon, 2012). Ce résultat est confirmé plus tard par l'équipe de Biologie Cellulaire des Archées (I2BC) qui montre que Sua5 de *Pyrococcus abyssi* catalyse en présence de tous les substrats (ATP, L-thréonine, bicarbonate) l'hydrolyse de l'ATP en AMP (Perrochia *et al.*, 2013).

La première réaction aboutissant à la formation de l'intermédiaire TC-AMP est ainsi catalysée par les enzymes de la famille YrdC/Sua5 (TsaC<sub>1</sub>/TsaC<sub>2</sub>) (figure 30 A) et la thréonylcarbamoylation de l'adénosine 37 des ARNt substrats est réalisée par les enzymes de la famille Kae1/TsaD/Qri7 (figure 30 B)

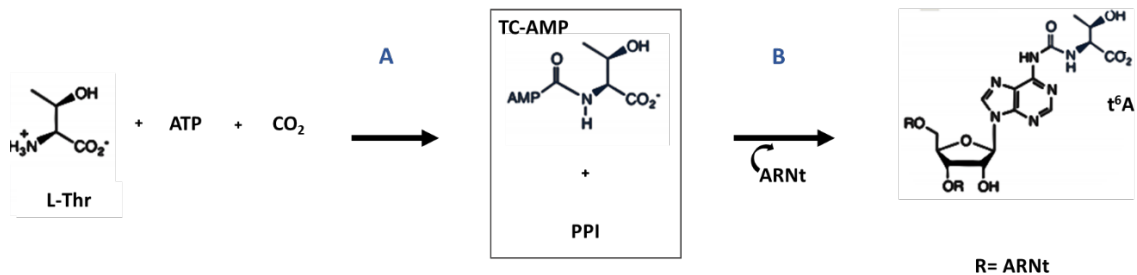


Figure 30 Voie de biosynthèse de la modification t<sup>6</sup>A. A : La première étape qui catalyse la formation du TC-AMP en présence de la L-thréonine, de l'ATP et du bicarbonate est assurée par les enzymes de la famille YrdC/Sua5 (TsaC<sub>1</sub>/TsaC<sub>2</sub> chez les procaryotes). B : Les enzymes de la famille Kae1/TsaD/Qri7 transfèrent le groupement thréonyl-carbamoylé sur les ARNt substrats et peuvent nécessiter la consommation d'ATP.

Il a été montré aussi que le TC-AMP, très labile, se dégrade en quantité équimolaire en L-thréonine et en oxazolidinone (Lauhon, 2012). D'autres composés de dégradation sont identifiés tels que l'isocyanate et le N-carboxyanhydride (figure 31).

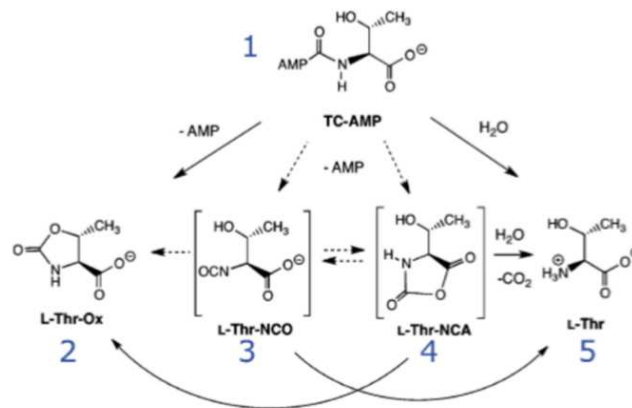


Figure 31 Le TC-AMP (1) se dégrade en AMP et en quantité équimolaire de L-thréonine (5) et d'oxazolidinone (2). D'autres composés intermédiaires sont identifiés ; l'isocyanate (3) et le N-carboxyanhydride (4). D'après Lauhon, 2012.

La réaction chimique de formation du TC-AMP se déroule en deux étapes. L'acide aminé L-thréonine doit être déprotoné sur sa fonction ammonium pour attaquer le CO<sub>2</sub>, résultant en un anion carbamate. L'ion carbamate attaque le phosphate α de l'ATP pour former le TC-AMP (Lauhon, 2012).

#### 4.3. Structure des enzymes t<sup>6</sup>A synthèses.

##### 4.3.1. Structure tridimensionnelle des protéines composants la famille YrdC/Sua5 ou TsaC<sub>1</sub>/TsaC<sub>2</sub>.

La structure cristallographique de la protéine TsaC<sub>1</sub> d'*Escherichia coli* (code PDB 1HRU) montre qu'elle se replie en une structure compacte α/β composée de 7 hélices α et 7 brins β parallèles et antiparallèles (figure 32 B). L'enzyme présente une large cavité positive, pouvant constituer une zone d'interaction avec les acides nucléiques (Teplova *et al.*, 2000). La structure de la protéine Sua5 de *Sulfolobus tokodaii* (Agari *et al.*, 2008) met en évidence, tout comme son homologue bactérien TsaC<sub>2</sub>, un domaine N-terminal nommé "domaine YrdC" pourvu d'une cavité électropositive, relié par un peptide linker au domaine supplémentaire C-terminal, nommé "domaine Sua5". Au sein de la cavité, une molécule d'AMP est fixée par un motif de 4 acides aminés (K/RxR-SxN). Ce motif est conservé chez tous les orthologues de cette famille (Teplova *et al.*, 2000) (figure 32 A). Plus récemment, une nouvelle structure a été obtenue, avec au sein de la cavité du domaine YrdC une molécule de TC-AMP (code PDB 4E1B). Le domaine Sua5, quant à lui, montre une similitude structurale avec certaines GTPases et adopte un pli du type "Rossmann fold" mais dépourvu de résidus impliqués dans la liaison

avec des nucléotides (Agari *et al.*, 2008). Enfin, la structure de Sua5 de *Pyrococcus abyssi* en présence de bicarbonate et de L-thréonine (code PDB 6F89) ou avec une molécule de pyrophosphate (code PDB 6F8Y) dans le site actif a permis d'observer le peptide linker entre les deux domaines YrdC et Sua5. Ce peptide se replie en épingle à cheveux et entre dans le site actif du domaine TsaC.

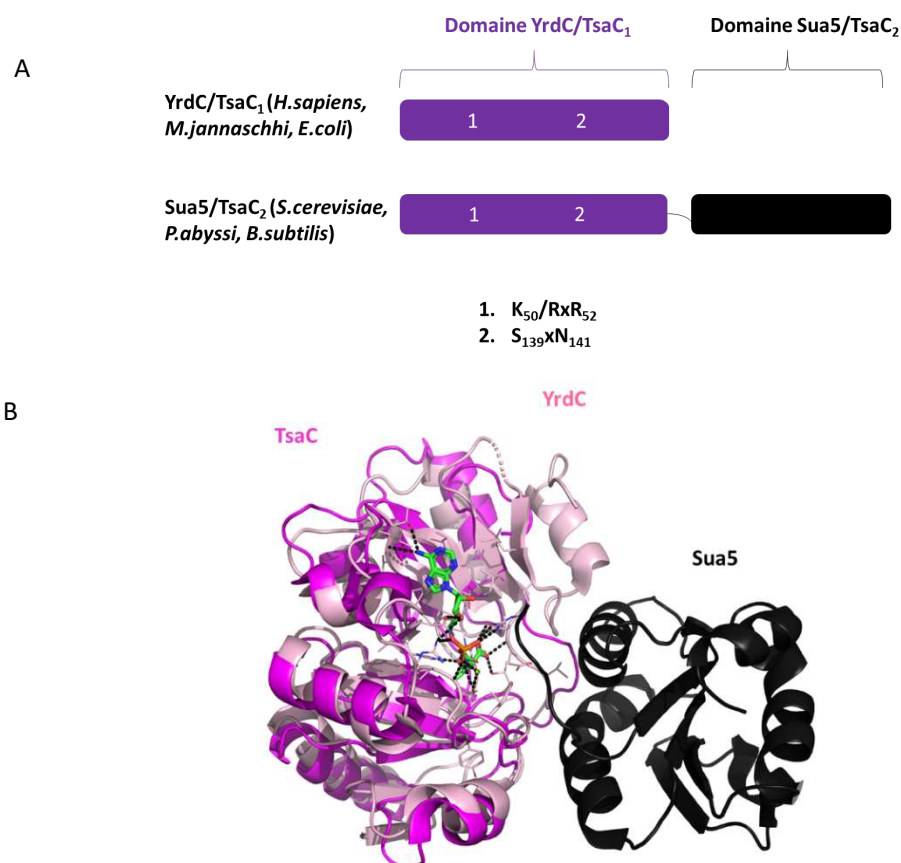


Figure 32 Structure des enzymes de la famille YrdC/TsaC/Sua5. A. Répartition et nomenclature des différents domaines YrdC/ TsaC<sub>1</sub> (en violet) et Sua/TsaC<sub>2</sub> (en noir). Le motif K/RxR-SxN est conservé au sein du domaine YrdC. B. Superposition des domaines TsaC<sub>1</sub> (magenta) de *Escherichia coli* (code PDB 1HRU) et du domaine YrdC (rose) de la protéine Sua5 de *Sulfurisphaera tokodaii* (code PDB 4E1B). Le domaine Sua5 est localisé à l'entrée du site actif du domaine YrdC. L'intermédiaire TC-AMP (en vert) est dans le site actif du domaine YrdC de la protéine Sua5 de *Sulfurisphaera tokodaii*.

Cette conformation pourrait jouer le rôle de "couvercle" fermant le site actif afin de créer des conditions favorables à la réaction enzymatique (Pichard-Kostuch *et al.*, 2018). Les homologues YrdC/TsaC<sub>1</sub>, qui ne possèdent ni de domaine Sua5 ni de peptide linker, présente une queue C-terminale flexible qui pourrait fermer le site actif (Teplova *et al.*, 2000). Ces structures mettent en lumière que malgré une répartition aléatoire des protéines YrdC/TsaC<sub>1</sub> ou Sua5/TsaC<sub>2</sub> au sein des organismes, le domaine YrdC présente un site actif qui interagit

avec la L-thréonine et les nucléotides (figure 32 B). Il faut noter que le rôle du domaine Sua5 reste à élucider.

Bien que la plupart des organismes ne présentent qu'une version de YrdC/Sua5 ou TsaC<sub>1</sub>/TsaC<sub>2</sub>, certains organismes possèdent un autre membre de cette famille : la protéine YciO identifiée chez *Escherichia coli*. La structure de cette protéine (Jia *et al.*, 2002) montre une analogie structurale avec TsaC, mais il existe des différences notables au niveau du site actif : en effet, la tétrade KRSN impliquée dans la fixation du nucléotide chez TsaC<sub>1</sub>/TsaC<sub>2</sub> est absente. Des expériences de complémentation génétique ont montré que ce paralogue ne semble pas remplir la même fonction que TsaC et représenterait une famille de protéines au rôle différent (El Yacoubi *et al.*, 2009; Gerdes *et al.*, 2011).

#### **4.3.2. La structure tridimensionnelle des membres de la famille Kae1/TsaD/Qri7.**

La première structure de Kae1 issu de *Pyrococcus abyssi* est résolue en 2007 (Hecker *et al.*, 2007). La protéine Kae1 est composée de deux domaines similaires, contenant cinq brins  $\beta$  antiparallèles encadrés par trois hélices  $\alpha$ . Les deux domaines englobent une cavité qui forme le site actif de l'enzyme. Un métal de fer ( $\text{Fe}^{2+}$ ) est coordonné par un motif de chaînes latérales que l'on retrouve chez les métalloprotéases. En effet, l'ion ferreux est stabilisé par les azotes N<sub>2</sub> des deux histidines H<sub>107</sub> et H<sub>111</sub> et par le carboxylate du résidu D<sub>285</sub>, conservés chez tous les homologues de cette famille de protéine. De plus, une liaison hydrogène est détectée entre le groupement OH de la tyrosine 137 et l'ion  $\text{Fe}^{2+}$ . Cette tyrosine est conservée uniquement chez les orthologues eucaryotes et archées, absente chez les orthologues procaryotes. Lors de la fixation d'un analogue non hydrolysable de nucléotide comme l'AMPPNP, le phosphate  $\gamma$  participe également à la coordination du métal (Hecker *et al.*, 2007 ; 2008).

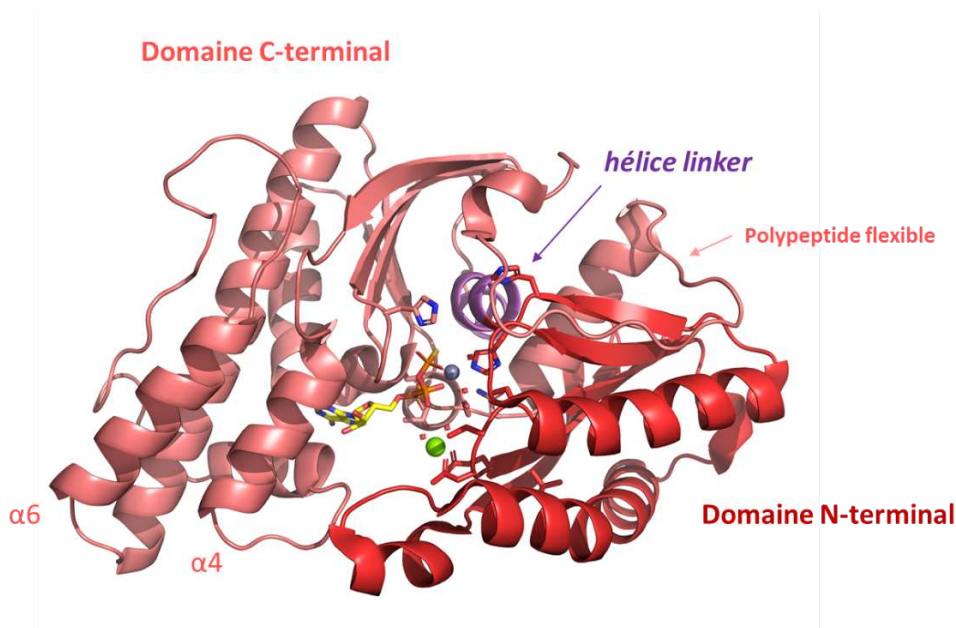


Figure 33 Structure tridimensionnelle de Tsad de la bactérie *Salmonella thyphirurium* (code PDB 3ZEU). Tsad est organisée en deux domaines : le domaine N-terminal (en rouge) et le domaine C-terminal (en rose) reliés par une hélice linker en violet. Le polypeptide flexible (mentionné dans le § 3.4.2.) est indiqué et se referme sur le site actif. L'ATPγS est représenté dans le site actif en jaune. Un ion  $Zn^{2+}$  (en gris) est fixé dans le site actif et coordonné par les résidus du motif HxxH et CDN, représenté en sticks. Un second ion  $Mg^{2+}$  (en vert) est également fixé dans le site actif, coordonné par les résidus S, C, D, E et T représentés en stick.

La structure de Tsad de *Salmonella thyphirurium* (Nichols *et al.*, 2013) met en évidence une architecture très similaire à celle de Kae1. Tsad et Kae1 se superposent avec une déviation (RMSD) de 1,61 Å (Z score = 16.1 sur 313 Cα ; identité de séquence = 34.1 %) (Nichols *et al.*, 2013). La protéine est organisée également en deux domaines colorés en rouge et rose respectivement sur la figure 33. Le domaine N-terminal est constitué de cinq brins β antiparallèles et de trois hélices α externes. Le domaine C-terminal présente une topologie identique, pourvu de deux hélices α<sub>4</sub> et α<sub>6</sub> supplémentaires. Une hélice est présente entre les deux domaines, servant de "linker". Pour finir, une hélice α et un peptide flexible sont localisés à l'extrémité C-terminale de Tsad, au-dessus de la cavité située entre les deux domaines (figure 33). Cette cavité présente les résidus conservés comme le motif HxxH. Selon la condition de cristallisation, deux ligands et deux métaux ont pu être modélisés dans la cavité : une molécule d'ATPγS et de zinc et une molécule d'AMP et de cadmium (respectivement code PDB 3ZEU et 3ZET) (Nichols *et al.*, 2013). Les métaux sont coordonnés par les acides aminés H<sub>111</sub>, H<sub>115</sub>, D<sub>300</sub> ainsi que par l'oxygène O'<sub>3</sub> du phosphate α de l'AMP. L'O'<sub>1</sub> du phosphate α de l'AMP forme une liaison hydrogène avec la S<sub>270</sub>. Les O'<sub>2</sub>' et O'<sub>3</sub>' du ribose forment des liaisons hydrogènes avec l'A<sub>167</sub> et l'atome N<sub>1</sub> de l'adénine forme une liaison hydrogène avec



l'asparagine N<sub>272</sub>. La fixation de l'ATPyS et de son métal Zn<sup>2+</sup> correspondant est similaire à celle de l'AMP et du Cd<sup>2+</sup>, excepté que l'O<sub>1</sub> du phosphate α forme une liaison hydrogène avec le métal et l'H<sub>139</sub> coordonne l'O<sub>3</sub> du phosphate γ. Pour finir un magnésium (présent dans la condition de cristallisation) est positionné au site nucléotidique auquel il est indirectement lié par des molécules d'eau au glutamate E<sub>12</sub> en présence de ATPyS. Le mutant E<sub>12</sub>A fixe toujours l'ATP mais perd la capacité à fixer l'ion Mg<sup>2+</sup> (code PDB 4WQA) supprimant l'activité de la modification t<sup>6</sup>A sur les ARNt substrats (Zhang *et al.*, 2015).

#### 4.4. Similitude avec les enzymes de carbamoylation HypF et TobZ.

Ces familles de protéines YrdC et Kae1 sont orthologues aux domaines de HypF et TobZ. Même si l'ordre des domaines diffère au sein de TobZ et HypF (figure 34) (El Yacoubi *et al.*, 2009; 2012) ces protéines supportent une réaction de carbamoylation. Ainsi, l'analogie structurale avec les domaines des protéines HypF et TobZ a permis d'affiner la compréhension du mécanisme de la modification t<sup>6</sup>A catalysée par les enzymes YrdC/Sua5 et Kae1/OSGEP/TsaD/Qri7.

HypF transfère le groupement carbamoylé à partir d'un intermédiaire carbamoylé-adénylate Ca-AMP sur la cystéine de l'hydrogénase HypE, la cystéine carbamoylée participe à

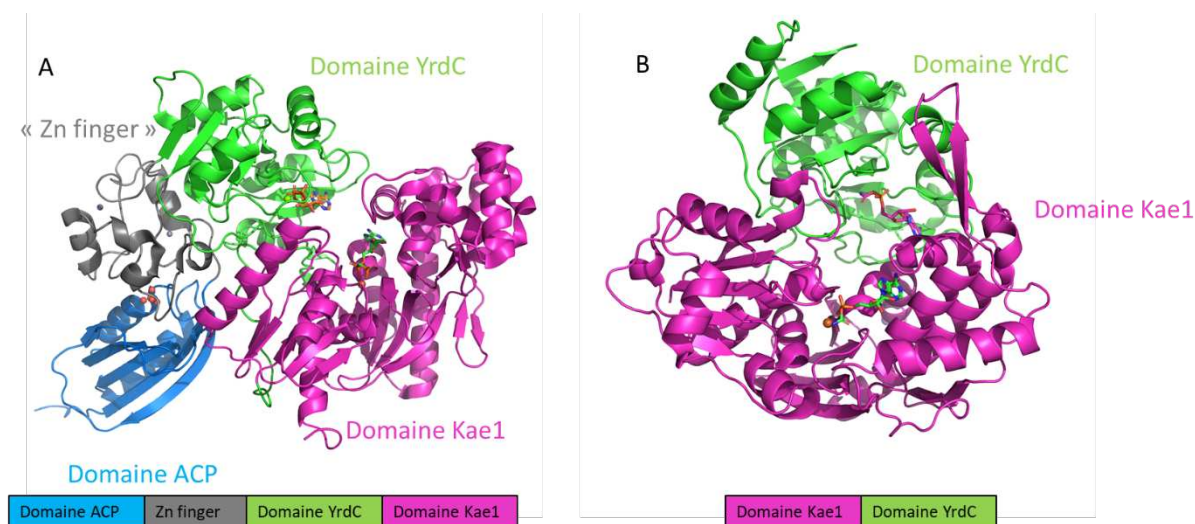


Figure 34 Structure tridimensionnelle des deux enzymes de carbamoylation : HypF de *Caldanaerobacter subterraneus subsp. tengcongensis MB4* et TobZ de *Streptoalloteichus tenebrarius*. A : HypF possède 4 domaines, chacun représenté par une couleur (code PDB 3VTH). Le domaine ACP, en bleu, fixe une molécule de phosphate (en rouge). Le domaine " Zn finger " fixe des ions Zn<sup>2+</sup> (sphères grises). Le domaine YrdC, en vert, fixe une molécule d'AMPCPP, en orange, dans son site actif. Le domaine YrdC hydrolyse le carbamoyl-phosphate et forme le carbamoyl-AMP en présence d'ATP. Le domaine Kae1, en magenta, fixe également une molécule d'AMPCPP, en vert, dans son site actif. Ce dernier domaine catalyse la carbamoylation de la cystéine de la protéine HypE B : TobZ possède 2 domaines (code PDB 3VER). Le domaine YrdC, en vert, fixe une molécule de Ca-AMP dans son site actif. Après diffusion et fixation du Ca-AMP, représenté en stick vert, dans le site actif du domaine Kae1 (en magenta), une réaction de carbamoylation est catalysée sur le substrat tobramycine. Cet tobramycine carbamoylé constitue l'antibiotique : nébramycine.

la fixation du cluster  $\text{NiFe}(\text{CN})_2(\text{CO})$  (Paschos *et al.*, 2002; Blokesch *et al.*, 2004; Petkun *et al.*, 2011; Shomura et Higuchi, 2012) (figure 34 A). TobZ catalyse la réaction de O-carbamoylation de la tobramycine pour former un antibiotique naturel, la nébramycine (Parthier *et al.*, 2012; Xiao *et al.*, 2014) (figure 34 B).

Bien que ces deux enzymes catalysent une réaction de carbamoylation, leurs mécanismes diffèrent. HypF possède 4 domaines différents représentés par des couleurs différentes (figure 34 A). Le domaine ACP (en bleu) fixe la molécule carbamoyl-phosphate et est suivi d'un domaine en doigt de zinc ("Zn finger") (en gris). Ensemble, le domaine ACP et "Zn finger" déphosphorylent le carbamoylphosphate (figure 35 A). Le carbamate est ensuite transféré au domaine YrdC, qui hydrolyse l'ATP et forme l'intermédiaire carbamoyl-AMP (figure 35 B). Le carbamoyl-AMP est enfin transféré au site actif du domaine Kae1 où est catalysé la S-carbamoylation de la cystéine C-terminale de l'hydrogénase HypE (figure 35 C). La cystéine carbamoylée participe à la stabilisation du cluster  $\text{NiFe}(\text{CN})_2(\text{CO})$  nécessaire à l'activité de HypE (Shomura et Higuchi, 2012).

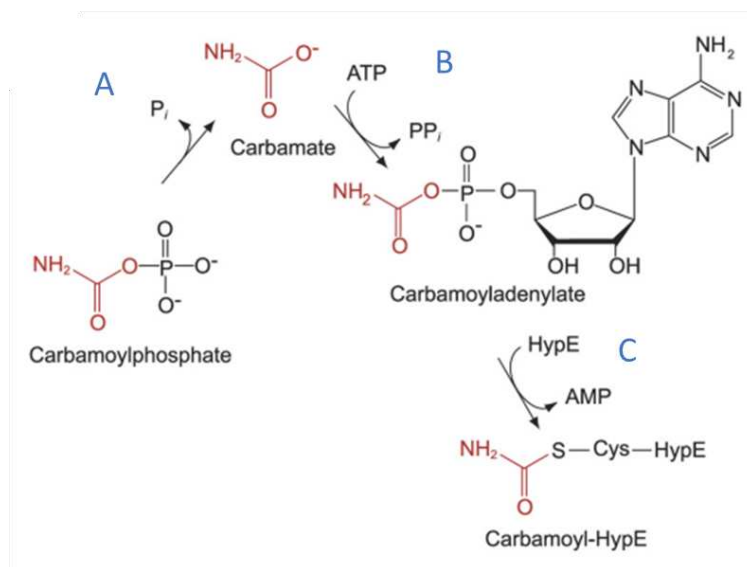


Figure 35 Réaction de S-carbamoylation de l'enzyme HypF. A. Déphosphorylation du carbamoylphosphate. B. Le carbamate en présence d'ATP forme l'intermédiaire carbamoyladenylate (Ca-AMP). C. La cystéine de HypE est carbamoylée.

En absence des domaines APC et " Zn finger " de la protéine TobZ, le domaine YrdC, en vert, déphosphoryle le carbamoyl-phosphate et hydrolyse l'ATP pour former l'intermédiaire carbamoyl-AMP (figure 36 A). Le site de fixation du carbamoyl-phosphate de TobZ diffère de celui de HypF. En effet, ce site est localisé au sein du domaine YrdC assurant les deux étapes intermédiaires aboutissant à la formation du carbamoyl-AMP (figure 36 B). L'intermédiaire carbamoyl-AMP est, ensuite, transféré au site actif du domaine Kae1 qui catalyse la O-

carbamoylation de la molécule tobramycine en nébramycine (figure 36 B) (Parthier *et al.*, 2012).

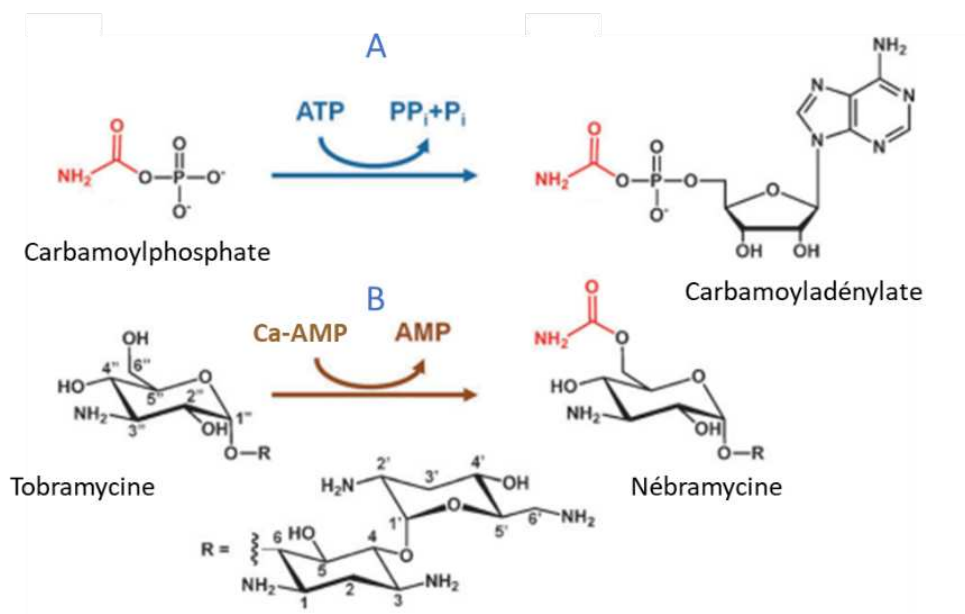


Figure 36 Réaction de O-carbamoylation. A. Réaction de déphosphorylation et de hydrolyse de l'ATP qui aboutit à la formation de l'intermédiaire Ca-AMP. B. Transfert du groupement carbamoylé sur la fonction 6''OH du tobramycine.

#### 4.5. La modification t<sup>6</sup>A chez les procaryotes.

##### 4.5.1. La biosynthèse de la modification t<sup>6</sup>A par les protéines TsaC, TsaB, TsaD et TsaE.

Les protéines TsaC et TsaD ne suffisent pas, *in vitro*, à catalyser seules le transfert du thréonyl-carbamoylé sur les ARNt substrats (El Yacoubi *et al.*, 2011), suggérant que des protéines partenaires sont impliquées dans la réaction. Également, il a été montré que TsaD ne complète pas la modification t<sup>6</sup>A dans les levures délétées du gène *KAE1* (El Yacoubi *et al.*, 2011). Ainsi, les protéines partenaires de TsaD ne sont probablement pas présentes au sein du domaine eucaryote.

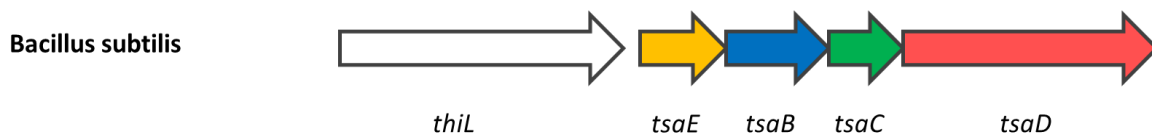


Figure 37 Proximité des gènes *tsaE*, *tsaB*, *tsaC* et *tsaD* de *Bacillus subtilis*.

Tout comme le gène *tsaD*, deux autres gènes essentiels à la survie de *Escherichia coli* sont identifiés : *tsaB* et *tsaE* (Butland *et al.*, 2005; Gil *et al.*, 2004). De plus, l'étude du contexte génomique chez *Bacillus subtilis* a mis en évidence la proximité, au sein d'un opéron, de gènes codants TsaC, TsaD, TsaB et TsaE, suggérant une fonction commune (Handford *et al.*, 2009) (figure 37). L'interaction de TsaB, TsaD et TsaE est confirmée par des expériences de "pull-down" (Handford *et al.*, 2009).

La formation *in vitro* de la modification t<sup>6</sup>A chez *Escherichia coli* a été dosée sur certains ARNt non modifiés en présence des protéines de *Escherichia coli* TsaC, TsaD, TsaB et TsaE, utilisant l'incorporation de C<sup>14</sup> L-thréonine en présence d'ATP, de bicarbonate et de différents ARNt substrats non modifiés : l'ARNt<sup>Thr</sup>, l'ARNt<sup>Lys</sup>, le bras AC de l'ARNt<sup>Lys</sup> d'*Escherichia coli* et également un ARNt de *Saccharomyces cerevisiae* isolé à partir de souches ΔSua5 pour que celui-ci soit dépourvu de nucléotide modifié t<sup>6</sup>A. Le taux d'incorporation de t<sup>6</sup>A est maximum sur les ARNt<sup>Thr</sup> et ARNt<sup>Lys</sup> d'*Escherichia coli*. En revanche, aucun nucléotide modifié n'est observé en utilisant le bras AC de l'ARNt<sup>Lys</sup> comme substrat (Deutsch *et al.*, 2012). Ce résultat concorde avec les expériences de microinjection d'ARNt préalablement transcrits *in vitro* et injectés dans les ovocytes de *Xenopus*, où la structure entière en forme de L des ARNt était supposé nécessaire à la biosynthèse de la modification t<sup>6</sup>A (Morin *et al.*, 1998). Sur les ARNt de *Saccharomyces cerevisiae*, la modification t<sup>6</sup>A est significativement diminuée (Deutsch *et al.*, 2012), suggérant que le substrat d'une autre espèce est moins efficace.

#### 4.5.1.1. TsaB est le paralogue tronqué de TsaD.

TsaB est également conservée au sein des procaryotes et appartient au cluster orthologue : COG1214. En plus d'être essentiel pour la survie d'*Escherichia coli*, le gène *tsaB* est également essentiel aux organismes *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Bacillus subtilis* et *Francisella novicida* (Freiberg *et al.*, 2001).

La protéine TsaB de *Salmonella typhimurium* est composée de 231 acides aminés contre 337 pour TsaD (Nichols *et al.*, 2006). Les protéines de cette famille sont considérées

comme des paralogues des protéines TsaD. La première structure de TsaB de *Salmonella typhimurium* (code PDB 2GEM et 2GEL) met en évidence un pli du type " HSP70 actine like " et, tout comme TsaD, comporte deux domaines (Nichols *et al.*, 2005) (figure 38).

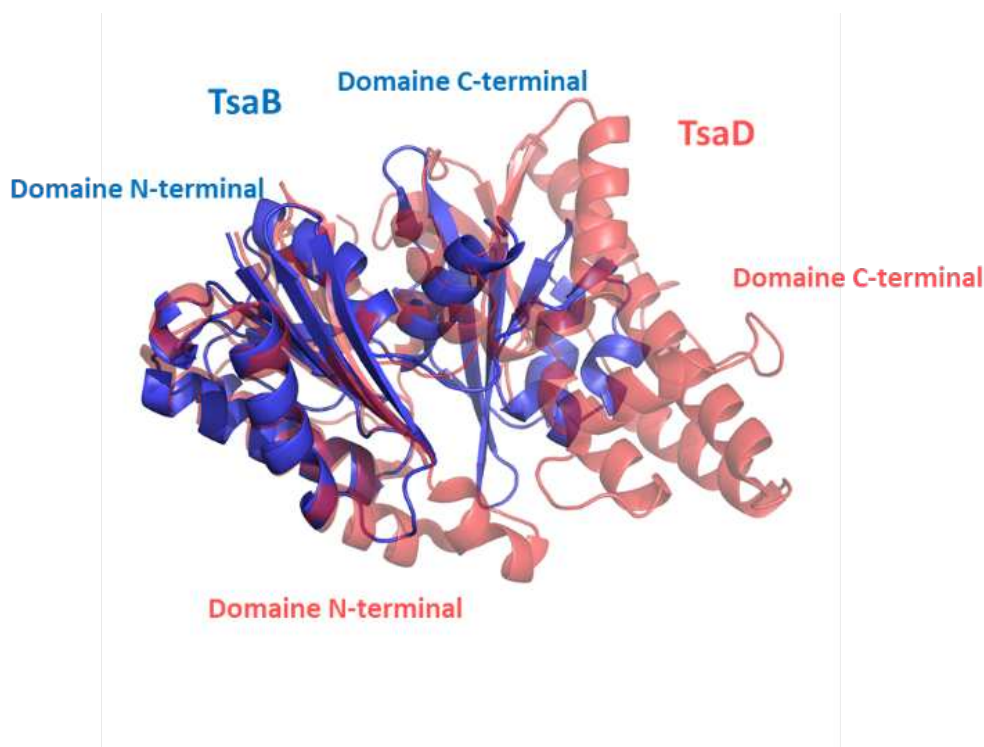


Figure 38 Superposition de la structure de TsaB de *Salmonella typhimurium* (code PDB 2GEM), en bleu, et de la structure de TsaD de *Salmonella typhimurium* (code PDB 3ZEU), en rouge. Les domaines N-terminaux sont similaires, cependant, les domaines C-terminaux diffèrent.

La superposition de la protéine TsaB sur la protéine TsaD met en évidence une similitude structurale (RMSD = 3.2 Å sur 172 C $\alpha$ , Z-score =6.4, identité de séquence = 17 %) (figure 38). Les domaines N-terminaux sont très similaires (RMSD = 1.19 Å sur 85 C $\alpha$ , Z-score = 9, identité de séquence = 28 %), en revanche, les domaines C-terminaux diffèrent fortement. En effet ce dernier est tronqué de 100 acides aminés et ne permet pas de former un site actif. TsaB ne possède donc ni de site de liaison à l'ATP, ni au métal, ni de motif HxxH ni de site actif (Nichols *et al.*, 2013) (figure 38).

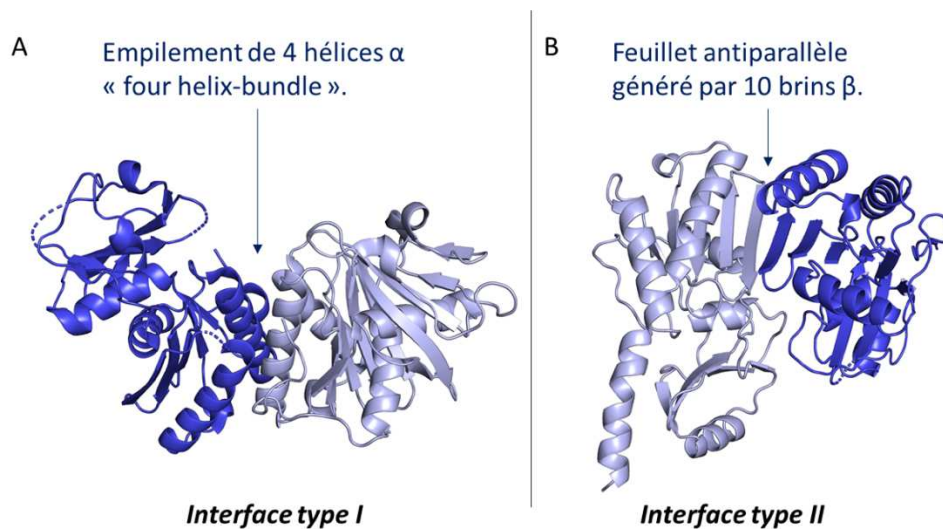


Figure 39 Oligomérisation du partenaire TsaB. A : Deux molécules de TsaB interagissent via leurs hélices  $\alpha 1$  et  $\alpha 2$  respectives formant un empilement de 4 hélices (4 helix bundle), stabilisés par des interactions hydrophobes (interface du type I) B : ou par leurs brins  $\beta 3$  respectifs, formant un long feuillet anti parallèle (interface du type II).

Deux formes de dimères sont observées dans certains cristaux de TsaB (Aydin *et al.*, 2011). La dimérisation de TsaB est confirmée, *in vitro*, en solution. La première surface de dimérisation (type I) est réalisée entre les domaines N-terminaux, stabilisée par des résidus hydrophobes hautement conservés (L<sub>40</sub>, L<sub>47</sub>, I<sub>74</sub>, I<sub>78</sub>, L<sub>82</sub>) (figure 39 A) formant un motif "4 helix bundle". La seconde interface de dimérisation (type II) est réalisée entre les brins  $\beta_3$  de chaque monomère, formant un long feuillet antiparallèle (Aydin *et al.*, 2011; Nichols *et al.*, 2013). L'interface de type II a été à ce jour observé uniquement chez TsaB de l'organisme *Thermotoga maritima* (figure 39 B) suggérant une spécificité propre aux organismes hyperthermophiles.

#### 4.5.1.2. TsaE, une ATPase comme troisième partenaire.

Le gène *tsaE* de la bactérie *Bacillus subtilis* code une petite protéine de 153 acides aminés, hautement conservée et unique chez les procaryotes mais absente notamment chez certaines espèces de Mollicutes (El Yacoubi *et al.*, 2011; Teplyakov *et al.*, 2002). La première structure cristallographique résolue de TsaE d'*Haemophilus influenzae* montre un repliement  $\alpha/\beta$  typique des GTPases possédant un motif P-loop, retrouvé dans de nombreuses protéines qui interagissent avec les nucléotides (Teplyakov *et al.*, 2002) (figure 40).

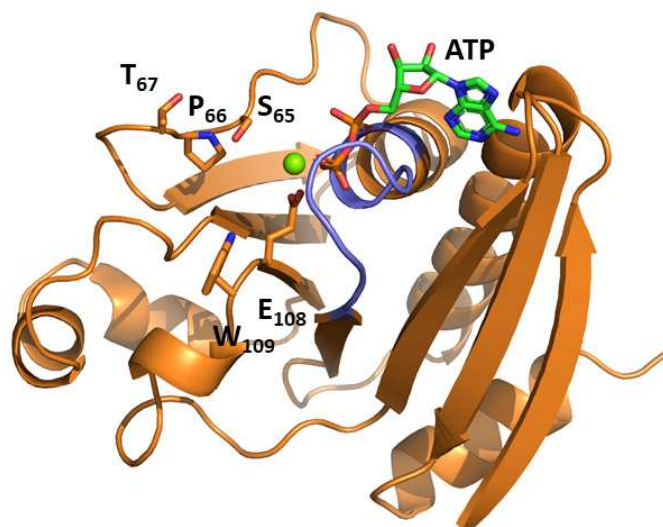


Figure 40 Le troisième partenaire TsaE de la machinerie t<sup>6</sup>A au sein des procaryotes. Structure tridimensionnelle de TsaE d'*Haemophilus influenzae*, une molécule d'ADP et un ion Mg<sup>2+</sup> en vert sont liés par les résidus de TsaE. Le motif P-loop est représenté en violet. Les acides aminés W109 et E108 sont en sticks, ainsi que les acides aminés du motif SPT (code PDB 1HTW).

Le motif P-loop génère une cavité hydrophobe permettant de stabiliser le nucléotide (Teplyakov *et al.*, 2002). En effet, la protéine TsaE d'*Haemophilus influenzae* lie un ADP et un Mg<sup>2+</sup>, les phosphates sont coordonnés par les thréonines T<sub>47</sub> et T<sub>48</sub>, les O<sub>2'</sub> et O<sub>3'</sub> du ribose partagent une liaison hydrogène avec le glutamate E<sub>11</sub>. Le métal est, quant à lui, coordonné par le glutamate E<sub>113</sub> et le tryptophane W<sub>114</sub> qui sont conservées (figure 40). Les GTPases de la même classe, les " petites protéines G ", sont caractérisées par la présence de " switches " moléculaires, des régions proches de la P-loop qui ont la capacité de subir des changements conformationnels en réponse à l'hydrolyse du GTP. Malgré la ressemblance structurale avec ces GTPases aucun changement conformationnel a été observé pour TsaE en absence ou présence d'ATP. En revanche, le motif SPT est conservé dans TsaE.

De plus, l'activité ATPase de TsaE a été caractérisée *in vitro* et est dépendante du Mg<sup>2+</sup> (K<sub>M</sub>= 38 μM, k<sub>cat</sub>=0,42 min<sup>-1</sup> k<sub>cat</sub>/K<sub>M</sub>= 180 s<sup>-1</sup>.M<sup>-1</sup>) (Teplyakov *et al.*, 2002). *In vitro*, le site de liaison au nucléotide a été analysé par l'étude de mutants de TsaE de la bactérie *Escherichia coli*, notamment le mutant E<sub>108</sub>A qui n'est ni capable de lier l'ADP, ni le Mg<sup>2+</sup> (Zhang *et al.*, 2015)

Récemment, une étude de TsaE de *Bacillus subtilis* montre que cette dernière possède aussi deux motifs HxD et SPT conservés et retrouvés dans les protéines kinases de type Hanks (Hanks *et al.*, 1988; Nguyen *et al.*, 2017). La spectrométrie de masse a permis d'identifier deux

résidus S<sub>60</sub> et T<sub>62</sub> phosphorylés sur le motif SPT de TsaE, montrant sa capacité de phosphoryler TsaD chez *Bacillus subtilis*, *in vitro*. Le rôle de cette phosphorylation n'est pas encore élucidé.

#### 4.5.1.3. Le complexe TsaBDE chez les procaryotes.

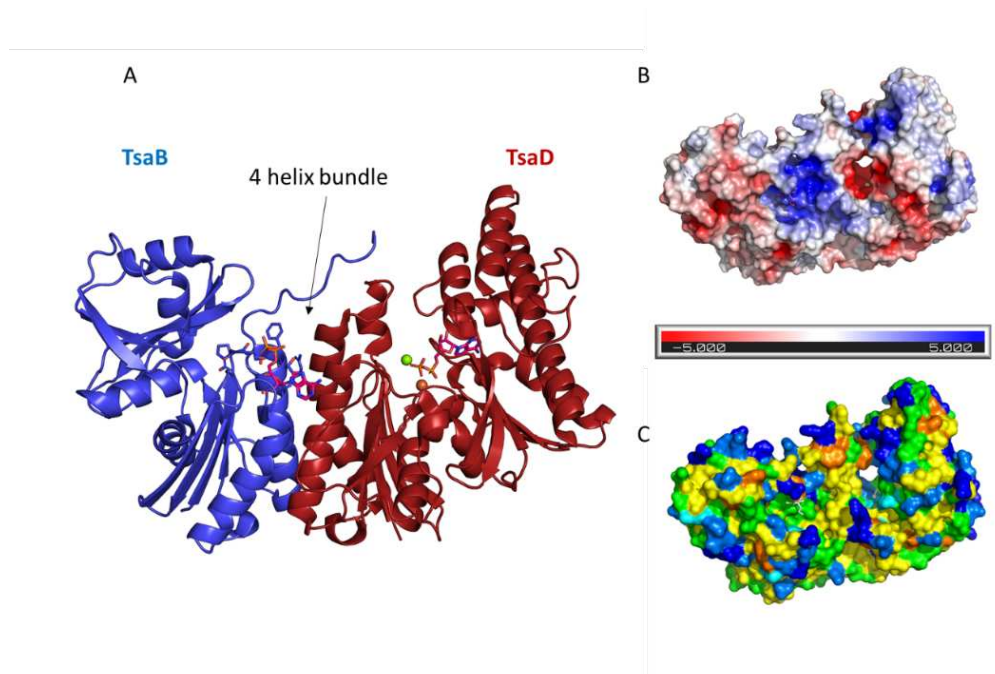


Figure 41 Propriétés structurales de TsaB-TsaD. A : Structure quaternaire de l'hétérodimère TsaB-TsaD (en bleu et rouge briques respectivement). TsaD fixe une molécule d'ADP et de Mg<sup>2+</sup> dans son site actif. L'interface de dimérisation est stabilisée par des résidus hydrophobes des hélices  $\alpha$ 1 et  $\alpha$ 2 de TsaB et des hélices  $\alpha$ 2 et  $\alpha$ 4 de TsaD, reconstituant l'agencement en " 4 helix bundle " observé chez l'homodimère TsaB (interface de type I). A cette interface, une molécule d'ADP est observée (en stick rose et bleu), les acides aminés qui les stabilisent sont représentés en sticks ci-contre. B : la surface électrostatique de TsaBD, C : la surface du degré d'hydrophobie sont représentées (les acides aminés hydrophobes sont représentés en jaune et orange et les acides aminés polaires en bleu et vert).

La structure de l'hétérodimère TsaD-TsaB montre que TsaB interagit avec TsaD par leurs domaines N-terminaux respectifs formant l'interface de type I observée dans l'homodimère TsaB (figure 41 A) reconstituant le motif "4 helix bundle". En solution, des expériences de micro calorimétrie mesurent une constante de dissociation ( $K_D$ ) entre les deux sous-unités de 0,29  $\mu$ M et une stœchiométrie 1 : 1 entre ces deux sous-unités confirmant leur interaction (Nichols *et al.*, 2013; Zhang *et al.*, 2015). Un motif conservé de TsaB (<sup>64</sup>GPG(S/Y/F)TG(I/L/V)R<sup>72</sup>) et des acides aminés de TsaD permettent la fixation d'une molécule d'ADP à l'interface. Ainsi, la structure de TsaD - TsaB de *Escherichia coli* arbore un second site de liaison aux nucléotides (Zhang *et al.*, 2015). Des mutants des acides aminés de TsaB et de TsaD empêchant la fixation de cet ADP, n'affectent pas la cohésion de l'hétérodimère TsaB – TsaD (Zhang *et al.*, 2015). L'hétérodimère TsaBD arbore une surface électrostatique majoritairement positive (figure 41



B) et apolaire (figure 41 C) suggérant qu'elle pourrait être une surface d'interaction pour l'ARNt substrat.

*In vitro*, il a été montré que l'activité ATPase de TsaE est nécessaire pour que les enzymes catalysent la modification t<sup>6</sup>A (Zhang *et al.*, 2015). Des analyses de retard sur gel en condition native ont montré que le mutant E<sub>108</sub>A, qui empêche l'activité ATPase de TsaE ne pouvait plus interagir avec TsaBD, suggérant que la fixation du nucléotide stabilise la formation du complexe hétérotrimérique. En absence de TsaE ou en présence du mutant E<sub>108</sub>A TsaE l'activité t<sup>6</sup>A est réduite de plus de 60% et un seul cycle catalytique de modification t<sup>6</sup>A est réalisée et l'activité ATPase de TsaE permet plusieurs cycles de modification t<sup>6</sup>A (Luthra *et al.*, 2018; Zhang *et al.*, 2015).

Aucune structure quaternaire du complexe TsaBDE n'a été résolue permettant de comprendre la stabilité et l'assemblage de ce complexe. Malgré de nombreuses données biochimiques concernant le mécanisme de la catalyse de la modification t<sup>6</sup>A, la résolution de la structure quaternaire du complexe TsaBDE est nécessaire pour mieux comprendre le rôle de la sous-unité TsaE.

#### **4.6. La modification t<sup>6</sup>A chez les archées et les eucaryotes.**

##### **4.6.1. Découverte du complexe KEOPS.**

Lors d'une étude génomique fonctionnelle de *Saccharomyces cerevisiae* conduite dans le but d'isoler de nouveaux gènes impliqués dans l'homéostasie des télomères, des partenaires de Kae1 ont été identifiés : Cgi121, Bud32 et Gon7 (yGon7) (Downey *et al.*, 2006). A la même époque, il a été montré que ces 3 protéines et une petite protéine nommée Pcc1 sont impliqués dans la transcription de gènes essentiels (Kisseleva-Romanova *et al.*, 2006). Ces protéines interagissent et forment un complexe nommé KEOPS (Kinase Endopeptidase et Other Proteins of small Size) ou EKC (Endopeptidase-like Kinase Chromatin-associated) (Kisseleva-Romanova *et al.*, 2006) (figure 42). A l'exception de yGon7, unique chez les champignons, les protéines qui interagissent avec la protéine universelle Kae1 sont conservées chez les eucaryotes et les archées (indiquées par une étoile noire sur la figure 42).

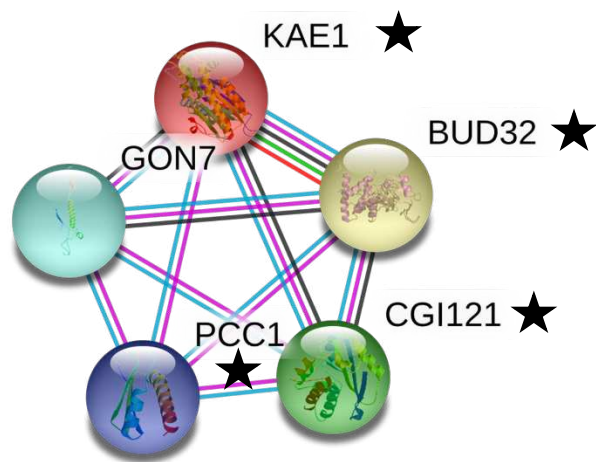


Figure 42 Réseau d'interaction de Kae1 chez *Saccharomyces cerevisiae*. Les protéines conservées dans les domaines des archées et des eucaryotes sont annotées par une étoile noire.

#### 4.6.2. Les partenaires de Kae1 dans le complexe KEOPS.

##### 4.6.2.1. La protéine Bud32, une kinase atypique.

Chez certaines archées, comme la *Methanocaldococcus jannashii*, le gène *kae1* est fusionné au gène *bud32* confirmant une fonction commune (Hunter *et al.*, 1997 ; Koonin *et al.*, 1998 ; Marcotte *et al.*, 1999 ; Lopraeito *et al.*, 2004 ; Naor *et al.*, 2012). La résolution de la structure de la protéine de fusion de *Methanocaldococcus janaaschii* a révélé la première structure de Bud32 et celle-ci est similaire à celles des protéines de la famille des kinases du type RIO2 (Hecker *et al.*, 2008; LaRonde, 2014; Mao *et al.*, 2008). La protéine Bud32 est agencée en deux domaines : un domaine C-terminal qui présente un repliement typique des kinases du type RIO2 composé de deux brins  $\beta$  et cinq hélices  $\alpha$  et un domaine N-terminal réduit par rapport aux domaines N-terminaux des kinases du type RIO, composé de 4 brins  $\beta$  ainsi que d'une hélice  $\alpha$  (Hecker *et al.*, 2008; Mao *et al.*, 2008) (figure 43 A). Aucun nucléotide n'a pu être co-cristallisé avec Bud32, en présence de Kae1, et dans ces conditions la P-loop de Bud32 est désordonnée (Hecker *et al.*, 2008; Mao *et al.*, 2008).

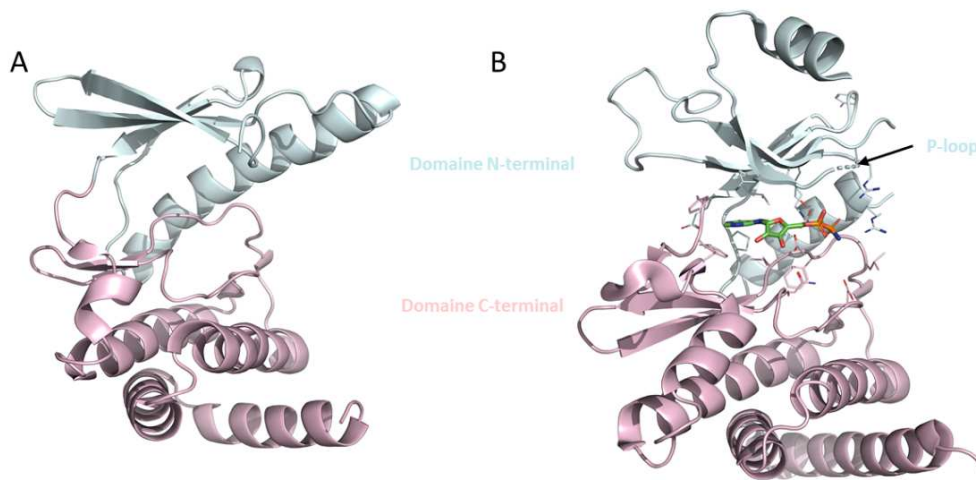


Figure 43 Structure tridimensionnelle de Bud32. A : Structure du domaine Bud32 de *Methanocaldococcus jannashii* en fusion avec Kae1 (non représentée sur la figure) (code PDB 2VWB). B : Structure tridimensionnelle de Bud32 de *Saccharomyces cerevisiae* (code PDB 4ww5). Les domaines N -terminaux et C-terminaux de Bud32 sont en bleu clair et rose respectivement. Entre les domaines N-terminaux et C-terminaux se forme une poche où se fixe une molécule d'AMP, stabilisée par les acides aminés représentés en stick dans Bud32 de *Saccharomyces cerevisiae*.

La structure de l'hétérodimère Bud32-Cgi121 de *Saccharomyces cerevisiae* montre la présence d'une molécule d'AMP dans le site actif de Bud32 (Zhang *et al.*, 2015) (figure 43 B). Le domaine N-terminal possède un brin  $\beta$  supplémentaire formant, avec les résidus du linker entre les deux domaines, une poche hydrophobe qui stabilise l'adénosine monophosphate. Un métal divalent  $Mg^{2+}$  est coordonné par le phosphate  $\beta$  du nucléotide et par les acides aminés D<sub>161</sub>, N<sub>166</sub>, D<sub>180</sub>, D<sub>182</sub>. Le phosphate  $\beta$  interagit quant à lui avec la lysine 52, acide aminé hautement conservé chez les protéines de la famille des kinases RIO (Zhang *et al.*, 2015).

Au sein des eucaryotes, la fonction de la famille des kinases RIO est reliée à la synthèse des ribosomes notamment dans le processing de l'ARNr 18S (Knüppel *et al.*, 2018). La lysine de la p-loop conservée chez les RIO2 kinases joue un rôle important dans la liaison avec la particule pré-40S du ribosome (Ferreira-Cerca *et al.*, 2012). Au sein de Bud32, cette lysine coordonne le phosphate  $\beta$  du nucléotide (Zhang *et al.*, 2015). Le mutant ponctuel K<sub>52</sub>A chez Bud32 empêche l'activité d'auto-phosphorylation (Mao *et al.*, 2008) et conduit à des défauts de croissance cellulaire chez la levure (Kisseleva-Romanova *et al.*, 2006). Il a été déjà montré que les kinases RIO ont une activité ATPase 50 à 100 fois plus élevée que leur activité d'auto-phosphorylation (Ferreira-Cerca *et al.*, 2012). Le motif composé des acides aminés H<sub>160</sub>G<sub>161</sub>D<sub>162</sub>, est aussi conservé et est retrouvé au sein de la famille Bud32/TP53RK (annoté

par le rectangle rose sur [figure 44](#)). Dans les kinases RIO2, ce motif serait responsable de l'attaque nucléophile de l'aspartate sur le phosphate  $\delta$  de l'ATP (LaRonde, 2014).

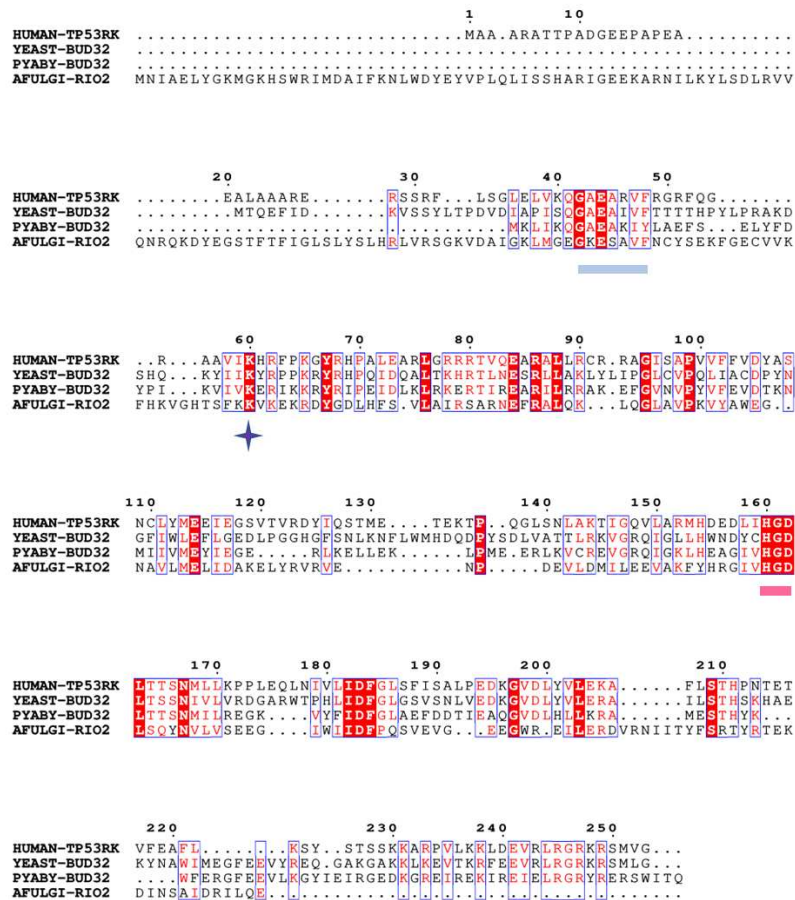


Figure 44 Alignement des séquences de la famille Bud32/TP53RK d'*Homo sapiens* (Human), de *Saccharomyces cerevisiae* (Yeast), de *Pyrococcus abyssi* (PYRAB) avec la séquence de la protéine kinase RIO2 *Archaeoglobus fulgidus* (AFULGI). La lysine conservée est identifiée par une étoile, le motif HGD par une barre rose et le motif P-loop par une barre bleue.

#### 4.6.2.2. Cgi121

Cgi121 est conservée chez les eucaryotes et les archées. La structure tridimensionnelle de cette protéine consiste en un feuillet central de brins  $\beta$  antiparallèles, entouré de deux et cinq hélices  $\alpha$  de chaque côté ([figure 45](#)) (Mao *et al.*, 2008 ; Zhang *et al.*, 2015). A ce jour, aucune autre structure de protéine homologue de Cgi121 n'a été résolue. Sa contribution lors de la synthèse de la modification t<sup>6</sup>A n'est pas encore bien comprise, cependant son absence diminue le taux de modification t<sup>6</sup>A sur les ARNt substrats (Perrochia *et al.*, 2013) et altère la croissance de *Haloferax volcanii* (Naor *et al.*, 2012).

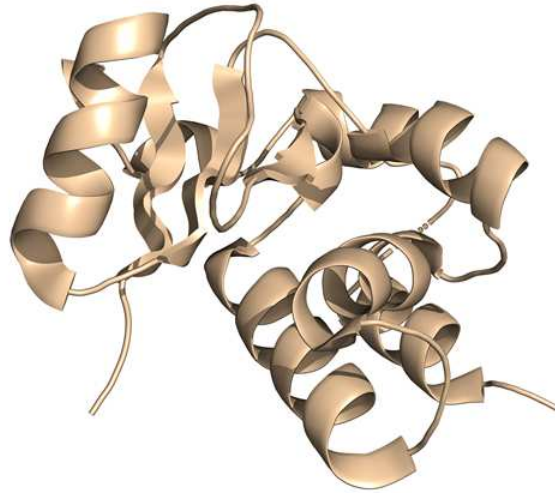


Figure 45 Structure tridimensionnelle de Cgi121 chez l'Homme (code PDB 3ENP).

#### 4.6.2.3. La sous-unité Pcc1.

La structure tridimensionnelle de Pcc1 montre que cette protéine est composée de trois brins  $\beta$  antiparallèles et de deux hélices  $\alpha$  (Mao *et al.*, 2008) et se dimérise (figure 46 A). Les brins  $\beta_1$  ainsi que les hélices  $\alpha_2$  de chaque sous-unité stabilisent l'interaction formant un long feuillet  $\beta$  de six brins (Mao *et al.*, 2008) majoritairement hydrophobe (figure 46 B). La structure tridimensionnelle de Pcc1 est similaire à celle des domaines KH, retrouvés notamment dans les protéines de liaison aux ARN, ADN double brin et ADN en épingle à cheveux (Grishin, 2001; Nicastro *et al.*, 2015; Valverde *et al.*, 2008). Cette famille de protéines présente un motif GxxG de fixation aux acides nucléiques. Ce motif est absent chez Pcc1, cependant, il a été mis en évidence que des domaines KH, ayant un motif GXXG altéré, peuvent fixer des acides nucléiques (Brykailo *et al.*, 2007).

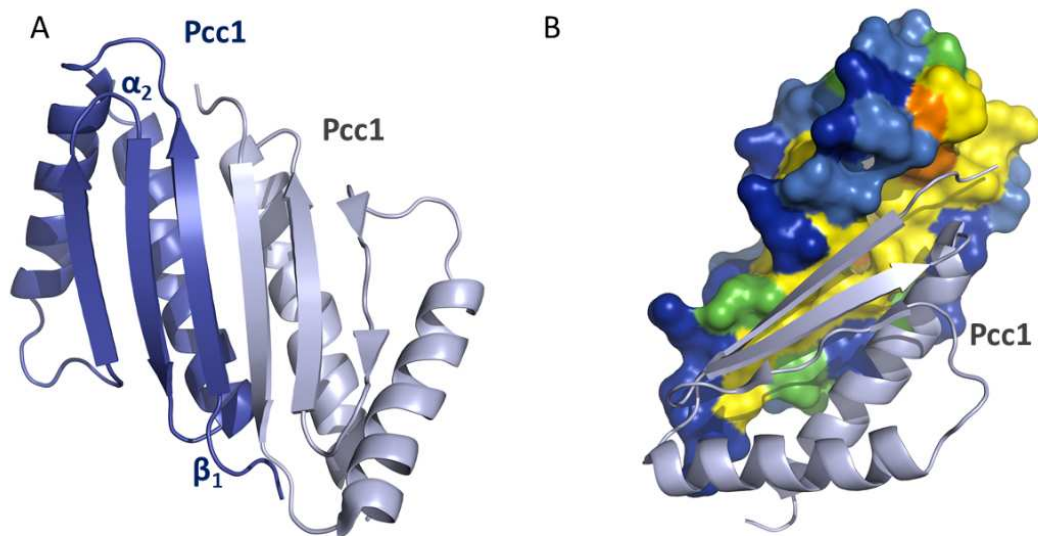


Figure 46 A : Structure de l'homodimère Pcc1 de *Pyrococcus furiosus* (code PDB 3ENC). Chaque sous-unité interagit par leurs brins  $\beta_1$  respectifs et par leurs hélices  $\alpha_2$ . B : Une des sous-unités de Pcc1 est représenté selon la polarité des acides aminés (les acides aminés hydrophobes sont en jaune et les acides aminés polaires sont en bleu), montrant que l'interaction entre les deux sous-unités est majoritairement hydrophobe.

#### 4.6.3. Les différentes sous-unités Kae1, Bud32, Cgi121 et Pcc1 s'agencent de façon linéaire.

La structure tridimensionnelle de Cgi121-Bud32-Kae1 montre que le complexe s'agence de façon linéaire, aucune interaction n'est détectée entre Cgi121 et Kae1 (Mao *et al.*, 2008) (figure 47). Aucun changement conformationnel drastique de Bud32 n'a été observé en interaction avec Kae1 ou Cgi121. Seule la boucle "p-loop" devient ordonnée en présence du nucléotide. Bud32 interagit avec Kae1 via ses deux domaines, alors que pour Kae1, l'interaction implique principalement le domaine C-terminal. Une région appelée insert spécifique 1 est localisée dans le domaine C-terminal de Kae1 et joue le rôle d'une plateforme d'interaction avec Bud32. De manière intéressante, la zone de contact avec Bud32 de cet insert est structurée seulement chez les orthologues d'archées et d'eucaryotes, et flexible chez les orthologues bactériens, chez qui Bud32 est absent. (Hecker *et al.*, 2008). En comparaison avec la structure de Kae1 de *Pyrococcus abyssi* (Hecker *et al.*, 2007), il semble que la présence de Bud32 ne modifie pas la conformation globale de Kae1. Enfin, une autre région d'interaction est particulièrement intéressante : les structures révèlent que la queue C-terminale de Bud32 se rapproche du site catalytique de Kae1, proche du centre de coordination du fer. La délétion de cette région de Bud32 ne perturbe pas l'interaction avec

Kae1, mais affecte fortement la croissance chez la levure, indiquant que cette région est nécessaire à la fonction de Bud32 ou pour la catalyse de la modification t<sup>6</sup>A (Mao *et al.*, 2008).

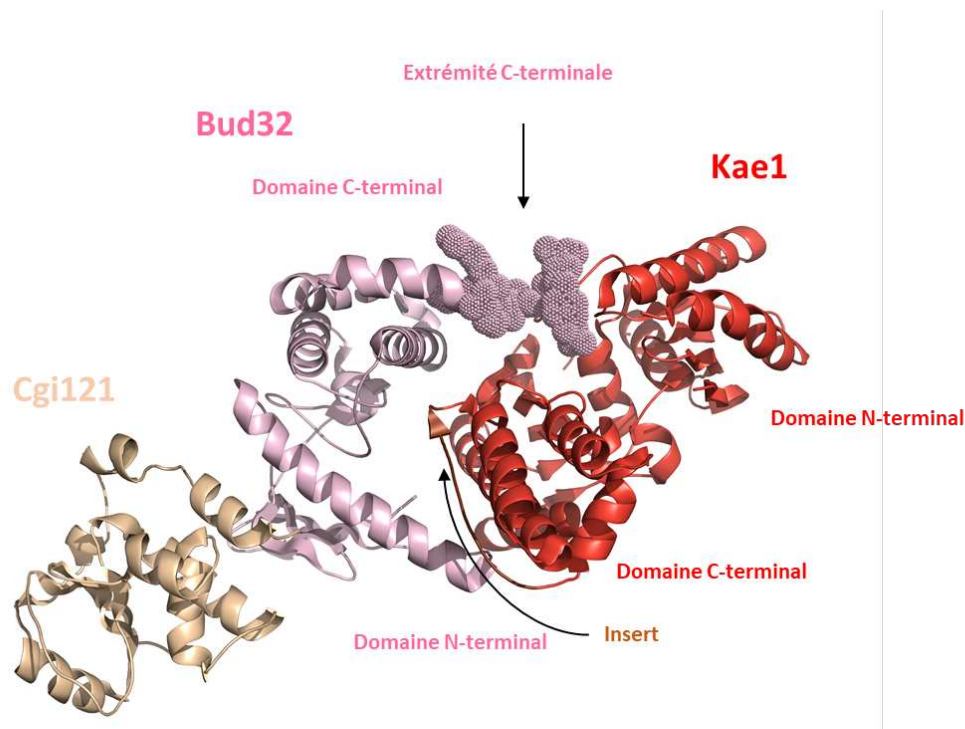


Figure 47 Structure quaternaire (code PDB 3ENH) de la protéine de fusion Kae1-Bud32 (en rouge et en rose respectivement) en interaction avec Cgi121 (en beige). Bud32 interagit avec le domaine C-terminal de Kae1. L'extrémité C-terminale de Bud32 représentée en boule se positionne à l'entrée du site actif de Kae1. L'insert de Kae1 en marron est situé à l'interface avec Bud32.

La comparaison des structures de Cgi121 seule et en complexe avec Bud32 ne montrent pas de changements conformationnels significatives. Cgi121 interagit exclusivement avec le domaine N-terminal de Bud32 (Zhang *et al.*, 2015). L'interface Bud32-Cgi121 est stabilisée par une combinaison d'interaction hydrophobe, de liaisons hydrogènes et d'interactions électrostatiques.

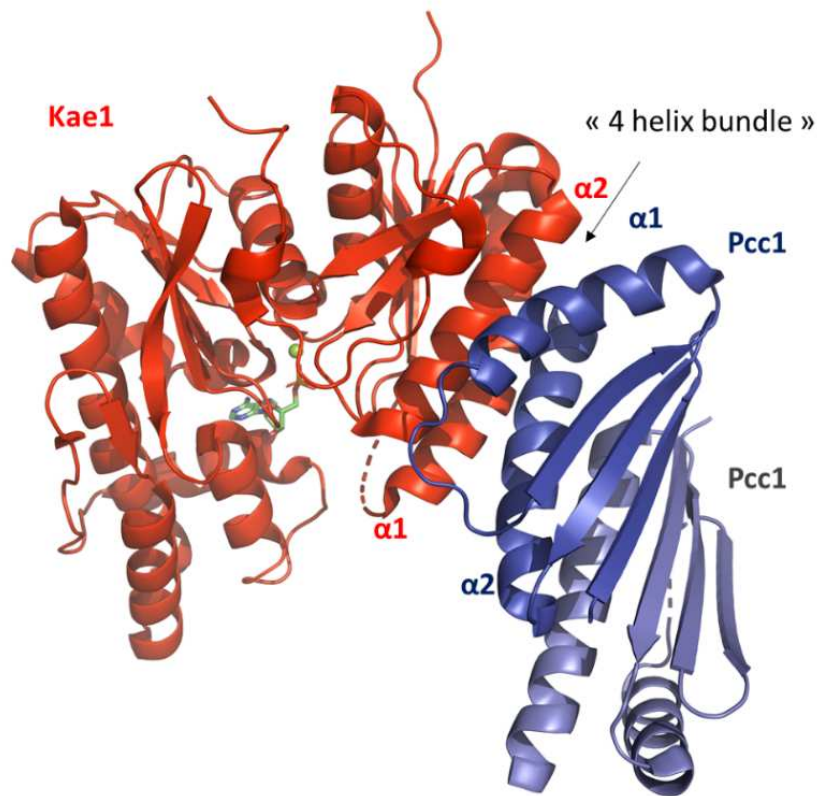


Figure 48 Structure quaternaire de Pcc1 en interaction Kae1 (code PDB 5jmv). Deux sous-unités de Pcc1 (bleu et bleu-gris) interagissent l'une avec l'autre via les brins  $\beta 1$  et les hélices  $\alpha 2$ . Un monomère de Pcc1 (en bleu) interagit avec Kae1. L'interface est stabilisée par des interactions hydrophobes de l'hélice  $\alpha 1$  -  $\alpha 2$  de Pcc1 et  $\alpha 2$  -  $\alpha 4$  de Kae1, formant un empilement de quatre hélices ("4 helix bundle"). Une molécule d'AMP et de  $Mg^{2+}$  (en vert) sont fixées dans le site actif de Kae1.

La structure du complexe d'archée montre que Pcc1 interagit avec le domaine N-terminal de Kae1 ; les hélices  $\alpha 1$  et  $\alpha 2$  de Kae1 et de Pcc1 stabilisent cette interface, formant un empilement de quatre hélices ("4 helix bundle") (figure 48) (Mao *et al.*, 2008). La majorité des interactions entre ces deux sous-unités sont de nature amphiphile. De façon intéressante, le mode d'interaction est similaire à celui de TsaBD, malgré l'absence de similitude structurale entre les deux protéines TsaB et Pcc1.

Pcc1-Kae1, dans le cristal, forme un complexe avec une stœchiométrie 2 :1. En revanche, en solution le complexe Pcc1-Kae1 d'archée et de levure forment un hétérotétramère avec une stœchiométrie 2 :2 (Mao *et al.*, 2008 ; Wan *et al.*, 2016 ; Zhang *et al.*, 2015). La surface d'interaction entre les deux sous-unités de Pcc1 dans le complexe trimérique Pcc1-Kae1 est identique à celle des deux sous unités Pcc1 de l'homodimère (Mao *et al.*, 2008). Ce résultat



suggère que la présence de Kae1 n'altère pas la surface d'interaction entre les deux sous-unités de Pcc1.

Pour finir, il a été mis en évidence que les sous-unités Kae1, Cgi121, Bud32 et Pcc1 chez les archées forment un complexe en solution de stœchiométrie 2 :2 :2 :2. Grâce aux différentes structures quaternaires des sous-complexes résolues jusqu'à présent, un modèle a été proposé où Cgi121 interagit avec Bud32, Bud32 interagit avec Kae1, et enfin Kae1 interagit avec Pcc1. Les deux sous-unités de Pcc1 se dimérisent assurant la stœchiométrie 2 :2 :2 :2 formant un octamère (Mao *et al.*, 2008). Cet agencement est également observé chez l'Homme (Wan *et al.*, 2017).

#### 4.6.3.1. Une cinquième sous-unité compose le complexe KEOPS chez les eucaryotes.

Une cinquième sous-unité, nommée  $\gamma$ Gon7, composant le complexe KEOPS, est retrouvée chez les levures (Kisseleva-Romanova *et al.*, 2006; Lopreiato *et al.*, 2004).  $\gamma$ Gon7 de *Saccharomyces cerevisiae* est une petite protéine de 123 acides aminés et est désordonnée seule en solution (Zhang *et al.*, 2015). La première structure cristallographique de  $\gamma$ Gon7 a été résolue en complexe avec Pcc1 (code PDB 4WXA) (Zhang *et al.*, 2015). Au contact de Pcc1,  $\gamma$ Gon7 se replie en deux brins  $\beta$  antiparallèles et en deux hélices  $\alpha$ . L'interaction entre Pcc1 et  $\gamma$ Gon7 est réalisée par les brins  $\beta_1$  de chaque sous-unité formant un long feuillet antiparallèle à 5 brins. Une seconde surface d'interaction implique les hélices  $\alpha_1$  et  $\alpha_2$  de  $\gamma$ Gon7 et Pcc1, respectivement (figure 49). Ce mode d'interaction est similaire à celui de l'homodimère Pcc1 (Zhang *et al.*, 2015) et est stabilisé majoritairement par des interactions hydrophobes via les acides aminés des hélices  $\alpha$  et polaires via les acides aminés des brins  $\beta$ .

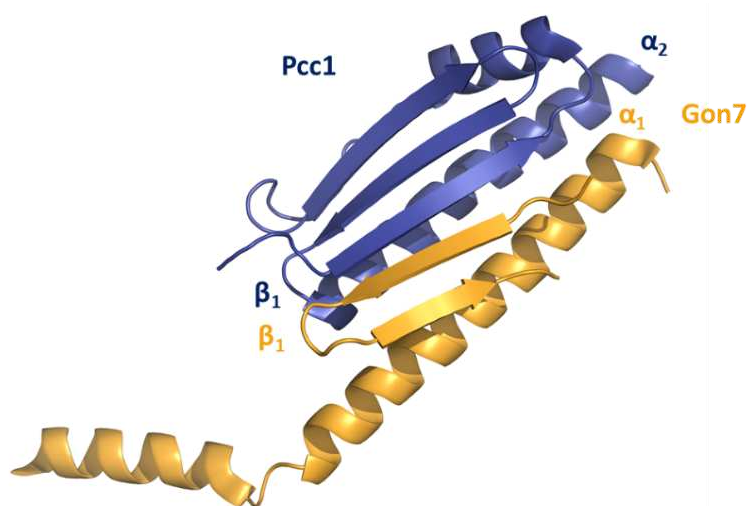


Figure 49 Structure quaternaire du complexe Pcc1 (en bleu) et  $\gamma$ Gon7 (en orange) de *Saccharomyces cerevisiae* (code PDB 4WXA). L'interface est stabilisée entre les résidus des hélices  $\alpha_2$  et  $\alpha_1$  de Pcc1 et  $\gamma$ Gon7, respectivement, ainsi que les résidus des brins  $\beta_1$  de Pcc1 et de  $\gamma$ Gon7, formant, ainsi, un long feuillet antiparallèle de cinq brins.

L'identification de l'homologue de  $\gamma$ Gon7 au sein de tous les eucaryotes par l'alignement de séquences fut un échec. En 2012, lors d'une étude des voies cellulaires activées par l'antigène tumoral humain PRAME (preferentially expressed antigen in melanoma), il a été mis en évidence que les sous-unités du complexe KEOPS interagissent avec la ligase Cullin2 (Costessi *et al.*, 2012). A partir de l'analyse de l'interactome des protéines du complexe KEOPS et de la Cullin2, de nombreux partenaires potentiels ont été identifiés dont une protéine nommée C14orf142. C14orf142 est prédite comme étant de petite taille (100 acides aminés) dépourvue d'éléments de structures secondaires stables (Costessi *et al.*, 2012) suggérant qu'elle pourrait être la cinquième sous-unité du complexe KEOPS humain. En 2016, l'équipe de Franck Sicheri analyse les séquences de Gon7 des champignons et découvre une similitude de séquence entre C14orf142 de l'Homme et Gon7 de *Candida glabrata* (identité de séquence = 19,7 %) (Wan *et al.*, 2016). De plus, des expériences de " pull-down " confirment que C14orf142 interagit directement avec la protéine LAGE3 suggérant que C14orf142 est l'analogue de  $\gamma$ Gon7 (Wan *et al.*, 2016). La structure de C14orf142 en complexe avec son partenaire LAGE3 (l'homologue humain de Pcc1) est nécessaire pour lever les incertitudes concernant l'analogie structurale de cette protéine avec la protéine  $\gamma$ Gon7 de la levure.

#### 4.6.4. La structure quaternaire du complexe KEOPS chez les eucaryotes et les archées.

C14orf142 empêche la dimérisation du complexe KEOPS et, *in vitro*, sa présence augmente l'activité de modification  $t^6A$  sur les ARN<sup>t<sup>Lys</sup>UUU</sup> d'un facteur 3 à 4 (Wan *et al.*, 2016). Les orthologues de C14orf142, que l'on renommara GON7 dans cette thèse par la suite, ont été retrouvés au sein des métazoaires élargissant le complexe de quatre à cinq sous-unités au sein des champignons et des métazoaires (figure 50). Un des objets de cette thèse sera la caractérisation structurale de la sous-unité GON7 chez l'Homme (c.f chapitre III).

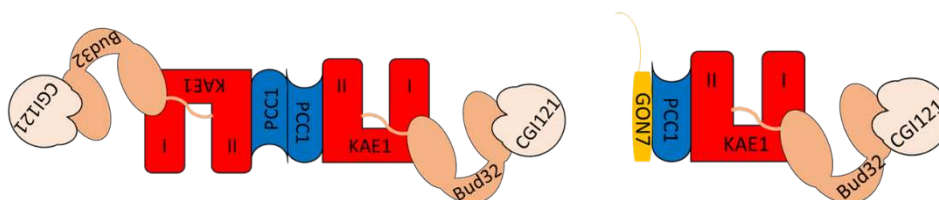


Figure 50 Représentation schématique des structures quaternaires des complexes KEOPS chez les archées (à gauche) et les eucaryotes (à droite). Le complexe archée est composé des sous-unités Pcc1, Kae1, Bud32 et Cgi121. Pcc1 forme une plateforme d'homodimérisation, conduisant à la formation d'un octamère linéaire et symétrique de 8 sous-unités. En présence de  $\gamma$ Gon7 (ou C14ORF142 chez les métazoaires), le complexe eucaryote forme un pentamère.

Un modèle de la structure tridimensionnelle du complexe KEOPS entier de la levure pentamérique proposé à partir de structures cristallographiques des sous complexes montre un arrangement linéaire où Pcc1 interagit avec Kae1 qui interagit avec Bud32 qui interagit avec Cgi121 (Zhang *et al.*, 2015). *In vivo* la délétion du gène *GON7* chez *Saccharomyces cerevisiae* conduit à une diminution drastique de la modification t<sup>6</sup>A (Wan *et al.*, 2016). Le complexe KEOPS des archées et des eucaryotes diffère par la présence de la cinquième sous-unité, qui jusqu'à présent n'a toujours pas été identifiée chez les archées. Un des objets de cette thèse portera sur la recherche de cette cinquième sous-unité chez les archées ([cf. chapitre IV](#)).

#### **4.6.5. Fonction des sous-unités du complexe KEOPS chez les eucaryotes et les archées.**

Après la découverte de l'équipe de Valérie De Crécy, plaçant Kae1 au centre de la biosynthèse de la modification t<sup>6</sup>A, de nombreuses études ont permis d'éclaircir le rôle des différentes sous-unités de KEOPS dans la biosynthèse de la modification t<sup>6</sup>A chez les archées et les eucaryotes. La délétion des gènes *BUD32*, *PCC1* et *KAE1* chez *Saccharomyces cerevisiae* conduit à une diminution significative de la modification t<sup>6</sup>A sur les ARNt<sup>lle</sup><sub>AAU</sub> ; en revanche, le rôle de Cgi121 semble moins primordial (Srinivasan *et al.*, 2011). De plus, des mutants ponctuels thermosensibles de Kae1 et Pcc1 chez *Saccharomyces cerevisiae* présentent une diminution de la quantité de t<sup>6</sup>A dans les conditions non permissives (Daugeron *et al.*, 2011). Chez l'archée *Haloferax volcanii*, l'absence de Pcc1 conduit également à une légère diminution de la quantité de t<sup>6</sup>A au niveau des ARNt et n'affecte pas significativement sa croissance, suggérant que cette protéine pourrait jouer un rôle accessoire dans la synthèse de la modification t<sup>6</sup>A (Naor *et al.*, 2012). Mao *et al.* ont montré que Kae1 est la cible de phosphorylation de Bud32 (Mao *et al.*, 2008). Et une activité ATPase de Bud32 est détectée et confirmée *in vitro* chez *Pyrococcus abyssi* (Perrochia *et al.*, 2013) et des mutations dans le site actif de Bud32 affecte la croissance des levures et des archées (Mao *et al.*, 2008; Perrochia *et al.*, 2013). Cependant au sein du complexe KEOPS, la fonction de Bud32 est à ce jour encore méconnue.

Les résultats obtenus par les précédentes études semblent indiquer que les sous-unités du complexe KEOPS participent à de nombreux processus cellulaires fondamentaux. Pendant de nombreuses années, la pléiotropie des phénotypes observés ne permettait pas d'établir un

rôle biologique clair pour le complexe KEOPS. En effet, il a été montré, en 2006, que Kae1 présente un rôle dans l'homéostasie des télomères (Downey *et al.*, 2006) et dans la régulation de la transcription des gènes essentiels (Kisseleva-Romanova *et al.*, 2006). L'hétérogénéité de ces phénotypes observés, suite à un défaut de TsaD/Kae1/Qri7, montrent que cette famille universelle présente un rôle crucial au sein de la machinerie cellulaire des différents organismes.

#### **4.7. Un système minimaliste dans les mitochondries.**

##### **4.7.1. Qri7, l'orthologue mitochondrial de Kae1/TsaD.**

Les eucaryotes possèdent une seconde version de Kae1, nommée Qri7 chez *Saccharomyces cerevisiae* ou OSGEPL chez l'Homme. Les études phylogénétiques ont montré que la version Qri7/OSGEPL est présente chez les génomes eucaryotes, exceptés dans les organismes dépourvus de mitochondries (tels que *Microsporidomycota* ou *Entamoebidae*) et émerge de la branche bactérienne (Huh *et al.*, 2003 ; Galperin et Koonin, 2004 ; Hecker *et al.*, 2007). Cette hypothèse est confirmée par les données expérimentales attribuant une localisation cytoplasmique à Kae1 et mitochondriale à Qri7 (Hecker *et al.*, 2007 ; Oberto *et al.*, 2009).

*QRI7* de *Saccharomyces cerevisiae* complémente l'extinction du gène KO *tsaD* en restaurant la viabilité de la souche bactérienne. En revanche, le gène *KAE1* de la levure ne complémente pas l'extinction du gène *tsaD* (Oberto *et al.*, 2009) suggérant que Qri7 seul peut assurer l'activité de modification t<sup>6</sup>A et au contraire que Kae1 seul ne soit pas capable en dehors du complexe KEOPS de restaurer l'activité de modification t<sup>6</sup>A. Le mutant  $\Delta KAE1$  de *Saccharomyces cerevisiae* présente une quantité résiduelle de t<sup>6</sup>A, alors qu'un double mutant  $\Delta KAE1/\Delta QRI7$  ne présente plus aucun nucléotide modifié t<sup>6</sup>A (Srinivasan *et al.*, 2011) confirmant le rôle de Qri7 dans la modification t<sup>6</sup>A (Wan *et al.*, 2013 ; Thiaville *et al.*, 2014).

##### **4.7.2. La formation d'un homodimère Qri7 assure la biosynthèse t<sup>6</sup>A sur les ARNt mitochondriaux.**

La structure de Qri7 est organisée en deux domaines et montre qu'elle est un homologue structural de Kae1 et TsaD (Wan *et al.*, 2013 ; Tominaga *et al.*, 2014). En solution, Qri7 forme un homodimère (Wan *et al.*, 2013) retrouvé également dans l'unité asymétrique de la structure cristallographique (code PDB 3WUH) (figure 51). Les deux monomères interagissent par leurs domaines N-terminaux respectifs et l'interface est formée par un faisceau de 4

hélices ("4-helix bundle"). La dimérisation de Qri7 est stabilisée par des liaisons hydrogènes, des interactions de van der Waals et apolaires entre les hélices  $\alpha 1$  et  $\alpha 2$  de chaque sous-unité (Tominaga *et al.*, 2014) (figure 51).

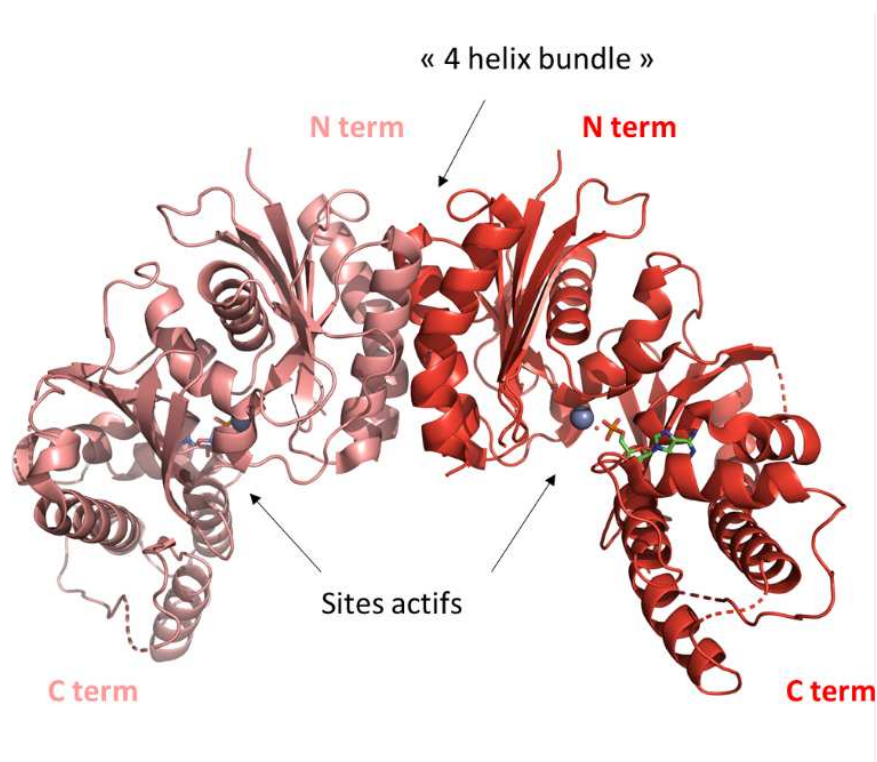


Figure 51 Structure quaternaire de l'homodimère Qri7 de *Saccharomyces cerevisiae* (PDB ascension 3WHU). Les deux sous-unités Qri7 (en rouge et rose) interagissent par leur domaine N terminal respectif via un motif d'amas d'hélices  $\alpha$  ("4 helix bundle"). Une molécule d'AMP est retrouvée dans les sites actifs des deux sous-unités (stick vert) et un ion  $Zn^{2+}$  (sphère grise).

Une étude de mutagenèse dirigée des résidus impliqués dans l'interaction a montré que la dimérisation de Qri7 est nécessaire à la biosynthèse de la modification  $t^6A$  (Wan *et al.*, 2013). Entre les domaines N et C-terminaux de Qri7, qui forment la cavité du site actif, une molécule d'AMP est retrouvée. Les groupement hydroxyles 2' et 3' OH du ribose sont coordonnés par l'aspartate 203. L'adénosine est stabilisée par les acides aminés D<sub>221</sub>, S<sub>328</sub> et N<sub>329</sub>, tandis que le phosphate est coordonné par G<sub>325</sub>, G<sub>173</sub> et S<sub>360</sub>. Le  $Zn^{2+}$  est coordonné par le motif HHD et par une molécule d'eau qui coordonne également le phosphate. La région comprise entre les acides aminés 72-79 est désordonnée bien que les acides aminés de cette boucle soient très conservés. Chez l'orthologue lointain TobZ, la boucle équivalente interagit avec le substrat tobramicyne et les acides aminés de cette boucle participe à la déprotonation de l'hydroxyle du substrat (Parthier *et al.*, 2012). Cette comparaison a permis de proposer que cette boucle

flexible de Qri7 pourrait jouer un rôle dans l'interaction avec l'ARNt (Tominaga *et al.*, 2014). A ce jour, aucune structure d'un membre de la famille Kae1/TsaD/Qri7 avec son substrat ARNt n'a été résolue et leur mode d'interaction reste inconnu.

Il est intéressant de noter que l'interface des deux sous-unités Qri7 est similaire aux interfaces observées pour les complexes entre TsaD-TsaB et Kae1-Pcc1 (Wan *et al.*, 2013). L'alignement des séquences des homologues de la famille Ka1/OSGEP/TsaD/Qri7 est exposé en annexe 2 ([cf. page 326](#)).

#### **4.8. Les t<sup>6</sup>A synthèses à travers les trois domaines du vivant.**

La caractérisation des machineries de biosynthèse de t<sup>6</sup>A au sein des différents domaines du vivant montre que ce sont deux protéines universelles, YrdC/Sua5 et Kae1/YgjD/Qri7 qui sont responsables de cette réaction.

En revanche des partenaires annexes non universels ont été sélectionnés selon les organismes et sont strictement requis ([figure 52](#)). Malgré les différences selon les espèces, la caractérisation structurale des différents complexes permet de trouver des analogies. En effet, les hélices du domaine N-terminal des membres de la famille Kae1/TsaD/Qri7 interagissent toujours avec deux hélices de leurs partenaires, reconstituant le motif "4-helix bundle". La conservation de ce type d'interaction au cours de l'évolution illustre probablement un besoin de stabilisation de l'enzyme catalytique. Les partenaires Bud32 et Cgi121 sont uniques au sein des eucaryotes et des archées. Ces protéines interagissent du côté du domaine C-terminal de Kae1. Hecker *et al.*, avait mis en évidence un peptide nommé "insert 1" de Kae1 qui interagit avec Bud32 localisé au domaine C-terminal et structuré uniquement chez les eucaryotes et archées. (Hecker *et al.*, 2008). Cet élément qui diverge chez les procaryotes pourrait être le résultat d'une spécialisation de Kae1 au sein des archées et des eucaryotes. Malgré la nécessité des protéines Bud32 et de Cgi121, leurs fonctions exactes restent toujours peu connues.

Pour finir, l'existence d'un système minimaliste fonctionnel à deux composants Sua5 et Qri7 dans la mitochondrie indique qu'il est possible que le dernier ancêtre commun possédait lui aussi un système à deux composants uniquement : Sua5 et Qri7 suffisant pour assurer la modification t<sup>6</sup>A.

Dans le schéma suivant, les protéines qui assurent la catalyse de la modification t<sup>6</sup>A sont récapitulées.

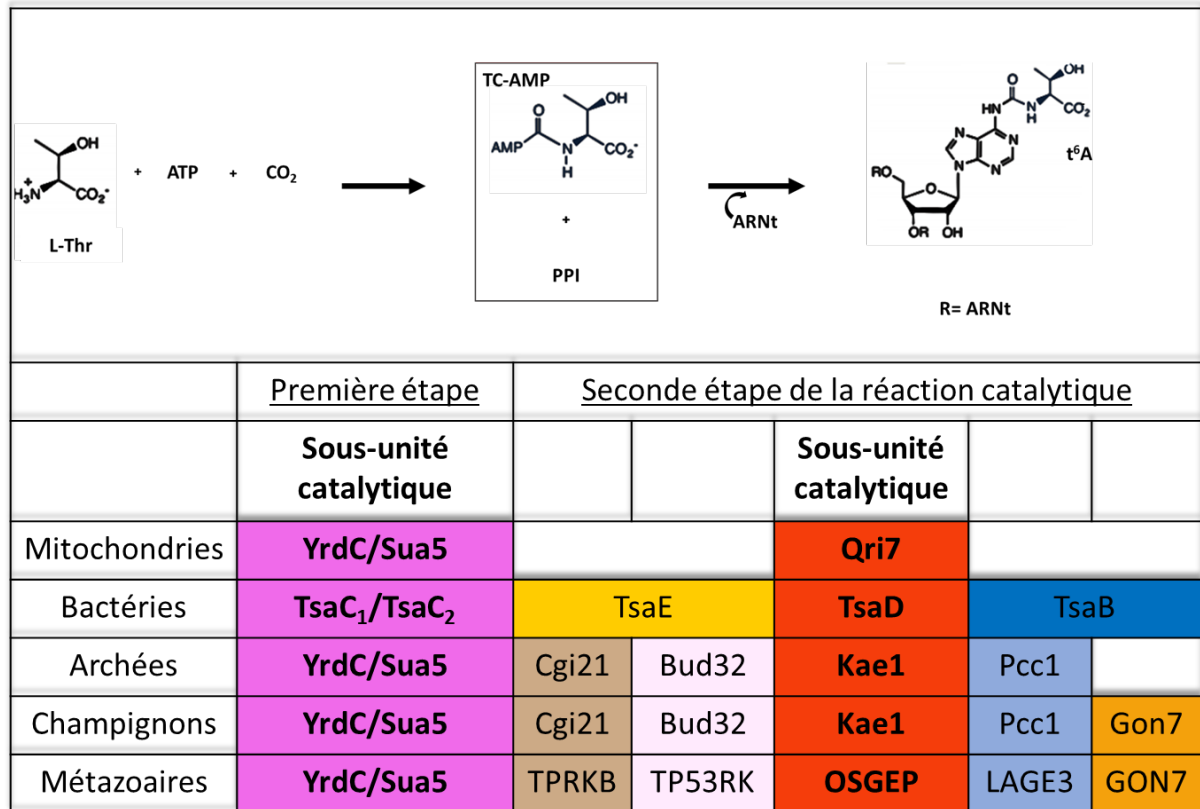


Figure 52 Schéma récapitulatif des protéines assurant la catalyse de la modification t<sup>6</sup>A.

#### 4.9. Le défaut de la modification t<sup>6</sup>A affectent différemment les espèces.

La modification t<sup>6</sup>A est essentielle à la vie cellulaire de la majorité des bactéries, à l'exception de *Deinococcus radiodurans*, *Thermus thermophilus*, *Synechocystis* PCC6803 et *Streptococcus mutans*, chez qui le défaut de t<sup>6</sup>A est compatible avec la survie (Thiaville *et al.*, 2015; de Crécy-Lagard et Jaroch, 2020). Des phénotypes pléiotropes selon les organismes sont observés suite à un défaut ou à l'absence de la modification t<sup>6</sup>A : des défauts de la division cellulaire, une augmentation des protéines de glycation, l'accumulation de la cyanophycine, une répartition inhabituelle de l'ADN, une perte du nucléoïde et une morphologie cellulaire modifiée (Allali-Hassani *et al.*, 2004; Handford *et al.*, 2009; Oberto *et al.*, 2009). Chez *Deinococcus radiodurans*, le défaut t<sup>6</sup>A semble sensibiliser davantage la bactérie à la mytomycine C, et induit un pontage (cross-link) au niveau de l'ADN. De plus, l'étude du protéome, met en évidence une augmentation significative d'expression des protéines en réponse au stress dans les souches  $\Delta$ TsaC (Onodera *et al.*, 2013). Bien que non-essentielle chez *Deinococcus radiodurans*,

l'absence de la modification t<sup>6</sup>A perturbe l'expression protéique (103 protéines sont sur-exprimées et 66 protéines sont sous-exprimées) (Thiaville *et al.*, 2015).

Chez les eucaryotes, la modification t<sup>6</sup>A n'est pas essentielle *stricto sensu*. Par exemple, chez *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe*, l'absence de la modification t<sup>6</sup>A provoque une croissance des cellules non optimales (Kisseleva-Romanova *et al.*, 2006). Des études ont montré que les mutants  $\Delta KAE1$  chez *Saccharomyces cerevisiae*, conduisent à des décalages de lecture et des défauts d'initiation de la traduction (El Yacoubi *et al.*, 2011; Lin *et al.*, 2010; Daugeron *et al.*, 2011). Également, une étude du " ribosome profiling " a mis en évidence que le défaut de t<sup>6</sup>A ne conduit pas à une altération globale du processus de traduction, en revanche, des erreurs de lecture du codon d'initiation et des décalages de lecture des codons de gènes spécifiques sont observés (Thiaville *et al.*, 2016). Aussi, le défaut de la modification t<sup>6</sup>A altère significativement le contrôle de la vitesse de lecture des codons, suggérant que la modification t<sup>6</sup>A présente un rôle dans la reconnaissance des différents codons ANN. Une traduction infidèle pourrait conduire à l'obtention de fragments de protéines non-fonctionnelles potentiellement toxiques pour la cellule (Thiaville *et al.*, 2016). Chez *Arabidopsis thaliana*, la délétion de *QR17* provoque un arrêt de la croissance de la plante (Haussehel *et al.*, 2009). Chez *Saccharomyces cerevisiae* et *Candida elegans*, l'absence de *Qri7* a pour conséquence une fragmentation du génome mitochondrial ainsi qu'une morphologie aberrante de cet organite (Oberto *et al.*, 2009).

Malgré son universalité, l'absence de la modification t<sup>6</sup>A n'a pas les mêmes conséquences chez tous les organismes. Il a été vu (§ 2.5.1.2.) que l'absence de la modification t<sup>6</sup>A empêche la lysylation de de cytosine 34 (k<sup>2</sup>C) chez les procaryotes qui permet la reconnaissance des codons AUA. Très récemment, il a été démontré que la modification hn<sup>6</sup>A, un variant de la modification t<sup>6</sup>A, peut être catalysée par les enzymes t<sup>6</sup>A synthases sur les ARNt substrats de *Escherichia coli* et *Thermotoga maritima* (Swinehart *et al.*, 2020). L'effet drastique du défaut des enzymes t<sup>6</sup>A synthases dans les bactéries pourrait être relié à ces modifications supplémentaires.



#### 4.10. Effet du défaut t<sup>6</sup>A sur les organismes pluricellulaires.

Le syndrome de Galloway-Mowat (GAMOS) est une maladie autosomique récessive rare dont les mécanismes restent encore à élucider (Galloway et Mowat, 1968). Il se caractérise par un syndrome néphrotique cortico-résistant, des anomalies du système nerveux central entraînant une microcéphalie, des retards mentaux et des dysfonctionnements rénaux conduisant à une greffe du rein dans la majorité des cas. La plupart des patients atteints du syndrome GAMOS meurent avant 20 ans (Braun *et al.*, 2017, 2018). Ces dernières années, un lien fort entre des dysfonctionnements du système de modifications des ARNt et des pathologies humaines a été mis en évidence (Power *et al.*, 2015 ; Kapur *et al.*, 2018). Chez l'Homme une mutation ponctuelle récessive dans OSGEP a été séquencée chez des patients présentant un retard global du développement qui se manifeste par une protéinurie et une hypomagnésémie (Edvardson *et al.*, 2017). Récemment, un séquençage exhaustif de l'exome de patients atteints de GAMOS a permis d'identifier des mutations sur les gènes codant pour 4 sous-unités du complexe KEOPS : OSGEP, LAGE3, TP53RK et TPRKB (Braun *et al.*, 2017). Le séquençage d'exome de patients a aussi permis d'identifier des mutations sur le gène codant la protéine putative WDR73, dont la fonction n'est pas bien déterminée mais pourrait avoir un rôle dans la régulation des microtubules d'une part, (Colin *et al.*, 2014) ainsi que le gène codant la protéine WDR4, d'autre part. De façon intéressante, cette dernière est requise pour catalyser la méthylation de la position 7 de la guanosine 46 (modification m<sup>7</sup>G) des ARNt, localisée sur la boucle variable (Alexandrov *et al.*, 2002; Leulliot *et al.*, 2008).

Également, le défaut de la modification t<sup>6</sup>A a été étudié chez la drosophile (*Drosophila melanogaster*) (Lin *et al.*, 2015). L'étude rapporte que les mutants homozygotes  $\Delta KAE1$  ont significativement moins d'ARNt possédant la modification t<sup>6</sup>A, et conduisent à un développement anormal des tissus et une absence totale des disques imaginaux, aboutissant à l'arrêt de la métamorphose des larves (Lin *et al.*, 2015 ; Rojas-Benítez *et al.*, 2017).

Au bilan, les défauts des modifications post-transcriptionnelles des ARNt peuvent présenter des phénotypes spécifiques selon les organismes, majoritairement ces défauts impactent les tissus neuronaux et rénaux chez l'Homme (Kapur *et al.*, 2018). Une traduction infidèle semble être la cause de ces syndromes, en revanche la vulnérabilité spécifique de certains types de tissus (reins et système nerveux chez l'Homme notamment) lors de l'embryogénèse reste incomprise.

La plupart des pathologies autosomiques rares à caractère héréditaire n'ont actuellement aucun traitement. Ainsi la compréhension de ces voies de modifications des ARNt, dont celles de la modification t<sup>6</sup>A, présente un intérêt crucial car elle pourrait permettre d'envisager des approches thérapeutiques dans le but de corriger les hypo et hyper-modifications des ARNt ou de réguler l'expression des enzymes de modification des ARNt. Pour le moment, l'identification des mutations sur les gènes codant les protéines responsables de ces voies de modifications pourrait être un grand pas pour le diagnostic précoce *in utero*. La compréhension du mécanisme des enzymes de la modification t<sup>6</sup>A pourrait contribuer, également, à obtenir plus d'informations concernant les modifications post-transcriptionnelles des ARNt.

## 5. Problématique et objectifs de la thèse.

En première approximation, la formation d'une molécule d'ARNt fonctionnelle suit un schéma global simple (de la transcription à la maturation). Dès la seconde moitié du 20<sup>e</sup> siècle, des modifications chimiques sont découvertes sur les acides nucléiques et notamment sur l'ARNt. Le rôle des modifications localisées sur la boucle anticodon de l'ARNt s'avèrent être essentielles pour l'efficacité et la fidélité de la traduction avec des phénotypes qui varient selon les organismes. Afin de comprendre le mécanisme de la biosynthèse de la modification t<sup>6</sup>A nous nous sommes concentrés sur l'étude des enzymes de la modification t<sup>6</sup>A. Une caractérisation structurale et une étude fonctionnelle de ces enzymes de modification a été entreprise dans les trois règnes du vivant.

L'objectif de ce travail de thèse était d'apporter des éléments de réponse concernant :

- (i) Le mécanisme du complexe TsaBDE chez les procaryotes. Les questions que nous nous posions étaient les suivantes :
  - Comment interagit la protéine TsaE avec TsaBD ? Quels sont les acides aminés impliqués dans cette interaction ? Quel est le rôle joué par TsaE ? La fixation de l'ARNt est-elle compatible avec la fixation de TsaE ?
  - Quels sont les acides aminés de TsaD, la sous-unité catalytique, impliqués dans la fixation de l'intermédiaire TC-AMP et ceux impliqués dans la catalyse de la

modification t<sup>6</sup>A ? A partir de ces connaissances peut-on cibler le site actif de TsaD pour développer des antibiotiques ?

- (ii) Le mécanisme du complexe KEOPS pentamérique chez l'homme responsable de la modification t<sup>6</sup>A.
  - Comment interagit la cinquième sous-unité GON7 avec les autres sous-unités ? Et quel est son rôle ? Comment s'assemble le complexe KEOPS ?
  - Peut-on, à partir de la structure tridimensionnelle du complexe KEOPS, comprendre les mutations conduisant au syndrome Galloway-Mowat ?
  - Comment interagit l'ARNt avec le complexe KEOPS ?
- (iii) Existe-t-il une cinquième sous-unité protéique au sein du complexe KEOPS chez les archées.
  - Peut-on identifier une cinquième sous-unité chez *Pyrococcus abyssi* ? Quelle est sa fonction ? Comment s'assemble le complexe KEOPS chez les archées ?

Nous verrons au chapitre I que nous nous sommes intéressés à l'étude du complexe bactérien en résolvant la structure tridimensionnelle du complexe ternaire TsaBDE de la bactérie *Thermotoga maritima*.

Dans le second chapitre nous avons, en collaboration avec l'équipe de Charles Lauhon, étudié l'effet d'un inhibiteur de la sous-unité catalytique TsaD de la bactérie d'*Escherichia coli* (chapitre II).

Nous nous sommes également intéressés au complexe KEOPS humain en collaboration avec l'équipe de Corine Antignac (de l'institut Imagine). La structure du complexe OSGEP/LAGE3/GON7 a été résolu par cristallographie et a permis de confirmer l'existence de GON7 dans le complexe KEOPS humain. L'effet de mutants de cette sous-unité au sein du complexe et de la protéine YrdC dans le syndrome GAMOS a été étudié. Ces résultats seront abordés au chapitre III.

Enfin dans le dernier chapitre, une étude, en collaboration avec Tamara Basta, Marie-Claire Daugeron et Violette Da Cunha de l'équipe de Jacques Oberto, ayant permis l'identification de la cinquième sous-unité du complexe KEOPS des archées sera développée.

Ce cinquième partenaire présenterait une fonction, au sein du complexe, similaire à celle de Gon7 chez les eucaryotes.



# Chapitre I

## I. Etude du rôle de TsaE dans la modification t<sup>6</sup>A chez les bactéries.

### I.1. Présentation de l'article.

Chez les bactéries, la modification t<sup>6</sup>A s'effectue donc en deux étapes successives catalysées par 4 protéines (TsaC1 ou TsaC2, TsaB, TsaD et TsaE). TsaC1 (ou TsaC2, l'orthologue possédant un domaine Sua5 C-terminal supplémentaire) forme l'intermédiaire TC-AMP à partir de la L-thréonine, de bicarbonate et d'ATP (figure 53 A). Lors d'une seconde étape, le complexe TsaBDE catalyse la thréonylcarbamoylation des ARNt substrats (figure 53 B). Si les protéines TsaBDE sont absentes, l'intermédiaire TC-AMP est labile et vite dégradé (figure 53 C) (Lauhon, 2012).

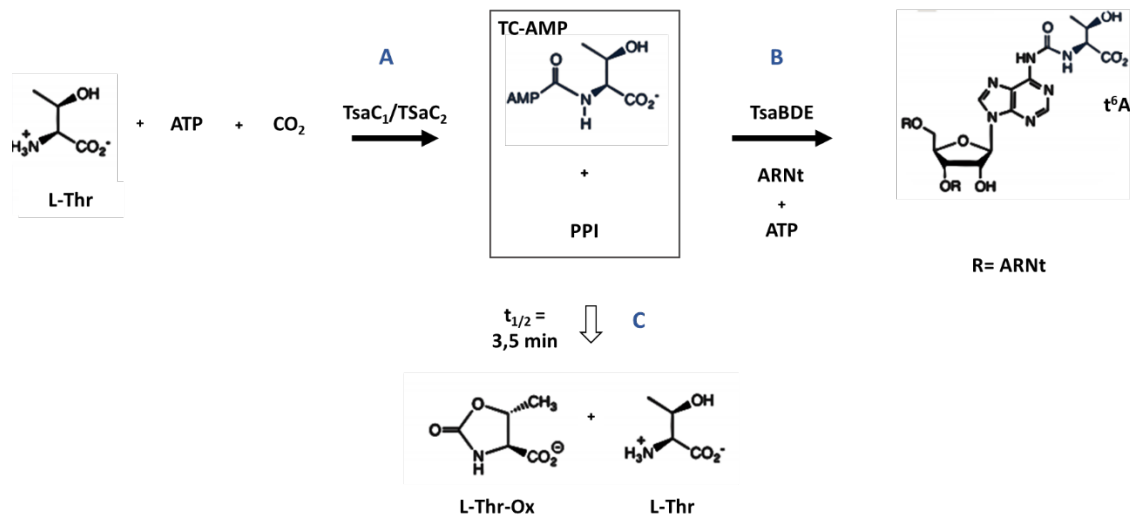


Figure 53 La réaction de modification t<sup>6</sup>A se déroule en deux étapes successives. A : Chez les procaryotes, TsaC1 ou TsaC2 utilise la L-thréonine, l'ATP et le carbonate pour synthétiser l'intermédiaire TC-AMP. B : Puis le complexe enzymatique TsaBDE catalyse la réaction de thréonylcarbamoylation de l'adénosine 37 des ARNt substrats. C : En absence de TsaBDE, le TC-AMP se dégrade en AMP et en L-thréonine ou en oxazolindione (L-thr-Ox). D'après Lauhon, 2012.

Les études de Wenhua Zhang dans l'équipe ont montré que l'activité ATPase de TsaE est absolument nécessaire pour l'activité t<sup>6</sup>A de TsaBD. TsaE appartient structurellement à la large famille des petites protéines G mais ne possède pas d'activité GTPase. Par contre elle a une très faible activité ATPase qui s'active en présence de TsaBD (K<sub>M</sub> = 1,4 mM et 0.644 mM respectivement en présence et en absence de TsaBD) (Teplyakov *et al.*, 2002; Zhang *et al.*, 2015).

L'ATP, ou des analogues non hydrolysables comme l'AMPPNP ou l'AMPCPP, sont indispensables pour la formation d'un complexe ternaire TsabDE. De façon surprenante, W. Zhang dans l'équipe avait également remarqué que le complexe TsabDE n'avait pas d'affinité pour l'ARNt (manuscrit de thèse). Afin de comprendre les interactions entre les différents partenaires pendant la modification du t<sup>6</sup>A, une étude structurale du complexe trimérique TsabDE de la bactérie *Thermotoga maritima* (le complexe de *Escherichia coli* n'a pas pu être cristallisé) a été entamée et les résultats sont présentés dans l'article " The structure of Tsab/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t<sup>6</sup>A tRNA-modification " publié dans Nucleic Acid Research en 2018. La structure du sous complexe de TsabD sans TsaE de la bactérie *Escherichia coli* (EcTsabD) (code PDB 4YDU) avait été résolue précédemment au sein du laboratoire et sera mise en perspective à la notre (Zhang *et al.*, 2015). Les résultats présentés seront ensuite comparés à ceux obtenus dans une étude biochimique et structurale qui a été menée en parallèle par le groupe de Manal Swarjo (Luthra *et al.*, 2018, 2019).

Des cristaux du complexe TsabDE de *Thermotoga maritima* (TmTsabDE) n'ont pu être obtenus qu'en présence d'AMPCPP. La structure quaternaire de TmTsabDE a été résolue initialement à l'aide d'un jeu de données à une résolution de 3,14 Å et dans un second temps un nouveau jeu de données a été enregistré à une résolution de 2,9 Å. Elles ont mis en évidence les points suivants :

- Le complexe TsabDE se dimérise et forme un hexamère de stœchiométrie 2 :2 :2 chez la bactérie *Thermotoga maritima*. La dimérisation est stabilisée par le feuillet de cinq brins β antiparallèles du domaine C-terminal de Tsab, formant ainsi un long feuillet antiparallèle (figure 5A de l'article). Différentes structures cristallographiques de Tsab de différents organismes ont mis préalablement en évidence l'homodimérisation de Tsab par le domaine N-terminal via une interface du type I (cf. § 4.5.1.1. page 61) et seul Tsab de *Thermotoga maritima*, à ce jour, présente un second type d'interaction (interface de type II), identique à celui observé dans l'hexamère TsabDE (Aydin *et al.*, 2011). Il est probable que cette dimérisation soit spécifique à certaines bactéries hyperthermophiles en réponse aux environnement plus extrêmes (figure 5 de l'article page 98).

- La présence de TsaE n'affecte pas de façon importante l'interface du complexe TsaBD ([figure 3 de l'article page 96](#)). Le mode d'interaction entre TsaB et TsaD est identique à celui du complexe TsaBD seul d'*Escherichia coli* (code PDB 4YDU) ([figure 2 de l'article page 94](#)).
- TsaE se positionne entre TsaB et TsaD. Elle engage majoritairement des interactions avec TsaD et à un degré moindre avec TsaB. L'AMPCPP et un ion  $Mg^{2+}$  sont fixés dans une poche à l'interface de TsaE et des hélices du domaine C-terminal de TsaD (les hélices  $\alpha_4$  et  $\alpha_6$ ). Ce site ne ressemble pas à un site de fixation de nucléotides classique et expliquerait pourquoi le nucléotide est nécessaire pour une interaction efficace entre TsaBD et TsaE.
- TsaE, utilise sa P-loop (GAGKTT) pour former des liaisons hydrogènes avec les phosphates du nucléotide. Le tryptophane  $W_{109}$  forme également une liaison hydrogène avec le phosphate  $\gamma$  de l'AMPCPP. Les acides aminés  $D_{158}$ ,  $E_{162}$ ,  $K_{166}$ ,  $K_{213}$ , localisés sur les hélices  $\alpha_4$  et  $\alpha_6$  de TsaD, forment aussi des liaisons hydrogènes avec le phosphate  $\gamma$  de l'AMPCPP. L'ion  $Mg^{2+}$  est, quant à lui, stabilisé par des liaisons hydrogènes formées avec les acides aminés  $T_{42}$ ,  $S_{61}$ ,  $T_{63}$ ,  $D_{80}$ ,  $E_{108}$  de TsaE et avec les acides aminés  $D_{158}$ ,  $E_{162}$ ,  $K_{166}$  de TsaD. Le site de fixation du nucléotide est ainsi partagé entre TsaE et TsaD ([figure S4 de l'article page 109](#)).
- L'interaction de TsaE avec TsaBD induit des changements de conformation importants de la sous unité TsaD ([figure 3 de l'article page 96](#)). En effet, en présence de TsaE, la position relative des deux domaines N et C-terminaux créant la cavité du site actif sont considérablement écartés, mettant en évidence une flexibilité intrinsèque de l'enzyme. En effet, si on superpose les deux conformations de TsaD sur leurs domaines C-terminaux respectifs, les hélices  $\alpha_2$  sont décalées de près de 11 Å, ouvrant ainsi la cavité du site actif. La structure de TmTsaBDE permet d'observer une conformation ouverte de la sous unité catalytique TmTsaD contrairement à celle fermée observée dans le complexe EcTsaBD. Également, la région des acides aminés 31 à 47 de TsaD en présence de TsaE est totalement désordonnée. La région correspondante dans le complexe de TsaBD d'*Escherichia coli* est repliée en une hélice  $\alpha$ . Dans cette conformation cette région rentrerait en clash stérique avec TsaE. Une seconde région désordonnée est



localisée entre les résidus 292 et 294 de TsaD en présence de TsaE (figure S5 de l'article page 110). A proximité de cette région, l'acide aminé D<sub>296</sub>, conservé, fournit le groupement carboxylate qui participe à la coordination de l'ADP et du métal Fe<sup>2+</sup> dans le site actif de TsaD en conformation fermée (TsaD de *Escherichia coli*). Le site actif, en conformation fermée, présente également un second métal Mg<sup>2+</sup> à proximité du phosphate β de l'ADP qui est coordonné par les acides aminés D<sub>11</sub> et E<sub>12</sub>. De façon surprenante, le mutant E<sub>12</sub>A abolit l'activité *in vitro* t<sup>6</sup>A mais l'ATP est toujours lié au site actif (Zhang *et al.*, 2015). Ce résultat suggère que l'ion Mg<sup>2+</sup> est nécessaire à la réaction de thréonylcarbamoylation. La fixation de TsaE induit des perturbations du site actif de la sous-unité catalytique TsaD qui affectent significativement la position des acides aminés (D<sub>296</sub>, H<sub>109</sub>, H<sub>113</sub>, D<sub>11</sub>, E<sub>12</sub>) impliqués dans la coordination des métaux et des phosphates des nucléotides.

- Des données cristallographiques à 2,9 Å de résolution du complexe TmTsaBDE ont permis de modéliser une seconde molécule d'ATP non hydrolysable (AMPCPP) dans le site actif de TsaD que la densité électronique de la précédente structure cristallographique ne nous avait pas permis d'identifier. Etant donné la meilleure qualité du second jeu de données et l'identification d'un nucléotide localisé au niveau du site actif de TsaD, un *corrigendum* a été publié dans Nucleic Acid Research en 2019 (cf. page 114). L'interaction de l'AMPCPP est stabilisée par les résidus du domaine C-terminal de TsaD (S<sub>134</sub>, G<sub>135</sub>, G<sub>161</sub>, D<sub>165</sub>, G<sub>178</sub>, D<sub>182</sub>, G<sub>264</sub>, N<sub>268</sub>). En revanche, aucune densité électronique ne permet de modéliser un métal dans le site actif. Les résidus impliqués dans la coordination du métal et des phosphates du nucléotide présentent toujours des réarrangements structuraux dans la conformation ouverte de la sous unité TsaD : (i) H<sub>109</sub> et H<sub>113</sub> sont décalés de plus de 3 Å, (ii) les acides aminés 290 à 295 à proximité de D<sub>296</sub> sont toujours désordonnés et (iii) D<sub>11</sub> et E<sub>12</sub> sont également décalés de 3 Å. Cette nouvelle structure ne contredit pas les conclusions du premier papier.
- Pour finir TsaE est majoritairement électro-négative (figure S6 de l'article page 111) et occupe l'entrée du site de TsaD. Il a été montré que la présence de TsaE empêche la fixation de l'ARNt (Luthra *et al.*, 2018). Sa position, ses propriétés

électronégatives et les données biochimiques expliquent l'incompatibilité de fixation de l'ARNt avec TsaE.

Une ouverture de l'enzyme TsaD est donc induite par la présence de TsaE. TsaE qui occupe l'entrée du site actif de TsaD empêche la fixation de l'ARNt dans le site actif de TsaD, suggérant que l'ouverture de TsaD, bien que compatible avec la coordination d'un nucléotide par les acides aminés du domaine C-terminal, est concomitante avec une inactivité catalytique (perturbation des acides aminés impliqués dans la fixation des métaux).



# The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t<sup>6</sup>A tRNA-modification

Sophia Missouri<sup>1</sup>, Stéphane Plancqueel<sup>1</sup>, Ines Li de la Sierra-Gallay<sup>1</sup>, Wenhua Zhang<sup>1</sup>, Dominique Liger<sup>1</sup>, Dominique Durand<sup>1</sup>, Raoudha Dammak<sup>1</sup>, Bruno Collinet<sup>1,2,\*</sup> and Herman van Tilbeurgh<sup>1,\*</sup>

<sup>1</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS UMR 9198, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif sur Yvette Cedex, France and <sup>2</sup>Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, UMR7590 CNRS/Sorbonne-Université, UPMC, Paris, France

Received March 12, 2018; Revised April 09, 2018; Editorial Decision April 11, 2018; Accepted April 17, 2018

## ABSTRACT

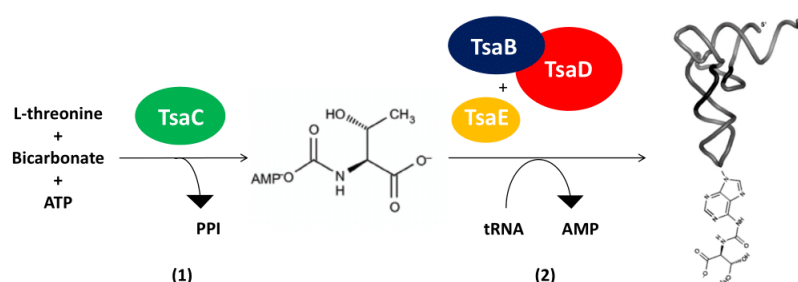
The universal N<sup>6</sup>-threonylcarbamoyladenine (t<sup>6</sup>A) modification at position A37 of ANN-decoding tRNAs is essential for translational fidelity. In bacteria the TsaC enzyme first synthesizes an L-threonylcarbamoyladenylate (TC-AMP) intermediate. In cooperation with TsaB and TsaE, TsaD then transfers the L-threonylcarbamoyl-moiety from TC-AMP onto tRNA. We determined the crystal structure of the TsaB–TsaE–TsaD (TsaBDE) complex of *Thermotoga maritima* in presence of a non-hydrolysable AMP-PCPP. TsaE is positioned at the entrance of the active site pocket of TsaD, contacting both the TsaB and TsaD subunits and prohibiting simultaneous tRNA binding. AMP-PCPP occupies the ATP binding site of TsaE and is sandwiched between TsaE and TsaD. Unexpectedly, the binding of TsaE partially denatures the active site of TsaD causing loss of its essential metal binding sites. TsaE interferes in a pre- or post-catalytic step and its binding to TsaD is regulated by ATP hydrolysis. This novel binding mode and activation mechanism of TsaE offers good opportunities for antimicrobial drug development.

## INTRODUCTION

Transfer RNAs (tRNAs) are subjected to considerable post-translational processing (1) and today >90 chemical modifications in tRNAs have been characterized (2). The N<sup>6</sup>-threonylcarbamoyladenine (t<sup>6</sup>A) modification at position 37 of ANN-decoding tRNAs is one of the few modifications that are found in the three domains of life (3–5).

Studies using oocytes demonstrated that A37 and U36 are strictly necessary for t<sup>6</sup>A formation (6). The t<sup>6</sup>A base in *Escherichia coli* tRNA<sup>Lys</sup> stacks with its adjacent A38 and forms a cross-strand stack with the first base of the codon on the mRNA, contributing to the translational fidelity (7). Mutations within the t<sup>6</sup>A modification pathway compromise anticodon–codon interaction, creating erroneous selection of start codons and aberrant frameshifts. They result in pleiotropic phenotypes and some mutations in human were recently associated with severe neurodegenerative and renal-neurological diseases (8,9). Although the modification is known for >40 years, the t<sup>6</sup>A biosynthesis enzymes and pathways have only recently been discovered (10–15). t<sup>6</sup>A biosynthesis is composed of two main steps (schematized in Figure 1): in the first, members of the universal TsaC/TsaC2 (YrdC/Sua5) protein family synthesize an unstable threonylcarbamoyladenylate (TC-AMP) intermediate from L-threonine, bicarbonate and adenosine triphosphate (ATP); the second step consists of the transfer of the threonylcarbamoyl (TC) moiety from TC-AMP onto A37 of substrate tRNA. In bacteria, the TC-transfer is carried out by the universal TsaD enzyme (YgjD in *E. coli*) assisted by two other proteins that are present in almost all bacteria: TsaB (YeaZ in *E. coli*) and TsaE (YjeE in *E. coli*). In archaea and eukaryotes, the TC-transfer reaction is carried out by the KEOPS multi-protein complex, composed of Kae1 (the TsaD orthologue), Bud32 (a small protein kinase), Cgi121 and Pcc1, complemented by a fifth protein (16–20) (Gon7 in yeast and C14ORF142 in human). Although TsaD/Kae1 is carrying the TC-transfer activity, the presence of the other subunits are mandatory for the reaction (exception made for Cgi121). In contrast, the yeast mitochondrial TsaD orthologue, Qri7, is capable of, if provided with TC-AMP, modi-

\*To whom correspondence should be addressed. Email: herman.van-tilbeurgh@u-psud.fr  
Correspondence may also be addressed to Bruno Collinet. Email: Bruno.collinet@i2bc.paris-saclay.fr  
Present address: Wenhua Zhang. School of Life Sciences, Lanzhou University, 730000 Lanzhou, China.



**Figure 1.** Reaction scheme for t<sup>6</sup>A synthesis in bacteria. TsaC uses L-threonine, bicarbonate and ATP for the synthesis of a very unstable TC-AMP intermediate. TsaBD and TsaE then collaborate to transfer the TC-moiety from TC-AMP onto the A37 of tRNA substrates.

fyng tRNA *in vitro* without assistance of ancillary proteins (12).

The biochemical function of the ancillary proteins involved in the t<sup>6</sup>A pathway is largely unknown, especially the contributions of the TsaE ATPase (bacteria) and the Bud32 kinase (archaea/eukaryotes) to the t<sup>6</sup>A synthesis reaction remains enigmatic. TsaB is a paralogue of TsaD that has lost its catalytic function and TsaE belongs to a large class of nucleotide hydrolases. TsaD physically interacts with TsaB and TsaE and their genes are often clustering (15,21–23). The crystal structures of the heterodimeric TsaBD (22,24) showed that TsaB and TsaD associate via hydrophobic packing of 2 two pairs of N-terminal helices into a helical bundle, mimicking the structure of homodimeric TsaB (25,26) and Qri7 (12). Although Pcc1 is unrelated to TsaB, a similar helical bundle is also found at the Kae1–Pcc1 heterodimer interface (27). Except for some intracellular or symbiotic organisms (e.g. *Mycoplasma*), TsaE is present in all bacteria (3). TsaE has an intrinsically weak ATPase activity that is strongly activated by its binding to the TsaBD complex regardless of the presence of other substrates for the biosynthesis of tRNA t<sup>6</sup>A (22–24). The need of ATP-hydrolysis for t<sup>6</sup>A biosynthesis in bacteria is presently unclear. The structure of TsaE in complex with ADP revealed the presence of a Walker A motif and two switch regions that are characteristic of P-loop GTP hydrolases (28,29). In presence of non-hydrolysable ATP analogues, *Ec*TsaB, *Ec*TsaD and *Ec*TsaE form a *Ec*TsaBDE complex with 1:1:1 stoichiometry (22). From small angle X-ray scattering (SAXS) measurements and biochemical data, it was proposed that *Ec*TsaE binds at the *Ec*TsaBD interface (22). In a very recent manuscript, Luthra *et al.* reported the characterization of the TsaBDE threonylcarbamoyl transfer complex of *T. maritima* (*Tm*TsaBDE) (23). Their SAXS analysis revealed that *Tm*TsaBDE forms a symmetric hexameric quaternary assembly in solution with 2:2:2 stoichiometry (*Tm*TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub>). This hexamer turned out to be a dimer of *Tm*TsaBDE heterotrimers, with *Tm*TsaB acting as a dimerization module, reminiscent of Pcc1 in eukaryotic and archaeal KEOPS (30,31). *Tm*TsaB and *Tm*TsaD were further shown to form a *Tm*TsaB<sub>2</sub>D<sub>2</sub> complex that was capable to bind one tRNA molecule, and accommodate a single TsaE subunit. Quantitative activity measurements demonstrated that *Tm*TsaB<sub>2</sub>D<sub>2</sub> alone can catalyse

only a single round of t<sup>6</sup>A synthesis (23). An ensemble of kinetic and binding experiments led to the hypothesis that *Tm*TsaE-catalyzed ATP hydrolysis occurs after the release of the t<sup>6</sup>A-modified tRNA (23).

The essential character of TsaE for bacterial survival, the presence of its encoding gene in all pathogenic bacterial genomes and the absence of human orthologues, makes TsaE an attractive antibacterial drug target (29,32). However, its role in the t<sup>6</sup>A process remains poorly understood, mainly because there is no high resolution structural information on its interactions with TsaBD. In this manuscript we present the X-ray structure of the *Tm*TsaBDE complex bound to AMPCPP at 3.1 Å resolution. *Tm*TsaE is bound at the interface of *Tm*TsaBD, making asymmetric contacts with *Tm*TsaB and *Tm*TsaD. The AMPCPP nucleotide was found sandwiched between *Tm*TsaE and *Tm*TsaD with its γ-phosphate group being coordinated by side chains from both subunits. Crystallographic symmetry generates the hexameric *Tm*TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub> complex that was proposed to be present in solution (23). Surprisingly, *Tm*TsaE partially denatures the active site of *Tm*TsaD and prevents simultaneous substrate tRNA binding, suggesting that *Tm*TsaE is required for a reaction step before or after the TC-transfer from TC-AMP to tRNA.

## MATERIALS AND METHODS

### Expression and purification of the *Tm*TsaBDE subunits

We prepared the *Tm*TsaBD binary complex and the *Tm*TsaE protein individually. A bicistronic expression vector (named BC14) aimed to co-express *Tm*TsaD and *Tm*TsaB-His was designed as described (33). All synthetic gene constructs were optimized according to *E. coli* codon usage using the EuGene software (34) (Supplementary Table S2) and obtained from Genscript (Piscataway, USA). The DNA sequences coding for *Tm*TsaD and *Tm*TsaB containing a 6His-tag at the C-terminus have been cloned into a pET21a backbone using the NdeI and XhoI restriction sites. The second vector (named BC15) aimed to express *Tm*TsaE-Nt-his was cloned into a pET24d backbone between NcoI and XhoI restriction sites. This plasmid, was used to express *Tm*TsaE with a N-terminal 6His-tag followed by a TEV proteolytic recognition site to remove the 6his-tag if necessary. *Tm*TsaD and *Tm*TsaB-His have been

co-expressed in *E. coli* Rosetta pLysS strain transformed with the BC14 plasmid and grown at 37°C in 2xYT liquid medium supplemented with ampicillin and chloramphenicol until mid-log phase. Expression was induced by adding 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside and the cells were incubated for three more hours at 37°C and then harvested by centrifugation, re-suspended in lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl), and stored at -20°C until purification. The same protocol was followed for the expression of *TmTsaE*, using plasmid BC15 and tetracycline antibiotic instead of ampicillin.

For purification, cells were thawed at 25°C and lysed by sonication (Branson Sonifier 250) on ice-cold water and centrifuged at 20 000 g for 30 min. Supernatant was applied onto NiIDA resin (Qiagen) previously equilibrated with Tris-HCl buffer (25 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM 2-mercaptoethanol). Resin was washed with 20 ml of the same buffer and the proteins were eluted using three fractions of 4 ml of the previous buffer supplemented with 100, 200 and 400 mM imidazole. Fractions containing the proteins of interest were pooled and concentrated by ultra-filtration and loaded on superdex 75 (GE-Healthcare) equilibrated in HEPES buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5 mM 2-mercaptoethanol).

Finally, fractions containing the proteins of interest were pooled and concentrated to 50 mg.ml<sup>-1</sup> or 60 mg.ml<sup>-1</sup> for *TmTsaE* and *TmTsaBD* respectively. Purity was checked by SDS-PAGE.

#### Crystal structure determination

Purified *TmTsaBD* and *TmTsaE* were mixed at a 1:1:1 ratio of the three subunits up to a final concentration of 12 mg.ml<sup>-1</sup> (0.148 mM) in HEPES buffer 20 mM pH 7.5, 200 mM NaCl, 5 mM 2-mercaptoethanol supplemented with 3 mM AMPCPP and 6 mM MgCl<sub>2</sub>. Crystals of *TmTsaBDE* were obtained at 293 K using the sitting-drop vapour diffusion method by mixing 100 nL of a protein solution with 100 nL of 13% PEG 8000, 0.1 M Imidazole pH 7.35. Crystals were cryoprotected with reservoir solution supplemented with 30% glycerol and 3mM AMPCPP, flash-frozen and stored in liquid nitrogen for data collection. X-ray diffraction data collection was carried out on beamline Proxima2 at the SOLEIL Synchrotron (Saint-Aubin, France) at 100K. Data were processed, integrated and scaled with the XDS program package (35). *TmTsaBDE* crystals belonged to space group  $P2_12_12_1$  with unit cell parameters of  $a = 84.31$  Å,  $b = 113.94$  Å,  $c = 177.62$  Å. The structure of *TmTsaBDE* was solved by molecular replacement using the PHASER module (36) implemented in the CCP4 software package (37). Search models for *TmTsaB* and *TmTsaD* were obtained by the MODELLER program (38) using the crystal structure of the *EcTsaBD* heterodimer (PDB ID: 4YDU) and *TmTsaE* was modelled from the *HiTsaE* structure (PDB ID: 1FL9). The initial structure was refined using the PHENIX program (39) and completed by interactive and manual model building using COOT (40). Two copies of the *TmTsaBDE* trimer were present in the asymmetric unit. AMP-CPP molecules and MgCl<sub>2</sub> ions could be modeled into the residual Fo-Fc electron density contoured at

3.0 $\sigma$ . Data collection and refinement statistics are gathered in Table 1.

#### tRNA docking

The *EcTsaDB*-tRNA complex model was generated by rigidly docking a tRNA molecule onto the *EcTsaBD* complex using the HADDOCK Server (version 2.2) (41). The tRNA structure was extracted from the crystal structure of threonyl-tRNA synthetase-tRNA complex (PDB : 1QF6) (42) in which the *E. coli* tRNA<sup>Thr</sup> UGC contains a m<sup>6</sup>t<sup>6</sup>A at position 37 in the anticodon loop. For docking, the m<sup>6</sup>t<sup>6</sup>-moiety and other modifications were removed from the *E. coli* tRNA<sup>Thr</sup> UGC. The docking protocols and parameters follow the default settings for protein-RNA complexes in HADDOCK. As for the direct contact sites, critically, His111 and His115 in *EcTsaD* were specified as active site residues in the receptor molecule whereas A37 was specified as substrate in tRNA. By default, the active sites-surrounding residues were specified as passive. Resulting models were ranked based on scoring performance and clustered. The first model from the top-ranked cluster was chosen to represent the structures.

#### Sequence alignment and structure comparison

The sequence conservation was analysed using the program CONSURF (43) with *TmTsaBDE* as the query. The frequency of the occurrence of amino acids on the consensus sequence was produced by WebLogo (44). Structural comparison was performed with the PDBFold web server (45). All graphic structure representations were produced with PYMOL (46). The electrostatic potential representation were rendered and produced by PDB2PQR (47) and APBS in the Chimera visualization system (48).

## RESULTS AND DISCUSSION

#### Overall structure of the *TmTsaBDE* complex

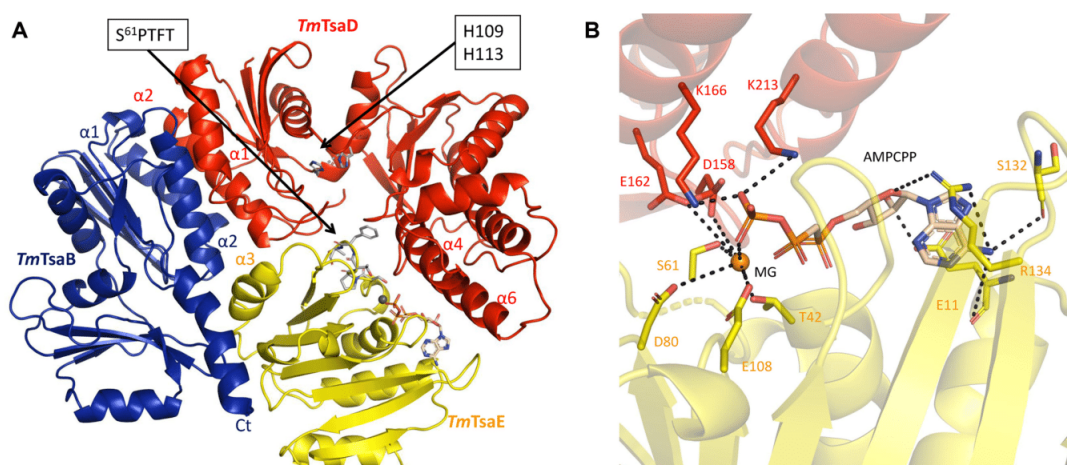
We obtained diffracting crystals of *TmTsaBDE* only in the presence of AMPCPP and the structure was solved at 3.14 Å resolution. The asymmetric unit of the crystals contains two copies of a *TmTsaBDE* trimer, whose structures are very similar (all RMSD values for superimposed structures are gathered in Supplementary Table S1).

The overall structures of *TmTsaB*, *TmTsaD* and *TmTsaE* in the complex are the same as those of the individual proteins (*TmTsaB*) or orthologues (*H. influenzae* (*Hi*)- and *B. subtilis* (*Bs*)/*TsaE* and *St* (*S. typhimurium*), *Ec* (*E. coli*)/*TsaD* and *Kae1*) (Supplementary Table S1) (28,29) (49). Sequence alignments with superposed secondary structure assignments of a few *TsaB*, *TsaD* and *TsaE* orthologues are represented in Supplementary Figure S1. As illustrated in Figure 2a, *TmTsaB* and *TmTsaD* embrace *TmTsaE* to form a compact ternary complex, in agreement with SAXS measurements on *Ec*- and *TmTsaBDE* in solution (22,23). *TmTsaE* is held in a grip between the helical insertion of the C-terminal domain of *TmTsaD* on one side and *TmTsaB* on the other. *TmTsaE* establishes two main contacts with *TmTsaBD*. The first involves the  $\alpha$ 3-helix (residues 90–100)

**Table 1.** Crystallographic data-collection and refinement statistics of TmTsaBDE

X-ray source	PROXIMA 2
Wavelength (Å)	0.9801
Temperature (K)	100
Unit-cell parameters (Å, °)	$a = 84.31, b = 113.94, c = 177.62, \alpha = \beta = \gamma = 90.0$
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution limits <sup>a</sup> (Å)	48.44 – 3.14 (3.33–3.14)
Number of observations <sup>a</sup>	138146 (21963)
Number of unique reflections	30297 (4627)
R-meas <sup>a</sup> (%)	19.1 (149.7)
Completeness <sup>a</sup> (%)	99.1 (95.2)
1/σ (I) <sup>a</sup>	7.1 (1.03)
CC (1/2) <sup>a</sup> (%)	99.2 (43.9)
Number of non-hydrogen atoms (proteins/ligands)	10418/77
R/R <sub>free</sub> (%)	22.58/29.37
R.M.S.D. Bonds (Å)/angles (°)	0.010/1.258
Average B-factor (proteins/ligands)	80.67/80.91

<sup>a</sup>Values in parentheses refer to the highest resolution shell.



**Figure 2.** Structure of *TmTsaBDE*. (A) Ribbon diagram of *TmTsaBDE* (*TmTsaB*: blue, *TmTsaD*: red, *TmTsaE*: yellow). The main structural elements involved in protein interfaces are indicated following the colour codes of the individual proteins (*TmTsaBD*:  $\alpha 1$  and  $\alpha 2$  from *TmTsaB* and *TmTsaD*; *TmTsaDE*:  $\alpha 4$  and  $\alpha 6$  from *TmTsaD* and loops from *TmTsaE*). The two histidines (His109 and His113) in the active site of *TmTsaD* responsible for metal binding and the S<sup>61</sup>PTFT motif of *TmTsaE* are indicated by an arrow. AMPCPP at the interface between *TmTsaD* and *TmTsaE* is represented by sticks. The Mg-ion bound to AMPCPP is represented as a grey sphere. (B) AMPCPP binding site. *TmTsaE* in yellow and *TmTsaD* in red, AMPCPP is in sticks, Mg-ion as orange sphere. Side chains in H-bond interaction with the AMPCPP  $\gamma$ P group are represented as sticks.

of *TmTsaE* that binds to the N-terminal end of the four-helical bundle of the TsaBD interface. The  $\alpha 3$ -helix is partially disordered in *BsTsaE* and forms two short perpendicularly oriented helices in *HiTsaE* (illustrated in Supplementary Figure S2). This helix and the connection between  $\beta 5$  and  $\beta 6$  (residues 115–120) contact both N- and C-terminal domains of *TmTsaB*. The second main contact between *TmTsaE* and *TmTsaBD* is provided by the loops emanating from the C-terminal end of the parallel section of the *TmTsaE*  $\beta$ -sheet and the helical insertion in the C-terminal domain of *TmTsaD*, engulfing the AMPCPP ligand (Figure 2A and Supplementary Figure S8). The totally conserved *TmTsaE* peptide S<sup>61</sup>PT(F/Y)T, dips into the active site of *TmTsaD*, containing two essential histidines (His109 and His113, highlighted in Figure 2A). This peptide contacts  $\alpha 4$  and the hairpin connection between  $\beta 6$

and  $\beta 7$  of *TmTsaD*, forming H-bonds between well conserved residues (Supplementary Figure S1). The loop between  $\beta 6$  and  $\beta 7$  sandwiches the  $\gamma$ P-group of ATP in its complex with *SiTsaBD* (24). Since the S<sup>61</sup>PT(F/Y)T motif is also involved in a H-bond network with the  $\gamma$ P-group of AMPCPP, this loop could be a communicator between the compound bound in the active site of TsaD and the ATP binding site on TsaE (see discussion in mechanism section).

The *TmTsaB* C-terminus is elongated and disordered beyond residue 194 in one copy of the asymmetric unit, while in the other it forms a helix comprised between residues 190 and 203. We showed that the C-terminal peptide of *EcTsaB* is not required for formation of the binary *EcTsaBD* complex (22), but it might contribute to the stability of *EcTsaBDE*. The regions involved in these contacts

are very poorly conserved, suggesting that they are not a major determinant for TsaE binding.

Although TsaB and TsaD have the capacity *in vitro* of forming homodimers whose quaternary structures are very similar to the TsaBD heterodimer, individually they do not interact with TsaE (22,23). Although there is no structure of the TsaD homodimer available, it likely resembles its mitochondrial orthologue Qri7, whose homodimer assembly is very similar to that of TsaB (12). *TmTsaE* is engaged in a totally asymmetric interaction with *TmTsaB* and *TmTsaD*, explaining why TsaE binds neither TsaB nor TsaD homodimers (22).

#### AMPCPP binds at the interface between *TmTsaD* and *TmTsaE*

After initial refinement of the molecular replacement solution, a consistent cloud of residual electron density was identified at the expected nucleotide-binding site of *TmTsaE*. As shown in Supplementary Figure S4, this density could be fitted by AMPCPP, whose presence was mandatory for obtaining diffracting crystals. Residual electron density was also present for a  $Mg^{2+}$  ion, coordinated by the  $\beta$ - and  $\gamma$ -phosphate groups of AMPCPP, Thr42<sup>O $\gamma$</sup>  and Glu108<sup>COO<sup>-</sup></sup>. AMPCPP is located at the C-terminal edge of the parallel strands of the central  $\beta$ -sheet. Most remarkably, AMPCPP is sandwiched between *TmTsaE* and the helical insertion of the C-terminal domain of *TmTsaD* that partially seals off its polyphosphate. The AMPCPP nucleotide adopts an unusual syn-conformation and overlaps with ADP in complex with *HtTsaE* and *BsTsaE* (28,29) (Supplementary Figure S2). *TmTsaE* binds AMPCPP in a manner globally observed in other P-loop proteins (Figure 2B). The N<sup>6</sup> atom of adenine hydrogen bonds to main chain atoms of Ser132 and Glu11 and the base-ring is stacked between the side chains of Glu11 and Arg134, which also hydrogen bonds with the ribose O<sup>4</sup>. The ribose O<sup>2</sup> hydrogen bonds with the Glu11 carboxylate and the ribose O<sup>3</sup> establishes a H-bond with the main chain carbonyl of *TmTsaD* Phe116. The phosphate moieties of AMPCPP make classic hydrogen contacts with the *TmTsaE* P-loop (G<sup>38</sup>AGKTT) at the N-terminus of  $\alpha$ 2. *EcTsaE* does not bind to *EcTsaBD* in presence of ADP (22) and the affinity of *TmTsaE* for *TmTsaBD* drops fourfold in absence of ATP (23). It was also shown that the basal low ATPase activity of *EcTsaE* is boosted by the presence of *EcTsaBD* (22). These observations can be explained by the fact that the position of the AMPCPP  $\gamma$ P-group is stabilized by salt bridge formation with the conserved Lys166 and Lys213 from *TmTsaD* (Figure 2B). Lys166<sup>N $\epsilon$</sup>  also hydrogen-bonds with Thr63<sup>O $\gamma$ H</sup> that is part of the totally conserved SPT motif in contact with the *TmTsaD* active site. The carboxylate group of the totally conserved *TmTsaE* Asp80 is at  $\sim 5$  Å from the  $\gamma$ P of AMPCPP, ideally positioned to act as a general base for ATP hydrolysis. It was noticed that this Asp is part of a conserved HxD(L/V)YR motif that was reminiscent of Hanks type kinases (29,51). Mutation of the homologous Asp80 in *BsTsaE* annihilated its (auto)phosphorylation activity.

Different effects of ATP nucleotides on the binding of TsaE to TsaBD have been observed for different organisms. *TmTsaE* is able to bind *TmTsaBD* in absence of added nu-

cleotides with an affinity constant of 1.3  $\mu$ M, which decreases about 4-fold in presence of ATP, while ADP has no effect (23). Two active-site mutants defective in ATP binding, *TmTsaE*<sup>T42A</sup> and *TmTsaE*<sup>E108A</sup>, preserved their binding to *TmTsaBD* (23). *EcTsaE* does not bind to *EcTsaBD* in absence of ATP but binds with 1:1:1 stoichiometry and a 0.6  $\mu$ M  $K_d$  in presence of  $Mg^{2+}$  and AMPPNP (22). *EcTsaE*<sup>T43A</sup> and *EcTsaE*<sup>E108A</sup> mutants (also defective in binding to ATP or ADP) did not interact with *EcTsaBD* (22). The strength and ATP dependence of the binding of TsaE to TsaDB probably is species dependent and variations might exist between for instance hyperthermophilic and mesophilic bacteria.

Comparison of TsaE with small GTPases suggested that Tyr82 and Trp109 are part of putative switch regions (Supplementary Figure S2) (22,50). Trp109<sup>N $\epsilon$ 1</sup> (switch I) hydrogen-bonds with the  $\gamma$ P of AMPCPP and Tyr82 (switch II) stacks against TsaD Lys166. *EcTsaE*<sup>Y82A</sup> and *EcTsaE*<sup>W109A</sup> had  $\sim 7$ -fold diminished affinity for ATP $\gamma$ S but *EcTsaE*<sup>W109A</sup> kept WT binding capacity to *EcTsaBD* in presence of AMPPNP. Although the *EcTsaE*<sup>W109A</sup> mutant had diminished ATPase activity it was active in t<sup>6</sup>A biosynthesis. It should be noticed that the conformation of switch regions in small GTPases is usually strongly influenced by the nature of the bound nucleotide (GTP versus GDP) (50), but this is not the case for *TmTsaE*.

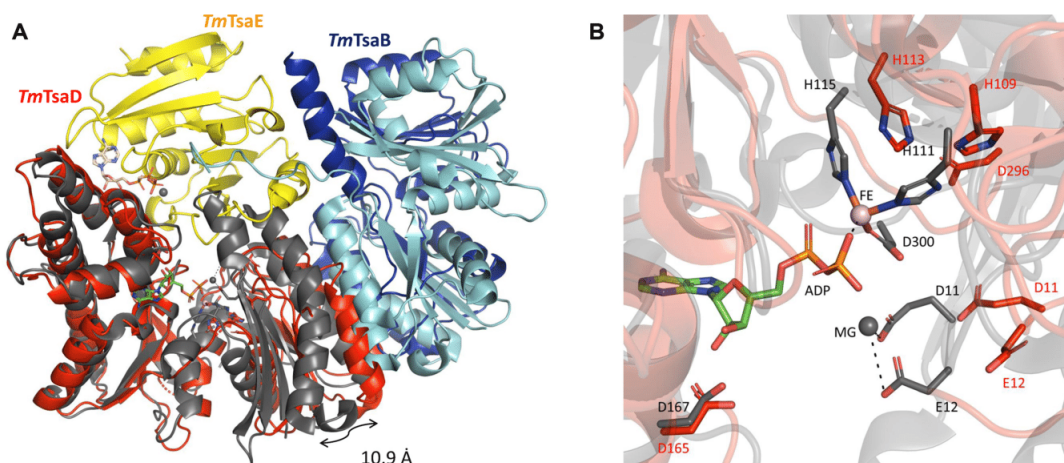
To conclude, AMPCPP binding in *TmTsaBDE* is conferred by residues both from *TmTsaD* and *TmTsaE*. Residues from *TmTsaD* complete the coordination of the  $\gamma$ P group of ATP and hence are likely responsible for the activation of the hydrolytic activity of *TmTsaE* upon binding to *TmTsaBD*. Our structure clearly explains in general why ATP is needed for (or increases) the binding of TsaE to TsaBD and why this complex boosts the ATPase activity of TsaE.

#### *TmTsaE* binding induces an inactive conformation of *TmTsaD*

Global superposition of *Tm*- and *EcTsaD* indicates that the relative positions of the N- and C-terminal domains changes considerably upon formation of the *TmTsaBDE* complex (Figure 3A). The  $\alpha$ 3 and  $\alpha$ 9 helices of *TmTsaD* at the bottom of the active site shift their positions by  $\sim 4$  Å forming the hinge of a rotation that separates the C- and N-terminal domains and opens up the active site cleft. At its periphery the N-terminal domain of *TmTsaD* moves by  $\sim 11$  Å. These observations corroborate reported SAXS data obtained from *TmTsaB<sub>2</sub>D<sub>2</sub>* in solution, which were best fitted by an equilibrium population between an open and closed form of *TmTsaD* (23). The distance between the two *TmTsaD* units in the 'open form' of *TmTsaB<sub>2</sub>D<sub>2</sub>* was estimated to be 85 Å, a value that agrees with the present crystal structure of *TmTsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub>*. The closed form of *TmTsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub>* likely corresponds with the structure of TsaD as observed in the *St*- and *EcTsaBD* binary complexes.

Apart from pushing the two domains of *TmTsaD* apart, *TmTsaE* binding partially 'melts' its active site pocket. Electron density was missing for residues 31–47 of *TmTsaD* for both copies in the asymmetric unit, suggesting they became





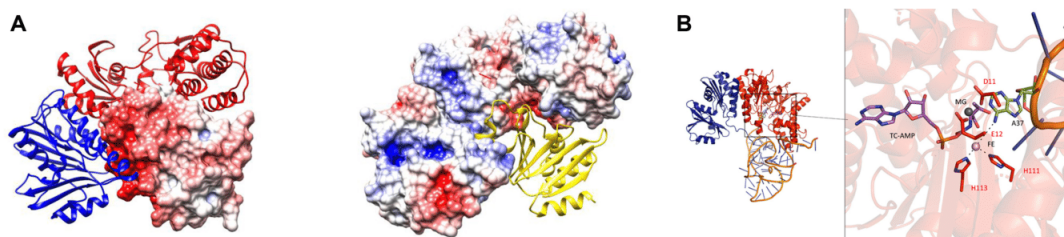
**Figure 3.** Structure comparison between *EcTsaBD* and *TmTsaBDE*. (A) Superimposition of *EcTsaBD* (PDB ID 4YDU; *EcTsaB* in cyan and *EcTsaD* in gray) and *TmTsaBDE* (same colour code as for Figure 2A). The C-terminal domains of TsaD were superimposed. The ADP (bound to *EcTsaBD*) and the AMPNP bound to *TmTsaE* are shown as sticks. As indicated the N-terminal domain position of TsaD shifts by ~11 Å between *EcTsaBD* and *TmTsaBDE*. (B) Zoom into the TsaD active site. Superposition of *EcTsaBD* (gray shadow) and *TmTsaBDE* (red shadow). Fe- and Mg-binding side chains are shown using the same color code. Fe- and Mg-ions are in pink and grey spheres respectively and ADP is in sticks.

disordered. This peptide region has a well-defined conformation in binary TsaBD complexes: residues 31–39 form a helix ( $\alpha 1$ ) that is collinear with strand  $\beta 3$  and residues 41–47 form the two first helical turns of  $\alpha 2$ . The superimposition of *EcTsaBD* and *TmTsaBDE* onto the TsaD N-terminal domains (Supplementary Figure S5), shows that the structure of this peptide as present in *EcTsaBD* would create an important steric clash with *TmTsaE*. A second region of *TmTsaD* that becomes disordered in *TmTsaBDE*, is contained between residues 292 and 294. The corresponding peptide in *EcTsaBD* provides a carboxylate ligand ( $^{Ec}Asp300$ ) to the Fe-ion. The induction of partial disorder and conformational changes into the *TmTsaD* active site by the presence of *TmTsaE* significantly affects the positions of many of the crucial residues involved in AMP and metal binding. Figure 3b illustrates ADP bound in the active site of *EcTsaBD* (22,24) superposed onto the active site of *TmTsaD*. The  $\alpha$ - and  $\beta$ -phosphates are ligands of a metal ion (Fe in *EcTsaBD*), which is further coordinated by His111 ( $^{Tm}His109$ ), His115 ( $^{Tm}His113$ ) and Asp300 ( $^{Tm}Asp296$ ). A Mg-ion is positioned at 4 Å away from the  $\alpha$ - and  $\beta$ -phosphates of ADP and is coordinated by the Asp11 ( $^{Tm}Asp11$ ) and Glu12 ( $^{Tm}Glu12$ ) carboxylates. The opening of the active site pocket and the induced disorder render the metal binding sites of *TmTsaD* ineffective (Figure 3B): (i) His109 and His113 shifted over >3 Å and are no longer in a configuration capable of binding a metal ion; (ii) the peptide containing the third metal-coordinating residue Asp296 became disordered; (iii) the two Mg-coordinating residues (Asp11 and Glu12) also moved by ~3 Å away from the polyphosphate moiety position of the nucleotide. Mutations of the homologous histidines in Kae1, renders yeast strains hardly viable and abolishes the  $t^6A$ -synthesizing activity (27,31). The *EcTsaD*<sup>E12A</sup> mutant is totally inactive, although it is still capable of bind-

ing *EcTsaE* and *EcTsaB* (22), suggesting that the integrity of the Mg-binding site of TsaD is essential for activity but not for TsaE binding. We conclude that the binding of *TmTsaE* opens and partially melts the active site of *TmTsaD*, resulting in an inactive conformation of *TmTsaD*.

#### TmTsaE blocks the active site entrance for a tRNA substrate

The biochemical characterization of tRNA binding to *TmTsaBDE* was recently reported (23). Both *TmTsaB<sub>2</sub>D<sub>2</sub>* and *TmTsaB<sub>2</sub>D<sub>2</sub>E* are able to bind tRNA with  $\mu M$  affinity, but no tRNA binding was observed for *TmTsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub>*. These data corroborated our observations that tRNA did not bind to *EcTsaBDE* (22). Our structural data explain why TsaE and tRNA are competing for binding to TsaBD: the position of TsaE prevents the access of the anti-codon loop of the bulky tRNA substrate to the TsaD carbamoylation site (centered on the Fe/Mg cluster). The surface projection of the electrostatic potential (Figure 4A) shows that *TmTsaBD* contains an extensive positively charged groove surrounding the active site pocket, obviously a good candidate for tRNA binding. We modeled a complex of a rigid tRNA substrate with *EcTsaBD* using the HADDOCK software (41). The tRNA in the HADDOCK model extensively interacts with the positive surface patch present on TsaBD (Figure 4B) and most importantly sterically overlaps with *TmTsaE* in the *TmTsaBDE* complex. Electrostatic surface calculation of *TmTsaE* reveals a negatively charged surface patch centered on the  $\alpha 3$  helix that is positioned onto a positively charged patch at the *TmTsaBD* interface. Although the residues at the interface are not completely conserved, the electronegative character of this region is maintained in the *Ec*-, *Bs*- and *TmTsaE* homologues (Supplementary Figure S6). TsaE can be considered as an electrostatic competitor of the tRNA substrate.



**Figure 4.** *TmTsaE* and tRNA binding sites overlap. (A) Left panel: the electrostatic potential surface of *TmTsaE* and ribbon presentation of *TmTsaBD*; right panel: the electrostatic potential surface of *TmTsaBD* and ribbon presentation of *TmTsaE* (negatively and positively charged surfaces are in red and blue respectively). Ribbon presentation uses same colour code as Figure 2A. (B) Left panel: the HADDOCK model of *EcTsaBD* (in ribbon) bound to tRNA; right panel: view of the active site of the modelled *EcTsaBD*-tRNA-TC-AMP complex. TC-AMP is in sticks. tRNA-A37 is in green sticks and some of the metal ligands are in red sticks. Fe-ion in pink and Mg-ion in gray sphere.

The competition between tRNA and TsaE also explains why the strong ATPase activity upon binding of TsaE to TsaBD is observed independently of tRNA (22,23). In agreement with conclusions from biochemical data (22,23), the *TmTsaBDE* structure definitely confirms that TsaE must operate in a reaction step that is unrelated to the proper carbamoyl-transfer from TC-AMP onto tRNA A37.

#### Quaternary structure of *TmTsaBDE*

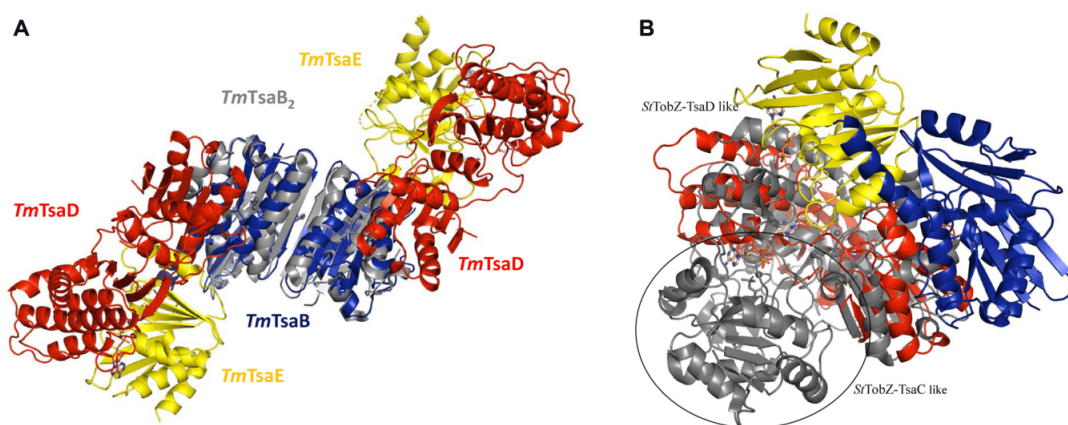
The structure of the *TmTsaB* dimer was previously determined by a structural genomics consortium (25). The crystal packing in that study suggested two possible dimerization modes for *TmTsaB*. In the first, the  $\beta 3$  strands of the two N-terminal domains align to form an antiparallel extended  $\beta$ -sheet. In the second dimer type, the  $\alpha 1$  and  $\alpha 2$  helices of the N-terminal domains of each subunit are packed into a four-helical bundle, closely mimicking the helical arrangement observed at the TsaBD interface. Although the latter dimerization has a smaller contact area mode, it was found exclusively in all other TsaB homologue structures (26,52). SAXS analysis of *TmTsaBDE* demonstrated it forms a heterohexamer in solution consisting of two copies of each subunit (TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub>) (23). The two copies of the *TmTsaBDE* trimer in the asymmetric unit do not associate into a hexamer. However, *TmTsaB* forms a crystal symmetry generated dimer that involves the  $\beta 3$  strands of the two N-terminal domains. Using *TmTsaB* as the dimerizing subunit, we generated a TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub> heterohexamer (Figure 5A) that is very close to the one proposed from SAXS data (23). We confirm that the quaternary structure of the *TmTsaBDE* complex is different from trimeric *Ec*- and *SrTsaBDE* (22–24). Structure imposition reveals that *Ec*- and *SrTsaB* have a 19 amino acid insertion in the C-terminal domain compared to *TmTsaB*, that forms an extra strand and a long helical connection. This insertion is positioned close to the *TmTsaB* interface region observed in the structures of *TmTsaB* and *TmTsaBDE*. Superimposition of the *Ec*- and *SrTsaB* monomer structures onto the *TmTsaB* dimer creates a considerable steric clash between the inserted regions of the two subunits (not shown). Therefore, *Ec*- and *SrTsaB* would not be able to adopt the *TmTsaB* type dimer structure, explaining why *Ec*- and *SrTsaBDE* are trimers and not hexamers in solution. This is probably the case for the majority of species, since most TsaB sequences have this C-

terminal insertion and we expect that the trimeric TsaBDE complex is the standard configuration.

As discussed above, bacterial TsaBDE adopts different quaternary assemblies according to species. This type of heterogeneity was also observed in the archaeal and eukaryotic systems. Archaeal KEOPS is a dimer of heterotetramers (Pcc1-Kae1-Bud32-Cgi121) (30) while in eukaryotes, a fifth KEOPS subunit (Gon7 in yeast) prevents Pcc1 dimerization and the functional unit is a heteropentamer. In humans, C14ORF142 plays a similar role as Gon7 in yeast and prevents the dimerization of the LAGE3 (the Pcc1 orthologue) subunit (19). In absence of C14ORF142, LAGE3-OSGEP-TPRK-TPRB forms a dimer of heterotetramers.

#### Activation mechanism of the t<sup>6</sup>A activity by TsaE

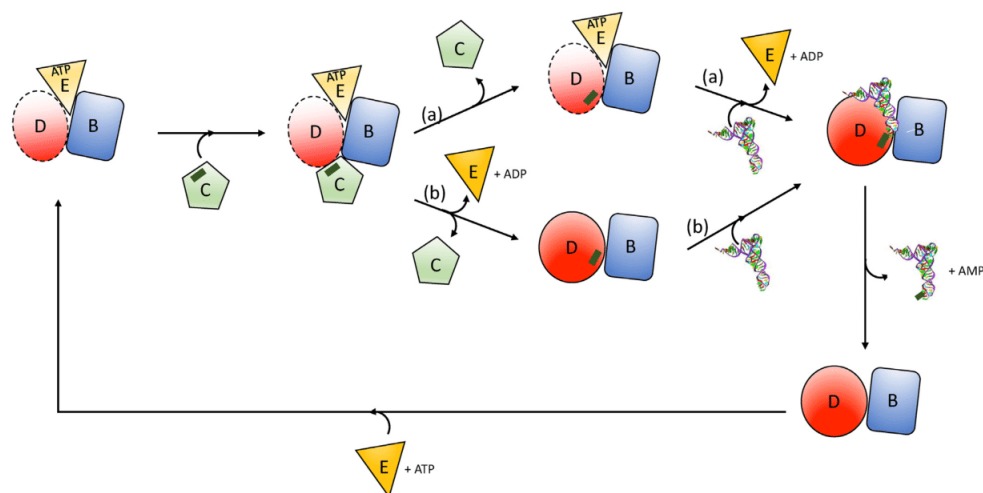
Kinetic experiments suggested that, in absence of *TmTsaE*, *TmTsaB<sub>2</sub>D<sub>2</sub>* is able to catalyse a single turnover. However, the presence of *TmTsaE* is required to attain multiple cycles of t<sup>6</sup>A activity (23). *TmTsaE* mutants that have lost their ATPase or *TmTsaBD* binding activities are incapable of sustaining multiple-round catalysis of the *TmTsaD* enzyme. On the other hand, the present structure of *TmTsaBDE*, supported by recent biochemical data (23), proves that *TmTsaE* and tRNA-substrate cannot simultaneously bind to *TmTsaBD*. *TmTsaE* must therefore carry out an essential step on the enzymatic pathway occurring before or after the transfer of the threonylcarbamoyl-moiety from TC-AMP onto A37. Luthra *et al.* suggested that after synthesis of t<sup>6</sup>A-tRNA, *TmTsaD* might be left in an inactive state and *TmTsaE*-mediated ATP hydrolysis might ‘reset’ *TmTsaD* for the next reaction cycle (23). Obviously, to provide access for the tRNA substrate to the active site of *TmTsaD*, *TmTsaE* must bind and dissociate from *TmTsaBD* during the reaction cycle, a process that is likely regulated by its ATPase activity. The affinity of *TmTsaE* for *TmTsaBD* is in the micromolar range, and diminishes in absence of ATP (23). For *EcTsaE* we did not observe binding to *EcTsaBD* without ATP (22). Apart from playing a regulatory role for tRNA binding, *TmTsaE* also affects the active site conformation of *TmTsaD*. *TmTsaE* perturbs the integrity of the *TmTsaD* metal-binding sites which are essential for activity and hence induces an inactive state of the *TmTsaD* enzyme. It is not obvious why *TmTsaD* should go through a ‘denat-



**Figure 5.** Quaternary structure of *TmTsaBDE*. (A) Ribbon representation of the *TmTsaB*<sub>2</sub>*D*<sub>2</sub>*E*<sub>2</sub> hexamer, a dimer of *TmTsaBDE* trimers, that was extracted from the crystal packing. This assembly agrees with the hexamer which was proposed in solution, based on SAXS experiments (23) (*TmTsaB*: blue, *TmTsaD*: red, *TmTsaE*: yellow). *TmTsaB* forms the central dimerization unit, corresponding to one of the dimer configurations in the crystal structure of *TmTsaB* alone (superposed in grey, PDB ID 2A6A), but was not observed in crystal structures of homologous TsaB structures. (B) Superimposition of *S/TobZ* (PDB ID 3VEN; *S/TobZ*-TsaD like in grey and *S/TobZ*-YrdC like in black) and *TmTsaBDE* (same color code as for Figure 2A).

uration' step during the reaction cycle. This kind of step could possibly be related to the release of reaction products or intermediates. However, affinities of TsaD for nucleotides and tRNA are moderate (22–24) and there is no evidence for the formation of stable intermediate/product complexes that would need assistance to release them from the enzyme. TsaE must therefore be involved in another step of the reaction cycle, for instance binding of the TC-AMP substrate. Inspection of *TmTsaBDE* shows that TC-AMP has full access to the active site. Since the lability of the TC-AMP intermediate hampers its use in structural studies, we made a model of the *EcTsaBD*-TC-AMP complex by superposing its AMP-moiety onto the *EcTsaBD*-ADP complex (22) (Figure 4b). This model places the carbamoyl group of TC-AMP close to the Fe-ion. We then compared this model with substrate complexes of TobZ, which is a fusion protein of an TsaC- and a TsaD-like module (53). TobZ from *Streptoalloteichus tenebrarius* is responsible for the O-carbamoylation of tobramycin (an aminoglycoside related to kanamycin) to form nebramycin 5'. Its TsaC-module catalyses the formation of a carbamoyl-AMP intermediate, whose carbamoyl moiety is subsequently transferred to tobramycin by the TsaC-module. It was noticed that the active sites of TsaD and TobZ-TsaD are very similar (53). In TobZ complexes with carbamoyl-AMP and the acceptor tobramycin, the carbamoyl group is liganded to the Fe-ion and the tobramycin-OH group approached the Fe-bound phosphate (53). Superimposition of TobZ bound to carbamoyl-AMP and the acceptor tobramycin with our model of the TsaD/TC-AMP complex showed a perfect overlap between the carbamoyl-AMP moieties. The Fe-cluster can hence be proposed as a carbamoylation centre for all t<sup>6</sup>A TC-transfer systems. Our HADDOCK model of the tRNA-TsaBD complex positions the N<sup>6</sup>-amino group of the A37 anticodon base close to the carbamoyl group of TC-AMP and is well positioned for a nucleophilic attack of the TC-AMP carbonyl (Figure 4b). We hypothesize that

the altered conformation of TsaD induced by TsaE could still allow binding of TC-AMP but not its chemical conversion. Superposition of TobZ onto TsaBDE, shows that the TsaC-like module of TobZ is bound to the TsaD-like module at the opposite side of TsaD compared with TsaE. TobZ contains a large central cavity that connects the active sites of the TsaC- and TsaD-like modules. This tunnel prevents diffusion of the unstable carbamoyl-phosphate intermediate into the bulk solution and channels it directly to the active site of the TsaD-like module (53). Similarly, binding of TsaC to TsaBDE in a manner that resembles the association of the TobZ TsaC and TsaD modules, would create a complex that could allow efficient transfer of TC-AMP from the active site of TsaC to that of TsaD. The role of TsaE might be to 'prepare' TsaD for interaction with TsaC. Complex formation between the four t<sup>6</sup>A synthesis protein partners of *T. maritima* was observed by native gel shift experiments although a stable quaternary complex could not be isolated by gel filtration (23). For enzymatic purposes however, this complex would not need to be stable, but should only be sufficiently long-lived to permit TC-AMP transfer between TsaC and TsaD. Once the TC-AMP is bound, ATP could be hydrolysed by TsaE, promoting its dissociation from TsaBD, followed by binding of the tRNA substrate, renaturation of the TsaD metal binding sites and threonylcarbamoyl transfer to A37. The region around the active site that could be involved in TC-AMP channelling has conserved sequence patches (Supplementary Figure S7). The highly conserved SPT(F/Y)T loop of TsaE has its side chains in the coordination sphere of the  $\gamma$ P of AMPCPP and is close to the threonylcarbamoyl moiety of TC-AMP in the TsaD active site. It is therefore well positioned to serve as a sensor to coordinate ATP hydrolysis and TC-AMP binding. A scheme representing the hypothetical role of TsaE during the t<sup>6</sup>A reaction cycle is represented in Figure 6.



**Figure 6.** Hypothetic reaction scheme for TsaBDE. Binding of TsaE and ATP to TsaBD converts TsaD into an inactive conformation (symbolized by shady red colour). This conformation of TsaD then interacts with TsaC followed by transfer of TC-AMP (green stick) from TsaC to TsaD (pathways a or b). TsaE hydrolyses ATP and dissociates from TsaBD, liberating the binding site on TsaBD of the tRNA substrate. After transfer of the TC-moiety from TC-AMP to tRNA, the modified tRNA dissociates from TsaBD.

#### Comparison between bacterial TsaE and archaeal/eukaryotic Bud32

The threonyl-carbamoyl transferase multiprotein complexes in eukaryotes/archaea and bacteria have only the TC-transferring enzyme in common, Kae1 and TsaD respectively. Kae1 is flanked in the KEOPS complex by two subunits: Pcc1 binds to its N-terminal and Bud32 to its C-terminal domain (27,54). The helical bundle at the Pcc1-Kae1 interface mimics the TsaBD interface, despite the fact that Pcc1 and TsaB are unrelated. Pcc1 and TsaB likely play similar roles in the  $t^6A$  reaction pathway. Superposition of *TmTsaDE* onto Kae1-Bud32 (Supplementary Figure S8) shows that TsaE and Bud32 both interact with the helical insertion of the C-terminal domain of TsaD and Kae1 respectively. TsaE however binds at the opposing end of TsaD compared to Bud32 and Kae1. Moreover, Bud32 does not block the access of tRNA to the active site of Kae1. The role of Bud32 in the  $t^6A$  reaction pathway has not been clarified so far. Bud32, resembles the small protein kinases of the Rio family, involved in the biogenesis of ribosomes (55). An ATPase-dependent role of the Rio proteins was suggested to regulate their dynamic association with pre ribosomal subunits (56). Inactivation of the active site of Bud32 is deleterious for yeast cells and also for  $t^6A$  activity of the KEOPS complex (14). It remains an open question whether TsaE and Bud32 are functional analogues or whether they play a different role in the  $t^6A$  reaction cycle.

#### CONCLUSION

Our data provide a structural rationale for the biochemical observations collected on *Ec*- and *Tm*TsaBDE, e.g. the competition between tRNA and TsaE for binding to TsaD. The structure of *Tm*TsaBDE confirms TsaE is not directly in-

involved in the TC-transfer to tRNA, but suggests that it 'prepares' TsaD at the start of the reaction cycle. *Tm*TsaE converts *Tm*TsaB<sub>2</sub>D<sub>2</sub> into an open form that could accept the TC-AMP intermediate from *Tm*TsaC2. After ATP hydrolysis, *Tm*TsaE dissociates from *Tm*TsaBD which switches to a closed active form, capable of binding metals and tRNA substrate. Then, modified tRNA leaves the active site and *Tm*TsaBD enters into a new reaction cycle. The open (inactive) and closed (active) form of *Tm*TsaBD might be in equilibrium in absence of *Tm*TsaE as suggested by SAXS data (23). It is tempting to correlate these two states with the observation that *Tm*TsaBD, in absence of *Tm*TsaE, catalyses a single turn-over tRNA modification. This equilibrium probably is species-dependent since it was not observed for *Ec*TsaBD (22).

It is of note that the TsaBDE complex offers an attractive target for antimicrobial drug development. Chemical compounds that either inhibit TsaE to bind to TsaBD or significantly stabilize the TsaBDE complex would have deleterious effects on  $t^6A$  activity and bacterial survival.

#### DATA AVAILABILITY

The coordinates of the *Tm*TsaBDE complex have been deposited at the PDB under the code 6FPE.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

Sophia Missouri received a scholarship from the 'Fondation de la Recherche Médicale'. We thank the staff of the Proxima1 and 2 beamlines at the SOLEIL synchrotron for

assistance during the diffraction data collections. We are grateful to Marc Graille (Ecole Polytechnique, Palaiseau) and Tamara Basta (I2BC, Gif s/Yvette) for critical reading of the manuscript.

**Authors Contributions:** S.M., S.P., I.G., D.D., B.C., D.L., R.D. and W.Z. did the experimental work. S.M., B.C. and H.v.T. wrote the paper.

## FUNDING

This work received support from the French Infrastructure for Structural Biology [FRISBI ANR-10-INSB-0501]. Funding for open access charge: lab budget.

**Conflict of interest statement.** None declared.

## REFERENCES

- Grosjean, H. (2009) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, pp. 1–18.
- Machnicka, M.A., Milanowska, K., Oglou, O.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–D267.
- 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.* (2015) Essentiality of threonylcarbamoyl adenosine (t6A), a universal tRNA modification, in bacteria. *Mol. Microbiol.*, **98**, 1199–1221.
- 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.
- 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.
- Morin, A., Auxilien, S., Senger, B., Tewari, R. and Grosjean, H. (1998) Structural requirements for enzymatic formation of threonylcarbamoyl adenosine (t6A) in tRNA: an in vivo study with *Xenopus laevis* oocytes. *RNA*, **4**, 24–37.
- Murphy, F.V., Ramakrishnan, V., Malkiewicz, A. and Agris, P.F. (2004) The role of modifications in codon discrimination by tRNALysUUU. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
- 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 nephritic syndrome with primary microcephaly. *Nat. Genet.*, **49**, 1529–1538.
- 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.*, **25**, 545–551.
- 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.
- 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.
- 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.
- Lauhon, C.T. (2012) Mechanism of N6-threonylcarbamoyl adenosine (t6A) biosynthesis: isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry*, **51**, 8950–8963.
- 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.
- 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.
- Downey, M., Houlsworth, R., Maringe, L., Rollic, 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.
- 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.
- Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Barelle, J., Collinet, B., Van Tilbeurgh, H., Forterre, P. and Basta, T. (2013) In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.*, **41**, 1953–1964.
- 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.
- Zhang, W., Collinet, B., Graille, M., Daugeron, M.C., Lazar, N., Libri, D., Durand, D. and Van Tilbeurgh, H. (2015) 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.
- 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.
- Zhang, W., Collinet, B., Perrochia, L., Durand, D. and Van Tilbeurgh, H. (2015) 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.
- Luthra, A., Swinehart, W., Bayoos, S., Phan, P., Stec, B., Iwata-Reuyl, D. and Swairjo, M.A. (2018) Structure and mechanism of a bacterial t6A biosynthesis system. *Nucleic Acids Res.*, **46**, 1395–1411.
- 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.*, **22**, 628–640.
- 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 Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **66**, 1230–1236.
- Vecchiotti, D., Ferrara, S., Rusmini, R., Macchi, R., Milani, M. and Bertoni, G. (2016) Crystal structure of YeaZ from *Pseudomonas aeruginosa*. *Biochem. Biophys. Res. Commun.*, **470**, 460–465.
- 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.
- 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 Struct. Funct. Genet.*, **48**, 220–226.
- Nguyen, H.A., El Khoury, T., Guiral, S., Laaberki, M.H., Candusso, M.P., Galisson, F., Foucher, A.E., Kesraoui, S., Ballut, L., Vallet, S. *et al.* (2017) Expanding the kinome World: a new protein kinase family widely conserved in bacteria. *J. Mol. Biol.*, **429**, 3056–3074.
- Wan, L.C.K., Pillon, M.C., Thevakumaran, N., Sun, Y., Chakraborty, 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.
- Perrochia, L., Guetta, D., Hecker, A., Forterre, P. and Basta, T. (2013) Functional assignment of KEOPS/EKC complex subunits in the

- biosynthesis of the universal t6A tRNA modification. *Nucleic Acids Res.*, **41**, 9484–9499.
32. Lerner, C.G., Hajduk, P.J., Wagner, R., Wagenaar, F.L., Woodall, C., Gu, Y.-G., Searle, X.B., Florjancic, A.S., Zhang, T., Clark, R.F. *et al.* (2007) From bacterial genomes to novel antibacterial agents: discovery, characterization, and antibacterial activity of compounds that bind to HI0065 (YjeE) from *Haemophilus influenzae*. *Chem. Biol. Drug Des.*, **69**, 395–404.
  33. Collinet, B., Friberg, A., Brooks, M.A., van den Elzen, T., Henriot, V., Dziembowski, A., Graille, M., Durand, D., Leulliot, N., Saint André, C. *et al.* (2011) Strategies for the structural analysis of multi-protein complexes: lessons from the 3D-Repertoire project. *J. Struct. Biol.*, **175**, 147–158.
  34. Gaspar, P., Oliveira, J.L., Frommlet, J., Santos, M.A.S. and Moura, G. (2012) EuGene: Maximizing synthetic gene design for heterologous expression. *Bioinformatics*, **28**, 2683–2684.
  35. Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D*, **66**, 125–132.
  36. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
  37. 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. Sect. D Biol. Crystallogr.*, **67**, 235–242.
  38. Webb, B. and Sali, A. (2014) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.*, **2014**, 5.6.1–5.6.32.
  39. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W. *et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94–106.
  40. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
  41. 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.
  42. Sankaranarayanan, R., Dock-Bregeon, A.C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. and Moras, D. (1999) The structure of threonyl-tRNA synthetase-tRNAThr complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell*, **97**, 371–381.
  43. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
  44. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
  45. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (PDBFold), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.*, **D60**, 2256–2268.
  46. DeLano, W.L. (2014) The PyMOL molecular graphics system, version 1.8. *Schrödinger LLC*, doi:10.1038/hr.2014.17.
  47. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.
  48. 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.
  49. Hecker, A., Leulliot, N., Gabelle, 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 KaeI is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity in vitro. *Nucleic Acids Res.*, **35**, 6042–6051.
  50. Cherfils, J. and Zeghouf, M. (2013) Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.*, **93**, 269–309.
  51. Hanks, S., Quinn, A. and Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, **241**, 42–52.
  52. Aydin, I., Saijo-Hamano, Y., Namba, K., Thomas, C. and Roujeinikova, A. (2011) Structural analysis of the essential resuscitation promoting factor yeast suggests a mechanism of nucleotide regulation through dimer reorganization. *PLoS One*, **6**, e23245.
  53. 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.*, **51**, 4046–4052.
  54. Hecker, A., Lopreiato, R., Graille, M., Collinet, B., Forterre, P., Libri, D. and van Tilbeurgh, H. (2008) Structure of the archaeal KaeI/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex. *EMBO J.*, **27**, 2340–2351.
  55. LaRonde-LeBlanc, N. and Wlodawer, A. (2005) A family portrait of the RIO kinases. *J. Biol. Chem.*, **280**, 37297–37300.
  56. Knüppel, R., Christensen, R.H., Gray, F.C., Esser, D., Strauß, D., Medenbach, J., Siebers, B., MacNeill, S.A., LaRonde, N. and Ferreira-Cerca, S. (2018) Insights into the evolutionary conserved regulation of Rio ATPase activity. *Nucleic Acids Res.*, **46**, 1441–1456.

## Supplementary Material

### SUPPLEMENTARY TABLES

**Table S1**

**Structure similarity between the two copies of *TmTsaBDE* in the asymmetric unit (chains CEG and ADB) and between the *Tm* Tsa-subunits and their orthologues.**

<b><i>TmTsaBDE</i> complex</b>	<b>free proteins</b>	<b>RMSD (Å)</b>	<b>Z-score</b>	<b>% sequence identity</b>
<i>TmTsaB</i> (chain C)	<i>TmTsaB</i> (chain A)	0.73 (180 Cα)	15.8	95
<i>TmTsaE</i> (chain E)	<i>TmTsaE</i> (chain D)	0.98 (147 Cα)	12.2	83
<i>TmTsaD</i> (chain G)	<i>TmTsaD</i> (chain B)	0.90 (301 Cα)	18.9	95
<i>TmTsaD</i> (chain G)	<i>EcTsaD</i> (PDB ID 4YDU)	2.58 (291 Cα)	10.9	37
<i>TmTsaD</i> (chain G)	<i>ScTsaD</i> (PDB ID 3ZEU)	2.65 (291 Cα)	10.1	37
<i>TmTsaE</i> (chain E)	<i>HiTsaE</i> (PDB ID 1FL9)	1.97 (141 Cα)	9.0	36
<i>TmTsaE</i> (chain E)	<i>HiTsaE</i> (PDB ID 1HTW)	1.93 (141 Cα)	9.6	35
<i>TmTsaE</i> (chain E)	<i>BsTsaE</i> (PDB ID 5MVR)	1.69 (137 Cα)	10.1	31
<i>TmTsaB</i> (chain C)	<i>EcTsaB</i> (PDB ID 4YDU)	1.78 (179 Cα)	10.6	21
<i>TmTsaB</i> (chain C)	<i>EcTsaB</i> (PDB ID 4YOW)	1.76 (171 Cα)	11.4	23
<i>TmTsaB</i> (chain C)	<i>ScTsaB</i> (PDB ID 3ZEU)	1.88 (180 Cα)	9.4	17
<i>TmTsaB</i> (chain C)	<i>StTsaB</i> (PDB ID 2GEL)	1.87 (178 Cα)	11.6	19
<i>TmTsaB</i> (chain C)	<i>TmTsaB</i> (PDB ID 2A6A)	1.02 (175 Cα)	14.8	85
<i>TmTsaB</i> (chain C)	<i>VpTsaB</i> (PDB ID 3R6M)	1.92 (176 Cα)	9.2	22

**Table S2**

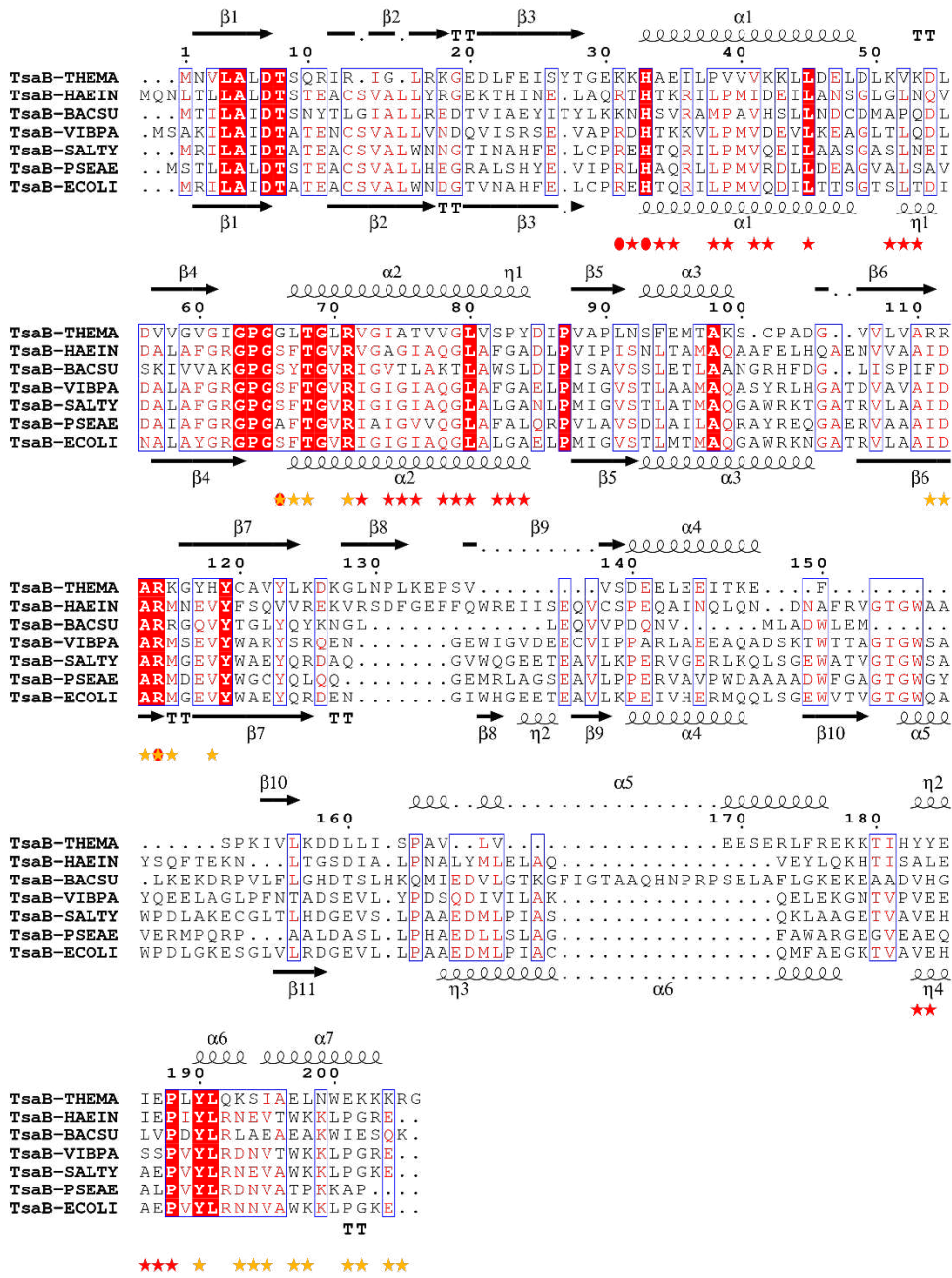
Sequences of the synthesized genes used in this study. In bold are restriction sites used for cloning into the corresponding expression vectors (i.e. pET21a for BC14 and pET24d for BC15).

Sequence (5'>3') of the synthesised genes
<p>&gt;BC14</p> <p><b>CAATATG</b>CGCGCTTCTTGGCATTGAAACCAGCTGTGACGAAACCGCTGTAGCAGTCTGGATGACGGTAAAAACGTTCTTGTAACTTTACCGT            TTCCAGATTGAGGTACACCAGAAAGTTCGGCGCGCTTGTTCGGAAAGTTGCTGCACGCCACCACCTTAAAAACCTGCCGATCCTTCTGAAGA            AAGCATTGAAAAAGTACCACCGGAGACCGTTGACGTAGTTGCTGCAACTTACGGCCCGGGTCTTATTGGTGCACTTCTTGTGGTCTGAGC            GCAGCAAAAGGCCCTGCAATCAGTCTTAAAAAACCGTTTGGTGGCGTTAACACAGTGGAGGCACACGTTTCCAGGCTGTTTCTGGCAAAACCC            GGACCTGAAACCACCGCTTGTGTCTGATGGTTAGCGGTGGCCACACCAGCTGATGAAAGTTGACGAGGATTACAGCATGGAAGTTCTCGG            GTGAAACCCCTTGACGACAGTGTGGCGAGGCATTTGACAAGGTTGCACGCCCTTCTGGCCTTGGTACCCGGGCGGTCCGGTTATTGATCCG            GTTGCAAAAGAAAGGCGACCCGAAAAAGTATTCCTTCCGCGTCCGATGCTGGACGACGATTTCTATAACTTTTCTTCGCTGGTCTGAAGAC            TTCCGTACTTTACTTTCTTCAGCGCGAAAAAGGTTATAAGGTGGAAGATGTTGCTGCATCCTTTCAGAAGGCTGTAGTTGACATCCTTGTG            AAAAACTTTTCGCCTGGCTCGCAACCTTGGCATTGCAAAATGCTTTTGTAGGCGGTGTGACGTAACCTCATGCTGCCGGAAGAGGTG            CGCAAAACCGCTGAACGCTGGAACATGAGGTTTTTTCCACCGCTGGAACCTTGTACCGATAACGCACCTGATGGTAGCAAGGCTGGTTA            TGA AAAAGCTAAACGCGCATGTTAGCCCGCTGTCTCTGACCGCAGACCCGAACCTGAACTTAACTCGAATTTCTAGAAAATTTTGT            AACTTTAAGAAGGAGATATAGTTATGAACGTTCTGGCACTTGTATACCTCCCAAGTATTCGTATCCGGCTGCGTAAAGGCGAAGACCTTTC            GAAATCTCTACACTGGTGTAGAAAAAACCCGAGAGATCTTCCGGTGTGTTAAAGAAGCTGTGGATGAACTGGACCTAAAGTTAAAGA            CCTTGACGTTGTTGGCGTTGGCATCGGCCGGCGGCTGACTGCTCTTCGCGTTGGCATTGCAACTCTTCTTGGCCTTGTTCGCCGTACG            ACATTCGGGTGCTCGCTGAACTCCTTTGAAATGACCCGAAAGTCTTGTCCGGCTGATGGTGTGACTTGTGCTCGTGTGCTCGCAAA            GGTTACCCTACTGTGCTGTTTACCTGAAGGATAAAGCCCTTAACCCGCTGAAAGAGCCGTCGGTTGATCCGATGAAGAACTGGAAGAGAT            TACTAAAGAGTTCTCCCGAAAAATGACTGAAGGATGATCTTCTTATTTCTCCGGCTGTTCTGGTTGAAGAGTCTGAGCGCTTTTTCGTG            AGAAAAAGACTATTCACTACTACGAAATGAAACCGCTGTACCTTCAAAAAATCCATTGCTGAACTGAACTGGGAAAAAAGAAACCGCGCCAC            CATCACCATCACCATTAACCTCGTAAAAATAATTTGTTTAACTTTAAGAAGGAGATATAGTCCATGGAATTTTAACT<b>CGAG</b></p> <p>&gt;BC15</p> <p><b>CCATGGG</b>GCATCACCATCACCATCACGAAAACTGTATTTTCAGGGCTATAATACTGTTGAAGAACA AAAAATGCGTCACCTGCGCTTCGAA            AACCTGACTGAAGAACAACCTGAAACGCTGGCTAAAACTCTGACTGAAAACCTGAAGGGCGGCAAGTTGTTATTCTGAGCGGCAACCTGGG            CGCAGGCAAACTACCTTCGTTAAGGCATGATTCTGCTATTGGTCTGGACGAAAAATGGTTAAATCTCCAACCTTTACTCTGATGAAG            TATACCCGGCTGAAAACTATTTATCATCTGACCTGTATCTGCTGCAGGATACCGATTTCCGCTCTGGACGTAGAAGATTTCTGGAA            GATGAAGATGGCATATGGTTGTAGAAATGGGGGACCTGTTGACGGCTTTTGGCCAGAAGATAGTATTAAGGTTAAGATGAAATTTGCTGA            TGAATCTCATCTGAATGTTGAAATCTGATTCGGGAAGAAGTAAATTTCTGGTAGAAAAATGAACTTATCGCAAGAACCTGCAGACA            CTTAACT<b>CGAG</b></p>

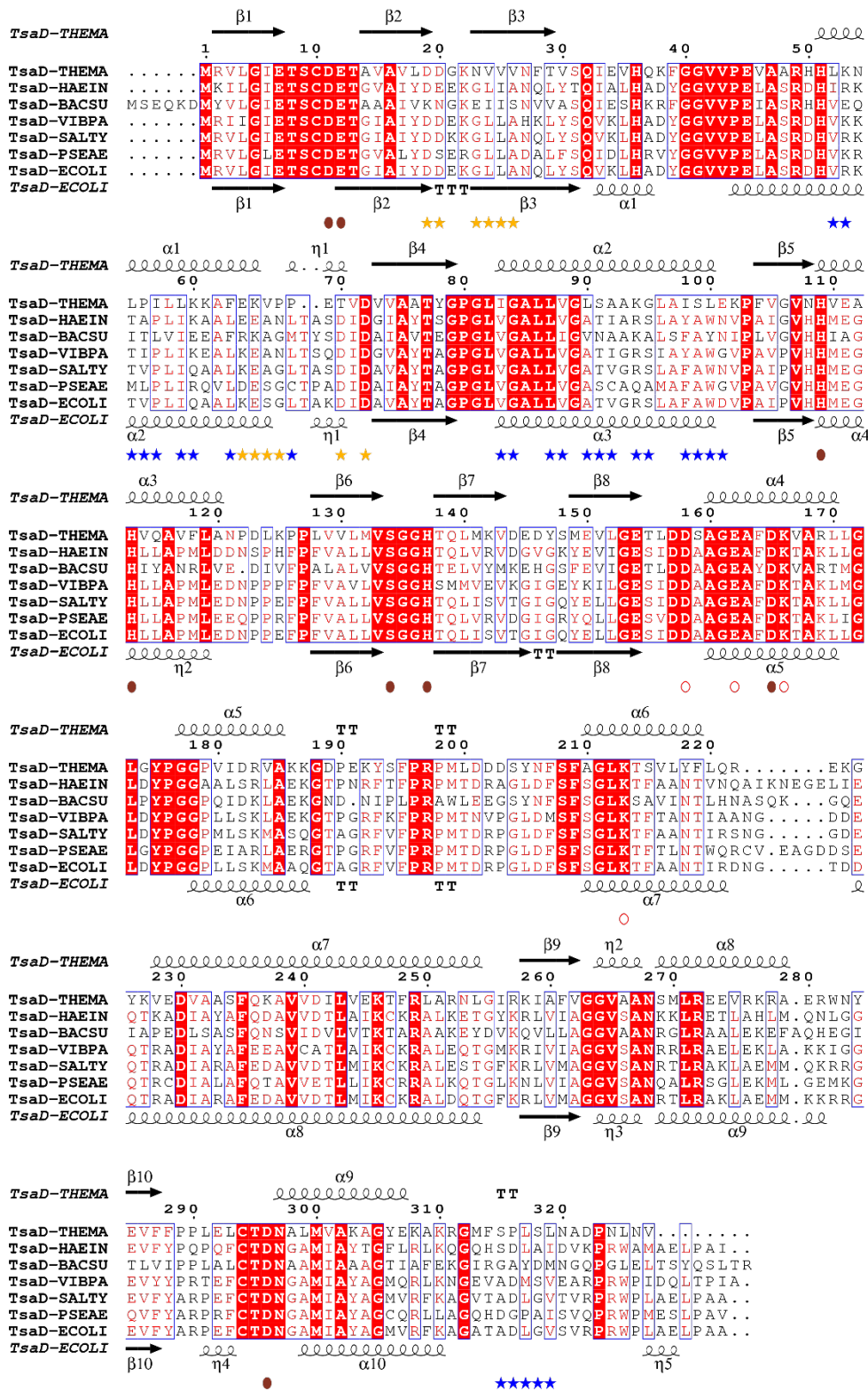


# Supplementary Figures

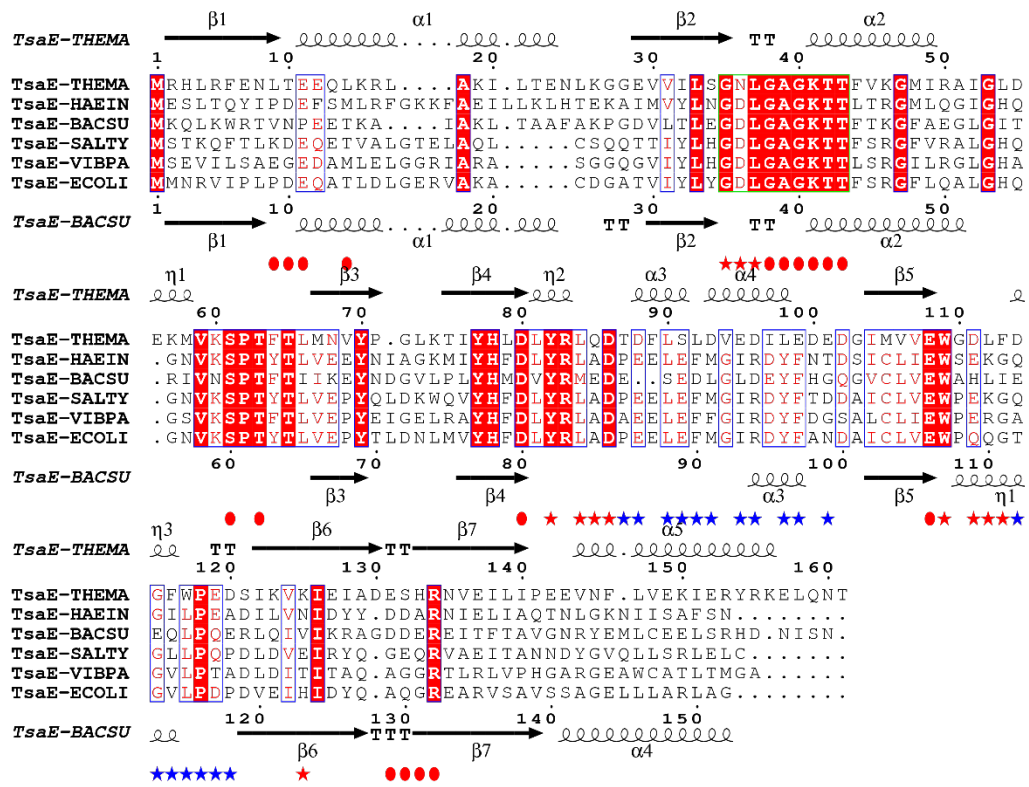
a.



b.



C.



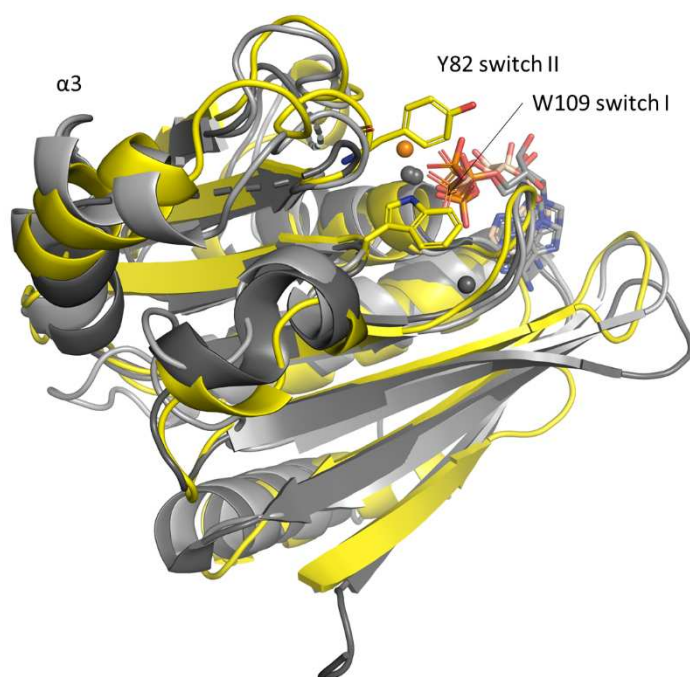
**Figure S1**

Structure based sequence alignment of the *TmTsaBDE* subunit homologues.

The sequence alignments of the subunits are superposed by the secondary structure assignment based on the crystal structure of *TmTsaBDE*. Conserved residues are boxed in red, well conserved residues are in red. THEMA: *T. maritima*; HAEGIN: *H. influenzae*; BACSU: *B. subtilis*; VIBPA: *V. parahaemolyticus*; SALTY: *S. typhimurium*; PSEAR: *P. aerogunosa*; ECOLI: *E. coli*

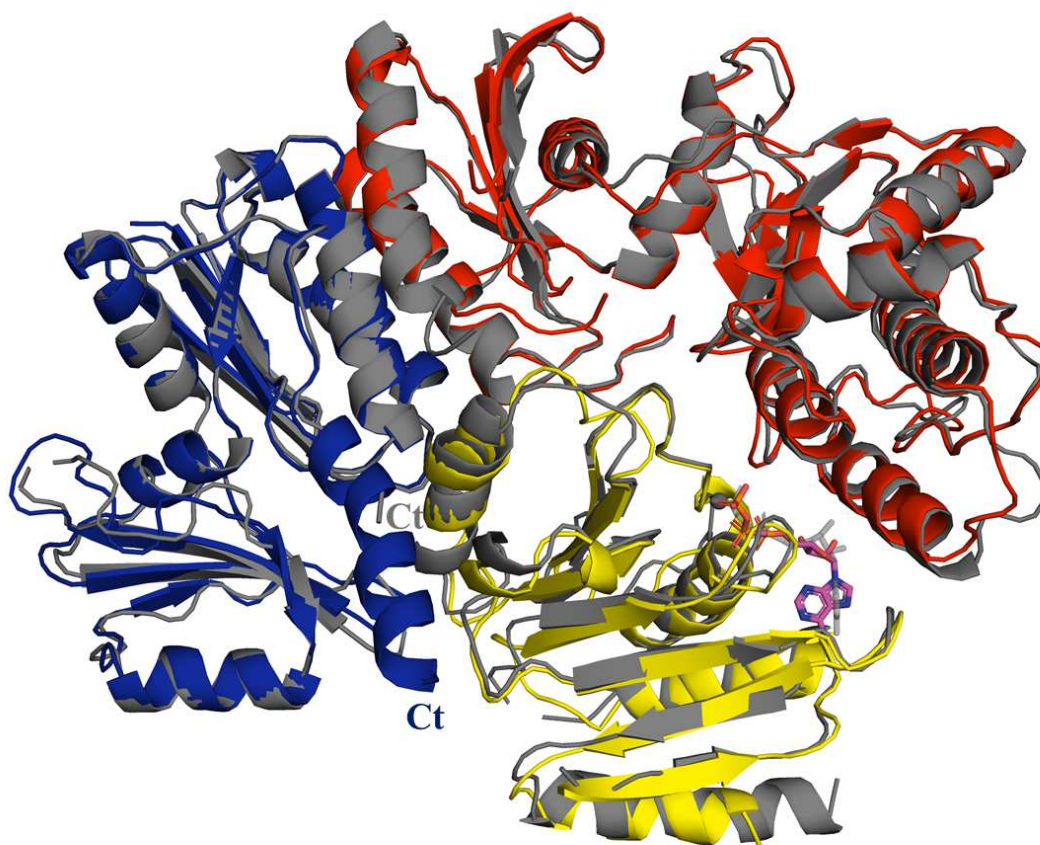
- a. Structure based sequence alignment of TsaB homologues. At the bottom: secondary structure of *EcTsaB*. Residues interacting with TsaD and TsaE are marked with red and yellow stars respectively. Residues involved in ADP binding as observed in the *EcTsaBD* crystal structure are marked as red ovals.

- b.** Structure based sequence alignment of TsuD homologues. At the bottom: secondary structure of *EcTsaD*. Residues interacting with *TmTsaB* and *TmTsaE* are marked with blue and yellow stars respectively. Residues involved in ADP binding of *EcTsaBD* are marked as brown filled ovals. Residues of *TmTsaD* involved in AMPCPP binding are marked as red empty ovals.
- c.** Structure based sequence alignment of TsaE homologues. At the bottom: secondary structure of *BsTsaE*. Residues interacting with *TmTsaB* and *TmTsaD* are marked with blue and red stars respectively. Residues of *TmTsaE* involved in AMPCPP binding are marked as red filled ovals.



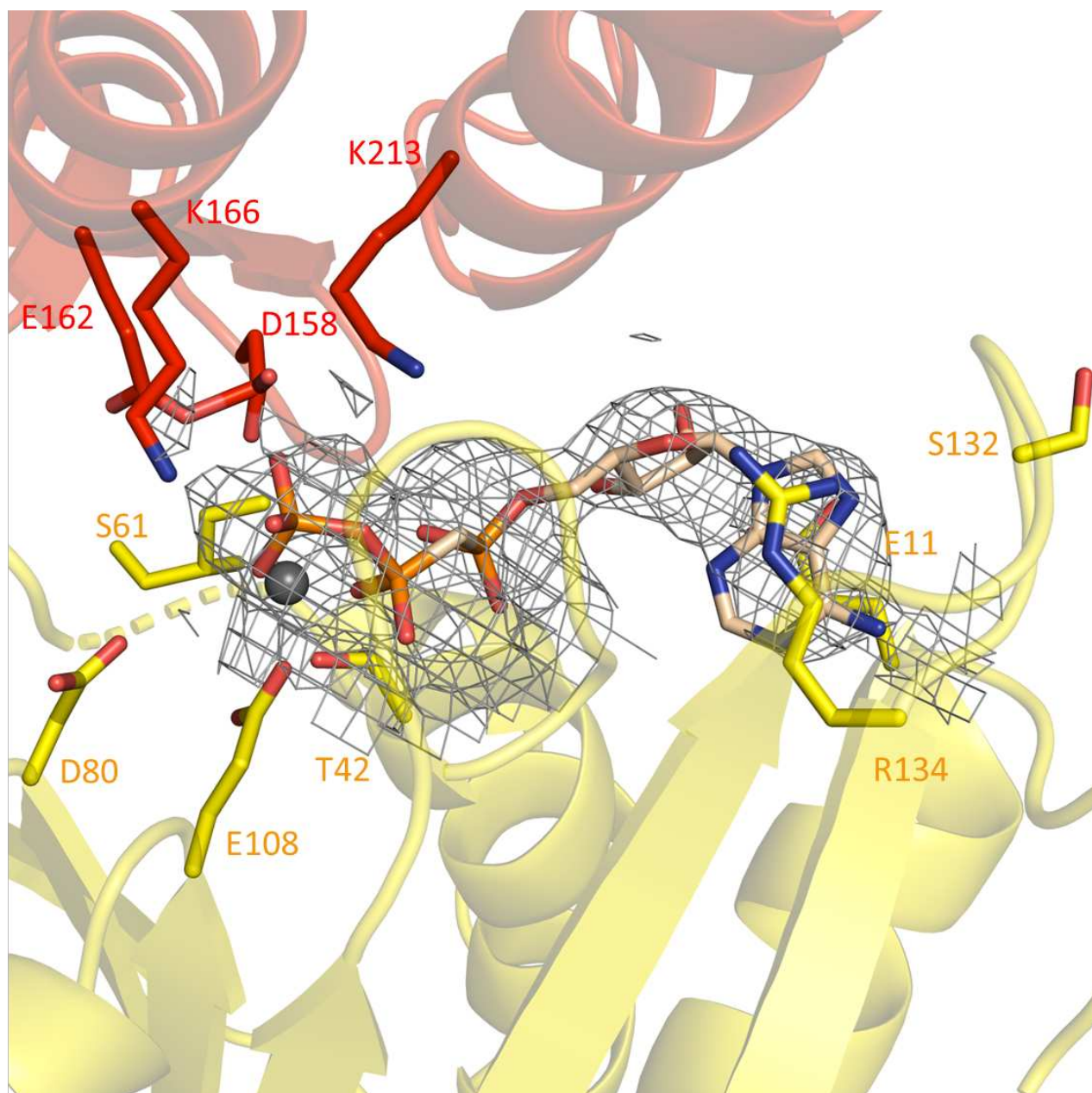
**Figure S2**

Superposition of *TmTsaE* (yellow), *BsTsaE* (light grey) and *HtTsaE* (grey). The ADPs (bound to *BsTsaE* and *HtTsaE*) and AMPCPP (bound to *TmTsaE*) are represented as sticks, Mg-ions from *BsTsaE* and *HtTsaE* as grey spheres, from *TmTsaE* as orange sphere.  $\alpha 3$  is the negatively charged helix that is bound near the *TmTsaBD* interface.



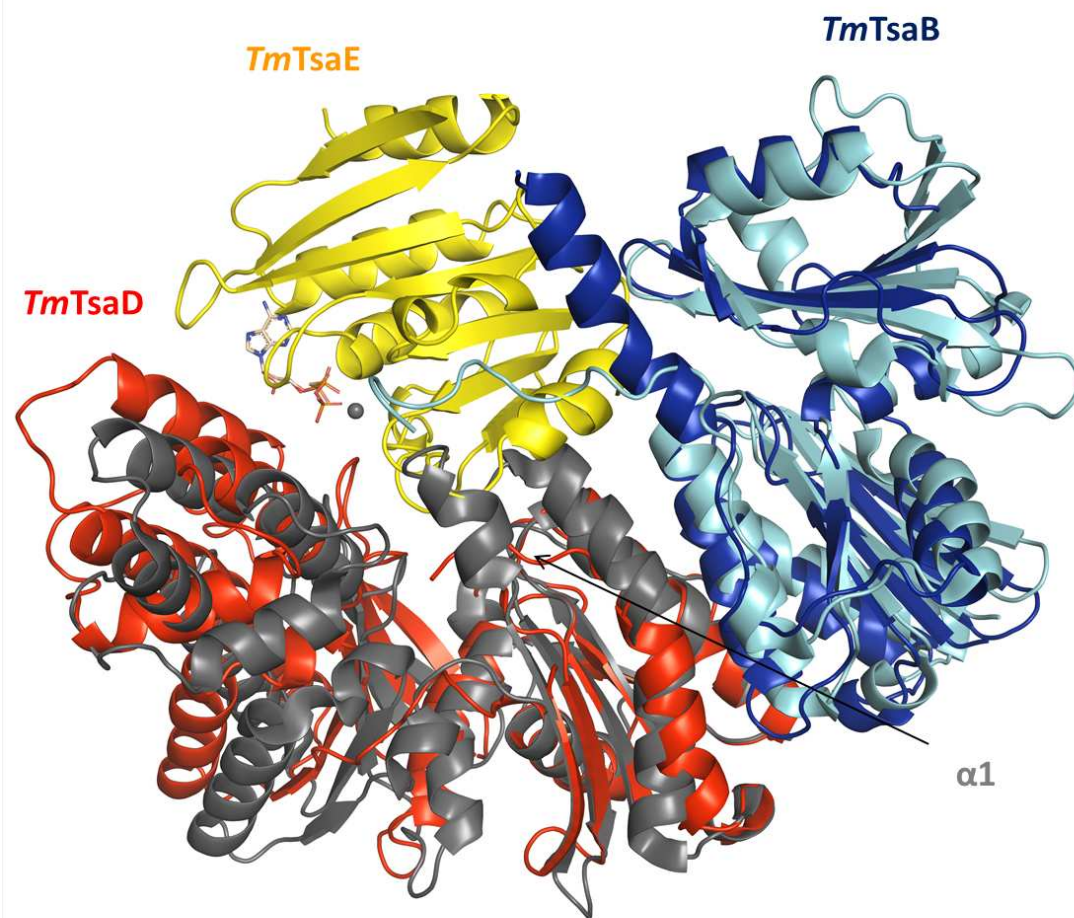
**Figure S3**

Superposition of the two *TmTsaBDE* copies in the asymmetric unit. One copy is in colour (*TmTsaB*: blue, *TmTsaD*: red, *TmTsaE*: yellow), the other is uniformly grey. The most significant structural difference is situated at the C-terminus (marked Ct) of TsaB: in one copy it is ordered as a helix (blue Ct), while in the other it is stretched and partially disordered (grey Ct).



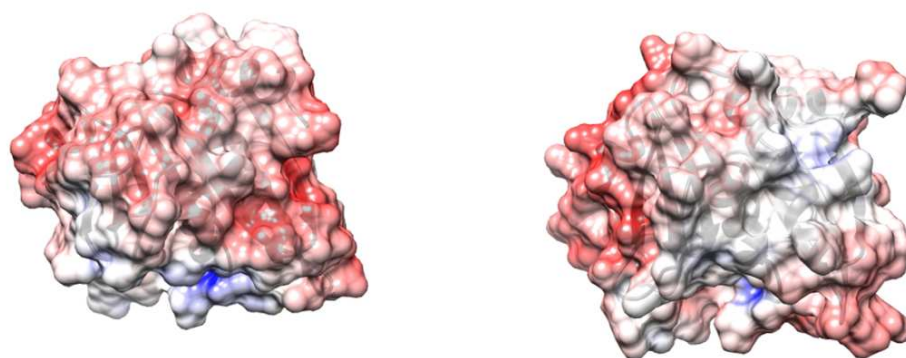
**Figure S4**

2Fo-Fc experimental residual electron density map centred around the AMPCPP ligand (sticks) and contoured at 1.5  $\sigma$ . Mg-ion represented as grey sphere, Mg-ligands of TsA are labelled (yellow sticks). TsA residues interacting with the 3'P of AMPCPP are shown as red sticks.



**Figure S5**

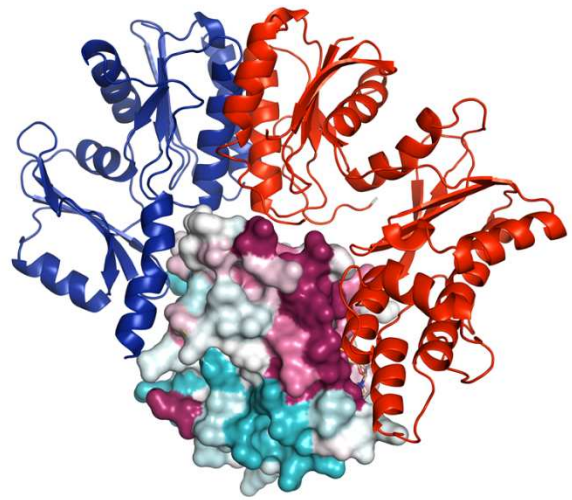
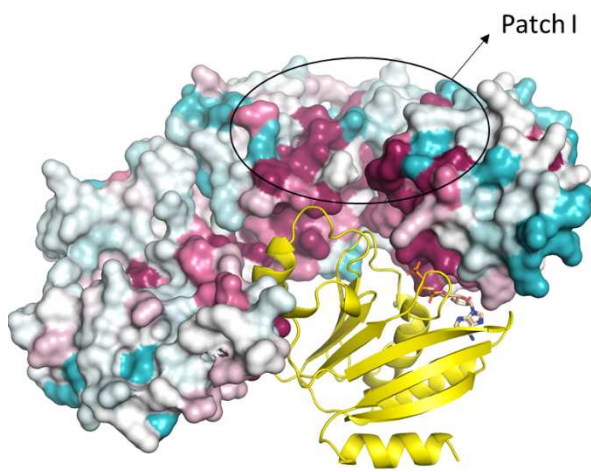
Overlay of *EcTsaBD* (PDB ID 4YDU; *EcTsaB* in cyan and *EcTsaD* in grey) and *TmTsaBDE* (same colour code as for panel a). The N-terminal domains of *TsaD* were used for superimposition. ADP (bound to the active site of *EcTsaBD*) and AMPNPP bound to *TmTsaE* are shown as sticks. The Mg- and Fe-ions bound to *EcTsaD* are shown as pink and grey spheres respectively.  $\alpha 1$  of *TmTsaD* which is disordered in the context of *TmTsaBDE* is indicated.



**Figure S6**

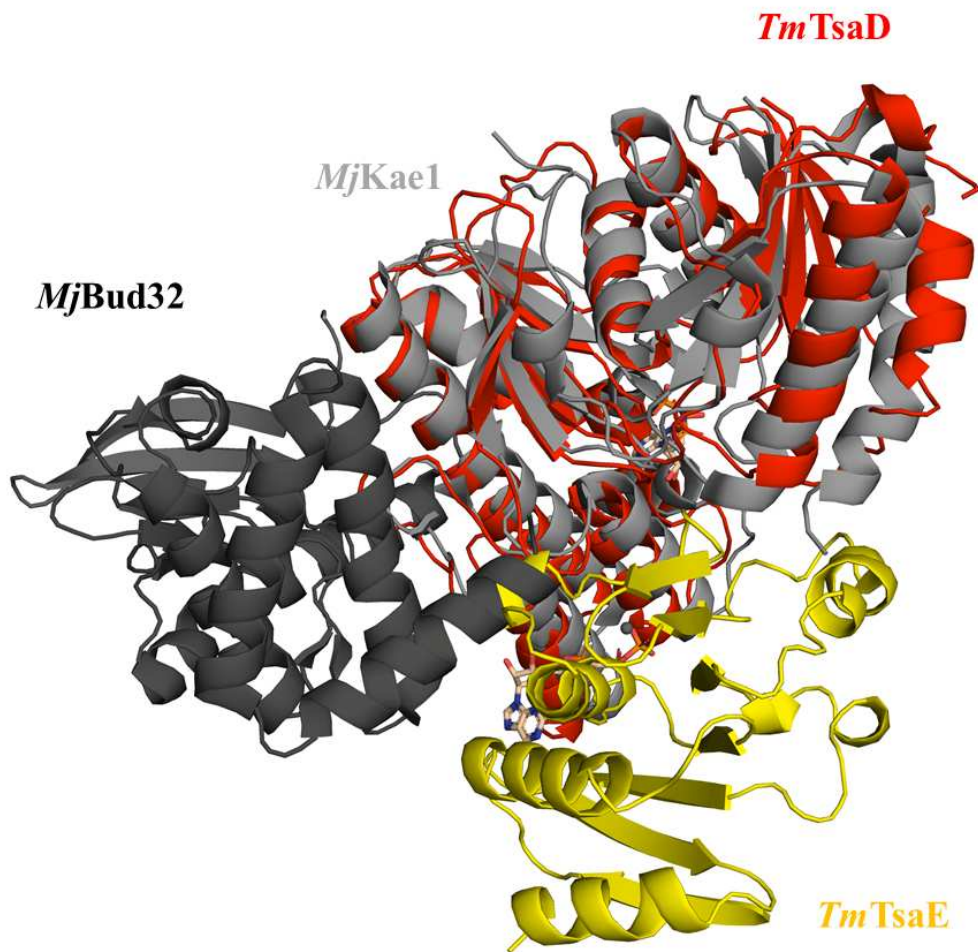
Electrostatic surface potential of *HtTsaE* (left) and *BsTsaE* (right). Colour scale from positively (blue) to negatively (red) charged.





**Figure S7**

The sequence conservation projected onto the surface. Left panel : surface presentation of TsabD with *TmTsaE* presented in ribbon (AMPCPP in sticks). The black circle represent here patch I whose preserved residues could involve in interaction with TsaC. Right panel : sequence conservation of TsaE with *TmTsaBD* presented in ribbon. The conservation scaled varies between variable (cyan) and conserved (raspberry).



**Figure S8**

Superposition of the archaeal *M. janaschii* Bud32/Kae1 complex<sup>2</sup> onto *TmTsaDE* (Bud32 dark grey, Kae1 light grey, TsaD red and TsaE yellow). Bud32 and TsaE both bind to the helical extension within the C-terminal domain of Kae1/TsaD but their binding modes (and structures) are unrelated.

## Corrigendum

# The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t<sup>6</sup>A tRNA-modification

Sophia Missouri<sup>1</sup>, Stéphane Plancqueel<sup>1</sup>, Ines Li de la Sierra-Gallay<sup>1</sup>, Wenhua Zhang<sup>1</sup>, Dominique Liger<sup>1</sup>, Dominique Durand<sup>1</sup>, Raoudha Dammak<sup>1</sup>, Bruno Collinet<sup>1,2,\*</sup> and Herman van Tilbeurgh<sup>1,\*</sup>

<sup>1</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS UMR 9198, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif sur Yvette Cedex, France and <sup>2</sup>Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, UMR7590 CNRS/Sorbonne-Université, UPMC, Paris, France

*Nucleic Acids Research*, 2018, 46(11): 5850–5860, <https://doi.org/10.1093/nar/gky323>

Subsequent to publication of this study, we obtained crystals that yielded slightly better diffraction data for the *Tm*TsaBDE complex (Table 1). These data resulted in improved electron density maps and allowed us to identify and correct a few errors in the previously deposited structure (originally deposited as PDB 6FPE). The latter (which is still available for download and comparison, at <https://www.rcsb.org/structure/removed/6FPE>) has been superseded by a new set of revised coordinates (PDB 6S84).

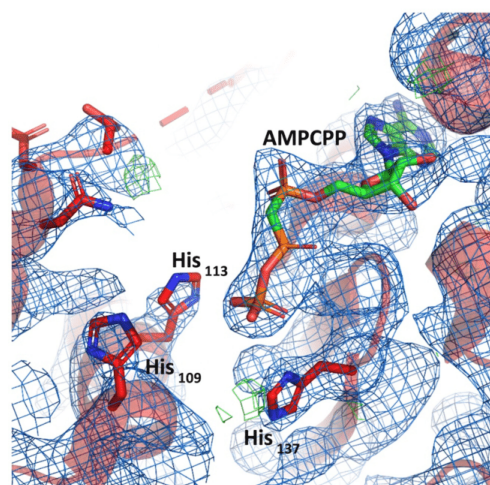
The most significant differences between the two coordinate sets consist of:

1. The C-terminus of one of the TsaB copies in the asymmetric unit: amino acids 194–205 that were missing in the 6FPE structure could be constructed into the density.
2. Modeling of a nucleotide bound at the active site of TsaD. We initially cautiously interpreted the residual electron density by glycerol and PEG moieties present in the crystal freezing liquor. The new maps clearly showed that a nucleotide was bound at this location (Figure 1). We could easily fit the density by AMPCPP, present as a ligand in the crystallization solution.

The new structure does not alter the primary conclusions of our manuscript:

- The C-terminal part of the active site of *Tm*TsaD remains well-structured and is still capable of binding a nucleotide in the context of the ternary *Tm*TsaBDE complex.
- The AMPCPP occupies exactly the same position as the carboxy-AMP compound present in the recent structure of *Tm*TsaBDE, reported by Swairjo *et al.* (1). We further confirm that in our structure the N-terminal part of the *Tm*TsaD active site remains partially disordered and that neither Zn nor Fe ions are bound. This contrasts with the structure reported by Swairjo *et al.*, which has an ordered metal binding site occupied by Zn. This latter structure was obtained in presence of ATP, and we suspect that the nature of the bound nucleotide (AMPCPP versus ATP) might play a role in the metal binding. The two structures of the TsaBDE complex (6N9A and 6S84) represent probably different snapshots along the catalytic pathway.

\*To whom correspondence should be addressed. Herman van Tilbeurgh. Email: [herman.van-tilbeurgh@u-psud.fr](mailto:herman.van-tilbeurgh@u-psud.fr)  
Correspondence may also be addressed to Bruno Collinet. Email: [bruno.collinet@i2bc.paris-saclay.fr](mailto:bruno.collinet@i2bc.paris-saclay.fr)  
Present address: Wenhua Zhang, School of Life Sciences, Lanzhou University, 730000 Lanzhou, China.



**Figure 1.**  $2F_o-F_c$  difference map (blue) of the AMPCPP surrounding at the active site of *TmTsaD*.  $F_o-F_c$  map: positive and negative densities are represented as green and red grids respectively. The metal coordinating histidines 109 and 113 are also shown. They are partially disordered and not in a configuration compatible with metal binding.

**Table 1.** New (vs. original) data collection and refinement statistics

	PDB 6S84	PDB 6FPE
Wavelength (Å)	0.978570	0.9801
Resolution range (Å)	46.14–2.90 (3.00–2.90)	48.44–3.14 (3.33–3.14)
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell a,b,c (Å)	a = 85.16, b = 108.21, c = 176.65	a = 84.31, b = 113.94, c = 177.62
Total reflections	324005	138146
Unique reflections	36849	30297
Completeness (%)	99.1 (94.8)	99.1 (95.2)
Mean I/sigma(I)	9.1 (0.87)	7.1 (1.03)
R-meas	0.23	0.19
CC1/2	99.7 (42.1)	99.2 (43.9)
R-work	0.209	0.23
R-free	0.282	0.29
Number of non-hydrogen atoms	10678	10418
RMSD bonds (Å)	0.003	0.011
RMSD angles (°)	1.011	1.258
Ramachandran favored (%)	94.77	95.3
Average B-factor (Å <sup>2</sup> )	89.66	80.67

Statistics for the highest-resolution shell are shown in parentheses.

## REFERENCE

- Luthra, A., Paranagama, N., Swinehart, W., Bayoos, S., Phan, P., Quach, V., Schiffer, J.M., Stec, B., Iwata-Reuyl, D. and Swairjo, M.A. (2019) Conformational communication mediates the reset step in t<sup>6</sup>A biosynthesis. *Nucleic Acids Res.*, **47**, 6551–6567.



## I.2 Discussion des résultats obtenus par le groupe de Manal Swarjo (Luthra *et al.*, 2019).

En parallèle à cette étude, une autre équipe aux USA a résolu une autre structure cristallographique de TmTsaBDE (Luthra *et al.*, 2019). La [figure 54](#) illustre les différents sites actifs de TsaD en conformation ouverte (code PDB 6S84 et 6N9A) et en conformation fermée (code PDB 4YDU). Les deux structures de TmTsaBDE (celle résolue dans l'équipe, code PDB 6S84 et celle de l'équipe de Manal Swarjo, code PDB 6N9A) sont très similaires mais montrent toutefois quelques différences intéressantes :

- L'équipe de Manal Swarjo observe une molécule d'ATP et un ion  $Mg^{2+}$ , localisés à l'interface de TsaE et TsaD stabilisant cette interaction. La présence de TsaE induit un changement de conformation de la sous unité catalytique TsaD, très comparable à notre structure. Les deux sous unités TsaD se superposent avec une déviation moyenne (RMSD) de 1.27 Å (Z-score 17.2 sur 304 positions C $\alpha$  alignées).
- Une molécule de carboxy-AMP (Cax-AMP), identifiée par l'équipe de M.Swarjo, est fixée dans le site actif de TsaD ([figure 42 B](#)). L'adénosine et le phosphate  $\alpha$  se superposent parfaitement avec l'adénosine et le phosphate  $\alpha$  de l'AMPCPP dans notre structure. L'origine de ce carboxy-AMP n'est pas claire et les auteurs de l'article n'expliquent pas pourquoi ce ligand pourrait bien se trouver dans le site actif de TsaD. Il reste à vérifier si la densité électronique de cette structure ne pourrait pas être expliquée par une molécule d'ADP (dont l'encombrement stérique est compatible avec la densité électronique) ou alors par une molécule de carbamoyl-AMP.
- En revanche, un ion  $Zn^{2+}$  est fixé dans le site actif de cette nouvelle structure. Les positions des acides aminés H<sub>109</sub>, H<sub>113</sub>, H<sub>136</sub> et D<sub>296</sub> sont déviées par rapport à notre structure ([figure 54 A](#)) et sont dans une conformation permettant la coordination du métal  $Zn^{2+}$  ([figure 54 B](#)). Cependant, la position relative du métal  $Zn^{2+}$  par rapport au métal  $Fe^{2+}$  présent dans EcTsaD est déviée de  $\sim 5$  Å et ne participe pas à la coordination du nucléotide dans la structure de Luthra *et al.*, 2019. Pour finir, les acides aminés D<sub>11</sub> et E<sub>12</sub> de EcTsaD coordonnent le second métal  $Mg^{2+}$  à proximité du phosphate  $\beta$  de l'ADP ([figure 54 C](#)) ; les positions relatives de ces acides aminés dans TmTsaD présentent une déviation de près de 11 Å et ne permettent pas la coordination du métal  $Mg^{2+}$ .

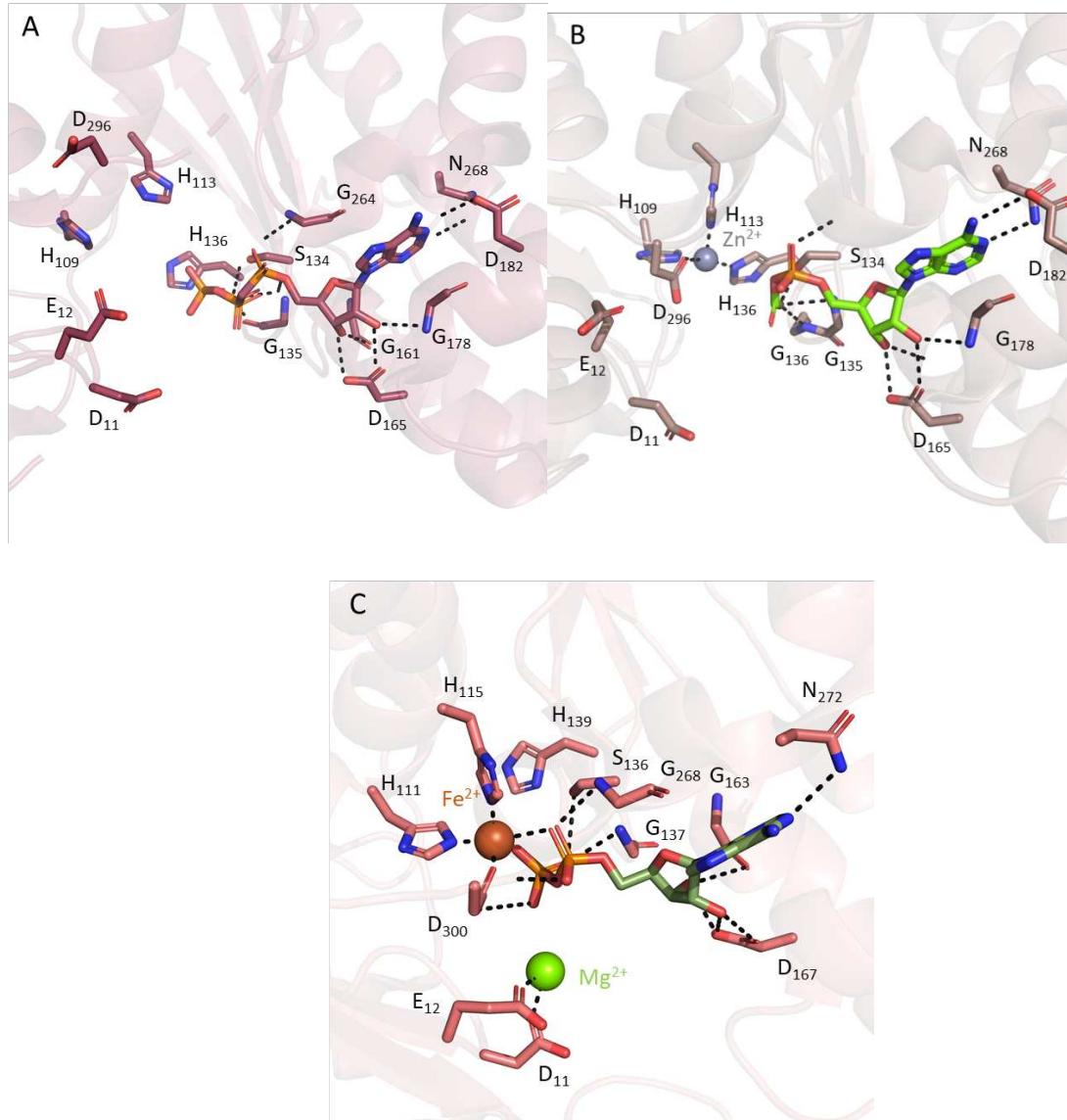


Figure 54 Représentation des sites actifs des différentes structures cristallographiques connues de TsuD. A : Site actif en conformation ouverte de TsuD de la bactérie *Thermotoga maritima* (code PDB 6S84). Un analogue non hydrolysable de l'ATP (AMPCPP) est stabilisée par les acides aminés du domaine C-terminal de TsuD représentés en stick. B : Site actif de TsuD de *Thermotoga maritima* (code PDB 6N9A). Une molécule de carboxy-AMP (Cax-AMP) est stabilisée par les acides aminés du domaine c-terminal de TsuD représentés en stick. Une molécule de  $Zn^{2+}$  est coordonnée par les acides aminés en stick. C : Site actif de TsuD de *Escherichia coli* en conformation fermée (code PDB 4YDU). Une molécule d'ADP et des ions  $Mg^{2+}$  et  $Fe^{2+}$ , en vert et orange respectivement, sont liés dans le site actif.

Cette nouvelle structure de TsaBDE de *Thermotoga maritima* (code PDB 6N9A) met en évidence que la position des deux domaines de TsaD présente toujours un changement de conformation drastique induisant une ouverture du site actif de l'enzyme en présence de TsaE. Les seules différences entre ces deux structures sont localisées aux résidus H<sub>109</sub>, H<sub>113</sub>, H<sub>136</sub> et D<sub>296</sub> qui présentent une orientation supportant la coordination du métal Zn<sup>2+</sup>. De façon inattendue, cette nouvelle structure cristallographique de TsaBDE montre que l'histidine 136 participe à la liaison du métal Zn<sup>2+</sup>, dont l'équivalent, l'histidine 139, ne participe pas à liaison du Fe<sup>2+</sup> dans EcTsaD en conformation fermée (déviations de ~ 3 Å du groupement N<sup>ε1</sup>). Cela suggère que la conformation ouverte et la présence du métal à proximité oriente l'H<sub>136</sub> de manière favorable à la coordination. Lors du passage vers la conformation fermée, la coordination serait rompue entre le métal Zn<sup>2+</sup> et l'H<sub>136</sub>, permettant la coordination du Zn<sup>2+</sup> au nucléotide. Il est probable que cette structure soit un autre état intermédiaire passant de la forme inactive à active.

### **I.3. Conclusion de l'article I.**

La catalyse de la modification t<sup>6</sup>A par les protéines TsaBDE implique 2 substrats : le premier étant l'intermédiaire TC-AMP et le second est l'ARNt.

Les structures TmTsaBDE supportent les données biochimiques mais laissent encore beaucoup de zones d'ombre en ce qui concerne la compréhension du mécanisme réactionnel. Même si TsaE est indispensable pour la réaction, sa position dans le complexe ternaire TsaBDE est incompatible avec une réaction en une seule étape. En effet, en présence d'un nucléotide non-hydrolysable, et sans doute aussi en présence d'ATP, le site actif de TsaD n'est pas accessible pour un substrat volumineux tel que l'ARNt en présence de TsaE. De plus, l'ouverture de TsaD et les réarrangements structuraux de son site actif induits par la présence de TsaE suggèrent que la sous-unité catalytique soit dans un état inactif dans le contexte du complexe ternaire TsaBDE fixant un nucléotide. Des expériences d'interaction ont montré que la présence de TsaE empêche la formation d'un complexe ARNt-TsaBDE, en revanche son absence permet de former le complexe nucléo-protéique ARNt-TsaBD (Luthra *et al.*, 2018).

Des expériences de cinétiques ont montré que TsaBD, seul, est capable de catalyser une réaction t<sup>6</sup>A ("single turn-over"), et que l'activité ATPase de TsaE est nécessaire pour atteindre plusieurs cycles de modification ("multiple turn-over") (Luthra *et al.*, 2018). La catalyse de la



modification t<sup>6</sup>A sur un ARNt substrat laisserait TsaD dans un état inactif que seule la présence de TsaE, accompagnée de l'hydrolyse d'une molécule d'ATP serait capable de restaurer.

Les différentes structures cristallographiques représentent probablement des conformations illustrant les différentes étapes ("snapshots") le long du chemin réactionnel et ont permis de proposer une fonction de la troisième sous-unité TsaE. Cependant, de nombreuses incertitudes demeurent quant à l'ordre d'interaction des différents substrats. Dans le complexe TsaBDE, un nucléotide (AMPCPP ou cax-AMP) est stabilisé dans le site actif de TsaD : l'intermédiaire TC-AMP pourrait être ainsi transmis au site actif de TsaD dans une conformation ouverte. L'ion Zn<sup>2+</sup> observé dans le site actif de TmTsaD de la structure du complexe TmTsaBDE de l'équipe de Manal Swarjo suggère un état pré-actif qui précéderait la dissociation de l'interaction avec TsaE et le rapprochement des deux domaines de la sous unité catalytique TsaD induisant, ainsi, la fermeture du site actif. Il est possible que la fermeture du site actif soit aussi induite après la fixation de l'ARNt substrat.

## Chapitre II

### I. Etude du site actif de TsaD chez les procaryotes.

#### II.1. Présentation de l'article II (soumis à NAR).

Comme énoncé précédemment, la catalyse de la modification t<sup>6</sup>A se déroule en deux étapes (figure 53 page 85). Les protéines TsaC<sub>1</sub> ou TsaC<sub>2</sub> forment un intermédiaire TC-AMP, ce dernier est pris en charge par le complexe trimérique TsaBDE pour modifier l'adénosine 37 des ARNt substrats. Le mécanisme de transfert du groupement thréonyl-carbamoylé sur les ARNt substrats est très peu comprise. Seule la structure tridimensionnelle de Sua5 a été résolue en présence de l'intermédiaire TC-AMP permettant d'identifier les résidus responsables de sa fixation (Agari *et al.*, 2008). Aucune structure tridimensionnelle de TsaD ou de ses homologues (Kae1, Qri7) n'a été résolue en présence de cet intermédiaire. Des structures en présence de nucléotides ATP, ADP ou AMP liés au site actif ont été obtenues mais permettent seulement d'émettre des hypothèses quant aux résidus impliqués dans la fixation de l'intermédiaire TC-AMP.

Notre article présente la structure de TsaBD de la bactérie *Escherichia coli* en présence d'un analogue TC-AMP non hydrolysable (nom de code BK951 ; figure 55). Le protocole de purification de TsaBD est exposé en annexe 1 (page 319). Ce composé a été synthétisé par Charles Lauhon de l'université du Wisconsin. A partir de la molécule 6-chloropurine riboside et de L-thréonine, une synthèse chimique a permis de substituer un oxygène du phosphate du TC-AMP par un carbone, ce qui le rend non hydrolysable et stable.

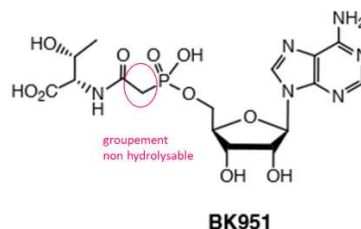


Figure 55 Produit finale de la synthèse de l'analogue non hydrolysable du TC-AMP nommé BK951, la synthèse a été réalisée par l'équipe de Charles Lauhon.

L'analyse complète de cette étude est illustrée par l'article en cours de rédaction et de publication intitulé " Structure of a reaction intermediate mimic in t<sup>6</sup>A biosynthesis bound in the active site of the TsaBD heterodimer from *Escherichia coli*" en collaboration avec Charles T.

Lauhon (Université de Madison, USA). Les résultats présentés seront comparés à ceux obtenus avec la structure de TsaBD de *Salmonella typhimurium* qui possède un analogue de l'ATP non hydrolysable dans son site actif (ATP $\gamma$ S) (code PDB 3ZEU) et d'*Escherichia coli* possédant un ADP dans son site actif (code PDB 4YDU). Pour finir, une dernière comparaison sera faite avec la structure tridimensionnelle de TobZ, protéine possédant les deux domaines homologues à TsaC<sub>1</sub> et TsaD et, au sein de son domaine TsaD, un adénylate carbamoylé et un ion Fe<sup>2+</sup> sont fixés (code PDB 3VER).

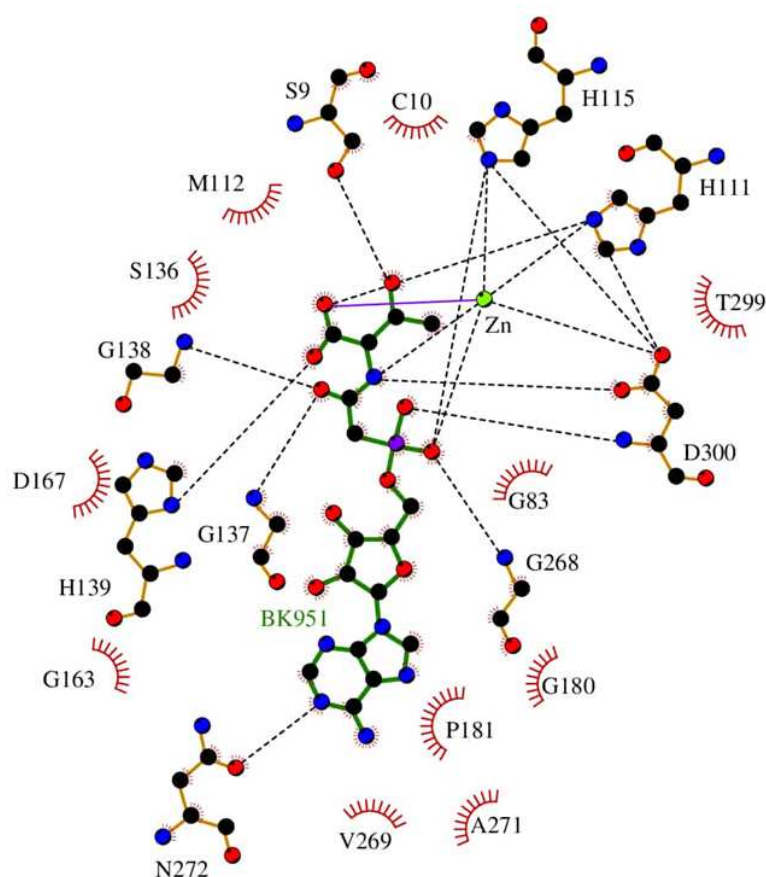


Figure 56 Site de fixation de la molécule BK951 (en vert) dans le site actif de TsaD. Les liaisons hydrogènes sont représentées par les pointillés noirs. La liaison de coordination entre le Zn<sup>2+</sup> et l'oxygène du carboxylate du BK951 est représenté en violet. Enfin les demi cercles rouges hachurés représentent les interactions de van der Waals.

Des analyses d'interaction ont montré que l'affinité de BK951 avec TsaBD est de l'ordre du micromolaire ( $K_D = 3.4 \pm 0.6 \mu\text{M}$  et une  $IC_{50}$  (concentration inhibitrice médiane) de  $1.3 \pm 0.5 \mu\text{M}$ ). Ces valeurs sont du même ordre de grandeur que celles retrouvées pour la fixation du TC-AMP ( $IC_{50} = 1.5 \pm 0.4 \mu\text{M}$ ). La structure du complexe TsaBD/BK951 a été résolue à une

résolution de 2.31 Å. Une densité électronique supplémentaire dans le site actif de TsaD a permis de placer la molécule BK951 ainsi qu'un ion  $Zn^{2+}$  dans le site actif de TsaD. Les structures de TsaBD, en présence et en absence de la molécule BK951, sont très similaires (RMSD  $\sim$  0.39 Å) montrant que la fixation de la molécule n'induit pas de changement conformationnel important. L'ion  $Zn^{2+}$  est coordonné par les acides aminés H<sub>111</sub>, H<sub>115</sub>, D<sub>300</sub> d'une part et par l'oxygène du phosphate  $\alpha$  et les oxygènes du carboxylate du BK951 de l'autre. La S<sub>9</sub> et l'H<sub>139</sub> forment des liaisons hydrogènes avec le groupement thréonyle et le groupement carboxylate respectivement. Les résidus N<sub>272</sub>, G<sub>268</sub>, G<sub>127</sub> et G<sub>138</sub> forment des liaisons hydrogènes avec l'adénosine, le phosphonate et le carbamate (figure 56).

Une étude de mutagenèse dirigée, réalisée par les collaborateurs, ciblant les acides aminés du site actif de TsaD de la bactérie *Escherichia coli* a été réalisée. Les mutants H<sub>111</sub>A, H<sub>115</sub>A et D<sub>300</sub>A ne possèdent plus d'activité t<sup>6</sup>A. A l'inverse, le mutant H<sub>139</sub>A ne montre pas de différence d'activité significative par rapport à l'enzyme sauvage suggérant que l'H<sub>139</sub> n'est pas un résidu catalytique. De façon surprenante, les mutants S<sub>9</sub>A et S<sub>9</sub>G diminuent significativement l'activité t<sup>6</sup>A de l'enzyme et le mutant S<sub>9</sub>T diminue au tiers l'activité. La cystéine voisine mutée, C<sub>10</sub>A, qui forme une interaction de van der Waals avec le groupement thréonyle, montre un effet moindre sur l'activité t<sup>6</sup>A. En revanche, le double mutant S<sub>9</sub>G et C<sub>10</sub>G abolit totalement l'activité de l'enzyme, suggérant un effet additif.



# Structure of a reaction intermediate mimic in t<sup>6</sup>A biosynthesis bound in the active site of the TsaBD heterodimer from *Escherichia coli*

Brett J. Kopina<sup>1,†</sup>, Sophia Missouri<sup>2,†</sup>, Bruno Collinet<sup>2,3</sup>, Mark G. Fulton<sup>1</sup>, Charles Cirio<sup>2</sup>, Herman van Tilbeurgh<sup>2,\*</sup> and Charles T. Lauhon<sup>1,\*</sup>

<sup>1</sup>Pharmaceutical Sciences Division, School of Pharmacy, University of Wisconsin, Madison, WI 53705, USA, <sup>2</sup>Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France and <sup>3</sup>Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (IMPMC), Sorbonne-Université, UMR7590 CNRS, MNHN, Paris, France

Received September 25, 2020; Revised December 21, 2020; Editorial Decision January 04, 2021; Accepted January 15, 2021

## ABSTRACT

The tRNA modification N6-threonylcarbamoyladenine (t<sup>6</sup>A) is universally conserved in all organisms. In bacteria, the biosynthesis of t<sup>6</sup>A requires four proteins (TsaBCDE) that catalyze the formation of t<sup>6</sup>A via the unstable intermediate L-threonylcarbamoyladenylate (TC-AMP). While the formation and stability of this intermediate has been studied in detail, the mechanism of its transfer to A37 in tRNA is poorly understood. To investigate this step, the structure of the TsaBD heterodimer from *Escherichia coli* has been solved bound to a stable phosphonate isosteric mimic of TC-AMP. The phosphonate inhibits t<sup>6</sup>A synthesis *in vitro* with an IC<sub>50</sub> value of 1.3 μM in the presence of millimolar ATP and L-threonine. The inhibitor binds to TsaBD by coordination to the active site Zn atom via an oxygen atom from both the phosphonate and the carboxylate moieties. The bound conformation of the inhibitor suggests that the catalysis exploits a putative oxyanion hole created by a conserved active site loop of TsaD and that the metal essentially serves as a binding scaffold for the intermediate. The phosphonate bound crystal structure should be useful for the rational design of potent, drug-like small molecule inhibitors as mechanistic probes or potentially novel antibiotics.

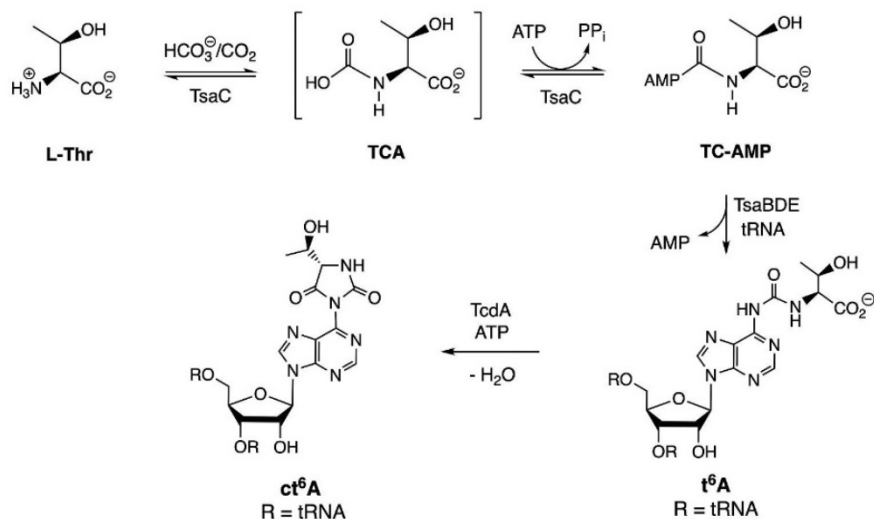
## INTRODUCTION

Transfer RNAs from all forms of life are subject to a large number of enzyme-catalysed post-transcriptional chemical

modifications that fine tune both its structure and function in the cell (1–4). The N6-threonylcarbamoyladenine (t<sup>6</sup>A) modification (5) is found at position 37 adjacent to the anticodon in tRNAs that decode 5'-ANN codons (6,7), and is universally conserved throughout life (8). Structurally, t<sup>6</sup>A enhances codon recognition by disrupting base pairing with U33 in the anticodon loop and by interacting with both anticodon bases as well as the first base of the mRNA codon (9). Mutants in both bacteria (10,11) and eukaryotes (11,12) that lack t<sup>6</sup>A typically have severe growth defects or, as in the case of *Escherichia coli*, are inviable (13). In many organisms t<sup>6</sup>A is further modified to the hydrolytically labile cyclized hydantoin ct<sup>6</sup>A, which is formed by ATP-catalysed cyclization of t<sup>6</sup>A (14,15) (Figure 1). This has prompted a call for re-evaluation of the structural explanation for its role in codon recognition. The catalytic machinery for the biosynthesis of t<sup>6</sup>A differs significantly between bacteria and eukaryotes/archaea. In eukaryotes, t<sup>6</sup>A is catalysed by the Tcs1 (SUA5) protein (10) in conjunction with the KEOPS complex (16–19), which contains five proteins. KEOPS was originally identified as a transcription complex (20) involved in telomere regulation (21). In addition to its requirement for t<sup>6</sup>A tRNA modification, KEOPS proteins in yeast have t<sup>6</sup>A-independent roles in telomere replication via SUA5p (22) and as an auxiliary sensor of DNA double strand breaks (23). KEOPS mutations have been recently implicated in neurodegeneration with nephrotic disease in humans (24–26). In mitochondria, a minimal t<sup>6</sup>A system operates, consisting of Tcs1 and the mitochondrial protein Qri7 (27). Expression of Qri7 in the cytoplasm can complement loss of KEOPS proteins in t<sup>6</sup>A formation. This result has implicated the KEOPS Qri7 homolog OSGEP (Tcs3) as the main catalytic component responsible for t<sup>6</sup>A syn-

\*To whom correspondence should be addressed. Tel: +1 608 262 3083; Email: clauhon@wisc.edu  
Correspondence may also be addressed to Herman van Tilbeurgh. Email: herman.van-tilbeurgh@universite-paris-saclay.fr  
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.  
Present address: Brett J. Kopina, Pfizer Inc. 275 N. Field Dr., Lake Forest, IL 60035, USA.

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)



**Figure 1.** Biosynthesis of  $t^6\text{A}$  in bacteria. In most bacterial tRNAs,  $t^6\text{A}$  is further cyclized to the hydantoin  $ct^6\text{A}$ , which is hydrolytically unstable to commonly used isolation procedures.

thesis. In human mitochondria,  $t^6\text{A}$  formation was recently found to function as a  $\text{CO}_2$  sensor, in that low  $\text{CO}_2$  levels result in low  $t^6\text{A}$  formation and reduced translation of mitochondrial proteins (28). Based on the dependence of  $t^6\text{A}$  in bacteria for normal growth, and the apparent divergence of the protein machinery between bacterial and eukaryotic  $t^6\text{A}$  synthesis, this modification has been identified as a potential antibiotic target.

As shown in Figure 1, the biosynthesis of  $t^6\text{A}$  in bacteria involves four distinct proteins (29). The first step in the biosynthesis is carried out by TsaC1(YrdC)/TsaC2 (SUA5), which are short/long bacterial homologs, respectively, of eukaryotic Tcs1/SUA5. These enzymes catalyse an L-threonine-dependent adenylation reaction, using bicarbonate (possibly via  $\text{CO}_2$ ), to form a putative L-threonine carbamate that is adenylated with ATP to give the unstable product TC-AMP (30). This mechanism was also shown to operate *in vitro* for both eukaryotes and archaea (31). There is evidence that although TsaC1 and TsaC2 catalyse the same reaction, they may operate by different mechanisms due to the involvement in TsaC2 of additional hinge and SUA5 domains (32). The TsaBDE proteins together in the presence of ATP can catalyse the addition reaction of exogenously supplied TC-AMP with the tRNA substrate to give  $t^6\text{A}$  and AMP as products (30). Similarly, in yeast it was shown that SUA5p and Qri7 can catalyse  $t^6\text{A}$  formation via TC-AMP diffusion without direct protein-protein interaction (33). The bacterial TsaD protein is a homolog to the ancestral (COG0533) family of carbamoyltransferase proteins that include KAE1/OSGEP/TobZ that utilize carbamoyl adenylate substrates or intermediates to carbamoylate nucleophiles (34,35). Previous structural and mechanistic studies in *Salmonella typhimurium* (36), *E. coli* (37) and *Thermotoga maritima* (38–41) have shown a functional requirement for the formation of a TsaBD heterodimer that binds tRNA (38) and likely binds TC-AMP. The role of the small TsaE protein in  $t^6\text{A}$  formation is more complicated. It

was characterized initially as an ATPase with low intrinsic activity (42–44) that is stimulated by the presence of TsaBD (29,37). As shown in Figure 1, the chemical transformation from TC-AMP to  $t^6\text{A}$  does not formally require ATP. Indeed, bacterial  $t^6\text{A}$  formation is observed *in vitro* in the absence of this protein (30); however, as explicitly shown in *T. maritima* (38), TsaE catalysed ATP-hydrolysis is required for multiple turnover.

A comparison of structural studies in bacteria show that the quaternary structure for the TsaBDE complex is species dependent. In *E. coli* and *S. typhimurium*, small angle X-ray scattering (SAXS) and isothermal calorimetry (ITC), respectively, were used to determine the ternary complex is a TsaB<sub>1</sub>D<sub>1</sub>E<sub>1</sub> heterotrimer, the formation of which is ATP-dependent (36,37). In *E. coli*, there is a novel additional ADP binding site at the TsaB–TsaD interface (37). The *T. maritima* proteins form a hexameric TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub> complex as a dimer of heterotrimers with TmTsaB as the dimerization domain (38,39). The role of the TsaE protein in both systems appears to be regulatory in nature, in that TmTsaE binding to the TmTsaB<sub>2</sub>D<sub>2</sub>E<sub>1</sub> complex has been shown to be competitive with tRNA (38). This is consistent with the observation in *E. coli* that EcTsaBDE does not bind tRNA (37), while EcTsaBD does (38). TsaE-catalysed ATP hydrolysis is required for additional turnover, possibly via a resetting of the active conformation of TsaBD (38). Missouri *et al.* (39,40) observed that TmTsaE binding to TmTsaBD induces a partial denaturing of the TsaD active site. Both studies are in agreement that TmTsaE binding precludes the binding of tRNA. Questions remain as to the exact mechanism and timing of the conformational effects elicited by TsaE during the catalytic cycle, as well as the timing of tRNA binding relative to TC-AMP loading onto TsaD and the potential role of TsaC1/C2 in the latter. Each of these steps may be specifically tailored in mesophiles versus thermophiles, due in part to the instability of TC-AMP and the need to avoid wasteful ATP utilization.

While a structure of TC-AMP bound to TsaC2 from *Sulfolobus tokodaii* has been reported (34,45), there are as yet no structures of TC-AMP bound to the TsaBD heterodimer. Luthra *et al.* (41) have recently reported the structure of *T. maritima* TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub> with the unstable ligand carboxy-AMP bound in the TsaD active site. This structure was used to model TC-AMP binding to the complex. Efforts at solving a structure of TsaBD with TC-AMP itself have been hampered by the known instability of this substrate (30). In an effort to circumvent this barrier, we report here the preparation of chemically stable analogs of TC-AMP and an assessment of their ability to inhibit the formation of t<sup>6</sup>A in the *E. coli* system. The most potent inhibitor is an isosteric phosphonate analog, in which the adenylate ester oxygen of TC-AMP is replaced by a methylene group. We show here that this stable analog binds to the EcTsaBD heterodimer with affinity in the low micromolar range and we report the crystal structure of the bound complex. Interestingly, the active site Zn ion interacts with the inhibitor by chelating oxygens from both the carboxylate of the L-threonine moiety as well as one of the oxygen atoms of the phosphonate. The structure suggests that TsaBD binds TC-AMP inside the catalytic pocket in an orientation that may be initially sterically unreactive but a simple conformational rearrangement may lead to a more reactive bound form upon binding of tRNA.

## MATERIALS AND METHODS

### Protein purification and preparation of site-directed mutants

Proteins for kinetic assays and binding analysis were overproduced and purified as previously described (30). Site-directed mutants of *E. coli* TsaD were prepared using Quikchange mutagenesis (Agilent Technologies) with *tsaD* cloned into pET28a as the template (30) and *pfu*Turbo DNA polymerase. Oligonucleotides for mutagenic PCR are from Integrated DNA Technologies, Inc. and are listed in Supplementary Table S1. Plasmids were purified using the QIAprep spin miniprep kit (QIAGEN) and both strands of the plasmid were sequenced across the entire gene to confirm the accuracy of the desired mutation (UW Biotechnology Center). For assays, proteins were quantified using the BioRad Bradford Protein Assay.

### Preparation of tRNA transcripts

Unmodified tRNA was prepared by *in vitro* runoff transcription using overlapping oligonucleotides essentially as described (46,47). T4 DNA polymerase Klenow fragment (Promega) was used to extend overlapping oligonucleotides containing a 12-bp overlapping region and a fused T7 promoter sequence. For *E. coli* tRNA<sup>Lys</sup>, the oligos utilized were the following: Plus strand: 5'-TAATACGACTCACTATAGGGTCGTTAGCTCA GTTGGTAGAGCAGTTGACTTTTAATC and minus strand: 5'-TGGTGGGTCGTGCAGGAATTGAACC TGCGACCAATTGATTAAGTCA (T7 promoter underlined and overlapping bases in italics). The extension reaction used 1 mM each oligo and 0.4 mM dNTPs with 10 units/ml polymerase in the manufacturer supplied buffer at 37°C for 2 h. DNA was isolated by phenol then chloroform

extraction and ethanol precipitation. The redissolved DNA was then used directly for *in vitro* transcription. Transcription reactions contained 1 μM template DNA in a reaction mixture consisting of 50 mM Tris, pH 7.8, 35 mM MgCl<sub>2</sub>, 1 mM spermidine, 5 mM of each NTP, 20 units/ml RNasin (Promega) and 1000 units of T7 RNA polymerase (Lucigen). After incubating at 37°C overnight, 10 units of RQ1 RNase-free DNase was added directly to the mixture and incubated at 37°C for 30 min. After addition of an equal volume of 7.5 M urea with 0.1% bromophenol blue, the RNA was purified by polyacrylamide gel electrophoresis using 8% slab gels and run at 20 W for 1.5 h. RNA bands were visualized by UV shadowing on a fluorescent TLC plate (Merck). The RNA was excised and eluted in 0.5 M NaCl, filtered and then precipitated with 2.5 vol ethanol at -20°C. After centrifugation at 14 000 × g at 4°C (Beckman Avanti-J 25.50 rotor), the RNA pellet was washed with cold 70% ethanol, air-dried briefly and resuspended in deionized water and stored at -80°C. Typical yields were 1.6 mg of tRNA transcript/ml of transcription reaction.

### *In vitro* assay for t<sup>6</sup>A formation

An HPLC assay was used to quantify t<sup>6</sup>A by UV detection at 260 nm. Full reaction mixtures (50 μl) contained 50 mM MOPS pH 7.5, 20 mM NaHCO<sub>3</sub>, 25 mM KCl, 50 mM L-threonine, 1 mM ATP, 10 μM unmodified *E. coli* tRNA<sup>Lys</sup>, and 1 μM each of TsaC, TsaB, TsaD and TsaE. (For substrates with apparent *K<sub>m</sub>* values < 10 μM, 0.5 μM enzyme concentrations were used.) For kinetic studies, one of the substrates was varied while all other substrates were kept at the concentrations stated above. Reactions were initiated with enzymes by adding TsaE, TsaB, TsaD and then TsaC as the last component followed by mixing by vortexing for 1 s. Reaction mixtures were incubated at 23°C for 5 min and quenched by adding 25 μl of the reaction mixture to a new tube containing 4 μl 0.25 M EDTA pH 8 followed by vortexing and placing on dry ice. Once a set of reactions was complete, tubes were thawed and placed in a microcentrifuge (Eppendorf) and spun at 14 000 rpm for 10 s. Small molecules were removed by addition of the quenched reaction mixtures to micro biospin columns (BioRad) containing 0.7 ml (settled volume) Sephadex G-50 (fine), which had been drained by centrifugation at 7000 rpm for 1 min using a microcentrifuge. After careful addition of the reaction mixture to the top of the gel, the column was eluted by centrifugation at 7000 rpm for 1 min. To the resulting RNA eluate was added 0.1 volume of 0.3 M NaOAc, 1 mM Zn(OAc)<sub>2</sub>, pH 6, and the tRNA was digested by addition of 1.5 units of nuclease P1 and 0.05 units of alkaline phosphatase for ≥ 2 h at 37°C. The digested RNA (~45 μl) was analysed directly by HPLC using a Supelco LC-18S column (250 × 4.6 mm) similar to reported methods (48,49). The HPLC analysis used an isocratic buffer system consisting of 5 mM ammonium acetate pH 5.3, containing 4% (v/v) acetonitrile. Under these conditions the nuclease P1 reaction was the rate determining enzyme in the digestion. If insufficient nuclease or digestion time was used for this step, an additional peak is observed at a longer retention time (24 min versus 21 min for t<sup>6</sup>A). Material isolated from this peak was consistent with the dinucleotide 5'-t<sup>6</sup>ApA-3' as determined by



enzymatic analysis. For kinetic analysis of TC-AMP as the substrate, TsuC, L-threonine and bicarbonate were omitted, and the reaction was initiated by addition of TC-AMP.

#### Synthesis of TC-AMP mimic BK951 (1)

Figure 3 shows the overall synthetic scheme. Experimental details for the synthesis of analogs 2–5 can be found in the Supplementary Material.

#### *N*α-(2-Chloroacetyl)-*O*-benzyl-L-threonine benzyl ester (7)

Chloroacetyl chloride (3.19 ml, 0.040 mol) was added drop wise to a solution of K<sub>2</sub>CO<sub>3</sub> 5.8 g (0.041 mol) and *O*-benzyl-L-threonine benzyl ester **6** (10.91 g, 0.036 mol) in DCM (73 ml) at 0°C. Following the addition, the ice bath was removed and the mixture was allowed to reach room temperature before heating to reflux. After stirring for 2.5 h at reflux, TLC verified that all of the starting material was consumed. The mixture was allowed to cool to room temperature and partitioned between DCM (300 ml) and H<sub>2</sub>O (300 ml). The organic layer was separated and set aside while the aqueous layer was extracted with DCM (3 × 200 ml). The combined organic layers were then washed with saturated NH<sub>4</sub>Cl solution (1 × 300 ml) and brine (1 × 300 ml), dried with MgSO<sub>4</sub>, and filtered. The solvent was removed under reduced pressure to give **7** as a white solid (12.84 g, 94%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.07–7.44 (m, 11H), 5.15, 5.09 (ABq, *J*<sub>AB</sub> = 12.20 Hz, 2H), 4.67 (dd, *J* = 9.16, 2.32 Hz, 1H), 4.52, 4.30 (ABq, *J*<sub>AB</sub> = 12.00 Hz, 2H), 4.19 (qd, *J* = 6.19, 2.32 Hz, 1H), 4.11 (s, 2H), 1.22 (d, *J* = 6.19 Hz, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 170.0, 166.9, 138.0, 135.4, 128.9, 128.8, 128.7, 128.1, 127.9, 74.3, 71.1, 67.7, 57.4, 42.8, 16.7. MS (ESI): *m/z* 398 (M+Na)<sup>+</sup>. HRMS calculated for C<sub>20</sub>H<sub>22</sub>ClNO<sub>4</sub> (M+Na)<sup>+</sup>: 398.112957; Found: 349.11309.

#### *N*α-[2-(Dimethoxyphosphinyl)acetyl]-*O*-benzyl-L-threonine benzyl ester (8a)

A round bottom flask equipped with a stir bar and condenser was charged with **7** (1.88 g, 0.005 mol) and trimethyl phosphite (5.14 ml, 0.03 mol), and heated to reflux at 130°C. The reaction was allowed to stir at these conditions overnight. The next morning TLC was used to verify the absence of starting material. Excess trimethyl phosphite was distilled off and the crude product was purified by column chromatography on silica gel, using 3:1 EtOAc:Hexanes to 100% EtOAc gradient to yield **8a** as a yellow oil (1.94 g, 86%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.22–7.35 (m, 8H), 7.16–7.22 (m, 2H), 7.04 (d, *J* = 8.30 Hz, 1H), 5.14, 5.09 (ABq, *J*<sub>AB</sub> = 12.20 Hz, 2H), 4.69 (dd, *J* = 8.30, 2.20 Hz, 1H), 4.51, 4.29 (ABq, *J*<sub>AB</sub> = 11.96 Hz, 2H), 4.17 (qd, *J* = 6.30, 2.20 Hz, 1H), 3.71–3.85 (m, 6H), 2.89–3.04 (m, 2H), 1.25 (d, *J* = 6.30 Hz, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 170.3, 164.6, 138.1, 135.5, 128.8, 128.6, 128.5, 127.9, 127.9, 74.5, 71.0, 67.4, 57.4, 16.5. <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 23.90. MS (ESI): *m/z* 472 (M+Na)<sup>+</sup>. HRMS calculated for C<sub>22</sub>H<sub>28</sub>NO<sub>7</sub>P (M+Na)<sup>+</sup>: 472.14956; Found: 472.14959.

#### *N*α-[2-(Methoxyhydroxyphosphinyl)acetyl]-*O*-benzyl-L-threonine benzyl ester (8b)

A round bottom flask equipped with a stir bar was charged with **8a** (1.94 g, 0.00432 mol) and CH<sub>3</sub>CN (85 ml). After the addition of LiBr (0.750 g, 0.00864 mol), the mixture was brought to 75°C and stirred for 3 h. Upon completion, the solvent was removed under reduced pressure and the crude oil was diluted with H<sub>2</sub>O (200 ml) and 1M HCl solution (50 ml). This aqueous layer was extracted with DCM (4 × 75 ml). The organic fractions were combined, washed with brine (1 × 100 ml), dried with MgSO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure to give **8b** as a hard, pale white film (1.5 g, 80%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 11.05 (br. s., 1H), 6.96–7.53 (m, 11H), 4.95–5.22 (m, 2H), 4.74 (d, *J* = 8.89 Hz, 1H), 4.45, 4.24 (ABq, *J*<sub>AB</sub> = 11.75 Hz, 2H), 4.11 (m, 1H), 3.65 (m, 3H), 2.85–3.10 (m, 2H), 1.01–1.36 (m, 3H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 170.2, 166.5, 138.0, 135.5, 128.6, 128.5, 128.5, 128.4, 127.8, 74.5, 70.8, 67.3, 57.2, 16.2. <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 21.48. MS (ESI): *m/z* 434 (M-H)<sup>-</sup>. HRMS calculated for C<sub>21</sub>H<sub>26</sub>NO<sub>7</sub>P (M-H)<sup>-</sup>: 434.137412; Found: 434.13863.

#### 5'-*O*-[(Methoxy)(2-oxo-2(*O*-benzyl-L-threoninebenzylester)ethyl)phosphinyl]-6-chloro-9-(2',3'-*O*-isopropylidene-D-ribofuranosyl)purine (9)

To **8b** (0.233 g, 0.54 mmol), **9** (0.175 g, 0.54 mmol) and PPh<sub>3</sub> (0.238 g, 0.00108 mol) under argon was added distilled THF (1.3 ml). Upon dissolution, DEAD (40% in toluene, 0.424 ml, 0.00108 mol) was added drop wise. Stirring was continued for 20 h at room temperature at which point nearly all starting material was consumed according to TLC. The solvent was removed under reduced pressure and the crude product was taken directly to purification by column chromatography on silica gel using a 2:1 EtOAc:Hexanes to 100% EtOAc gradient. Compound **10** was isolated as a mix of diastereomers (2:1 ratio) as a white foam (0.315 g, 78%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.78 (m, 3H), 8.41 (s), 8.28 (s), 7.22–7.35 (m), 7.13–7.20 (m), 6.99–7.05 (m), 6.86–6.93 (m), 6.19 (d, *J* = 2.67 Hz, 2H), 6.17 (d, *J* = 2.94 Hz, 1H), 5.31–5.38 (m), 5.04–5.17 (m), 4.62–4.70 (m), 4.57 (m), 4.48 (m), 4.30–4.43 (m), 4.21–4.30 (m), 4.10–4.20 (m), 3.77 (m), 3.72 (m), 3.68 (m), 2.90–2.99 (m), 2.82–2.90 (m), 1.59–1.67 (m), 1.39 (s), 1.16–1.26 (m). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 169.9, 164.2, 164.2, 152.1, 152.0, 151.1, 151.0, 151.0, 144.5, 144.3, 137.6, 137.6, 135.1, 132.1, 132.1, 132.0, 132.0, 131.9, 131.8, 128.5, 128.4, 128.3, 128.2, 127.7, 127.6, 127.5, 114.9, 114.7, 90.8, 85.2, 85.2, 85.1, 85.0, 84.2, 84.0, 81.0, 74.4, 70.9, 70.8, 67.2, 66.0, 66.0, 65.7, 65.7, 57.2, 53.2, 53.2, 35.4, 35.3, 34.4, 34.2, 27.3, 27.2, 25.5, 16.3. <sup>31</sup>P NMR (162 MHz, CDCl<sub>3</sub>) δ 23.97, 23.60. MS (ESI): *m/z* 766 (M+Na)<sup>+</sup>. HRMS calculated for C<sub>34</sub>H<sub>39</sub>ClN<sub>5</sub>O<sub>10</sub>P (M+Na)<sup>+</sup>: 766.201527; Found: 766.20110.

#### 5'-*O*-[(Hydroxy)(2-oxo-2(*O*-benzyl-L-threoninebenzylester)ethyl)phosphinyl]-6-azido-9-(2',3'-*O*-isopropylidene-D-ribofuranosyl)purine sodium salt (11)

A round bottom flask was charged with a diastereomeric mix of **10** (0.4 g, 0.54 mmol) and stir bar. After addition

of DMF (5.83 ml) and  $\text{NaN}_3$  (0.087 g, 1.35 mmol), the reaction was allowed to stir overnight at 60°C. TLC the following morning confirmed the absence of starting material. Following the removal of solvent under reduced pressure, the crude oil product was purified directly by column chromatography on silica gel using 10%  $\text{H}_2\text{O}$  in  $\text{CH}_3\text{CN}$  as a running solvent. **11** was isolated as a single diastereomer, resembling an off white solid (0.3 g, 73%). It should be noted that a small fraction (>10%) of the sample appears to exist as the N1, C6 tetrazole tautomer in  $\text{CD}_3\text{OD}$ .  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  9.81 (s, 1H), 8.81 (s, 1H), 7.10–7.28 (m, 10H), 6.40 (d,  $J = 2.64$  Hz, 1H), 5.39 (br. s., 1H), 5.18 (d,  $J = 4.70$  Hz, 1H), 5.04 (s, 2H), 4.66 (m, 1H) 4.55 (br. s., 1H), 4.47, 4.25 (ABq,  $J_{\text{AB}} = 11.75$  Hz, 2H), 4.06–4.18 (m, 3H), 2.68–2.93 (m, 2H), 1.56–1.65 (m, 3H), 1.38 (s, 3H), 1.17 (m, 3H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  147.2, 145.0, 143.7, 139.9, 137.4, 130.0, 129.9, 129.8, 129.8, 129.4, 129.1, 115.7, 93.8, 88.4, 88.3, 87.1, 83.9, 76.2, 72.5, 68.7, 66.8, 28.2, 26.3, 17.2.  $^{31}\text{P}$  NMR (202 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  14.17. MS (ESI):  $m/z$  735 ( $\text{M}^-$ ). HRMS calculated for  $\text{C}_{33}\text{H}_{36}\text{N}_8\text{O}_{10}\text{P}$  ( $\text{M}^-$ ): 735.22975; Found: 735.22854.

#### 5'-O-[(Hydroxy)(2-oxo-2-(O-benzyl-L-threoninebenzylester)ethyl)phosphiny]-adenosine sodium salt (**1**, BK951)

A round bottom flask equipped with a stir bar was charged with **11** (0.085 g, 0.112 mmol) and  $\text{PdCl}_2$  (0.02 g, 0.112 mmol), then subsequently flushed with nitrogen. MeOH (2.25 ml) and  $\text{H}_2\text{O}$  (2.25 ml) were added via glass syringe. Following dissolution, the round bottom flask was equipped with a hydrogen balloon and the reaction was allowed to stir for 26 h. At this point, crude NMR revealed that a majority of the reaction mixture was the desired product. The minor product present was the debenzylated 2', 3'-O-isopropylideneadenosine analogue. TFA (0.2 ml) was added and the reaction was stirred for an additional hour in order to convert the remaining isopropylidene analogue to the desired product. The solvent was then removed under reduced pressure and the crude reaction mixture was taken up in MeOH and filtered through celite. The filtered product was then further purified by HPLC using a Vydac C18 prep column (250  $\times$  30 mm) and the following gradient: A = 10 mM ammonium bicarbonate; B = 10 mM ammonium bicarbonate, 20% (v/v) MeOH. 0–12 min: 100% A. 12–40 min: 0  $\rightarrow$  100% A  $\rightarrow$  B to give (0.02 g, 36%) of **BK951 (1)** as a transparent foam.  $^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  8.48 (s, 1H), 8.22 (s, 1H), 6.13 (d,  $J = 5.74$  Hz, 1H), 4.77 (dd,  $J = 5.74, 4.31$  Hz, 1H), 4.52 (m, 1H), 4.41 (br. s., 1H), 4.12–4.30 (m, 4H), 2.80–3.02 (m, 2H), 1.16 (d,  $J = 6.25$  Hz, 3H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  176.3, 168.7, 155.0, 152.3, 148.5, 139.3, 118.1, 86.4, 83.5, 73.8, 70.0, 67.8, 63.6, 60.3, 36.3, 35.3, 18.9.  $^{31}\text{P}$  NMR (162 MHz,  $\text{D}_2\text{O}$ )  $\delta$  16.21. MS (ESI):  $m/z$  513 ( $\text{M}+\text{Na}^+$ ). HRMS calculated for  $\text{C}_{16}\text{H}_{23}\text{N}_6\text{O}_{10}\text{P}$  ( $\text{M}+\text{Na}^+$ ): 513.110548; Found: 513.10904.

#### Measurement of $\text{IC}_{50}$ values for inhibitors

Reaction mixtures and the procedure for measuring  $\text{IC}_{50}$  values for the overall reaction were the same as described

above, except the concentration of L-threonine was reduced to 1 mM and the inhibitors were preincubated with enzymes in the reaction mixture for 1 min. Reactions were initiated by the addition of ATP. For measuring  $\text{IC}_{50}$  values for the isolated second step of the reaction, the reaction mixtures contained 10  $\mu\text{M}$  TC-AMP, 1 mM ATP, 0.5  $\mu\text{M}$  of EcTsaBDE and variable amounts of BK951 in a total volume of 50  $\mu\text{l}$ . Reactions were initiated by addition of TC-AMP and incubated at 23°C for 5 min and worked up and analysed as described for the full reaction. Measurement of  $\text{IC}_{50}$  values for TsaC were performed with reaction mixtures that were the same as for the overall assays except TsaBDE and tRNA were omitted. TsaC was incubated with all components except for ATP for 2 min at 23°C before initiating the reaction by the addition of ATP. After incubating for 1 min at 23°C, the mixture was injected directly onto a Supelco LC-18T HPLC column (4.6  $\times$  250 mm). Isocratic elution was used with 50 mM potassium phosphate pH 6 and 1% acetonitrile and the amount of TC-AMP was measured by UV detection at 260 nm.

#### Enzymatic preparation of TC-AMP and SC-AMP

Reaction mixtures (80  $\mu\text{l}$ ) contained 50 mM MOPS pH 7.5, 10 mM  $\text{MgCl}_2$ , 25 mM KCl, 2 mM ATP, 50 mM L-threonine, 40 mM  $\text{NaHCO}_3$  and 42  $\mu\text{g}$  (37  $\mu\text{M}$ ) of EcTsaC. After incubating at 23°C for 10 min, the reaction mixture was injected onto a Supelco LC-18T column (250 mm  $\times$  4.6 mm) and eluted with an isocratic buffer of 50 mM potassium phosphate pH 6.0. Typical yields were 25 nmol at a concentration of 20–30  $\mu\text{M}$  for TC-AMP. Solutions could be carefully concentrated by rotary evaporation (water bath < 30°C) and used directly or aliquoted and stored at –80°C. (Approximately 10% degradation was typically observed for each freeze/thaw cycle.) The serine analog of TC-AMP (SC-AMP) was also prepared by substituting L-serine for L-threonine and purifying by an analogous method. Typical yields for SC-AMP were 10 nmol per reaction. The identity of SC-AMP was confirmed by accurate mass determination (HRMS Calcd for  $\text{MH}^+$  479.09222; Found 479.09214) and purity was assessed by HPLC (>95%, Supplementary Figure S5).

#### Data analysis for determining kinetically derived parameters

For kinetic analysis, concentrations for all fixed substrates were the same as described above for the full reaction mixture. Reactions were run in duplicate and values for  $K_m$  and  $k_{\text{cat}}$  were derived from fitting to the standard Michaelis–Menten model (Equation 1). For substrates with  $K_m$  values near the enzyme concentrations ( $\leq 5$ -fold), kinetic values were derived from Morrison's equation (50) for tight binding substrates (Equation 2). For kinetics that show substrate inhibition, the data were fit to a simple model for complete substrate inhibition in which the complex containing two bound substrates is inactive (Equation 3). Error values reported are the standard error for the fit. All data analysis was performed using Prism 8 (GraphPad) software. Values for  $k_{\text{cat}}$  in Tables 1 and 2 were obtained by dividing  $V_0$

( $V_{\max}$ ) values obtained from curve fitting by the known enzyme concentration.

$$v = \frac{V_0 [S]}{K_M + [S]} \quad (1)$$

$$v = V_0 \frac{([E_T] + [S_T] + K_M) - \sqrt{([E_T] + [S_T] + K_M)^2 - 4[E_T][S_T]}}{2[E_T]} \quad (2)$$

$$v = \frac{V_0 [S]}{K_M + [S] \left(1 + \frac{[S]}{K_i}\right)} \quad (3)$$

### Microscale thermophoresis (MST) binding studies

MST measurements were performed using a Monolith NT115.1 Pico instrument (Nanotemper Technologies). EcTsaD was fluorescently labelled with the Monolith Red-NHS second generation dye (cat. no. MO-L011) using the manufacturer's protocol. Briefly, purified His<sub>6</sub>-TsaD (20  $\mu$ M) in buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 5% glycerol) was mixed with an equal volume (100  $\mu$ l) of dye reagent (60  $\mu$ M) in same buffer containing 5% DMSO for 25 min at room temperature in the dark. Excess reagent was removed using gel exclusion chromatography with buffer described above and column provided in the kit to give the cyan-colored protein. Fluorescently labelled TsaD was analysed for t<sup>6</sup>A activity using the full reaction HPLC assay as described above and found to be 95% active relative to the unlabelled enzyme. Aliquots of the labelled protein were frozen in dry ice/acetone and stored at  $-80^\circ\text{C}$ . Binding measurements were conducted in t<sup>6</sup>A Buffer (equal to full assay buffer described above) containing PBS with 0.05% Tween 80. Full t<sup>6</sup>A activity was observed in this buffer. A stock concentration of 1  $\mu$ M EcTsaD: 2  $\mu$ M EcTsaB was prepared in binding buffer and dilutions were carried out to give aliquots of added to separate micro (PCR) striptubes containing 10 nM of the labelled TsaBD complex. After an initial addition of an equal volume (10  $\mu$ l) of 1 mM of the ligand BK951 was made to the first tube, serial dilutions were prepared to span the range of 0.1–500  $\mu$ M BK951 and 5 nM EcTsaBD complex. A glass capillary was then filled with the binding reaction mixture from each tube and placed in the MST tray. The capillaries were analysed for thermophoretic mobility. The instrument software analyses for fluorescence homogeneity in the samples and artefacts such as ligand induced autobleaching, as well as protein aggregation, which could lead to erroneous results. When suitable conditions are found, binding data is collected and analysed by the software package included with the instrument. Binding measurements were performed in triplicate.

### TCAMP stability studies in complex with EcTsaBD and EcTsaC

To study the effects of TsaBD on TC-AMP stability, HPLC-purified TC-AMP in 50 mM potassium phosphate was added to a solution of TsaBD in 50 mM HEPES pH 7.5, 0.3

M NaCl to give 100  $\mu$ l of 10  $\mu$ M TC-AMP with or without 40  $\mu$ M TsaBD. The mixtures were incubated at  $37^\circ\text{C}$  and 10  $\mu$ l timed aliquots were withdrawn and analysed by HPLC for AMP and TC-AMP by measuring peak areas using absorbance at 260 nm. Plots of  $\ln [\text{TC-AMP}]/[\text{TC-AMP} + \text{AMP}]$  vs time were analysed by linear regression (GraphPad Prism 8) to get a series of rate constants for first order TC-AMP decomposition to AMP. To measure the effect of TsaC on TC-AMP stability, EcTsaC was overproduced in BL21(DE3) cells from IPTG induction of the gene cloned in pET15b. The protein was purified using Ni-NTA chromatography and buffer exchanged with a 10 ml Sephadex G-50 column as previously described (30). This procedure typically gives 1 ml of 100–150  $\mu$ M EcTsaC, which contains 70–80% bound TC-AMP. The experiments described in this work used 130  $\mu$ M TsaC with 100  $\mu$ M (77%) bound TC-AMP as determined by HPLC and Bradford protein assay. The stability of this complex at different concentrations at  $37^\circ\text{C}$  in the same reaction buffer was analysed using the same HPLC protocol as described for TC-AMP:TsaBD mixtures to get a series of rate constants for each concentration of the complex.

### Fluorescence anisotropy binding measurements

The binding of the *E. coli* TsaBD complex to 3'Fluorescein-labelled tRNA<sup>Lys</sup><sub>UUU</sub> was monitored with a TECAN INFINITE F200PRO microplate reader by using a Nunc F96 microwell plate (Thermo Scientific). 40 nM of 3'Fluorescein-labelled tRNA probe was incubated in each well (final volume of 200  $\mu$ l) with increasing amounts of TsaBD (final concentrations between 0 and 3000 nM) in 20 mM HEPES pH 7.5, 4 mM NaCl and 2 mM DTT supplemented or not with BK951 (final concentration 50  $\mu$ M). The plate was incubated at  $22^\circ\text{C}$  in the dark for 30 min and then fluorescence anisotropy was recorded at 535 nm ( $\lambda_{\text{exc}} = 495$  nm). Titration curves have been obtained by fitting experimental values to the following hyperbolic saturation function:  $A = B_{\text{max}} \cdot [\text{TsaBD}] / (K_D + [\text{TsaBD}])$  using SigmaPlot (Systat) software. Measurements have been performed in triplicate and mean values have been used for fitting.

### Cloning, expression, purification and crystallization of TsaD and TsaB from *Escherichia coli*

Genes coding EcTsaD and EcTsaB have been cloned in pET9a expression plasmids and proteins have been produced and purified as previously described by Zhang *et al.* (37). TsaBD complex was concentrated to 10.1 mg.ml<sup>-1</sup> for crystallization trials. Crystals of heterodimer TsaBD from *Escherichia coli* were obtained by hanging-drop vapor diffusion after at  $18^\circ\text{C}$  by mixing 1  $\mu$ l of the protein solution pre incubated with 2.5 mM BK951 and 5 mM MgCl<sub>2</sub> with 1  $\mu$ l of 0.25 mM lithium sulfate 0.11 M HEPES pH 7.5 0.15 mM sodium acetate and 25% PEG 4K. The crystallization drop was equilibrated against a 500  $\mu$ l reservoir solution containing 0.25 mM lithium sulfate, 0.11 M HEPES pH 7.5, 0.15 mM sodium acetate and 25% PEG 4K. Crystals were cryo-protected by quick-soaking in crystallization buffer supplemented with 25% ethylene glycol, 3 mM BK951 and 6 mM MgCl<sub>2</sub> prior to flash freezing in liquid nitrogen.

**Table 1.** Summary of kinetically-derived parameters for t<sup>6</sup>A overall and partial reactions

Ligand	Overall reaction		TsaBDE reaction		TsaC reaction	
	$K_m$ ( $\mu$ M)	$k_{cat}$ ( $\text{min}^{-1}$ )	$K_m$ ( $\mu$ M)	$k_{cat}$ ( $\text{min}^{-1}$ )	$K_m$ (mM)	$k_{cat}$ ( $\text{min}^{-1}$ )
L-Thr	590 $\pm$ 90	0.42 $\pm$ 0.02	—	—	1.77 $\pm$ 0.26	9.0 $\pm$ 0.4
ATP	60 $\pm$ 6 ( $K_{is}$ = 2200) <sup>a</sup>	0.40 $\pm$ 0.4	14 $\pm$ 3 ( $K_{is}$ = 1800) <sup>a</sup>	0.12 $\pm$ 0.1	0.13 $\pm$ 0.01	7.9 $\pm$ 0.2
tRNA <sup>Lys</sup> <sub>UUU</sub>	0.5 $\pm$ 0.2 ( $K_d$ = 87 $\pm$ 11)	0.50 $\pm$ 0.03	nd <sup>b</sup>	nd		
TC-AMP			1.5 $\pm$ 0.4	0.31 $\pm$ 0.02		

<sup>a</sup>Calculated for observed substrate inhibition.<sup>b</sup>Not determined.**Table 2.** Summary of kinetically-derived inhibition and thermodynamically-derived binding constants of TCAMP and its analogs to t<sup>6</sup>A proteins

Ligand	IC <sub>50</sub> ( $\mu$ M)		
	Overall reaction <sup>a</sup>	TsaBDE reaction <sup>b</sup>	TsaC reaction <sup>c</sup>
BK951	1.3 $\pm$ 0.5	7.5 $\pm$ 1.1 ( $K_i$ = 2.5) <sup>d</sup> ( $K_d$ = 3.4 $\pm$ 0.6) <sup>e</sup>	>180
2	51.2 $\pm$ 1.4	123 $\pm$ 1.2 ( $K_i$ = 41) <sup>d</sup>	31 $\pm$ 1.1
3	45.5 $\pm$ 1.1	99 $\pm$ 1.1 ( $K_i$ = 33) <sup>d</sup>	97 $\pm$ 1.4
4	>1000	nd <sup>f</sup>	nd
5	>1000	Nd	nd
SC-AMP	9.3 $\pm$ 1.2	Nd	nd
5'-ADP	160 $\pm$ 20	Nd	nd
5'-AMP	113 $\pm$ 10	Nd	nd

<sup>a</sup>ATP = 1 mM, L-Thr = 0.6 mM.<sup>b</sup>ATP = 1mM.<sup>c</sup>ATP = 1.2 mM, L-Thr = 1.8 mM.<sup>d</sup>Calculated using Cheng-Prusoff equation (59).<sup>e</sup>Obtained using MST with fluorescently labelled EcTsaBD.<sup>f</sup>Not determined.

### Data collection and structure determination

X-ray diffraction data collection was carried out on beamline Proxima2 at the SOLEIL Synchrotron (Saint-Aubin, France) at 100 K. Data were processed, integrated and scaled with the XDS program package. TsaBD belonged to space group *P1* with unit cell parameters of  $a = 63.23$  Å,  $b = 69.25$  Å,  $c = 86.54$  Å;  $\alpha = 109.38^\circ$ ,  $\beta = 92.116^\circ$ ,  $\gamma = 116.94^\circ$ . The structure was solved by molecular replacement using the PHASER module implemented in the CCP4 software package (search model was EcTsaBD, PDB ID: 4YDU). Two copies of the TsaBD dimer were found into the asymmetric unit of the crystal. The  $2F_o - F_c$  electronic density indicated the presence of the BK951 inhibitor bound to a zinc ion. The initial structure was refined using the BUSTER program and completed by interactive and manual model building using COOT. Data collection and refinement statistics are gathered in Table 3. The refined structure was deposited at the Protein Data Bank (code 6Z81). All figures of structures were generated with PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

### Molecular docking of adenine into the active site of TsaD

To find out how A37 of tRNA substrate could be oriented into the active site, molecular docking was performed using AutoDockTools 4.2. Adenine was retrieved from the PDB and converted using the Babel format molecule converter (<http://openbabel.org/>). We removed TsaB from the present crystal structure (keeping TsaD bound to BK951)

**Table 3.** Data collection and refinement statistics

	EcTsaBD + BK951 + Zn
PDB code	<b>6Z81</b>
Number of copies in asymmetric unit	2
Space group cell parameters	$P1$ : $a = 63.23$ Å, $b = 69.25$ Å, $c = 86.54$ Å, $\alpha = 109.38^\circ$ , $\beta = 92^\circ$ , $\gamma = 116.94^\circ$
Resolution (Å)	42.47–2.31 (2.33–2.31)
No. of total reflections	181 321 (15 816)
No. of unique reflections	51 130 (1023)
Multiplicity	3.5 (3.4)
$R_{pim}$	0.77 (0.525)
Completeness (%)	96.1 (88.76)
$I/\sigma(I)$	7.75 (1.71)
$CC_{1/2}$	0.992 (0.75)
$R_{crist}$ (%)	18.84 (26.95)
$R_{free}$ (%)	24.31 (32.70)
RMSD bond lengths (Å)	0.014
RMSD bond angles ( $^\circ$ )	1.81
Average $B$ ( $\text{\AA}^2$ ) Protein Ligand Solvent	38.47 38.38 50.27 31.60
Clashscore <sup>a</sup> MolProbity score <sup>a</sup>	3.63 1.85
Ramachandran plot (%) Favoured Allowed Outliers	96.64 3.36 0.00

Values for the highest resolution shell are in parentheses.  $CC_{1/2}$  = percentage of correlation between intensities from random half-dataset.

<sup>a</sup>Calculated with MolProbity.

and added Kollman charges. The Lamarckian Genetic Algorithm (LGA) implemented in Autodock 4.2 was used for docking experiments. The docking parameters were as fol-

lows: 50 docking trials, population size 300 and 2 500 000 energy evaluations. We used a grid consisting of twenty 0.5 Å spacings in each dimension centered onto the active site.

## RESULTS

### Rationale and preparation of stable mimics of TC-AMP

Because TC-AMP is unstable, we sought to prepare chemically stable analogs that are close structural mimics in order to obtain inhibitors that could provide structural information on the bound conformation of TC-AMP itself when bound to TsABD. The structures of the mimics are shown in Figure 2. We began with the isosteric phosphonate BK951 (**1**). Simple replacement of the ester oxygen of the carbamate in TC-AMP gives a molecule that is stable to all of the observed TC-AMP degradation pathways that result in the formation of AMP. These include cyclization, direct hydrolysis or (potential) elimination pathways (30). For the synthesis (Figure 3), we envisioned a convergent strategy, wherein L-threonine and inosine would provide the sources of stereochemistry in the target molecule. The retrosynthesis employs a key disconnection between one of the phosphonate ester oxygens and the ribose 5'-carbon, which may be linked via Mitsunobu coupling (51). 6-Chloropurine riboside was used instead of an adenosine derivative to minimize the tendency of adenosine to cyclize at N3 during activation of the 5'-hydroxyl group (52). Protecting groups included an isopropylidene for the 2',3'-*cis*-diol of the riboside, while L-threonine was protected using benzyl groups at both the  $\beta$ -hydroxyl and carboxylic acid groups. The 6-chloro could be converted to a 6-azido functionality using azide, and the benzyl groups could then be removed during azide reduction to regenerate the adenosine moiety in a single step. As shown in Figure 3, the route begins with the starting dibenzyl protected L-threonine derivative **6**, which was reacted with chloroacetyl chloride to give 94% of the chloride **7**, which itself was converted to the dimethylphosphonate by Arbuzov reaction with trimethylphosphite in 86% yield. Dealkylation with LiBr in acetonitrile gave the phosphonate monoester, which was coupled with the isopropylidene protected 6-chloroinosine **10** in 78% yield. Reaction of **10** with sodium azide in DMF with heating gave the 6-azido purine derivative with concomitant demethylation to the phosphonic acid monoester. Global deprotection and reduction of the azide to the amine at N6 was achieved by hydrogenation using PdCl<sub>2</sub> in MeOH, followed by acidification with TFA to complete the removal of the isopropylidene group to give crude **1**. Purification of **1** by reversed phase HPLC gave the pure material in 36% yield from **10** on 100 mg scale. This material was >99% pure (Supplementary Figure S5) and used for both the enzymatic inhibition assays and structural analysis with TsABD.

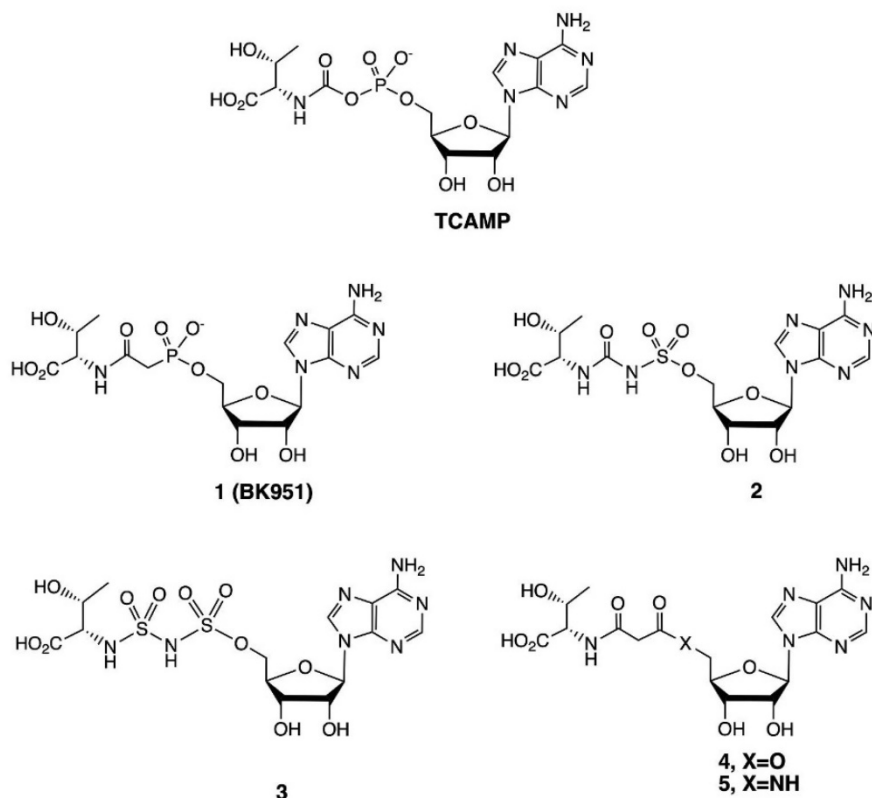
A series of other bioisosteres of TC-AMP were prepared (see Supplementary Data and Figures S1-S4 for experimental details). These include sulfamate isosteres **2** and **3**, in which either the phosphonate only or both phosphonate and the amide of the intermediate were replaced by sulfamoyl and sulfonylurea groups, respectively. Sulfamoyl groups are well described successful substitutions as phosphodiester isosteres for the preparation of adenylate mimics in the inhibition of aminoacyl-tRNA synthetases and

many other adenylate intermediates (53,54). For both **2** and **3**, the bridging nitrogen is expected to be deprotonated under physiological conditions. The alkylaminosulfonylsulfamate ester moiety of **3** is unprecedented and was found to be chemically stable. We chose this group due to the likelihood that the carbonyl carbon of the TC-AMP carbamate would form a tetrahedral sp<sup>3</sup>-hybridized geometry during attack of N6 of A37 of the tRNA in the carbamoyl transfer reaction catalysed by TsABDE. Such geometry could be selectively stabilized by TsABD binding. Compounds **4** and **5** employed malonates as the bridging group between the ribose and threonine and were designed to promote potential Zn binding to the active site of TsAD, via the enol tautomer of the beta-amido ester (**4**) or amide (**5**). Final products were purified by HPLC and characterized by NMR and High Resolution MS, which were consistent with the assigned structures (see Supplementary Material).

### Kinetic analysis of *E. coli* t<sup>6</sup>A reaction using *in vitro* assays

Before the TC-AMP mimics were tested for inhibition of t<sup>6</sup>A formation, we carried out an initial kinetic investigation of the overall reaction, as well as the half reaction for TC-AMP conversion to t<sup>6</sup>A, using a previously described HPLC assay (30). The assay is based on isolation and enzymatic digestion of the modified tRNA product and HPLC quantitation of t<sup>6</sup>A directly. This assay shows linear reaction rates (Supplementary Figure S6) and allowed us to approximate initial rates for fitting data to the Michaelis-Menten (MM) model (see Table 1). Because of the relatively weak affinity of the TsABDE ternary complex, protein concentrations of 0.5  $\mu$ M or above were required for the assays. We are thus reporting apparent kinetic parameters, due to the likely heterogeneity in protein complex formation and its effects on the initial rates for some substrate and inhibitor concentrations. For L-threonine in the overall reaction, fitting the kinetic data to the classical MM model gave an apparent value for  $K_m = 0.59 \pm 0.09$  mM (Figure 4A, Table 1). The kinetic data for ATP as the variable substrate show a good fit to MM kinetics up to a concentration of 1 mM with a value for  $K_m = 60 \pm 6$   $\mu$ M (Figure 4B). However, severe substrate inhibition was observed for higher ATP concentrations (Figure 4C). This inhibition shows a 50% reduction in the  $V_{max}$  at 5 mM ATP (90% at 10 mM). The physiological relevance of this inhibition is not clear. Although average intracellular ATP concentrations in *E. coli* are in the millimolar range (55,56), free ATP is unknown. Single cell analysis in *E. coli* has also shown a wide distribution of intracellular concentrations from 0.1 to 2 mM (57). The substrate inhibition observed is consistent with a similar observation recently reported by Swinehart *et al.* (58). Kinetic data with tRNA as the variable substrate was somewhat limited by the apparent low values of  $K_m$  for most tRNA transcripts. We were able to measure an apparent  $K_m = 0.5 \pm 0.2$   $\mu$ M for unmodified tRNA<sup>Lys</sup> (Figure 4D) by fitting data to the Morrison quadratic equation (50), which is used to account for substrate depletion in reactions in which the tRNA substrate is close to the enzyme concentration.

In addition to measuring t<sup>6</sup>A formation with the full reaction mixture, we could also isolate the second step of the



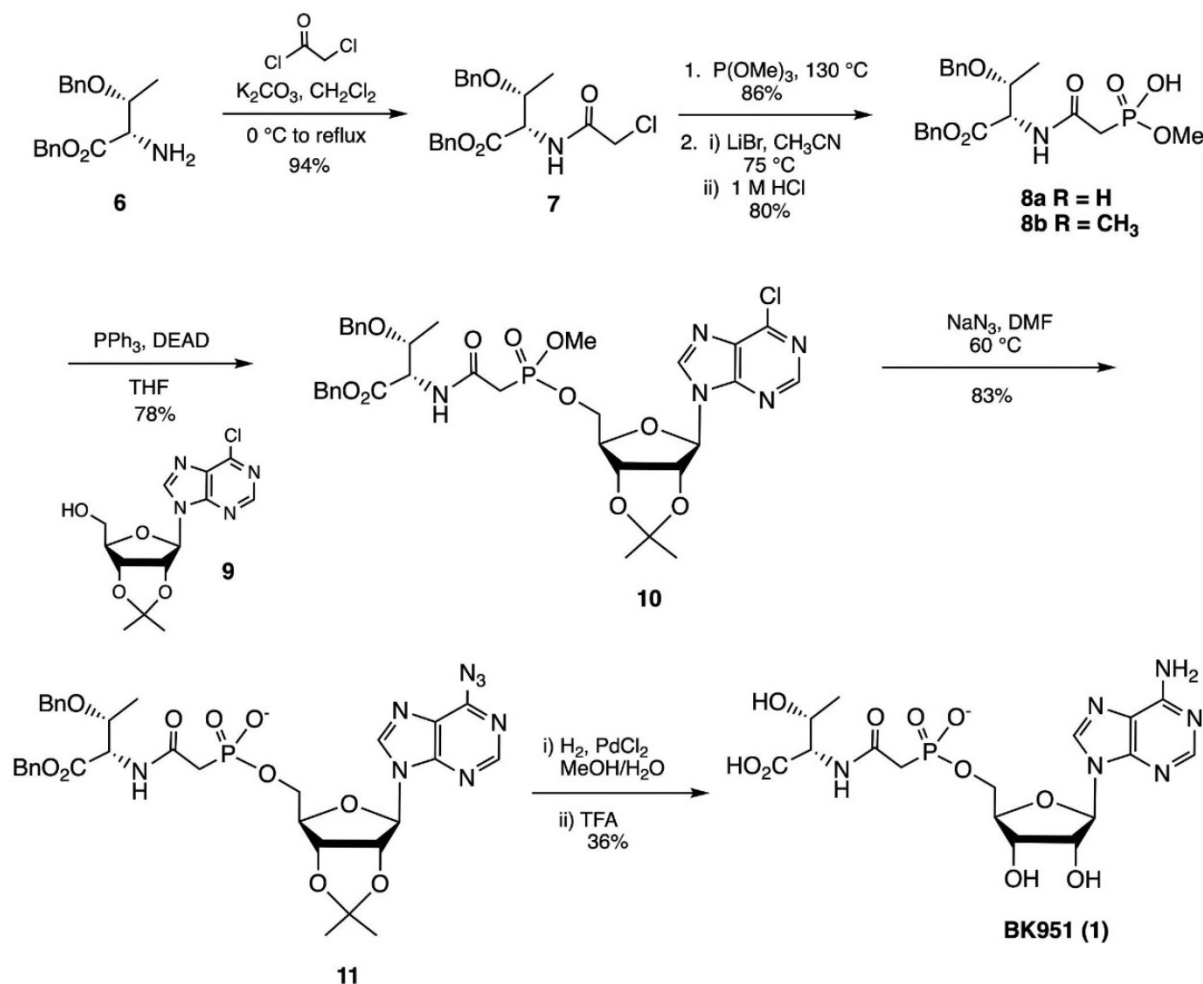
**Figure 2.** Chemical structures of TC-AMP and inhibitors prepared and analysed in this work.

modification reaction by using HPLC-purified TC-AMP as the substrate with TsaBDE in the presence of ATP (30). Using TsaBDE at a concentration of  $0.5 \mu\text{M}$ , we used the Morrison equation to fit kinetic data for reaction rates with TC-AMP in the concentration range of  $0.5\text{--}10 \mu\text{M}$ . Under these conditions, with  $\text{ATP} = 1 \text{ mM}$ , we observe an apparent  $K_m = 1.5 \pm 0.4 \mu\text{M}$  for TC-AMP (Table 1, Figure 4E). Furthermore, with ATP as the variable substrate and  $[\text{TC-AMP}] = 10 \mu\text{M}$ , we see similar substrate inhibition as observed in the overall reaction (Figure 4F). This suggests that the substrate inhibition observed in the overall reaction is due to ATP interference with the TsaBDE-catalysed transfer of the carbamoyl group to tRNA. Although ATP binding was not observed by ITC with EcTsaBD, likely due to hydrolysis, isosteric ATP analogs have been observed to bind to the heterodimer with low micromolar  $K_d$  values (37). By fitting the data to a substrate inhibition model (Equation 3 above), we measured a  $K_m = 14 \mu\text{M}$  and a  $K_{is} = 1.8 \text{ mM}$  for ATP in the second step with TC-AMP as substrate (Table 1). Figure 4G shows more clearly the fit of the data for variable ATP concentrations  $\leq 1 \text{ mM}$ . Interestingly, this  $K_m$  value is 40-fold lower than the  $K_m$  value ( $0.64 \text{ mM}$ ) for TsaE-catalyzed ATPase activity by the EcTsaBDE complex (37). Thus, ATP is utilized more efficiently for  $t^6\text{A}$  formation at low ATP concentrations, albeit with a turnover that is  $\sim 80$ -fold lower than the ATPase activity. It is not yet clear what the implications of these data are for the mech-

anistic role of TsaE in  $t^6\text{A}$  formation or its physiological relevance.

#### Inhibition of $t^6\text{A}$ formation by TCAMP analogs

In testing the TC-AMP mimics for inhibition of the  $t^6\text{A}$  reaction, we first used the full reaction assay to measure  $\text{IC}_{50}$  values with concentrations of ATP and L-threonine at  $1 \text{ mM}$  each. These substrate concentrations were necessary to achieve an optimum rate of  $t^6\text{A}$  formation in our *in vitro* assay and are close to the estimated physiological concentrations in *E. coli* (55). The ATP concentration was also chosen to avoid substrate inhibition. The  $\text{IC}_{50}$  values were obtained by varying the inhibitor concentration and measuring as close to the initial rates as possible, given limitations on the lower range of enzyme concentration that could be used. The  $\text{IC}_{50}$  values for the inhibitors are listed in Table 2. The most potent inhibitor was the phosphonate BK951, which had an  $\text{IC}_{50}$  value of  $1.3 \pm 0.5 \mu\text{M}$  (Figure 5A). The sulfamate derivative **2** ( $\text{IC}_{50} = 51 \mu\text{M}$ ) showed 40-fold weaker inhibition under these conditions, as did the aminosulfonysulfamate **3** ( $\text{IC}_{50} = 45.5 \mu\text{M}$ ). We also prepared the serine analog of TC-AMP, SC-AMP, which is a poor substrate, and gives an  $\text{IC}_{50} = 9.3 \mu\text{M}$  (Supplementary Figure S7A–C). This shows the relative importance of the methyl group of the threonine moiety in binding. Both the malonate ester derivative **4** and the malonate amide **5**



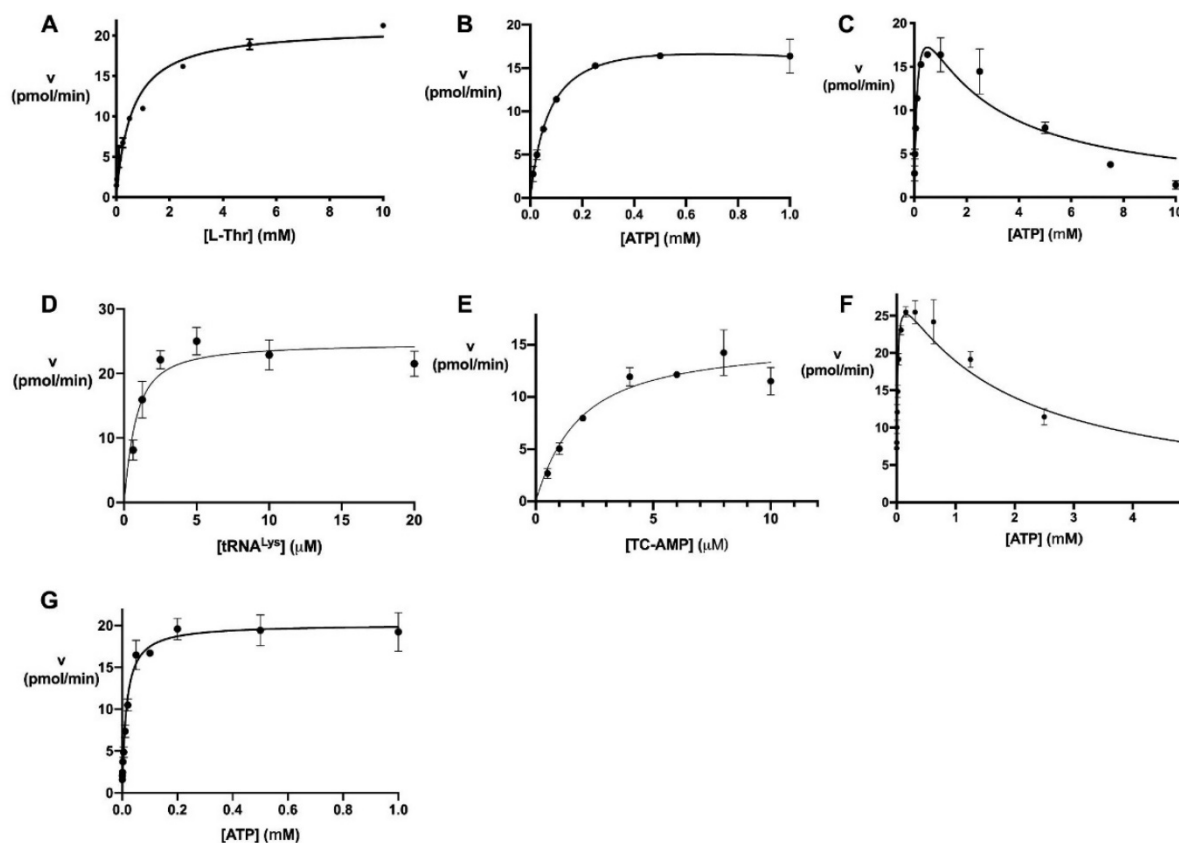
**Figure 3.** Synthesis of phosphonate inhibitor BK951. See Materials and Methods for experimental details.

showed no observable inhibition at concentrations as high as 1 mM (data not shown). The ester analog **4** was also particularly unstable during storage at  $-20^\circ\text{C}$  or in solution.

Since TC-AMP is a product of Tsac in the first step and a substrate for TsabDE in the second step of  $t^6\text{A}$  formation, we tested the analogs as inhibitors for each step of the reaction. For the phosphonate BK951, we found an  $\text{IC}_{50} = 7.5 \pm 1.2\ \mu\text{M}$  for the second reaction step with  $10\ \mu\text{M}$  TC-AMP and 1 mM ATP (Figure 5B). By assuming competitive inhibition with TC-AMP, the  $K_i$  for BK951 can be estimated using the Cheng–Prusoff equation (59) to give a calculated value of  $2.5\ \mu\text{M}$ . This  $K_i$  calculation assumes that ATP concentration does not affect inhibitor binding. We did observe that the ATP concentration does have a measurable effect on the  $\text{IC}_{50}$  value for BK951 in the overall reaction (Supplementary Figure S8A). At 0.1 mM ATP, the  $\text{IC}_{50}$  for BK951 is reduced to  $0.4\ \mu\text{M}$ , while increasing ATP to 5 mM raises the  $\text{IC}_{50}$  of BK951 to  $5.6\ \mu\text{M}$ . Thus, the  $K_i$  calculation is only an estimate due to ATP competition with TC-AMP in the active site. Similar measurements at 1 mM ATP with

analog **2** and **3** gave  $\text{IC}_{50}$  values of 123 and  $99\ \mu\text{M}$ , respectively, showing that the phosphonate is a better inhibitor of the TsabDE catalyzed second step in  $t^6\text{A}$  formation.

To provide an additional, direct measurement of BK951 binding to EcTsaBD, we used a recently described technique, microscale thermophoresis (MST), which measures the difference of thermal motion of fluorescently labelled targets as a function of ligand binding (60). The technique is highly versatile, and can be used to measure protein-protein interactions, as well as small molecule and ion binding to proteins or nucleic acids (61). For these experiments, we used fluorescently labelled EcTsaD, prepared using a probe supplied by the manufacturer that covalently reacts with accessible amines on the protein surface. We found that labelled EcTsaD showed 95% of the activity of the non-labelled protein (data not shown) in our  $t^6\text{A}$  assay. In addition, measurement of the affinity of fluorescently tagged EcTsaD to EcTsaB by MST gave a value of  $45 \pm 6\ \text{nM}$  (Figure 5C), which is in good agreement with the previously reported measurement of 120 nM by isothermal calorime-

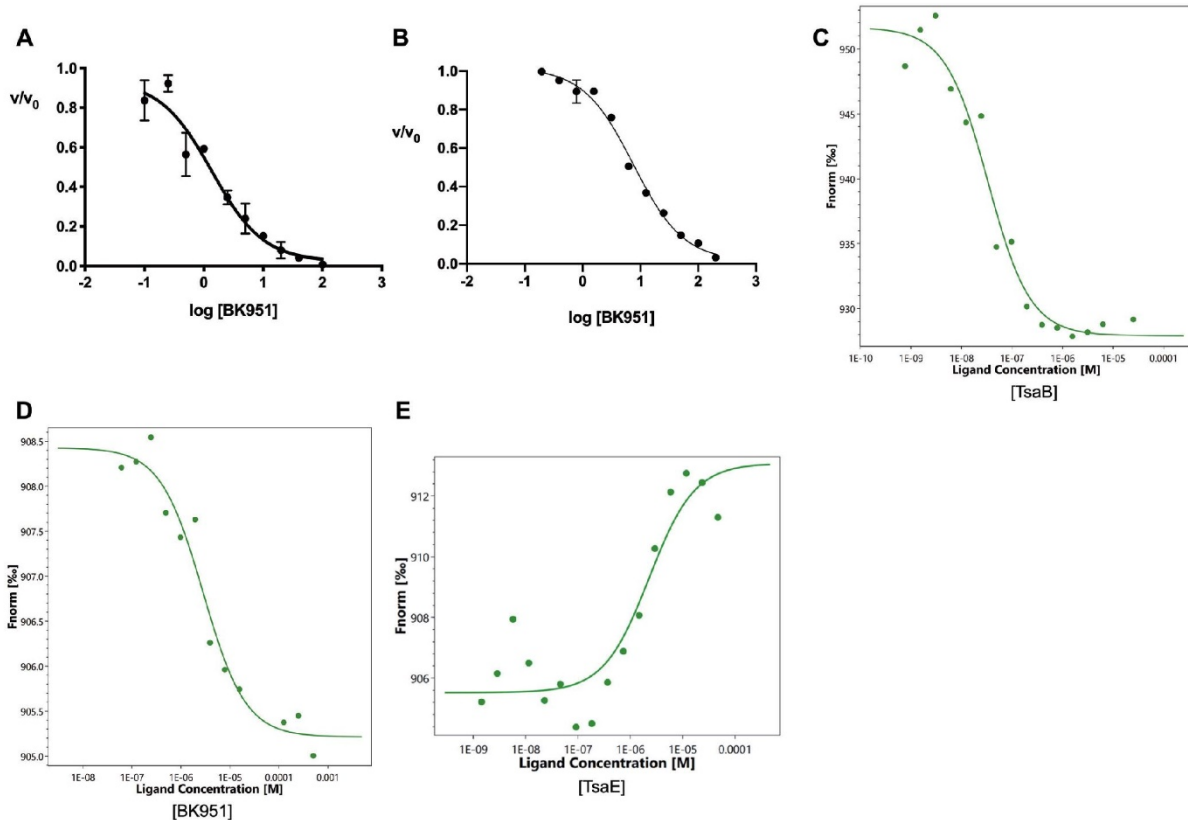


**Figure 4.** Kinetics of *E. coli*  $t^6A$  formation in the full reaction and isolated second step with purified TC-AMP as substrate. (A) Full reaction mixture with L-threonine as variable substrate fit to MM kinetics (Equation 1). (B) Full reaction mixture with ATP as variable substrate  $\leq 1$  mM. (C) Full reaction with ATP as variable substrate showing substrate inhibition at high concentrations (fit to Equation 3) (D) Full reaction with  $tRNA^{Lys}_{UUU}$  as the variable substrate fit to Morrison Equation (2). (E) Isolated second step of reaction with TC-AMP as variable substrate and 1 mM ATP fit to Morrison equation (Equation 2). (F) Second step of the reaction with TC-AMP (10  $\mu M$ ) as substrate and variable ATP displaying substrate inhibition as in (C). (G) Second step of the reaction with constant TCAMP (10  $\mu M$ ) and ATP as the variable substrate ( $\leq 1$  mM) fit to MM Equation (1).

try (34). Measurement of the binding affinity of BK951 with the preformed TsaBD heterodimer (using 0.5  $\mu M$  unlabelled TsaB) gave a  $K_d$  value of  $3.4 \pm 0.6 \mu M$  (Figure 5D). We also used MST to confirm the binding of EcTsaE to the EcTsaBD heterodimer ( $K_d = 2.2 \mu M$ , Figure 5E). An attempt to measure binding of BK951 to the EcTsaBDE ternary complex was unsuccessful and gave only scattered data points (data not shown). ATP (1 mM) was included in these experiments to promote ternary complex formation (37), although it was found to be hydrolyzed to ADP during the lifetime of the binding measurements, which was confirmed by HPLC. It is thus possible that the high ADP concentration could be inhibiting BK951 binding to EcTsaBDE. To answer this question, we measured the  $IC_{50}$  values for both ADP and AMP as inhibitors. As shown in Table 2 (and Supplementary Figure S8B,C), these nucleotides had  $IC_{50}$  values of 160 and 113  $\mu M$ , respectively. Although ADP is likely inhibiting BK951 binding in the MST experiments, it is unlikely that either AMP or ADP are formed in high enough concentrations during the kinetic assays to cause the ATP substrate inhibition we observe.

We also looked at the effect of high concentrations of TsaE on the binding of Bk951 in the kinetic assay of the overall reaction. We find that high relative concentrations of TsaE ( $> 2.5 \mu M$  with the other enzymes at 0.5  $\mu M$ ) strongly inhibit  $t^6A$  formation (Supplementary Figure S9A). However, this has little effect on BK951 binding ( $IC_{50} = 2.4 \pm 0.8 \mu M$  at 10  $\mu M$  TsaE, Supplementary Figure S9B). On the other hand, 10  $\mu M$  TsaE increases the  $K_m$  of  $tRNA^{Lys}$  over 20-fold to 12.0  $\mu M$  (Supplementary Figure S9C). This latter observation is consistent with the lack of observed  $tRNA$  binding by the EcTsaBDE ternary complex (37) and with the rapid loss of  $tRNA$  binding upon increasing TsaE to give the TsaB<sub>2</sub>D<sub>2</sub>E<sub>2</sub> complex of *T. maritima*. These results suggest that TsaE competes with  $tRNA$  and BK951, we also asked if the binding of BK951 affects  $tRNA$  binding to EcTsaBD. Preparation of a fluorescently labelled  $tRNA^{Lys}$  allowed us to use fluorescence anisotropy to measure the binding of  $tRNA$  to the EcTsaBD heterodimer. By titrating a solution of labelled  $tRNA$  with increasing con-





**Figure 5.** Characterization of BK951 inhibition of  $t^6A$  formation. (A)  $IC_{50}$  determination for BK951 with full reaction mixture. (B)  $IC_{50}$  determination for BK951 with 10  $\mu M$  TC-AMP, 0.5  $\mu M$  EcTsaBDE and 1 mM ATP. (C)  $K_d$  measurement of TsaB binding to fluorescently labelled TsaD using Microscale Thermophoresis (MST). (D)  $K_d$  measurement of BK951 binding to labelled TsaBD heterodimer using MST. (E)  $K_d$  measurement of TsaE to labelled TsaBD in the presence of 100  $\mu M$  ATP. For the MST studies, fluorescently labelled TsaD = 10 nM, TsaB = 0.5  $\mu M$  ( $10 \times K_d$ ) and TsaE = 20  $\mu M$  ( $\sim 10 \times K_d$ ).

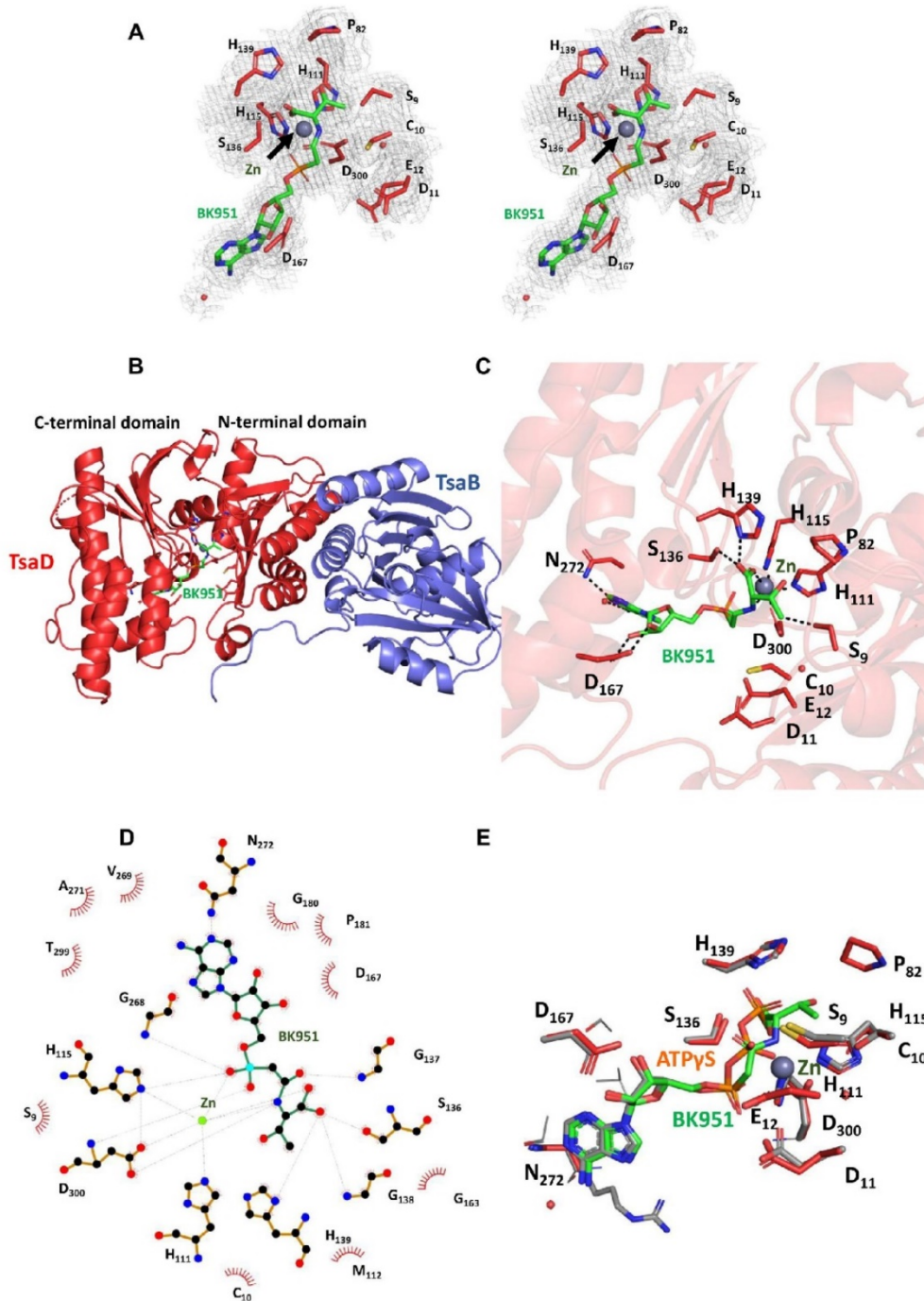
centrations of EcTsaBD, either alone or in the presence of 50  $\mu M$  BK951, we measured  $K_d$  values of  $87 \pm 11$  and  $112 \pm 13$  nM, respectively (Supplementary Figure S10). This shows that the presence of bound BK951 has no significant effect on the binding affinity of the tRNA substrate.

Finally, since TC-AMP has been observed as a bound product to TsaC during isolation (34, *vide infra*), we tested the most potent inhibitors of the overall reaction for the ability to inhibit EcTsaC as product analogs. Table 2 shows that BK951 surprisingly does not bind with significant affinity to this enzyme. This contrasts with the sulfamates 2 and 3, which bind with  $IC_{50}$  values of 31 and 97  $\mu M$ , respectively (Supplementary Figure S7D, E). Thus, the inhibition of the overall reaction by phosphonate BK951 is due solely to its binding to TsaBD, whereas for the sulfamate derivatives, it is a combination of inhibition of both enzymatic steps.

### Structure of BK951 bound to EcTsaBD

Knowledge of the interactions between the TC-AMP substrate and TsaD is essential for the understanding of the  $t^6A$  reaction mechanism. We therefore determined the crystal structure of EcTsaBD in complex with our highest affinity

inhibitor BK951. Crystallization of EcTsaBD in presence of 2 mM BK951 yielded crystals that diffracted to a resolution of 2.4  $\text{\AA}$ . The structure was solved by molecular replacement, revealing two identical copies of the EcTsaBD complex in the asymmetric unit. The structure of EcTsaBD is nearly identical to those of *apo*-form or nucleotide complexes of StTsaBD and EcTsaBD (36,37) (RMSD of about 0.39  $\text{\AA}$  for 337 superposed  $C\alpha$  positions). The presence of the inhibitor bound to EcTsaD was supported by an extended cloud of residual electron density near the metal sites of TsaD for the two copies in the asymmetric unit (Figure 6A). A complete BK951 molecule could be easily fitted into this density. At full occupancy, the refined  $B$ -factors of the BK951 atoms are on average comparable to those of the surrounding side chains. BK951 binds as expected in the active site groove between the N- and C-terminal domains of EcTsaD (Figure 6B). The inhibitor is bound by an extensive network of hydrogen bonding and van der Waals interactions between active site residues, the bound metal ion and the inhibitor (Figure 6C). The purine and ribose engage in identical interactions as those observed for the corresponding moieties of the ATP $\gamma$ S complex with StTsaD (36): the ribose oxygens are H-bonding with the Asp167 carboxy-



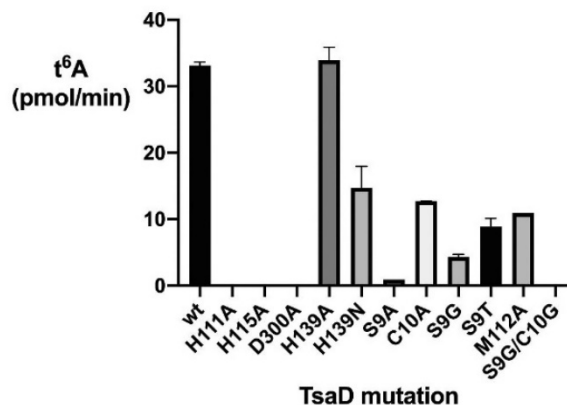
**Figure 6.** Structural characteristics of the BK951–EcTsaBD complex. (A) Stereoview of the residual 2Fo-Fc electron density contoured at 1σ levels for the bound BK951 (green sticks) with metal ion (gray sphere) and surrounding side chains (red sticks). (B) Cartoon presentation of the TsaBD heterodimer bound to BK951; TsaB in blue, TsaD in red, BK951 in green sticks and Zinc as a grey sphere. (C) BK951 (green sticks) bound in the active site of TsaD, Zinc ion as grey sphere. H-bonds and ligand-metal interactions: black dotted lines. (D) LigPlot+ (69) representation of the inhibitor BK951 bound into the active site of EcTsaD. (E) Superposition of ATPγS bound to StTsaD (orange sticks) onto BK951 bound to EcTsaD (green sticks). Only residues from EcTsaD are labelled.

late and the N1 of adenine interacts with the Asn272 side chain (Figure 6D). Figure 6E superposes ATP $\gamma$ S (StTsaD) onto BK951, showing that (i) their  $\alpha$ -PO $_4$  moieties superpose very well and (ii) the carbamoyl carbon of BK951 superposes onto the  $\beta$ -P position. The active site Zn coordination sphere is composed of a phosphonate-oxygen and the carboxylate of BK951, His111, His115 and Asp300. The interactions of the threonylcarbamoyl moiety of BK951 are discussed below.

#### Binding pocket of the L-threonine moiety bound to EcTsaBD

For the first time, the BK951-EcTsaBD structure allows a detailed look at the binding pocket for the threonine moiety. Asp300 binds to the Zn and is conformationally restricted by hydrogen bonding with Ser9, while also making a weak hydrogen bond with the carbamoyl NH of the phosphonate inhibitor. The threonine side chain specificity for the bound inhibitor appears to be the result of hydrogen bonding to the  $\beta$ -hydroxyl by Ser9 and a pocket formed by van der Waals interactions between the  $\beta$ -methyl group of the inhibitor with His111, Cys10, and a small loop containing residues Pro82, Gly83 and Leu84 (Figure 6C and D). Our kinetic studies show that ( $\pm$ )-3-hydroxynorvaline (hNV) is a good substrate with a  $K_m = 0.5$  mM to give the methyl homolog hn $^6$ A (Supplementary Figure S11), which has been observed in the tRNA of both bacteria (62) and archaea (63). *In vitro* substrate activity of hNV with SUA5 and KEOPS has been shown in the archaeal system by Perrochia, et al. (31) and also recently shown for *E. coli* and *T. maritima* by Swinehart et al. (58). In the structure, the binding pocket for the threonine side chain has an open face into the 'tunnel' formed in TsaD such that extended alkyl chains may be accommodated if they can project into the tunnel but not if they are restricted in the direction of Cys10. The insertion of a methyl group for one of the  $\beta$ -methyl hydrogens is likely tolerated by projecting the methyl into this open space.

The threonine carboxylate is both coordinated to the active site Zn atom and also bound by hydrogen bonds from Ser136 and His139. One of the possible strategies to catalyze the reaction would be an activation of the carbamoyl group by the bound metal ion. In the present configuration, the carbamoyl is ideally positioned for nucleophilic attack by the N6 of A37. One can imagine upon tRNA binding a simple rotation about the C $\alpha$ -amine C–N bond could position the carbonyl oxygen of the TC-AMP carbamoyl group to replace the carboxylate inside the coordination sphere of the Zn. This would allow Lewis acid activation of the carbonyl and facilitate acylation of the N6 of A37 of the tRNA by TC-AMP. The plausibility of such a conformational rotation is suggested by the comparison of the structure of BK951/EcTsaBD and ATP $\gamma$ S/StTsaBD (Figure 6E). The  $\alpha$  and  $\beta$ -phosphates of ATP $\gamma$ S superpose onto the  $\alpha$ P and carbamoyl carbon of BK951, but the dihedral configuration of the bonds is different. One can easily imagine a rotation of the C $\alpha$ -N bonds that leads to an orientation comparable to the P–O configurations of ATP $\gamma$ S to bring the carbamoyl into a catalytic competent orientation. Alternatively, the metal could impose a reactive conformation of the TC-AMP without actively participating in chemical cataly-



**Figure 7.** t $^6$ A activity of active site TsaD mutants using the full assay mixture (see Materials and Methods for details).

sis of the reaction pathway. Indeed, as discussed further, catalytic assistance is offered by active site residues other than the metal ion. By initially binding TC-AMP in a stabilizing conformation the enzyme could protect it from decomposition until the tRNA is bound. Further kinetic studies would be useful to determine if the binding of TC-AMP precedes tRNA binding which would lend support to this hypothesis.

#### Functional analysis of residues in the TsaD active site

We used site-directed mutagenesis to explore the functional significance of EcTsaD amino acid side chains in the active site binding region of the threonine moiety of the inhibitor (Figure 7). As in the previously reported structure of EcTsaBD bound to ADP, the Zn atom is coordinated by three universally conserved residues His111, His113 and Asp300 (37). These residues were each mutated to alanine and found to give purified proteins that were inactive in the *in vitro* t $^6$ A assay. We also mutated His139, because of its conservation in bacterial TsaD and its position as a Zn ligand in TmTsaBDE (41). In the EcTsaBD-ADP structure, His139 is not liganded to the Zn, and in the BK951 bound structure reported here, it is also not chelated to the Zn but it is making a relatively short hydrogen bond to the non-liganded oxygen of the carboxylate of L-threonine (Figure 6C). Surprisingly, both the His139Ala and His139Asn mutants retained activity, with the alanine mutant indistinguishable from wild type. It is possible that a compensatory interaction in the mutants makes up for the loss of this hydrogen bond. In the *T. maritima* carboxy-AMP bound TsaB $_2$ D $_2$ E $_2$  structure, the analogous His137 is a ligand for the active site Zn ion and found to be required for protein stability. These results likely reflect subtle differences in the organization of the active site in these quite different organisms and may have mechanistic implications.

We also investigated possible specificity determinants for the threonine moiety of BK951. Interactions between EcTsaD and the threonyl moiety of BK951 are largely made of weak hydrogen bonding interactions (>3.5 Å) with Ser9, His111 and the amide nitrogen of Gly83, as well as van der Waals interactions with the  $\beta$ -carbon of Cys10 and

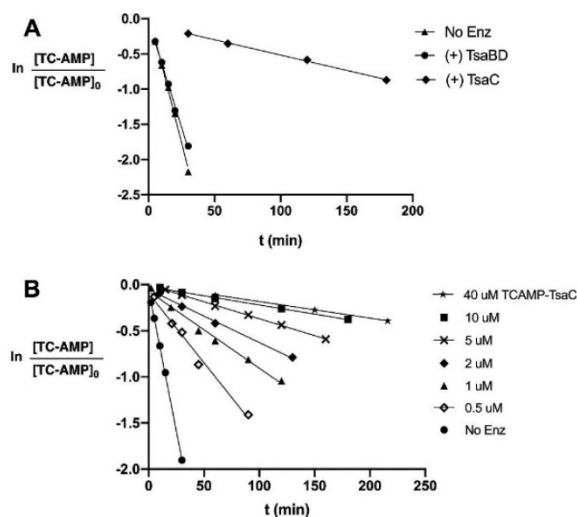
atoms of the Pro82 side chain (Figure 6C and D). Ser9 also makes a hydrogen bonding interaction (3.3 Å) with the unliganded imidazole nitrogen of His111 and likely acts to restrict its conformation for interaction with the threonine moiety of the inhibitor. The Ser9Ala and Ser9Gly mutants have significantly reduced activity, while the Cys10Ala mutant is reduced only by about a third. The Ser9Gly, Cys10Gly double mutant is inactive, which indicates additive effects of removing van der Waals interactions with the threonine moiety. These mutants suggest that the specificity for the L-threonyl part of TC-AMP is enforced by a number of weak interactions that together lead to selection of threonine/hydroxynorvaline over other potential amino acids and provides a structural explanation for the recent report of TsaD as the main guardian of such specificity (58).

### Modeling A37 of the tRNA substrate into the active site of EcTsaD

Assuming the BK951 adopts the position of the TC-AMP in the active site just before transfer of the threonylcarbamoyl group onto acceptor tRNA, we wondered how the incoming A37 of the latter could be positioned. We therefore modeled an adenine into the active site occupied by BK951 using the Autodock 2 software (64). The protein was kept rigid during the docking process and the distance between the N6 atom of the adenine and the carbamoyl carbon was restrained to 2.5 Å (Supplementary Figure S12). In this model the adenine ring is held between the loops encompassing residues 9–12, 40–44, 137–138 and 176–179. The adenine N9 atom points towards an empty space above the active site, leaving room for an attached tRNA molecule. Because tRNAs that are modified in their stem loops often undergo important conformational changes when bound in the active site of the corresponding enzymes, we did not attempt to model a complete tRNA into the active site of TsaD. Residues Cys10, Asp11, Val42 and Tyr177 are surrounding the adenine. Mutation of Cys10 into Ala causes a more than 50% drop in  $t^6A$  activity. To provide full access of a flipped-out adenine connected to a complete tRNA substrate, the 40–44 loop probably would have to move from its present position in the TsaBD–BK951 complex. Interestingly, this loop was shown to be very flexible and completely disordered in the context of the TmTsaBDE complex (39,40).

### Stabilization of TCAMP bound by EcTsaBD and EcTsaC

The inhibitor BK951 is bound to TsaBD in an extended conformation that likely precludes certain decomposition pathways. However, in this structure, the molecule appears to be readily exposed to solvent water. To understand more about the potential stabilizing effect of TsaBD on bound TC-AMP, we carried out a stability study. A mixture of 10  $\mu\text{M}$  purified TC-AMP was incubated with or without 20  $\mu\text{M}$  EcTsaBD at 37°C and timed aliquots were removed for HPLC analysis of TC-AMP conversion to AMP. As shown in Figure 8A, there is only a small protective effect ( $k_{\text{obs}} = 6.0 \pm 0.2 \times 10^{-2}$  versus  $7.3 \pm 0.2 \times 10^{-2} \text{ min}^{-1}$ ) at a concentration of EcTsaBD where TC-AMP should be



**Figure 8.** Stability of TC-AMP bound to either EcTsaBD or EcTsaC. (A) Comparison of rate of decomposition of 10  $\mu\text{M}$  TC-AMP at 37°C in 50 mM HEPES pH 7.5, 300 mM NaCl with 20  $\mu\text{M}$  TsaBD (filled circles); 13  $\mu\text{M}$  TsaC (filled diamonds); or in the absence of enzyme (filled triangles). (B) Concentration dependence of the rate of decomposition of a 1.3:1 molar ratio of TsaC:TC-AMP at varying concentrations under the same conditions as (A).

mostly bound. For comparison, the stabilising effect of 13  $\mu\text{M}$  TsaC is nearly an order of magnitude higher ( $k_{\text{obs}} = 4.3 \pm 0.1 \times 10^{-3} \text{ min}^{-1}$ ). The rate constants correspond to a  $t_{1/2} = 9.4$  min without enzyme, compared to  $t_{1/2} = 11.5$  min in the presence of TsaBD and  $t_{1/2} = 160$  min in the presence of TsaC. These data suggest that either TC-AMP is transferred to tRNA rapidly after binding to TsaBD, or additional binding events involving TsaC, or tRNA may help to protect bound TC-AMP from hydrolysis before it is transferred by shielding it from solvent water.

For the TsaC experiments, we utilized the TC-AMP bound form of EcTsaC that can be isolated from the purification of highly overexpressed protein. We find that the amount of TC-AMP bound to TsaC can be quantified directly by HPLC and typically approaches 70% occupancy after affinity and size exclusion chromatography (Supplementary Figure S13A). Bound TC-AMP can also be converted to ATP quantitatively by the addition of 1.5 equivalents of pyrophosphate in the presence of Mg ion (Supplementary Figure S13B) and subsequently removed by size exclusion chromatography to provide the unliganded form of the enzyme (Supplementary Figure S13C). Further investigation of TsaC stabilization of TC-AMP in these complexes shows a significant reduction in protection as the concentration of the complex is lowered (Figure 8B). This concentration dependence is likely due to an increased fraction of unbound TC-AMP. A plot of the  $t_{1/2}$  values for TC-AMP stability versus TC-AMP concentration gives a crude estimated half-maximal value of  $\sim 5 \mu\text{M}$  (data not shown), which may reflect a similar magnitude for the  $K_d$  of TC-AMP binding to EcTsaC. If this were true, it would suggest that unless the concentration of TsaC or TC-AMP in *E. coli*

is greater than 5  $\mu\text{M}$ , the majority of TsaC in the cell will be free of bound TC-AMP.

## DISCUSSION

### The phosphonate BK951 is a potent inhibitor of $t^6\text{A}$ biosynthesis *in vitro*

The phosphonate BK951 was the most potent analog in this study, with an  $\text{IC}_{50}$  value of 1.3  $\mu\text{M}$  for the overall  $t^6\text{A}$  reaction. We measured inhibitor  $\text{IC}_{50}$  values in the presence of L-threonine and ATP concentrations of 1 mM each. For *E. coli*, 1 mM ATP is in the lower part of the physiological range of up to 10 mM. The intracellular L-threonine concentration has been recently reported to be 1.3 mM (55). Substitution of the phosphonate with a sulfamoyl group in **2** results in a loss of inhibition in the kinetic assay by 40-fold, which is possibly due to the weaker binding of the sulfonyl oxygen with Zn compared to the fully ionized phosphonate. Further substitution of the carbamoyl group of TC-AMP with a tetrahedral sulfonyl gives an alkylaminosulfonyl-*N*-sulfamate **3**, which is an intended transition state mimic of the attack by N6 of A37. The unique pharmacophore structure of this inhibitor has surprisingly not been previously described in the literature. Similar to a few known alkylaminosulfonamides, this structure likely forms an 'inner salt', with the bridging nitrogen deprotonated and the amine of the L-threonine protonated to give a zwitterion structure. Despite the novelty of its structure, compound **3** has virtually the same  $\text{IC}_{50}$  as sulfamate analog **2**. Since the carbonyl oxygen of BK951 is making only one hydrogen bond in the EcTsaBD bound structure, there does not appear to be amino acid side chains in place to distinguish the  $\text{sp}^3$  orientation of the sulfonyl group in this area of the active site in this conformation of the protein. Finally, both the malonate ester and the malonate amide bridge between the adenosine ribose and threonine (**4** and **5**, Figure 2) failed to give molecules that showed any observable binding to EcTsaBD. Both molecules were designed to examine the hypothesis that the Zn could chelate the phosphonate and the carbamate carbonyl of TC-AMP, which in this structure is not observed. Hence, the lack of observed binding of the malonates supports the overall significance of the current structure.

If we assume the affinity of TC-AMP to EcTsaBD is similar in magnitude to the  $K_m$  value of 1.5  $\mu\text{M}$ , the binding data for the phosphonate analog suggests it is a good mimic that is taking advantage of most of the protein binding interactions. This conclusion is supported by the similarity of the conformation of atoms of the adenosyl 5'-phosphate/phosphonate group in the active site of TsaD of both BK951 and ATP. In the recently modelled binding of TC-AMP to TmTsaD based on the structure of bound carboxy-AMP, Luthra *et al.* proposed a similar binding of the threonine carboxylate moiety to the active site Zn (41). Their model differs from our structure in the lack of chelation between the phosphonate and the Zn. It is possible that there are species-specific differences in the orientation of some amino acids, as well as rearrangements of the substrates along the reaction coordinate. Thus, many of the various structures may each be relevant for a distinct part of the transformation.

### The sulfamate analogs are better mimics of TC-AMP binding to TsaC

Although BK951 is more potent in inhibiting the second step of  $t^6\text{A}$  synthesis, it surprisingly shows no inhibition of the first step of the reaction catalysed by TsaC. The sulfamates **2** and **3**, on the other hand, show significant inhibition of both enzymatic steps in  $t^6\text{A}$  formation. A re-examination of the crystal structure of TC-AMP bound to TsaC2 from *S. tokodaii* (34) shows a hydrogen bond between the backbone amide of conserved Ser144 to the adenylate ester oxygen of TC-AMP that is replaced by a methylene in BK951. There are also close van der Waals interactions likely between the backbone carbonyl and alpha carbons of Pro143 (usually a Thr in TsaC1). A possible steric clash between these carbon atoms and the methylene of the phosphonate analog, in addition to the loss of hydrogen bond would be sufficient to explain its lack of binding to TsaC. Conversely, the amide nitrogen of the sulfamate group in analogs **2** and **3** is acidic enough to be deprotonated, which makes it an excellent mimic of the adenylate ester oxygen as previously documented (53,54). Therefore, TsaC appears to use these interactions to closely select for an ester at this position and analogs that preserve the integrity of these interactions will be generally more potent as inhibitors.

### The binding conformation of BK951 suggests a precatalytic complex

The binding interactions between BK951 and the active site of the TsaD protein suggests that at one point in the reaction coordinate, the Zn interacts directly with the phosphonate oxygen and carboxylate oxygens of TC-AMP. The carbamoyl oxygen does not form a ligand to the metal ion, but establishes two H-bonds with mainchain NH groups of Ser<sup>136</sup>GlyGly<sup>138</sup>, the  $\beta$ -hairpin connection between strands  $\beta 6$ - $\beta 7$ . This is a similar binding mode as the carbamoyl adenylate bound to the homologous TobZ protein, which catalyses the carbamoylation of tobramycin (34). This position of the carbonyl suggests that the amide main chain NH groups of Ser<sup>136</sup>GlyGly<sup>138</sup> could provide an oxyanion hole during the nucleophilic attack of the A37 6-amino group during the transfer reaction. As shown in Figure 9, this loop is extremely well conserved in the TsaD/Kae1/Qri7 family, suggesting its importance for catalytic activity. This loop is also present in the related carbamoylating enzymes TobZ and HypF, reinforcing the hypothesis of its role in the stabilization of the intermediate. This means that the main role of the metal is to provide a scaffold for productive binding of the substrate. In the TobZ active site, a histidine is positioned near the carbamoyl group that could act as a general base to subtract a proton from the incoming tobramycin OH nucleophile. This is not the case for TsaD, probably because the A37  $\text{NH}_2$  nucleophile is deprotonated. Alternatively, the carbamoyl group of the TC-AMP could slightly rotate and engage in a Lewis acid interaction with the metal to increase its reactivity. The latter scenario seems less likely to us because it would render the carbon less accessible for an incoming nucleophile. Clearly, structural data in presence of the tRNA substrate are needed for a better understanding of the mechanism. Assuming that the binding mode of the BK951 carbonyl mimics that of the TC-AMP

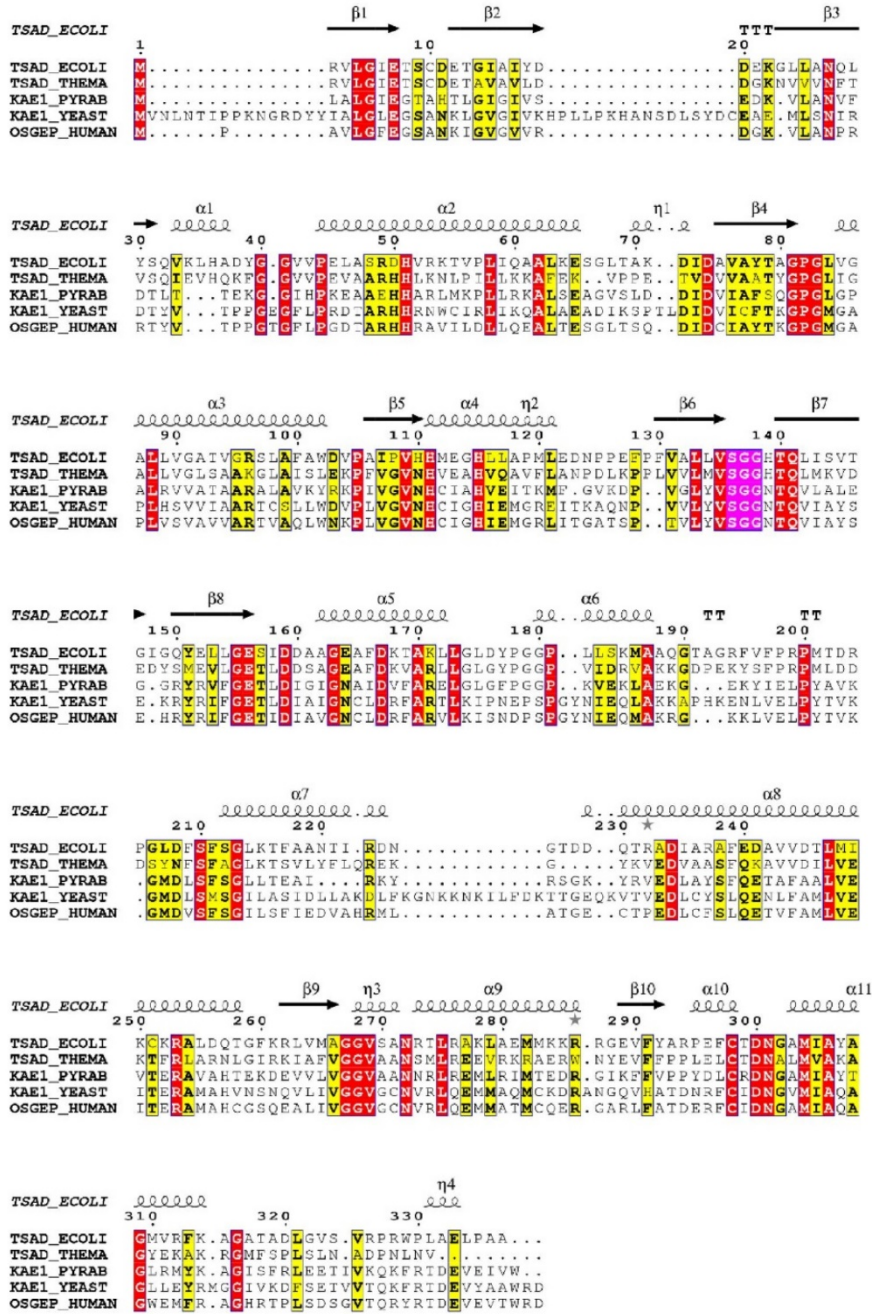


Figure 9. Multiple sequence alignment of the TsaD/KaeI/OSGEP family members was realized using the blosum 45 with MAFFT (online version 7) (70). Structure-based alignment was performed using the ESPRIPT web-server (71). The SGG peptide conserved in all homologs is highlighted in pink.

Downloaded from https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab26/6125664 by guest on 04 February 2021

substrate, its access by the bulky tRNA substrate would likely be hampered by some of the surrounding active site loops and helices, especially the region contained between residues 32 and 45. Interestingly this region becomes completely disordered upon TsaE binding (40), suggesting that flexibility plays a major role in the catalytic cycle.

### The ability of EcTsaBD and EcTsaC to stabilize bound TC-AMP differ significantly

Previous work showed that TC-AMP is unstable with a half-life of 3.5 minutes at pH 7.5 in the presence of 2 mM Mg ion (30). Both the instability and the magnitude of the  $K_m$  value for TC-AMP (1.5  $\mu$ M) for the isolated TsaBDE-catalysed t<sup>6</sup>A reaction are similar in magnitude for the structurally analogous primary metabolite carbamoyl phosphate (CP). The  $K_m$  values for CP for the enzymes aspartate and ornithine transcarbamoylases are ~4–10  $\mu$ M. Wang *et al.* (65) have shown that CP is stabilized by binding to these enzymes in the absence of the other substrates, and the effect is more pronounced for hyperthermophiles to the point of requiring substrate channelling (66). Such binding prevents the major mechanism of degradation of CP which involves intramolecular transfer of the carbamoyl NH proton to the phosphate and subsequent rapid formation of cyanate. In our previous analysis of TC-AMP decomposition, NMR data showed L-threonine and the cyclic oxazolidinone as the only two detectable final products produced in an equimolar ratio (30). The oxazolidinone forms via attack by the  $\beta$ -hydroxyl oxygen on the carbamoyl carbonyl carbon (or isocyanate intermediate) to release AMP and give the cyclic product. The structure reported here suggests that TC-AMP is bound by TsaBD in an extended conformation which would minimize the cyclization pathway. Although we did find that EcTsaBD stabilizes TC-AMP in solution, the effect is small. This is likely due to TC-AMP remaining accessible to solvent water when bound to TsaBD, which allows for direct hydrolysis. The fact that the half-life of TC-AMP bound to EcTsaBD is more similar to free TC-AMP in the absence of metal ion (16 min) than in its presence (3.5 min) is consistent with the conformation of bound BK951, in which the carbamoyl carbonyl is not chelated to the Zn. Retaining stability of TC-AMP may be another reason for avoiding such coordination. In other words, if the carbamoyl carbonyl were coordinated to the Zn, the enzyme could catalyze wasteful transfer of the carbamoyl group to water.

In contrast to EcTsaBD, we find an order of magnitude greater stabilizing effect on TC-AMP by binding to EcTsaC. Previous work has shown that overproduced EcTsaC retains bound TC-AMP after purification (30) and that crystallized TsaC2 from *S. tokodaii* (45) contained bound TC-AMP (34). In this structure, the threonylcarbamoyl group of TC-AMP is sequestered from solvent in a long, narrow binding pocket. The observed greater stabilization by TsaC may be used to prevent wasteful hydrolysis of TC-AMP in the cell (34). However, we find that as the concentration of EcTsaC is lowered into the low micromolar range, the stabilizing effect of the enzyme on TC-AMP is diminished significantly. It thus may be less of an imperative for mesophilic organisms to stabilize this metabolite which could be a relic

of enforced stabilization in hyperthermophilic organisms, considering the ancient nature of this enzyme system (34). Harris, *et al.* (67) have reported that the C-terminus of EcTsaC contains sites that are highly dynamic in structure and this region may be involved in induced fit of ATP binding, but also in protein-protein interaction with TsaD. This flexibility of the C-terminus is also consistent with the crystal structure of unliganded EcTsaC (68). Thus, a major question still exists of whether TsaC presents bound TC-AMP to the TsaBD heterodimer (in mesophiles) or if the intermediate diffuses freely as suggested by *in vitro* activity experiments (27,30). It will be interesting to see if the channelling that is observed for CP enzymes from hyperthermophiles is also operative for TC-AMP in t<sup>6</sup>A formation in organisms such as *T. maritima*. For example, if strict channelling of TC-AMP occurs in the hyperthermophilic organisms, the kinetic inhibition of BK951 may be observed to be very weak due to restricted access to the active site. The structure reported in this paper suggests that TsaBD-bound TC-AMP, if also protected from solvent water in a higher order complex, would likely provide considerable stabilization to the intermediate.

In summary, we report here the synthesis, biochemical characterization and cocrystal structure of a potent inhibitor of bacterial t<sup>6</sup>A formation. The inhibitor appears to bind in the active site of the heterodimeric TsaBD complex in a manner similar to the natural substrate TC-AMP and could therefore be useful as a mechanistic probe for the reaction in a variety of biological contexts. The crystal structure should be useful for the rational design of more potent, cell permeable inhibitors of t<sup>6</sup>A modification that may find application as biochemical probes in a variety of organisms given the universal importance of this modification. Inhibitors optimized for the bacterial enzymes may find use as potentially novel antibiotics.

### DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession number 6Z81.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

Sophia Missouri received a scholarship from the 'Fondation de la Recherche Médicale'. We thank the staff of the Proxima1 and 2 beamlines at the SOLEIL synchrotron for assistance during the diffraction data collection. This work received support from the French Infrastructure for Structural Biology (FRISBI ANR-10-INSB-0501).

### FUNDING

French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-0501 to H.v.T.]; University of Wisconsin School of Pharmacy and Wisconsin Alumni Research Fund (to C.L.). Funding for open access charge: University of Wisconsin School of Pharmacy.

Conflict of interest statement. None declared.

## REFERENCES

- Gustilo, E.M., Vendeix, F.A.P. and Agris, P.F. (2008) tRNAs modifications bring order to gene expression. *Curr. Opin. Microbiol.*, **11**, 134–140.
- Björk, G.R. (1995) Biosynthesis and function of modified nucleosides. In: Söll, D. and Rajbhandary, U. (eds) *tRNA: Structure, Biosynthesis, and Function*. American Society for Microbiology, Washington, DC, pp. 165–205.
- El Yacoubi, B., Bailly, M. and de Crécy-Lagard, V. (2012) Biosynthesis and function of post-transcriptional modifications of transfer RNAs. *Annu. Rev. Gen.*, **46**, 69–95.
- Urbonavicius, J., Qian, Q., Durand, J.M., Hagervall, T.G. and Björk, G.R. (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.*, **20**, 4863–4873.
- 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*, **8**, 3283–3289.
- Powers, D.M. and Peterkofsky, A. (1972) The presence of N-(purin-6-ylcarbamoyl)threonine in transfer ribonucleic acid species whose codons begin with adenine. *J. Biol. Chem.*, **247**, 6394–6401.
- Miller, J.P., Hussain, Z. and Schweizer, M.P. (1976) The involvement of the anticodon adjacent modified nucleoside N-(9-(beta-D-ribofuranosyl)-purin-6-ylcarbamoyl)-threonine in the biological function of *E. coli* tRNA<sup>Leu</sup>. *Nuc. Acids Res.*, **3**, 1185–1201.
- Thiaville, P.C., Iwata-Reuyl, D. and de Crécy-Lagard, V. (2015) Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t<sup>6</sup>A) a universal modification of tRNA. *RNA Biol.*, **12**, 1529–1539.
- Murphy, F.V., Ramakrishnan, V., Malkiewicz, A. and Agris, P.F. (2004) The role of modifications in codon discrimination by tRNA<sup>Lys</sup><sub>UUU</sub>. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
- 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.
- El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williams, J.R., Schimmel, P., Swairjo, M.A. and de Crécy-Lagard, V. (2009) The universal YrdC/Sua5 family is required for the formation of threonyl-carbamoyladenine in tRNA. *Nuc. Acids Res.*, **37**, 2894–2909.
- 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.
- Thiaville, P.C., El Yacoubi, B., Kohrer, C., Thiaville, J.J., Deutsch, C., Iwata-Reuyl, D., Bacusmo, J.M., Armengaud, J., Bessho, Y., Wetzel, C. et al. (2015) Essentiality of threonylcarbamoyladenine (t<sup>6</sup>A), a universal tRNA modification, in bacteria. *Mol. Microbiol.*, **98**, 1199–1221.
- Miyauchi, K., Kimura, S. and Suzuki, T. (2013) A cyclic form of N<sup>6</sup>-threonylcarbamoyladenine as a widely distributed tRNA hypermodification. *Nat. Chem. Biol.*, **9**, 105–113.
- Mateszewski, M., Wojciechowski, J., Miyauchi, K., Gdaniec, Z., Wolf, W.M., Suzuki, T. and Sochacka, E. (2017) A hydantoin isoform of cyclic N<sup>6</sup>-threonylcarbamoyladenine (ct<sup>6</sup>A) is present in tRNAs. *Nuc. Acids Res.*, **45**, 2137–2149.
- El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., Murzin, A.G. and de Crécy-Lagard, V. (2011) A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. *EMBO J.*, **30**, 882–893.
- 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, t<sup>6</sup>A. *EMBO J.*, **30**, 873–881.
- 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 t<sup>6</sup>A modification of tRNAs. *Nuc. Acids Res.*, **39**, 6148–6160.
- Zhang, W., Collinet, B., Graille, M., Daugeron, M.C., Lazar, N., Libri, D., Durand, D. and van Tilbeurgh, H. (2015) Crystal Structures of the Gon7/Pcc1 and Bud32/Cgi121 complexes provide a model for the complete yeast KEOPS complex. *Nuc. Acids Res.*, **43**, 3358–3372.
- 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.
- Downey, M., Houlsworth, R., Maringele, 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.
- Meng, F.L., Hu, Y., Shen, N., Tong, X.J., Wang, J., Ding, J. and Zhou, J.Q. (2009) SUA5p is a single-stranded telomeric DNA binding protein that facilitates telomere replication. *EMBO J.*, **28**, 1466–1478.
- He, M.-H., Liu, J.-C., Lu, Y.-S., Wu, Z.-J., Liu, Y.-Y., Wu, Z., Peng, J. and Zhou, J.-Q. (2019) KEOPS complex promotes homologous recombination via DNA resection. *Nuc. Acids Res.*, **47**, 5684–5697.
- Braun, D.A., Rao, J., Mollet, G., Schapiro, G., 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.
- 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 N<sup>6</sup>-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3(OGSEP) mutation manifest by neurodegeneration and renal tubulopathy. *Eur. J. Hum. Genet.*, **25**, 545–551.
- Arrandel, C., Missouri, S., Snoek, R., Patat, J., Menara, G., Collinet, B., Liger, D., Durand, D., Gribouval, O., Boyer, O. et al. (2019) Defects in t(6A) tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowatt syndrome. *Nat. Comm.*, **10**, 3967–3979.
- Wan, L.C., Mao, D.Y., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., Poda, G., Thevakumaran, N., Yuan, F. and Szilard, R.K. (2013) Reconstitution and characterization of eukaryotic N<sup>6</sup>-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nuc. Acids Res.*, **41**, 6332–6346.
- Lin, H., Miyauchi, K., Harada, T., Okita, R., Takeshita, E., Komaki, H., Fujioka, K., Yagasaki, H., Goto, Y., Yanaka, K. et al. (2018) CO<sub>2</sub>-Sensitive tRNA modification associated with human mitochondrial disease. *Nat. Comm.*, **9**, 1875–1892.
- Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V. and Iwata-Reuyl, D. (2012) Biosynthesis of threonylcarbamoyl adenosine (t<sup>6</sup>A), a universal tRNA nucleoside. *J. Biol. Chem.*, **287**, 13666–13673.
- Lauhon, C.T. (2012) Mechanism of threonylcarbamoyladenine (t<sup>6</sup>A) biosynthesis in bacteria: isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry*, **51**, 8950–8961.
- Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., van Tilbeurgh, H., Forterre, P. and Basta, T. (2013) *In vitro* biosynthesis of a universal t<sup>6</sup>A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.*, **41**, 1953–1964.
- Kostuch, A.P., Zheng, W., Liger, D., Daugeron, M.-C., Letoquart, J., Sierra-Galley, I.L.D.L., Forterre, P., Collinet, B., van Tilbeurgh, H. and Basta, T. (2018) Structure-function analysis of SUA5 protein reveals novel functional motifs required for the biosynthesis of the universal t<sup>6</sup>A tRNA modification. *RNA*, **24**, 926–938.
- Wan, L.C.K., Pillin, M.C., Thevakumaran, N., Sun, Y., Chakrabarty, A., Guarne, A., Kurinov, I., Durocher, D. and Sicheri, F. (2016) Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t<sup>6</sup>A biosynthesis. *Nucleic Acids Res.*, **44**, 6971–6980.
- Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Brüner, 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.
- Thiaville, P.C., El Yacoubi, B., Perrochia, L., Hecker, A., Prigent, M., Thiaville, J.T., Forterre, P., Namy, O., Basta, T. and de Crécy-Lagard, V. (2014) Cross kingdom function of the core universally conserved threonylcarbamoyladenine synthesis enzymes. *Eukaryot. Cell*, **13**, 1222–1231.
- 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* and proteins, YgjD and YeaZ and calorimetric evidence for the formation of a ternary YgjD-YeaZ-YjeE complex. *Protein Sci.*, **22**, 628–640.



37. Zhang, W., Collinet, B., Perrochia, L., Durand, D. and van Tilbeurgh, H. (2015) The ATP-mediated formation of the YgiD–YeaZ–YjeE complex is required for the biosynthesis of tRNA t<sup>6</sup>A in *Escherichia coli*. *Nuc. Acids Res.*, **43**, 1804–1817.
38. Luthra, A., Swinehart, W., Bayooz, S., Phan, P., Stec, B., Iwata-Reuyl, D. and Swairjo, M.A. (2018) Structure and mechanism of a bacterial t<sup>6</sup>A modification system. *Nuc. Acids Res.*, **46**, 1395–1411.
39. Missouri, S., Plancqueel, S., de la Sierra-Galley, I.L., Zhang, W., Liger, D., Durand, D., Dammak, R., Collinet, B. and van Tilbeurgh, H. (2018) The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t<sup>6</sup>A modification (erratum). *Nucleic Acids Res.*, **47**, 9464–9465.
40. Missouri, S., Plancqueel, S., de la Sierra-Galley, I.L., Zhang, W., Liger, D., Durand, D., Dammak, R., Collinet, B. and van Tilbeurgh, H. (2019) The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t<sup>6</sup>A modification. *Nucleic Acids Res.*, **46**, 5850–5860.
41. Luthra, A., Paraganama, N., Swinehart, W., Bayooz, S., Phan, P., Quach, V., Schiffer, J.M., Stec, B., Iwata-Reuyl, D. and Swairjo, M.A. (2019) Conformational communication mediates the reset step in t<sup>6</sup>A biosynthesis. *Nucleic Acids Res.*, **47**, 6551–6567.
42. 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.
43. 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.
44. Karst, J.C., Foucher, A.E., Campbell, T.L., DiGuilmi, A.M., Stroebel, D., Mangat, C.S., Brown, E.D. and Jault, J.M. (2009) The ATPase activity of an “essential” *Bacillus subtilis* enzyme, YdiB, is required for its cellular function and is modulated by oligomerization. *Microbiology*, **155**, 944–956.
45. Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., Karamitsu, S. and Shinkai, A. (2008) X-ray crystal structure of a hypothetical SUA5 protein from *Sulfolobus tokodaii* strain 7. *Proteins*, **70**, 1108–1111.
46. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
47. Peterson, E.T. and Uhlenbeck, O.C. (1992) Determination of recognition nucleotides for *Escherichia coli* phenylalanine tRNA synthetase. *Biochemistry*, **31**, 10380–10389.
48. Gehrke, C.W., Kuo, K.C., McCune, R.A. and Gerhardt, K.O. (1982) Ribonucleoside analysis by reversed phase high performance liquid chromatography. *J. Chromatogr.*, **230**, 297–308.
49. Pomerantz, S.C. and McCloskey, J.A. (1990) Analysis of RNA hydrolyzates by liquid-chromatography-mass spectrometry. *Methods Enzymol.*, **193**, 796–824.
50. Morrison, J.F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta - Enzymol.*, **185**, 269–286.
51. Kimura, J., Yagi, K., Suzuki, H. and Mitsunobu, O. (1980) Studies on nucleosides and nucleotides. VIII. Preparations and reactions of triphenylphosphorandynucleosides. *Bull. Chem. Soc. Jpn.*, **53**, 3670.
52. Klein, E., Mons, S., Valleix, A., Mioskowski, C. and Lebeau, L. (2002) Synthesis of enzymatically and chemically non-hydrolyzable analogs of dinucleoside triphosphates Ap<sub>3</sub>A and Gp<sub>3</sub>G. *J. Org. Chem.*, **67**, 146–153.
53. Duckworth, B.P., Nelson, K.M. and Aldrich, C.C. (2012) Adenylating enzymes in *Mycobacterium tuberculosis* as drug targets. *Curr. Top. Med. Chem.*, **12**, 766–796.
54. Lux, M.C., Standke, L.C. and Tan, D.S. (2019) Targeting adenylate-forming enzymes with designed sulfonyladenine inhibitors. *J. Antibiot.*, **72**, 325–349.
55. Park, J.O., Rubin, S.A., Xu, Y.F., Amador-Nogues, D., Fan, J., Schlomi, T. and Rabinowitz, J.D. (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.*, **12**, 482–492.
56. Buckstein, M.H., He, J. and Rubin, H. (2008) Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. *J. Bacteriol.*, **190**, 718–726.
57. Yaginuma, H., Kawai, S., Tabata, K.V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., Noji, H. and Imamura, H. (2014) Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.*, **4**, 6522–6529.
58. Swinehart, W., Deutsch, C., Sarachan, K.L., Luthra, A., Bacusmo, J.M., de Crecy-Lagard, V., Swairjo, M.A., Agris, P. and Iwata-Reuyl, D. (2020) Specificity in the biosynthesis of the universal tRNA nucleoside N6-threonylcarbamoyl adenosine (t<sup>6</sup>A) – TsaD is the gatekeeper. *RNA*, **26**, 1094–1103.
59. Cheng, Y. and Prusoff, W.H. (1973) Relationship between the inhibition constant (K<sub>i</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem. Pharmacol.*, **22**, 3099–3108.
60. Weinken, C.J., Baaske, P., Rothbauer, U., Braun, D. and Duhr, S. (2010) Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Comm.*, **1**, 100.
61. Amari, M., Ratih, R., Alhazmi, H.A. and Deeb, S.E. (2018) Thermophoresis for characterizing biomolecular interaction. *Methods*, **146**, 107–119.
62. 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. *Nuc. Acids Res.*, **20**, 5607–5615.
63. Yu, N., Jora, M., Solivio, B., Thakur, P., Acevedo-Rocha, C.G., Randau, L., de Crecy-Lagard, V., Addepalli, B. and Limbach, P.A. (2019) tRNA modification profiles and codon-decoding strategies in *Methanocaldococcus jannaschii*. *J. Bacteriol.*, **201**, e00690-18.
64. Morris, G.M., Goodsell, G.S., Huey, R. and Olson, A.J. (1996) Distributed automated docking of flexible ligands to proteins. Parallel applications of Autodock 2.4. *J. Comput.-Aided Mol. Des.*, **10**, 293–304.
65. Wang, Q., Xia, J., Guallar, V., Krilov, G. and Kantrowitz, E.R. (2008) Mechanism of thermal decomposition of carbamoyl phosphate and its stabilization by aspartate and ornithine transcarbamoylases. *Proc. Nat. Acad. Sci. U.S.A.*, **105**, 16918–16925.
66. Purcarea, C., Evans, D.R. and Herve, G. (1999) Channeling of carbamoyl phosphate to the pyrimidine and arginine biosynthetic pathways in the deep sea hyperthermophilic archeon *Pyrococcus abyssi*. *J. Biol. Chem.*, **274**, 6122–6129.
67. 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.
68. 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. *Prot. Sci.*, **9**, 2557–2566.
69. Laskowski, R.A. and Swindells, M.B. (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, **51**, 2778–2786.
70. Katoh, M. and Kuma, M. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
71. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new Endscript server. *Nucleic Acids Res.*, **42**, W320–W324.

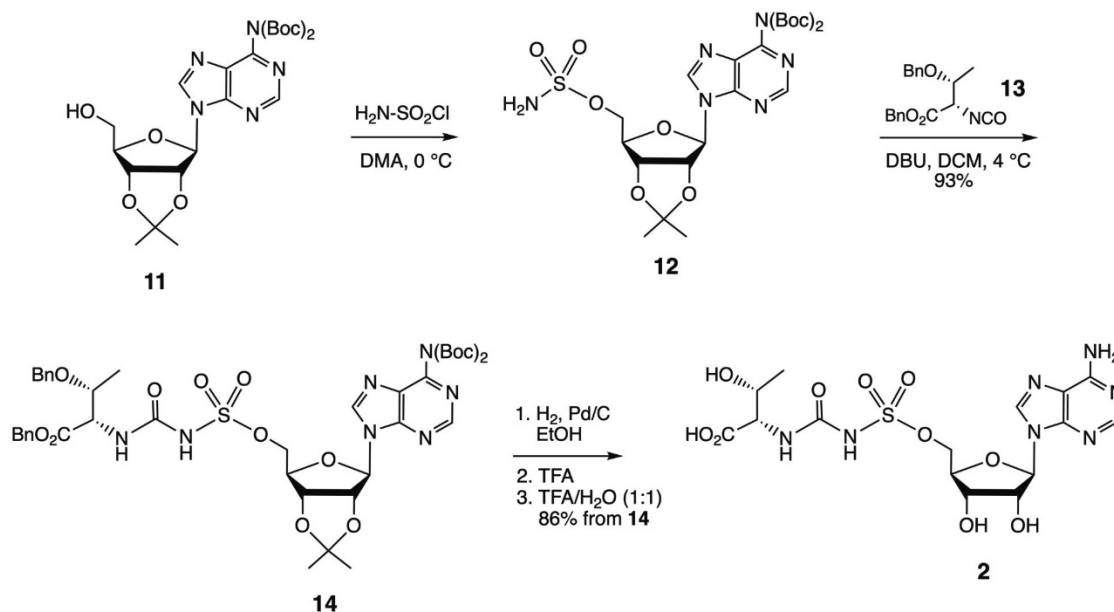
## SUPPLEMENTARY MATERIAL

### MATERIALS AND METHODS

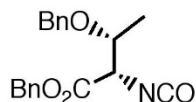
#### General.

<sup>1</sup>H NMR spectra were recorded on a Varian Innova 400 MHz spectrometer at ambient temperature. <sup>13</sup>C NMR spectra were recorded at 75.0 MHz at ambient temperature. Data for <sup>1</sup>H NMR are reported as follows: chemical shift, integration, multiplicity (app = apparent, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet) and J-coupling constants. All <sup>13</sup>C NMR spectra were recorded with complete proton decoupling. High-resolution mass spectra (HRMS) were obtained in the UW-Madison School of Pharmacy Analytical Instrumentation Center using a Bruker MaXis 4G QTOF instrument with an ESI source. Microscale Thermophoresis measurements were obtained using a Monolith NT.115 instrument (Nano Temper Technologies) in the UW School of Pharmacy Medicinal Chemistry Center. Analytical thin layer chromatography was performed on 0.25 mm silica gel 60-F plates. Chromatography was performed using 200-400 mesh silica gel (EMD). Yields refer to chromatographically and spectroscopically pure materials, unless otherwise stated. Dry DCM, THF, and DMF were obtained using a dry solvent dispensing system. All other solvents were commercial reagent grade. All reactions were carried out in oven-dried glassware under a nitrogen atmosphere unless otherwise noted. Compounds **11** and **12** were synthesized as previously described (ref. 1, 2).

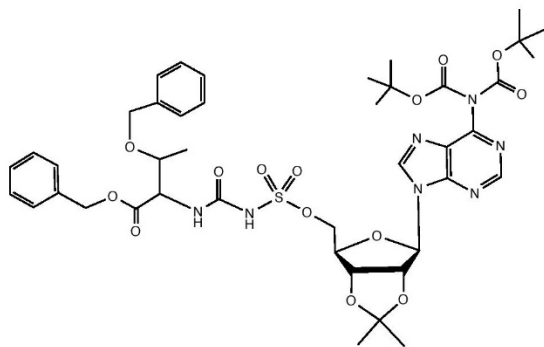
#### Synthesis of Compounds 2-5



**Supplementary Figure S1.** Synthetic route for the preparation of sulfamate inhibitor **2**.

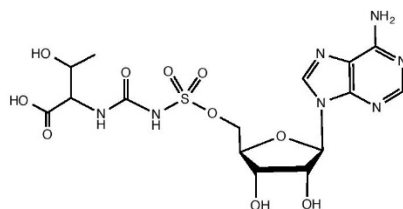


**Benzyl-(S)-2-isocyanato-(R)-benzyloxypropanoate (13)** (ref. 3, 4). To a solution of 510 mg (1.7 mmol) benzyl(O-benzyl)-L-threonine (**6**) in 7 mL DCM is added 7 mL saturated aqueous sodium bicarbonate. After stirring at 0 °C for 10 min, 167 mg (0.56 mmol) of triphosgene is added in one portion. The mixture is stirred vigorously for 25 min then poured into a separatory funnel. The bottom organic layer is removed and the top aqueous layer is extracted with 10 mL DCM, and the two organic layers are combined, dried (Na<sub>2</sub>SO<sub>4</sub>) and the solvent is removed to give 500 mg of the isocyanate as a yellow oil which solidifies at -20 °C. This material was typically used directly in the next reaction. <sup>1</sup>H NMR (CDCl<sub>3</sub>) δ 1.28 (d, 3H, J = 3 Hz), 3.88 (d, 1H, J = 2.5 Hz), 4.03 (m, 1H), 4.53,4.32 (ABq, 2H, J<sub>AB</sub> = 12 Hz), 5.13, 5.19 (ABq, 2H, J<sub>AB</sub> = 12 Hz), 7.16-7.31 (m, 10H). <sup>13</sup>C NMR (CDCl<sub>3</sub>) δ 63.0, 68.1, 70.8, 74.4, 127.1, 127.5, 127.7, 128.4, 128.5, 128.67, 128.71, 128.74, 134.9, 137.8, 169.6. HRMS calcd for C<sub>19</sub>H<sub>19</sub>NO<sub>4</sub>: 326.13869 (MH<sup>+</sup>); Found: 326.13847.

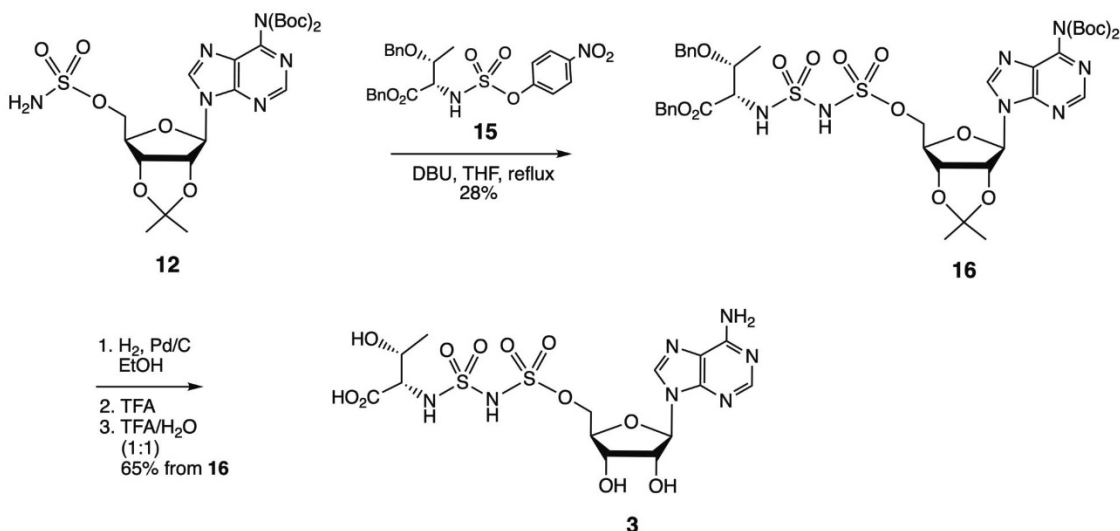


**6-N-Bis-t-butoxycarbonyl-2',3'-isopropylidene-5'-[O-benzyl-(O-benzyl)-L-threonylcarbamoyl]-sulfamoyladenine (14)**. To a solution of 185 mg sulfamate **12** (ref. 2) in 1.5 mL of THF at 0 °C was added 100 mg (0.31 mmol) benzyl-(O-benzyl)-L-threonine isocyanate **13** followed by 48 μL (0.31 mmol) of DBU via syringe. Mixture was stirred for 2h at 0 °C then stored at 4 °C overnight. The mixture was then diluted with CH<sub>2</sub>Cl<sub>2</sub>, washed successively with cold 10% aqueous AcOH, water 0.5 M NaHCO<sub>3</sub>, and brine. The organic layer was then dried with solid Na<sub>2</sub>SO<sub>4</sub>, filtered and solvent removed to give crude urea as an oil. Further purification by chromatography (1-6% methanol in CH<sub>2</sub>Cl<sub>2</sub>) gave 263 mg (93%) of pure **14** as a white foam. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 1.16 (d, 3 H, J = 6.4 Hz, 3H), 1.39 (s, 3H), 1.45 (s, 18H), 1.62 (s, 3H), 4.12 (q, J = 6.4 Hz, 1H), 4.25 (d, J = 12 Hz, 1H), 4.40 (dd, J = 5.6, 12.0 Hz, 1H), 4.47 (d, J = 12 Hz, 2H), 4.55 (m, 2H), 5.10 (ABq with overlapping 4'-H, J<sub>AB</sub> = 12.0 Hz, 3H), 5.49 (d, J = 5.6 Hz, 1H), 6.22 (s, 1H), 6.74 (d, J = 8.8 Hz, 1H), 7.16 (d, J = 7.2 Hz, 1H), 7.24-7.27 (m, 10H), 8.33 (s, 1H), 8.87 (s, 1H). <sup>13</sup>C NMR: (CDCl<sub>3</sub>): δ 16.6, 25.5, 27.3, 28.0, 58.2, 67.6, 70.8, 71.4, 73.9, 77.5, 84.2, 84.4, 84.6, 91.4, 114.9, 127.8, 127.9, 128.55, 128.57,

128.7, 128.8, 129.2, 135.3, 137.9, 144.3, 150.5, 151.0, 151.7, 152.6, 152.9, 170.3. HRMS calcd for C<sub>42</sub>H<sub>53</sub>N<sub>7</sub>O<sub>14</sub>S: 934.3263 (MNa<sup>+</sup>); Found: 934.3268.



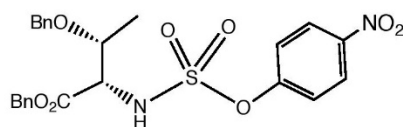
**5'-(L-Threonylcarbamoylsulfamoyl)-adenosine (2).** To a solution of 50 mg (92 μmol) of **14** in 2 mL of absolute ethanol was added 20 mg of 10% Pd/C. The mixture was stirred under a hydrogen atmosphere (balloon) overnight at rt, after which the reaction was complete as judged by TLC (CH<sub>3</sub>CN/H<sub>2</sub>O/HOAc- 20:4:1). The mixture was filtered through a plug of glass wool, then a sintered glass funnel and the solvent was removed to give 50 mg of the debenzylated product as a clear glass. To this material at 0 °C, was added 7 mL of TFA and 7 mL of water. The mixture was then stirred at rt for 6 h. The solvents were removed *in vacuo* and the residue twice dissolved in 2 mL of water and volatiles removed to give 43 mg (86%) of crude **2** which was 95% pure by HPLC. Final purification was achieved using HPLC on a Supelco LC-18 semiprep column (rt = 10 min. Buffer A: 10 mM ammonium bicarbonate pH 7.3; Buffer B: same as buffer A with 20% methanol. Gradient 0-30 min, from 0-100% Buffer B). <sup>1</sup>H NMR (D<sub>2</sub>O) δ 1.08 (d, J = 6.4 Hz, 3H), 3.95 (d, J = 3.2 Hz, 1H), 4.13 (m, 1H), 4.29 (m, 2H), 4.40 (m, 1H), 4.45 (m, 1H), 6.09 (d, J = 5.6 Hz, 1H), 8.21 (s, 1H), 8.42 (s, 1H). <sup>13</sup>C NMR (D<sub>2</sub>O) δ 19.1, 60.9, 67.7, 68.4, 70.3, 74.1, 82.7, 87.2, 110.0, 118.5, 139.7, 148.8, 152.6, 155.4, 161.2, 178.3. HRMS Calcd for C<sub>15</sub>H<sub>21</sub>N<sub>7</sub>O<sub>10</sub>S, 490.09978 (M-H<sup>+</sup>); Found, 490.10077.



**Supplementary Figure S2.** Synthetic route for the preparation of aminosulfonylsulfamate inhibitor **3**.

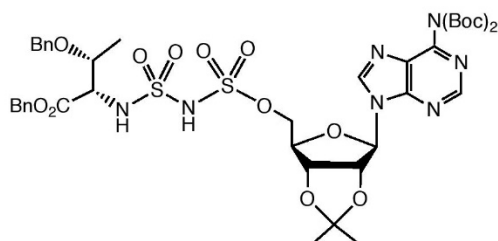
### ***p*-Nitrophenolsulfonylchloride**

A mixture of 2.78 g (20 mmol) *p*-nitrophenol and 1.62 mL (1.59 g, 20 mmol) anhydrous pyridine in 20 mL anhydrous ether was added dropwise to a mixture of 1.60 mL of sulfonyl chloride (2.7g, 20 mmol) in 20 mL of ether previously cooled to -78°C (dry ice/acetone), under a nitrogen atmosphere. After the addition (~ 5 min), the reaction mixture was allowed to warm up to room temperature and stirred for 18 h. The solvent was removed under vacuum, and the resulting oil was purified by chromatography using DCM to give 3.93 g of *p*-nitrophenylsulfonylchloride, as a pale yellow oil (82.5% yield) which solidified upon storage at -20 °C. Spectral data were in agreement with previously reported values (ref. 5).



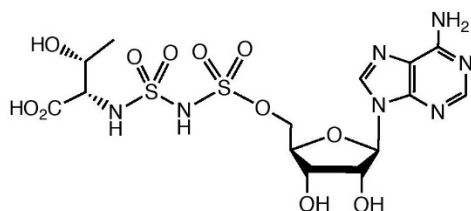
### **Benzyl-(O-benzyl)-L-threonine-N-sulfamoyl-*p*-nitrophenol (**15**)**

To a solution of 700 mg (2.34 mmol) benzyl-(O-benzyl)-L-threonine **6** (neutral form) in 9.7 mL DCM was added 1.98 mL (1.42 g, 14.0 mmol, 6 eq) of triethylamine, and 0.977 g (7.02 mmol, 3 eq) of *p*-nitrophenol. This mixture was added dropwise to a solution of 1.11 g (4.68 mmol, 2 eq) *p*-nitrophenylsulfonyl chloride dissolved in 2.4 mL of DCM previously cooled to -78°C under a nitrogen atmosphere with stirring. The mixture was stirred for one hour at -78 °C, and then concentrated under vacuum and purified by chromatography using a gradient of 4:1 to 2:1 ethyl acetate:hexane. Product **15** was obtained as a yellow solid (0.862 g, 75% yield). <sup>1</sup>H NMR (CDCl<sub>3</sub>) δ = 1.19 (d, J = 6.4 Hz, 3H), 3.99 (m, 1H), 4.16 (m, 2H), 4.39 (d, J = Hz, 1H), 5.02 (ABX pattern, 2H), 5.81 (bs, 1H), 7.03 (m, 2H), 7.06-7.11 (m, 8H), 7.24 (d, J = 9.2 Hz, 2H), 8.02 (d, J = 9.2 Hz, 2H). <sup>13</sup>C NMR (CDCl<sub>3</sub>) δ 16.4, 62.1, 68.14, 68.14, 68.18, 68.20, 71.1, 74.4, 122.6, 125.7, 128.0, 128.2, 128.7, 128.9, 128.97, 129.1, 135.0, 137.6, 146.1, 154.8, 169.6. HRMS Calcd for C<sub>24</sub>H<sub>24</sub>N<sub>2</sub>O<sub>8</sub>S (Na<sup>+</sup>): 523.11456; Found: 523.11478.



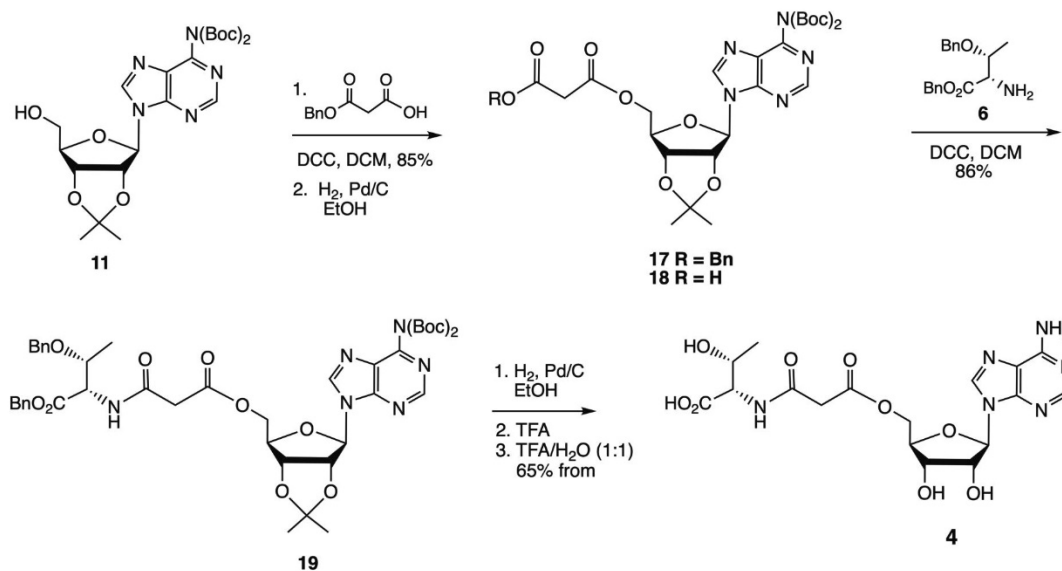
### 5'-[Benzyl-(O-benzyl)-L-N-threonylsulfonyl]-sulfamoyladenosine (**16**)

To 1.80 mL of THF was added 426 mg (0.73 mmol) **12**, 0.400 g (0.80 mmol, 1.1 eq) **14**, and 110  $\mu$ L (0.727 mmol, 1 eq) DBU at RT. This mixture was heated to 80°C, and refluxed overnight. Solvent was removed under reduced pressure and the residue was dissolved in DCM and purified by chromatography using ethyl acetate:hexane (1:1) as eluent to give 188 mg (28% yield) of **15** as a white solid.  $^1\text{H NMR}$  ( $\text{CDCl}_3$ )  $\delta$  1.21 (d,  $J = 6.4$  Hz, 3H), 1.38 (s, 3H), 1.43 (s, 18H), 1.59 (s, 3H), 4.01 (m, 1H), 4.23 (d, 1H,  $J =$ ), 4.19-4.35 (m, 4H), 4.42 (d, 1H,  $J = 12.4$  Hz), 4.58 (bs, 1H), 5.33 (d, 1H  $J = 2.4$  Hz), 5.01-5.15 (m, 3H), 5.68 (d, 1H,  $J = 5.4$  Hz), 6.27 (d, 1H,  $J = 2.4$ ), 7.05-7.30 (m, 10H), 8.68 (s, 1H), 8.82 (s, 1H). HRMS Calcd for  $\text{C}_{41}\text{H}_{53}\text{N}_7\text{O}_{15}\text{S}_2$  (M-H): 946.29628. Found: 946.29893.

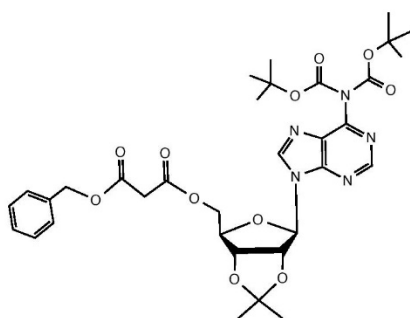


### 5'-(L-N-Threonylsulfonyl)-sulfamoyladenosine (**3**).

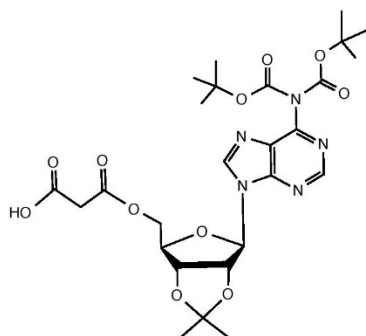
Hydrogenation of the benzyl protecting groups was performed similar to the preparation of (**2**) above. Removal of the Boc and isopropylidene groups was done with 50% TFA in water. After stirring for 3 hours, the solvent removed under vacuum and the residue was dissolved in  $\text{H}_2\text{O}$  and solvent was removed (2x) to give crude **3**. Final purification was by HPLC (Supelco LC-18 semiprep column).  $^1\text{H NMR}$  ( $\text{D}_2\text{O}$ )  $\delta$  1.18 (d,  $J = 6.4$  Hz, 3H), 3.95 (m, 2H), 4.38 (m, 1H), 4.40 (m, 1H), 4.49 (d,  $J = 4.4$  Hz, 1H), 6.07 (d,  $J = 5.6$  Hz, 1H), 8.17 (s, 1H), 8.36 (s, 1H). HRMS Calcd for  $\text{C}_{14}\text{H}_{21}\text{N}_7\text{O}_{11}\text{S}_2$  (M-H): 526.06677. Found: 526.06837.



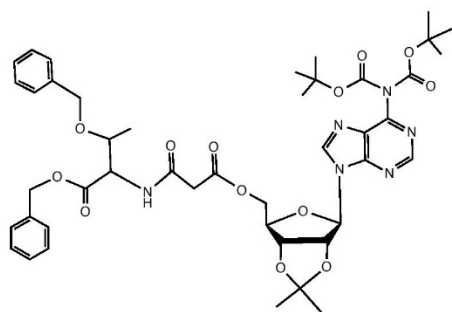
**Supplementary Figure S3.** Synthetic route for the preparation of malonic amide ester inhibitor 4.



**6-N-Bis-t-butoxycarbonyl-2',3'-isopropylidene-5'-benzylmalonyladenosine (17).** To a solution of 221 mg (0.43 mmol) of in 0.43 mL of  $\text{CH}_2\text{Cl}_2$  in an ice bath was added 91 mg (0.47 mmol) malonic acid monobenzyl ester followed by 107 mg (0.52 mmol) of DCC. The mixture was stirred 1 h then filtered on a sintered glass funnel and the solid was washed with 1 mL of  $\text{CH}_2\text{Cl}_2$ . The filtrate was concentrated to an oil which was purified by chromatography (ethyl acetate:hexanes, 1:2 to 1:1) to give 251 mg (85%) of as a white waxy solid.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  1.40 (s, 3H), 1.46 (s, 18H), 1.63 (s, 3H), 3.41, 3.37 (ABq,  $J = 16$  Hz, 2H), 4.36 (dd,  $J = 5.6, 12.0$  Hz, 1H), 4.43 (dd,  $J = 4.0, 12$  Hz, 1H), 4.50 (m, 1H), 5.02 (dd,  $J = 3.2, 6.0$  Hz, 1H), 5.16 (s, 2H), 5.39 (dd,  $J = 2.4, 6.0$  Hz, 1H), 6.17 (d,  $J = 2$  Hz, 1H), 7.33-7.36 (m, 5H), 8.17 (s, 1H), 8.87 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  25.2, 25.6, 25.9, 27.4, 28.0, 34.1, 41.4, 64.8, 67.6, 81.5, 84.1, 84.3, 84.7, 91.1, 115.1, 128.6, 128.7, 128.8, 129.7, 135.3, 143.9, 150.7, 150.9, 152.5, 152.6, 166.0, 166.3. HRMS Calcd for  $\text{C}_{33}\text{H}_{41}\text{N}_5\text{O}_{11}$ : 684.28753 (MH<sup>+</sup>); Found: 684.28764.



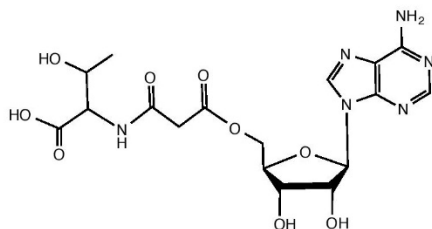
**6-N-Bis-t-butoxycarbonyl-2',3'-isopropylidene-5'-malonyladenosine (18).** To a solution of 118 mg (0.17 mmol) of in 4 mL of absolute ethanol was added 40 mg of 10% Pd/C. The mixture was stirred under a hydrogen atmosphere overnight at rt then filtered through a glass plug and the solvent removed to give 97 mg of a clear glass that was homogenous by TLC.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  1.39 (s, 3H), 1.47 (s, 18H), 1.64 (s, 3H), 3.32 (s, 2H), 4.38 (dd,  $J = 3.2, 12.0$  Hz, 1H), 4.50 (dd,  $J = 4.4, 12.0$  Hz, 1H), 4.60 (bs, 1H), 5.00 (dd,  $J = 2.8, 6.4$  Hz, 1H), 5.45 (dd,  $J = 2.0, 4.0$  Hz, 1H), 6.24 (d,  $J = 2.4$  Hz, 1H), 8.40 (s, 1H), 8.92 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  25.5, 27.4, 28.0, 41.3, 64.7, 77.5, 81.3, 84.5, 84.8, 91.3, 115.1, 129.0, 144.3, 150.3, 150.7, 152.7, 166.5, 169.0. HRMS Calcd for  $\text{C}_{26}\text{H}_{35}\text{N}_5\text{O}_{11}$ : 594.24058 (MH $^+$ ); Found: 594.24031.



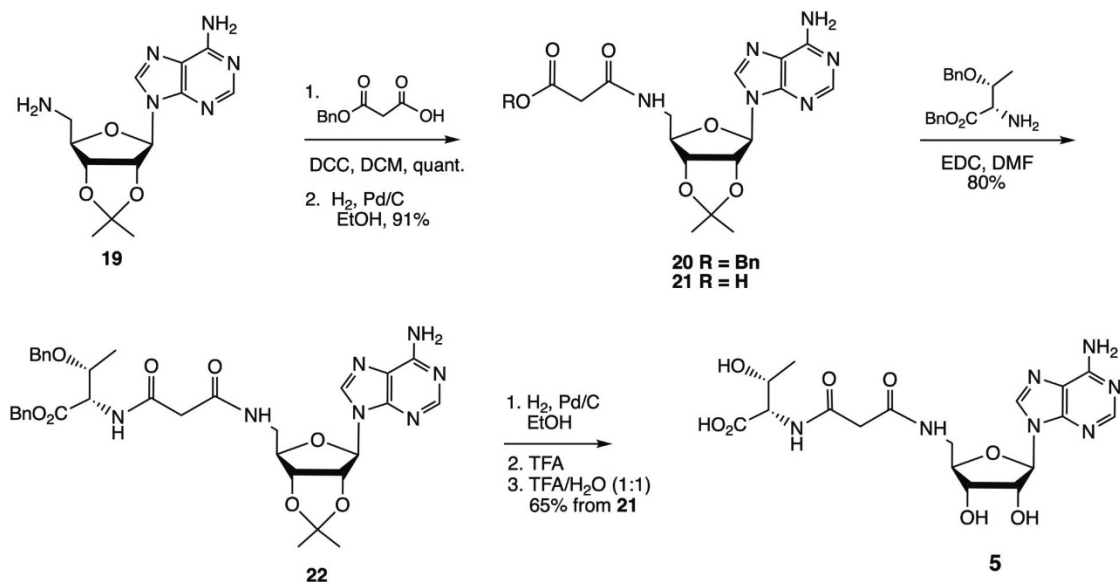
**6-N-Bis-t-butoxycarbonyl-2',3'-isopropylidene-5'-[benzyl-(O-benzyl)-L-threonyl-N-malonyl]adenosine (19).** To a solution of 233 mg (0.39 mmol) of in 1.0 mL of  $\text{CH}_2\text{Cl}_2$ , at 0 °C is added 129 mg (0.43 mmol) of amine followed by 89 mg (0.43 mmol) of DCC with stirring. The ice bath was removed and the mixture was stirred at rt for 1 h, diluted with 10 mL of ether and filtered on a sintered glass funnel. The filtrate was diluted with additional ether and washed successively with 20 mL of 5% cold aqueous acetic acid, water, saturated aqueous sodium bicarbonate and brine. The organic layer was dried with anhydrous sodium sulfate, filtered and the solvent was removed to give 322 mg of crude amide. This material was purified by chromatography using ethyl acetate/hexane (1:2 to 1:1) as eluent to give 298 mg (86%) of



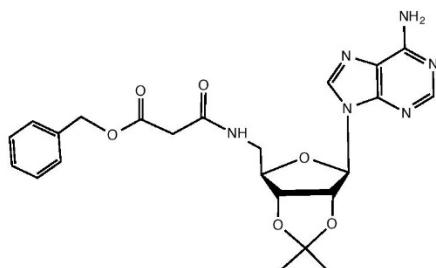
malonate amide as a clear oil.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  1.17 (d,  $J$  = 6.4 Hz, 3H), 1.40 (s, 3H), 1.46 (s, 18H), 1.63 (s, 3H), 3.27, 3.35 (ABq,  $J$  = 14.0 Hz, 2H), 4.14 (dq,  $J$  = 2.4, 6.4 Hz, 1H), 4.16 (dd,  $J$  = 2.4, 6.4 Hz, 1H), 4.32 (dd,  $J$  = 5.6, 12.0 Hz, 1H), 4.27, 4.49 (ABq,  $J$  = 12.0 Hz, 2H), 4.42 (dd,  $J$  = 4.0, 12.0 Hz, 1H), 4.53 (m, 1H), 4.66 (dd,  $J$  = 2.0, 9.0 Hz, 1H), 5.05 (dd,  $J$  = 2.8, 6.4 Hz, 1H), 5.07, 5.11 (ABq,  $J$  = 12.4 Hz, 2H), 5.42 (dd,  $J$  = 2.0, 6.4 Hz, 1H), 6.17 (d,  $J$  = 2.0 Hz, 1H), 8.23 (s, 1H), 8.87 (s, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  16.6, 25.6, 27.4, 28.0, 41.9, 57.2, 64.9, 67.4, 71.0, 74.3, 81.6, 84.1, 84.4, 84.7, 91.3, 115.1, 127.9, 128.0, 128.6, 128.64, 128.8, 129.7, 135.4, 138.0, 135.4, 138.0, 144.0, 150.7, 150.9, 152.5, 152.7, 165.5, 167.8, 170.2. HRMS Calcd for  $\text{C}_{44}\text{H}_{51}\text{N}_6\text{O}_{13}$ , 875.3822 (MH<sup>+</sup>); Found, 875.3828.



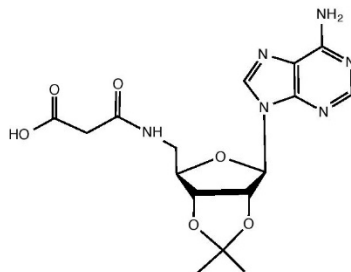
**5'-L-Threonyl-N-malonyadenosine (4).** To 80 mg of is added 3 mL of ethanol is added 30 mg 10% Pd/C under N<sub>2</sub>. The flask is equipped with a stopcock adapter attached to a balloon containing hydrogen gas and the reaction mixture is stirred overnight at RT. The mixture is filtered through glass wool and the solvent is removed. The residue is then dissolved in a mixture of 2 mL trifluoroacetic acid and 2 mL H<sub>2</sub>O and stirred at RT for 3 h. HPLC shows a mixture of desired product and adenosine (90:10). The volatiles were removed and the residue was dissolved in 2 mL of H<sub>2</sub>O and evaporated (3x). The crude product was twice dissolved in 2 mL of H<sub>2</sub>O and the solvent was removed to give a pale yellow oil. Final purification was by HPLC using a semiprep Supelco LC-18 column with 10 mM NH<sub>4</sub>OAc pH 6 as eluent to give 31 mg (75%) of pure **4** as the ammonium salt.  $^1\text{H}$  NMR ( $\text{D}_2\text{O}$ )  $\delta$  1.08 (d,  $J$  = 6.8 Hz, 3H), 3.94 (d,  $J$  = 2.8 Hz, 1H), 4.13 (dq,  $J$  = 2.4, 4.0 Hz, 1H), 4.29 (bs, 2H), 4.39 (d,  $J$  = 2.8 Hz, 1H), 4.44 (m, 1H), 4.70 (m, 1H), 6.07 (d,  $J$  = 5.6 Hz, 1H), 8.17 (s, 1H), 8.39 (s, 1H).



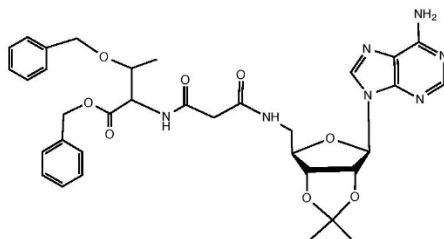
**Supplementary Figure S4.** Synthetic route for the preparation of malonic diamide inhibitor **5**.



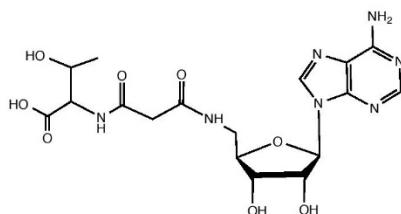
**5'-Benzylmalonylamino-5'-deoxyadenosine (20).** To a solution of 180 mg (0.58 mmol) **19** and 124 mg (0.63 mmol) of malonic acid monobenzyl ester in 1 mL of  $\text{CH}_2\text{Cl}_2$  at 0 °C was added 120 mg (0.63 mmol) of DCC. After stirring at rt 30 min, the mixture was filtered and the solid was washed with 1 mL of  $\text{CH}_2\text{Cl}_2$ . The filtrate was evaporated to give crude, which was purified by chromatography (ethyl acetate: methanol 95:5) to give 282 mg (quant.) of **20** as a white foam.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  1.36 (s, 3H), 1.62 (s, 3H), 3.32 (t,  $J = 7.5$  Hz, 1H), 3.48 (s, 2H), 4.13 (t,  $J = 7.5$  Hz, 1H), 4.49 (dd,  $J = 2.0, 6.0$  Hz, 1H), 4.91 (dd,  $J = 4.8, 6.0$  Hz, 1H), 5.11 (AB,  $J = 12.0$  Hz, 2H), 5.83 (d,  $J = 4.8$  Hz, 1H), 6.57 (bs, 2H), 7.17-7.28 (m, 5H), 7.29 (s, 1H), 8.26 (s, 1H), 8.71 (dd,  $J = 2.0, 8.4$  Hz, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  25.4, 27.6, 41.5, 43.0, 67.5, 77.6, 81.9, 82.5, 83.8, 92.7, 114.8, 121.2, 128.5, 128.6, 135.3, 140.6, 148.9, 152.8, 156.4, 166.1, 168.5. HRMS Calcd for  $\text{C}_{23}\text{H}_{26}\text{N}_6\text{O}_6$ : 483.1986 (MH<sup>+</sup>); Found: 483.1961.



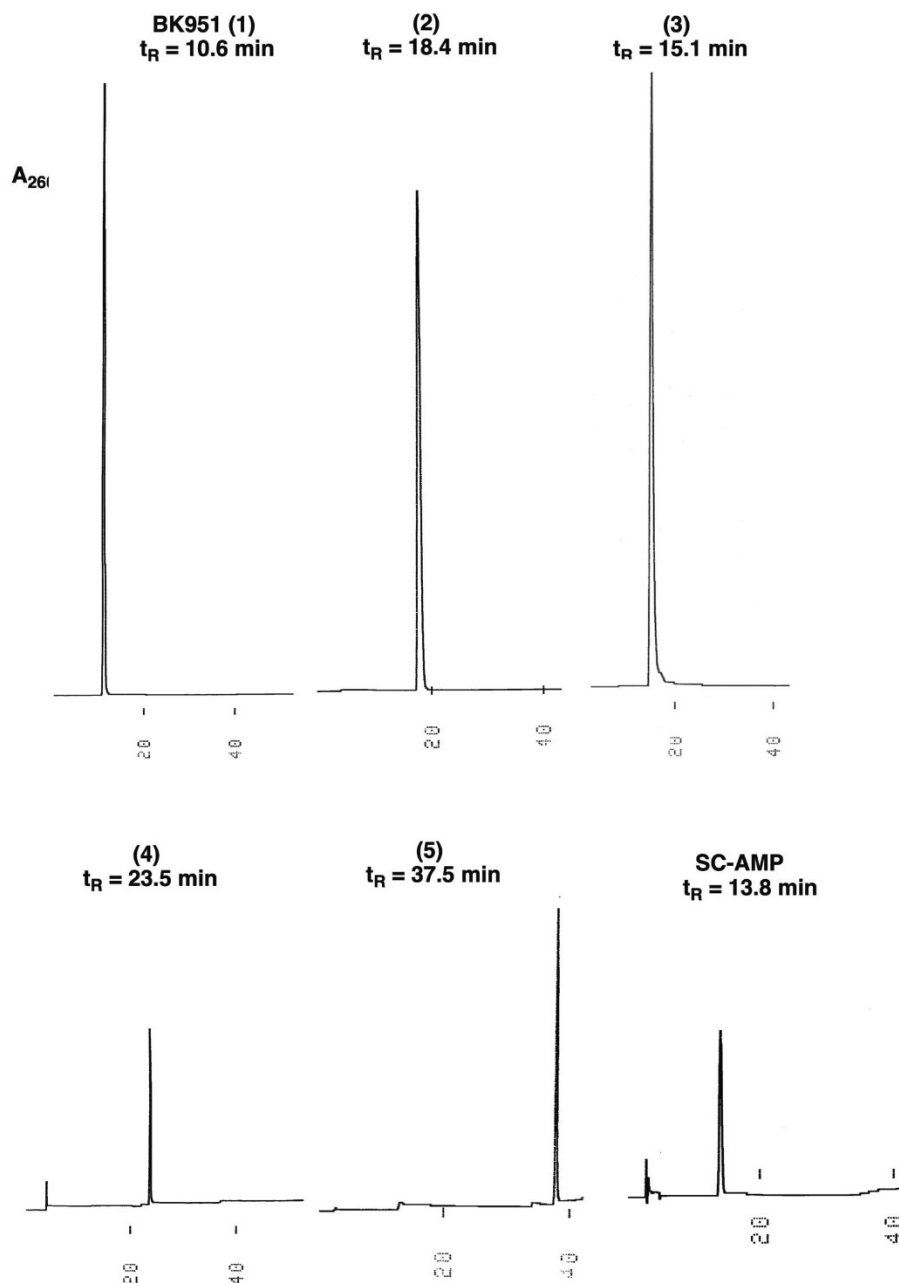
**5'-malonylamino-5'-deoxyadenosine (21).** To a solution of 284 mg (0.59 mmol) of in 10 mL of absolute methanol was added 90 mg of 10% Pd/C. After stirring under a hydrogen atmosphere (balloon) overnight at rt, the mixture was filtered and the solvent removed to give 211 mg (91%) of the debenzylated product as a colorless glass.  $^1\text{H NMR}$  ( $\text{CD}_3\text{OD}$ )  $\delta$  1.32 (s, 3H), 1.55 (s, 3H), 3.32 (s, 2H), 3.42 (dd,  $J = 4.4, 14.4$  Hz, 1H), 3.71 (dd,  $J = 5.2, 14.4$  Hz, 1H), 4.34 (dd,  $J = 5.2, 8.0$  Hz, 1H), 4.94 (dd,  $J = 3.2, 6.4$  Hz, 1H), 5.36 (dd,  $J = 3.2, 6.0$  Hz, 1H), 6.08 (d,  $J = 3.2$  Hz, 1H), 8.25 (s, 1H), 8.26 (s, 1H).  $^{13}\text{C NMR}$  ( $\text{CD}_3\text{OD}$ )  $\delta$  24.4, 26.3, 41.1, 81.9, 82.0, 83.4, 83.5, 84.5, 90.8, 90.9, 114.6, 119.6, 141.2, 148.8, 151.5, 155.4, 168.3, 170.4. HRMS Calcd for  $\text{C}_{16}\text{H}_{20}\text{N}_6\text{O}_6$ : 393.1517 (MH<sup>+</sup>); Found: 393.1516.



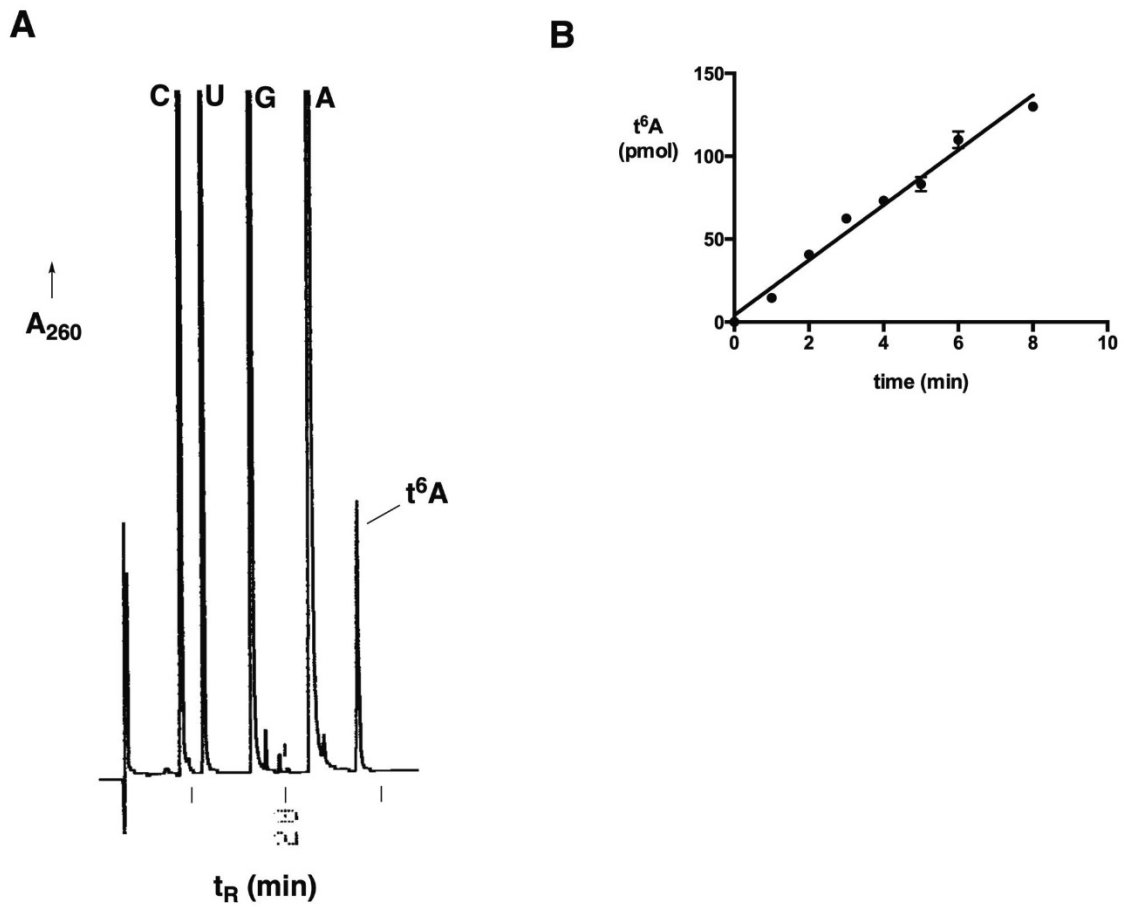
**Benzyl-(O-benzyl)-L-threonyl-N-malonylamino-5'-deoxyadenosine (22).** To a solution of 184 mg (0.47 mmol) of malonate in 1 mL of DMF was added 154 mg (0.51 mmol) of amine followed by 99 mg (0.51 mmol) of EDC. After stirring 2 h at rt the reaction was quenched by addition of 5 mL of cold 10% aq. acetic acid and extracted with 10 mL of  $\text{CH}_2\text{Cl}_2$ . The organic layer was successively washed with 10 mL of water, 0.5 M aq.  $\text{NaHCO}_3$ , and sat. aq.  $\text{NaCl}$ , then dried ( $\text{Na}_2\text{SO}_4$ ), filtered and the solvent removed to give crude as an oil. This material was purified by chromatography to give 252 mg (80%) of the final as a clear oil.  $^1\text{H NMR}$  ( $\text{CDCl}_3$ )  $\delta$  1.20 (d,  $J = 6.4$  Hz, 3 H), 1.34 (s, 3H), 1.60 (s, 3H), 3.38 (app. d,  $J = 13.0$  Hz, 1H), 3.44 (s, 2H), 4.04 (ddd,  $J = 3.2, 8.0, 13.0$  Hz, 1H), 4.16 (dq,  $J = 2.4, 6.4$  Hz, 1H), 4.27, 4.46, (ABq w/overlapping 4'-H,  $J = 12.0$  Hz, 3H), 4.74 (dd,  $J = 2.4, 9.2$  Hz, 1H), 4.87 (dd,  $J = 2.4, 6.4$  Hz, 1H), 5.07 (ABq,  $J = 12.4$  Hz, 2H), 5.27 ( $J = 4.4, 6.0$  Hz, 1H), 5.87 (d,  $J = 4.4$  Hz, 1H), 6.62 (s, 2H), 7.13-7.27 (m, 5H), 7.87 (s, 1H), 8.12 ( $J = 9.2$  Hz, 1H), 8.36 (s, 1H), 8.66 (d,  $J = 5.6$  Hz, 1H).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ )  $\delta$  16.6, 25.5, 27.7, 41.4, 43.5, 57.3, 67.4, 70.9, 74.3, 77.6, 81.8, 83.0, 83.9, 92.0, 114.9, 120.8, 127.8, 128.47, 128.53, 128.54, 128.72, 135.5, 138.0, 140.4, 149.0, 153.2, 156.3, 167.8, 168.2, 170.5. HRMS Calcd for  $\text{C}_{34}\text{H}_{39}\text{N}_7\text{O}_8$ : 674.2933 (MH<sup>+</sup>); Found: 674.2931.



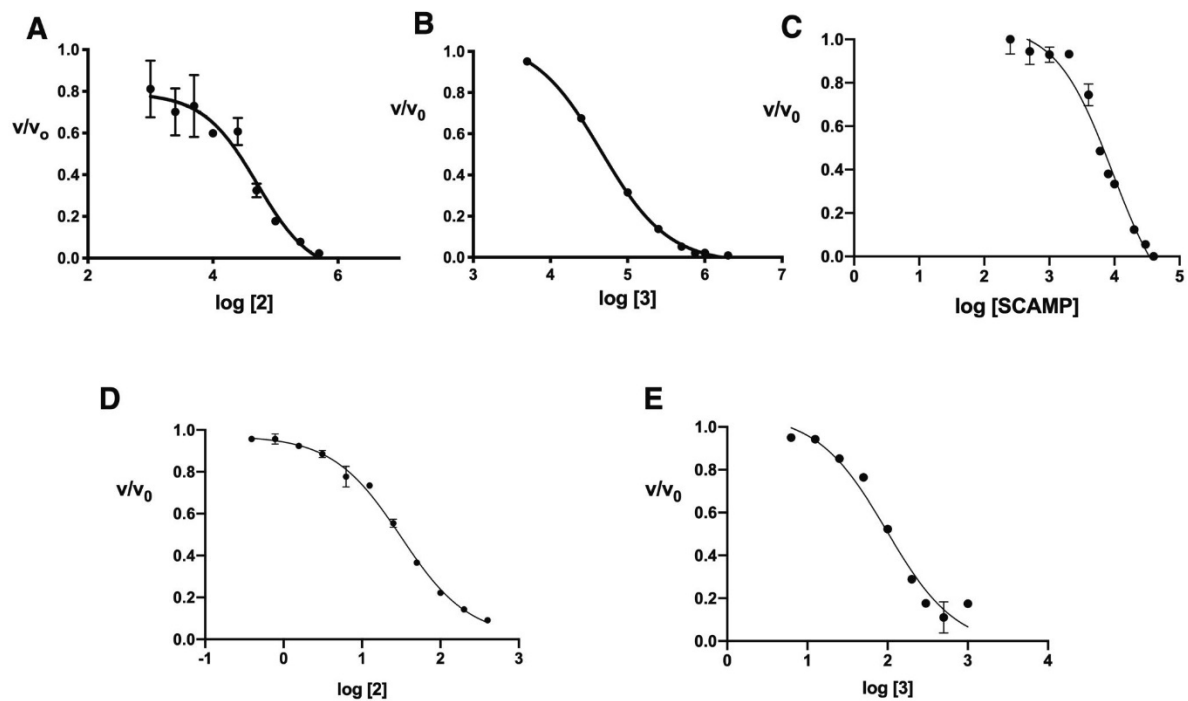
**5'-L-Threonyl-N-malonylamino-5'-deoxyadenosine (5).** To a solution of 238 mg (0.36 mmol) of **5** in 8 mL of absolute ethanol was added 80 mg of 10% Pd/C. The mixture was stirred overnight under a hydrogen atmosphere, then filtered and the solvent removed to give 131 mg 82% of the crude acid. A portion of this material ( mg, mmol) was dissolved in 4 mL of a 1:1 mixture of TFA:H<sub>2</sub>O at 0 °C, then stirred at rt for 5 h. The solvent was removed and redissolved in 4 mL of H<sub>2</sub>O and the solvent removed to give crude **5**. This material was purified by dissolving in 10 mL of H<sub>2</sub>O (pH was adjusted to 7.5 with 1N NaOH) and loading onto a column of DEAE-Sephadex A-25 (5 mL resin) that was equilibrated with 40 mM triethylammonium bicarbonate, pH 7.8. The column was eluted with a 50 mL step gradient of 50 - 200 mM (5 mL fractions). The desired product was found in fractions 7-8 by absorbance at 260 nm and the combined fractions were evaporated to give the triethylammonium salt of **5**. This material was dissolved in 5 mL of H<sub>2</sub>O and the solvent removed twice to remove buffer, then passed through a 1 mL column of Dowex (Na<sup>+</sup> form) to give 33 mg of **5** as the sodium salt. <sup>1</sup>H NMR (D<sub>2</sub>O) δ 1.07 (d, J = 6.4 Hz, 1H), 4.11 (d, J = 3.6 Hz, 1H), 4.19 (m, 1H), 4.72 (m, 1H), 5.93 (d, J = 2.4 Hz, 1H), 8.11 (s, 1H), 8.19 (s, 1H). <sup>13</sup>C NMR (D<sub>2</sub>O) δ 15.6, 18.9, 40.5, 60.1, 67.5, 70.6, 72.8, 82.8, 87.4, 140.0, 148.2, 152.2, 155.0, 169.1. HRMS Calcd for C<sub>17</sub>H<sub>23</sub>N<sub>7</sub>O<sub>8</sub>, 452.1535 (M); Found, 452.1547.



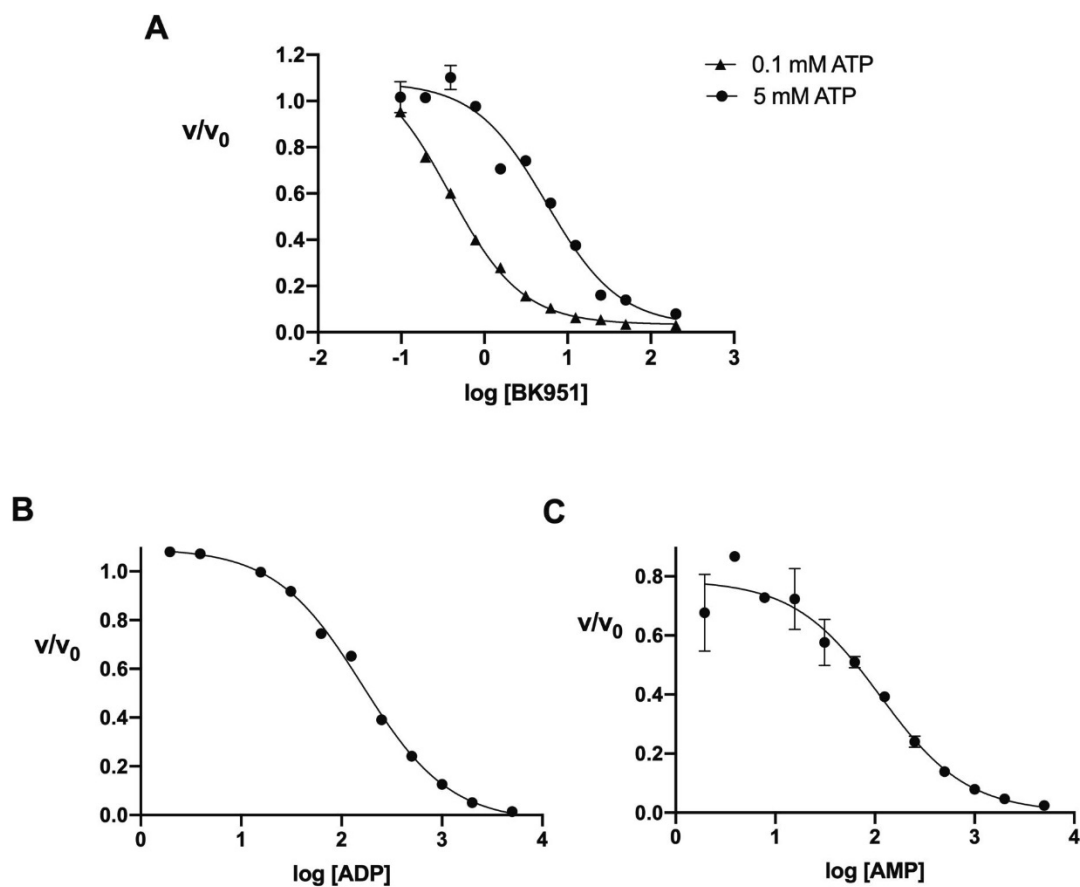
**Supplementary Figure S5.** HPLC traces of purified inhibitors using a Supelco LC-18S column, 5  $\mu$ M, 4.6 x 250 mm. For **1** and **5**, Buffer A = 10 mM Ammonium Bicarbonate; Buffer B = Buffer A with 20% Methanol. Elution used a linear gradient of 0-100% B over 30 min. For **2**, **3**, **SC-AMP**, Buffer = 50 mM potassium phosphate pH 6, 1% acetonitrile in water, isocratic elution. For **4**, Buffer A = 5 mM NH<sub>4</sub>OAc; Buffer B = 40% aq. acetonitrile, linear gradient 0-20% B over 45 min.



**Supplementary Figure S6.** A) Representative data for HPLC assay using the assay as described in Materials and Methods. The amount of  $t^6A$  is measured by integration of peak area and comparison to a standard curve. B) Time course for HPLC assay with full assay mixture.

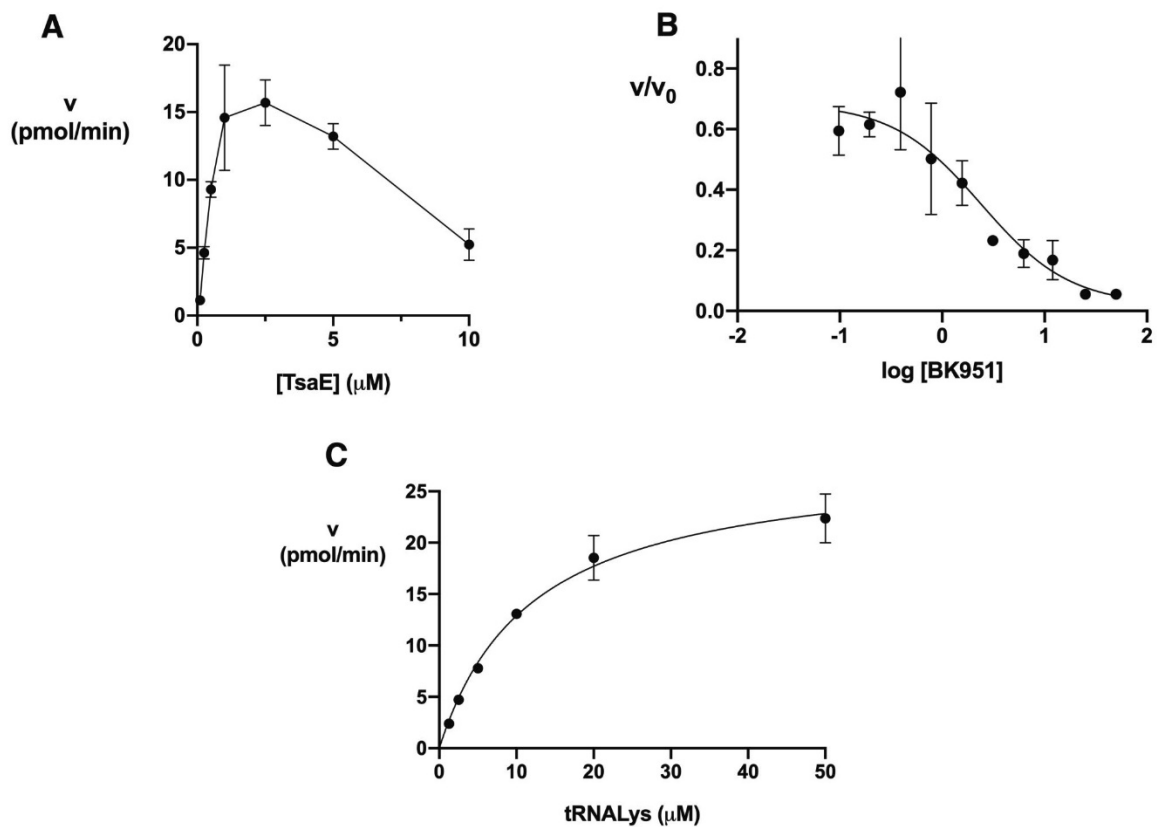


**Supplementary Figure S7. A-C:** IC<sub>50</sub> data for inhibition of overall t<sup>6</sup>A reaction. **(A)** sulfamate inhibitor **2**; **(B)** bis-sulfamoylimide inhibitor **3**; **(C)** serine analog SC-AMP. **D, E:** IC<sub>50</sub> data for inhibition of EctSaC: **(D)** sulfamate **2**; **(E)** aminosulfonylsulfamate **3**.

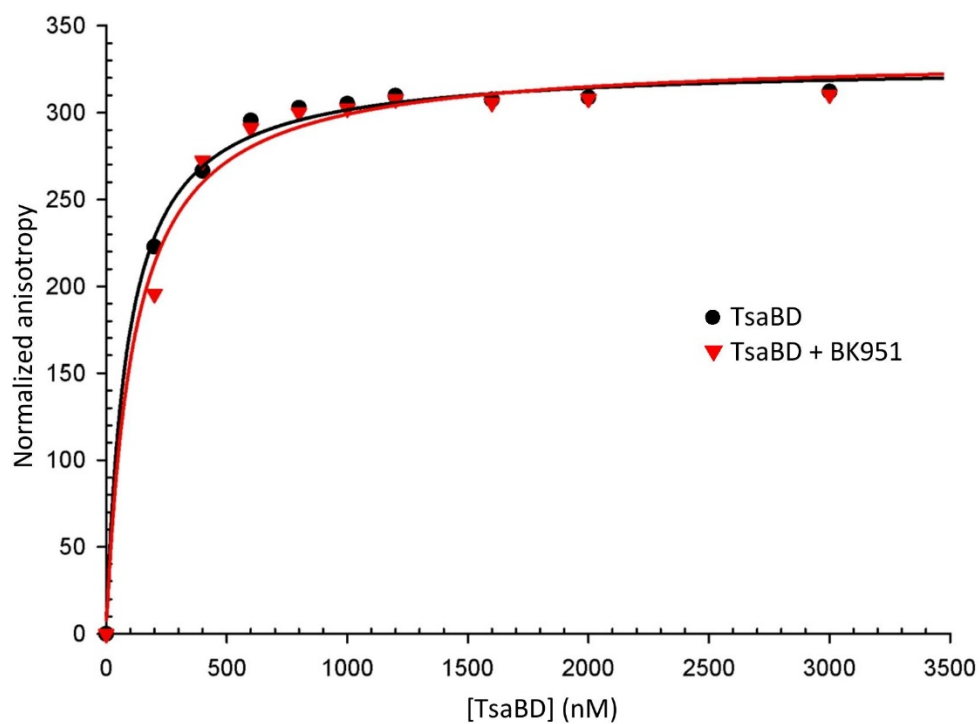


**Supplementary Figure S8.** (A) Inhibition of  $t^6A$  formation by BK951 in the presence of 0.1 mM ATP ( $IC_{50} = 0.4 \mu M$ ) (triangles) or 5 mM ATP ( $IC_{50} = 5.6 \mu M$ ) (circles). (B) Inhibition of  $t^6A$  formation by ADP ( $IC_{50} = 160 \mu M$ ). (C) Inhibition of  $t^6A$  formation by AMP ( $IC_{50} = 113 \mu M$ ).

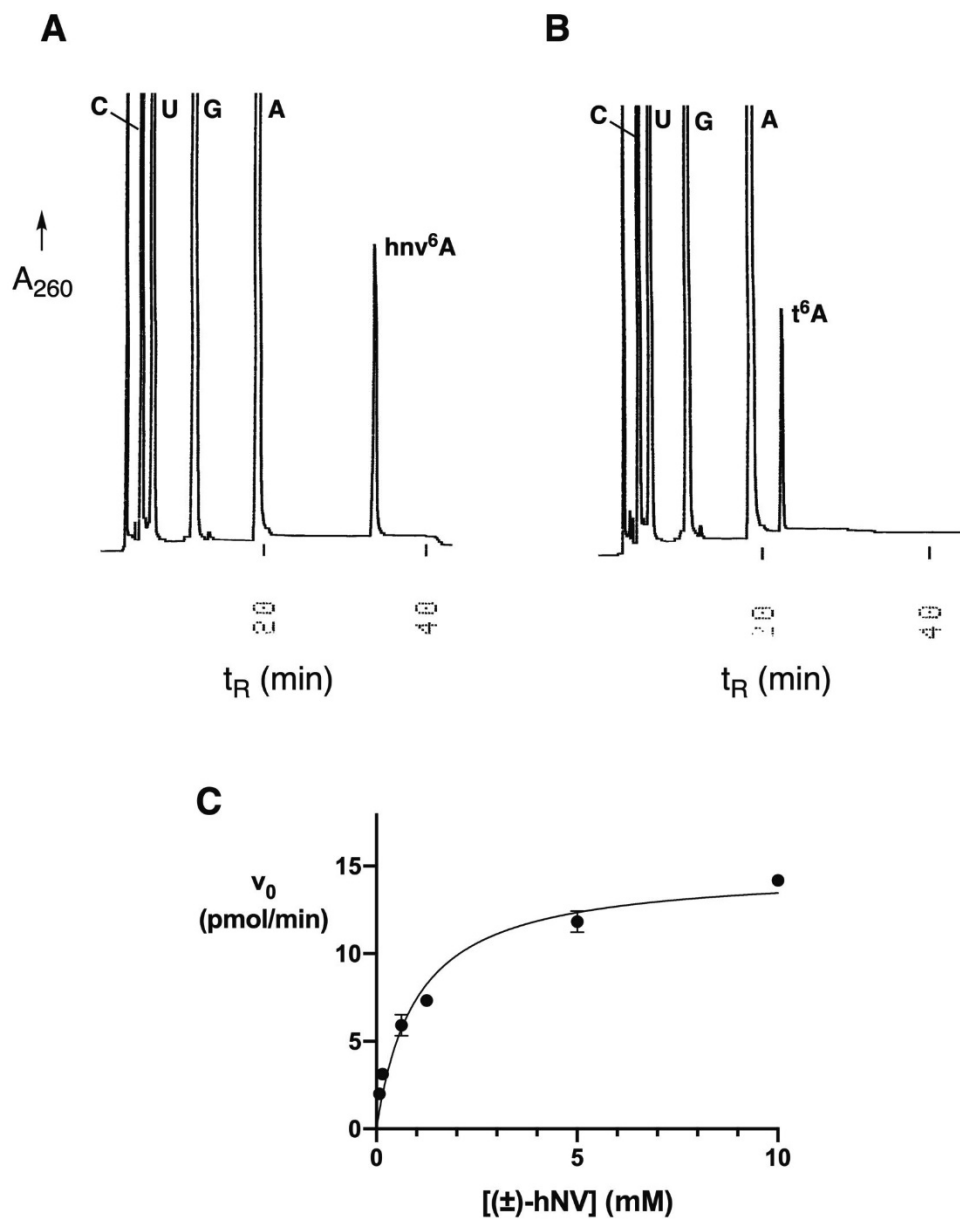




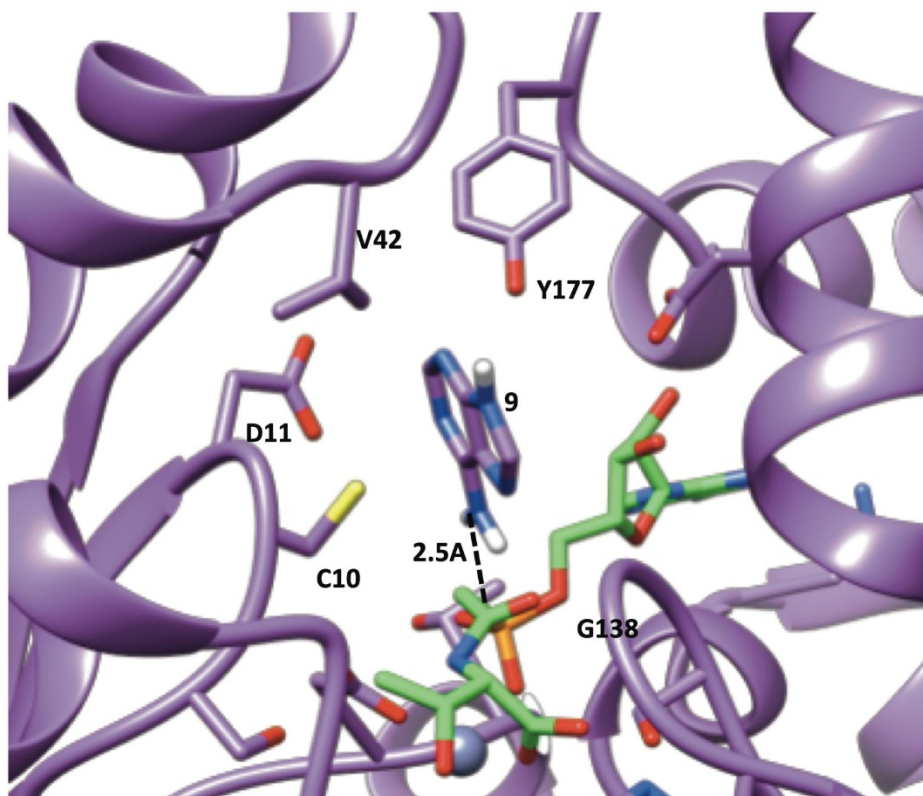
**Supplementary Figure S9.** (A) Effect of varying Tsae concentration in the full t<sup>6</sup>A reaction with TsaeBCD concentrations = 0.5  $\mu\text{M}$ . (B) IC<sub>50</sub> measurement of BK951 in full reaction with in the presence of 10  $\mu\text{M}$  Tsae (IC<sub>50</sub> = 2.5  $\mu\text{M}$ ) (C) Full reaction with variable tRNA<sup>Lys</sup> as substrate in the presence of 10  $\mu\text{M}$  Tsae.



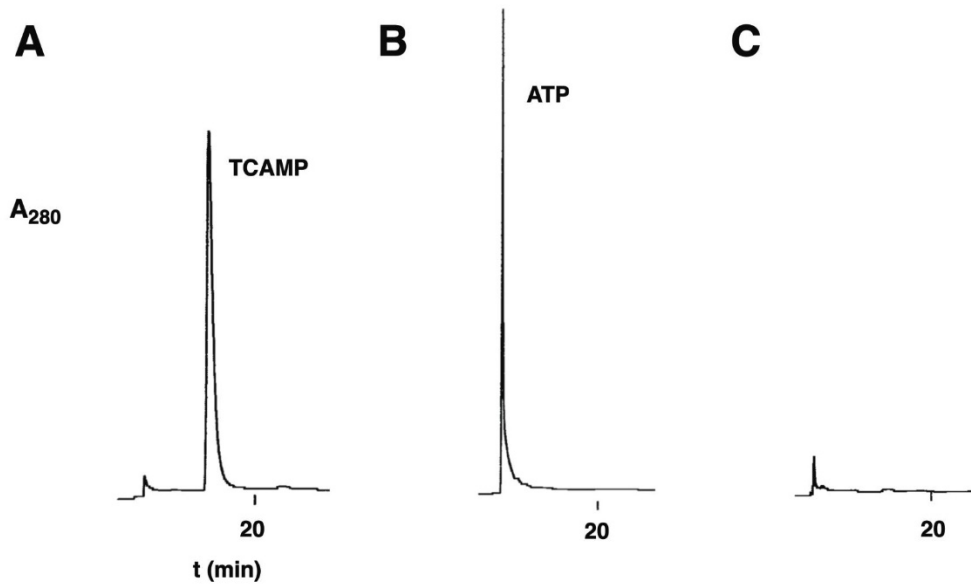
**Supplementary Figure S10.** Binding of fluorescein-3' labeled tRNA<sup>Lys</sup><sub>UUU</sub> to TsaBD followed by fluorescence anisotropy. 40 nM fluorescent tRNA<sup>Lys</sup><sub>UUU</sub> probe has been titrated with increasing concentrations of TsaBD alone (black dots) or TsaBD supplemented with 50 $\mu$ M of BK951 (red triangles) giving respective  $K_D$  values of  $87 \pm 11$  nM and  $112 \pm 13$  nM.



**Supplementary Figure S11.** In vitro activity of (±)-3-hydroxynorvaline ((±)-hNV) with EcTsaBDE and EcTsaC in the full assay. (A) HPLC trace of reaction with (±)-hNV (B) HPLC trace of reaction with L-threonine using the same HPLC system for comparison (C) Kinetics of full reaction with (±)-hNV ( $K_m = 0.5$  mM). The value for  $K_m$  assumes only the L-isomer is active.



**Supplementary Figure S12.** Model of adenine in the active site of TsaD (pink ribbon) bound to BK951 (carbons in green sticks). Some interacting residues of the surrounding loops are represented in sticks. Metal site (Zn) is a grey sphere. The incoming adenine N6 nucleophile and the carbamoyl Carbon are connected by a broken line. The N9 position to which the tRNA is attached is labeled "9".



**Supplementary Figure S13.** Purification and analysis of EcTsaC:TC-AMP complex. **(A)** HPLC analysis of 10  $\mu\text{L}$  aliquot of TC-AMP (100  $\mu\text{M}$ ) bound to TsaC (130  $\mu\text{M}$ ) after purification from *E. coli* BL21(DE3). **(B)** Reaction mixture of TsaC-TCAMP complex, as shown in (C), with 1.5 equiv. (150  $\mu\text{M}$ ) of  $\text{PP}_i$  in the presence of 0.5 mM  $\text{MgCl}_2$  in buffer for 30 min on ice. **(C)** Aliquot of reaction mixture after Sephadex G-50 spin column purification. Sample contained approximately 1 nmol TsaC by protein assay.

## Supplementary Tables

**Table S1.** List of oligonucleotides used for site directed mutagenesis of EcTsaD.

Oligo name	Sequence (5' -> 3')
EcTsaD-S9Afw	GTA CTG GGT ATT GAA ACT GCC TGC GAT GAA ACC GGC ATC
EcTsaD-S9Arev	GAT GCC GGT TTC ATC GCA GGC AGT TTC AAT ACC CAG TAC
EcTsaD-S9Gfw	GTA CTG GGT ATT GAA ACT GGC TGC GAT GAA ACC GGC ATC
EcTsaD-S9Grev	GAT GCC GGT TTC ATC GCA GCC AGT TTC AAT ACC CAG TAC
EcTsaD-C10Afw	CTG GGT ATT GAA ACT TCC GCC GAT GAA ACC GGC ATC GCC
EcTsaD-C10Arev	GGC GAT GCC GGT TTC ATC GGC GGA AGT TTC AAT ACC CAG
EcTsaD-H111Afw	CCG GCG ATC CCT GTA CAC GCG ATG GAA GGG CAT CTG TTA
EcTsaD-H111Arev	TAA CAG ATG CCC TTC CAT CGC GTG TAC AGG GAT CGC CGG
EcTsaD-H115Afw	GTA CAC CAT ATG GAA GGG GCG CTG TTA GCG CCG ATG CTG
EcTsaD-H115Arev	CAG CAT CGG CGC TAA CAG CGC CCC TTC CAT ATG GTG TAC
EcTsaD-H139Afw	CTG CTT GTT TCC GGC GGT GCG ACG CAG TTA ATC AGC GTG
EcTsaD-H139Arev	CAC GCT GAT TAA CTG CGT CGC ACC GCC GGA AAC AAG CAG
EcTsaD-H139Nfw	CTG CTT GTT TCC GGC GGT AAT ACG CAG TTA ATC AGC GTG
EcTsaD-H139Nrev	CAC GCT GAT TAA CTG CGT ATT ACC GCC GGA AAC AAG CAG
EcTsaD-D167Afw	GCC GCC GGG GAA GCG TTT GCG AAA ACC GCG AAG CTG CTG
EcTsaD-D167Arev	CAG CAG CTT CGC GGT TTT CGC AAA CGC TTC CCC GGC GGC
EcTsaD-D300Afw	CGT CCG GAA TTT TGT ACT GCG AAC GGC GCG ATG ATC GCC
EcTsaD-D300Arev	GGC GAT CAT CGC GCC GTT CGC AGT ACA AAA TTC CGG ACG
EcTsaD-M112Afw	GCG ATC CCT GTA CAC CAT GCG GAA GGG CAT CTG TTA GCG
EcTsaD-M112Arev	CGC TAA CAG ATG CCC TTC CGC ATG GTG TAC AGG GAT CGC

### Supplementary References

1. Ikeuchi, H., Meyer, M. E., Ding, Y., Hiratake, J., Richards, N. G. A critical electrostatic interaction mediates inhibitor recognition by human asparagine synthetase. *Bioorg. Med. Chem.* 2009, **17**, 6641– 6650
2. Ferreras, J. A., Stirret, K. L, Lu, X., Ryu, J.-S., Soll, C. E., Tan, D. S., Quadri, L. E. N. (2008) Mycobacterial phenolic glycolipid virulence factor biosynthesis: mechanism and small molecule inhibition of polyketide chain initiation *Chem. Biol.* **15**, 51-61.
3. Tsai, J. H., Takaoka, L. R., Powell, N. A., Nowick, J. S. (2002) Synthesis of amino acid ester isocyanates: Methyl-(S)-2-isocyanato-3-phenylpropanoate *Org. Syn.* **78**, 220.
4. Chheda, G. B., Hong, C. I. (1971) Synthesis of naturally occurring 6-ureidopurines and their nucleosides. *J. Med. Chem.* **14**, 748-753.
5. Fettes, K., J., Howard, N., Hickman, D. T., Adah, S., Player, M. R., Torrence, P. F., Micklefield, J. (2002) Synthesis and nucleic acid binding properties of sulfamide and 3'-N-sulfamate-modified DNA *J. Chem. Soc. Perk. Trans. I* **2002**, 485-495.







## II.2. Représentations schématiques du site actif de TsaBD et du domaine Kae1 de TobZ en présence de différents substrats.

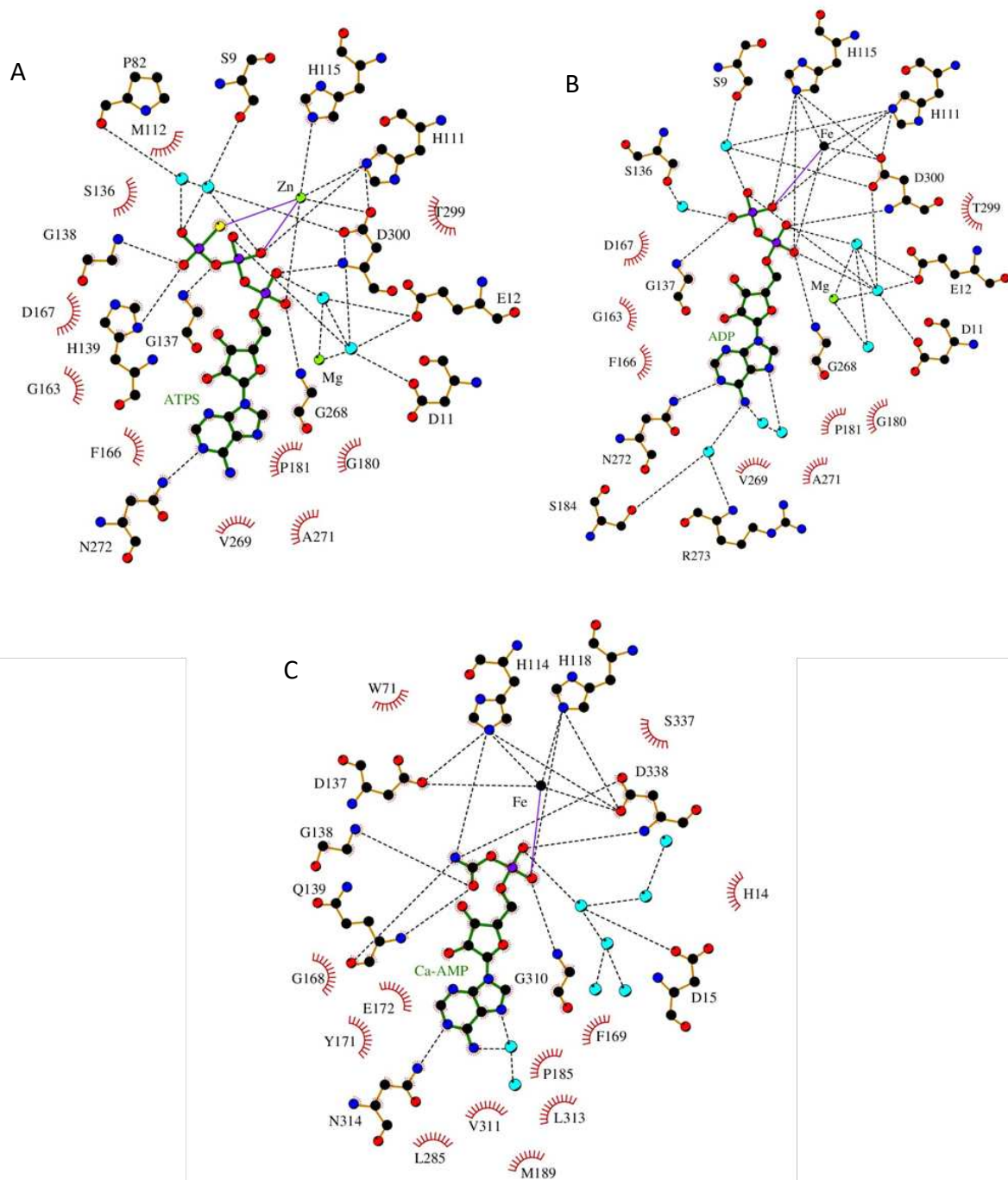


Figure 57 Les sites de fixation des différents substrats dans le site actif de TsaD de *Samolnella typhirunium* (panel A), de *Escherichia coli* (panel B) et de TobZ de *Streptoalloteichus tenebrarius* (panel A). Les nucléotides sont en vert. Les liaisons hydrogènes sont représentées en pointillées, les liaisons de coordination sont représentées en violet et enfin les interactions de van der Waals sont représentées en demi-cercle hachuré rouge. Les molécules d'eau sont représentées par des sphères cyan.

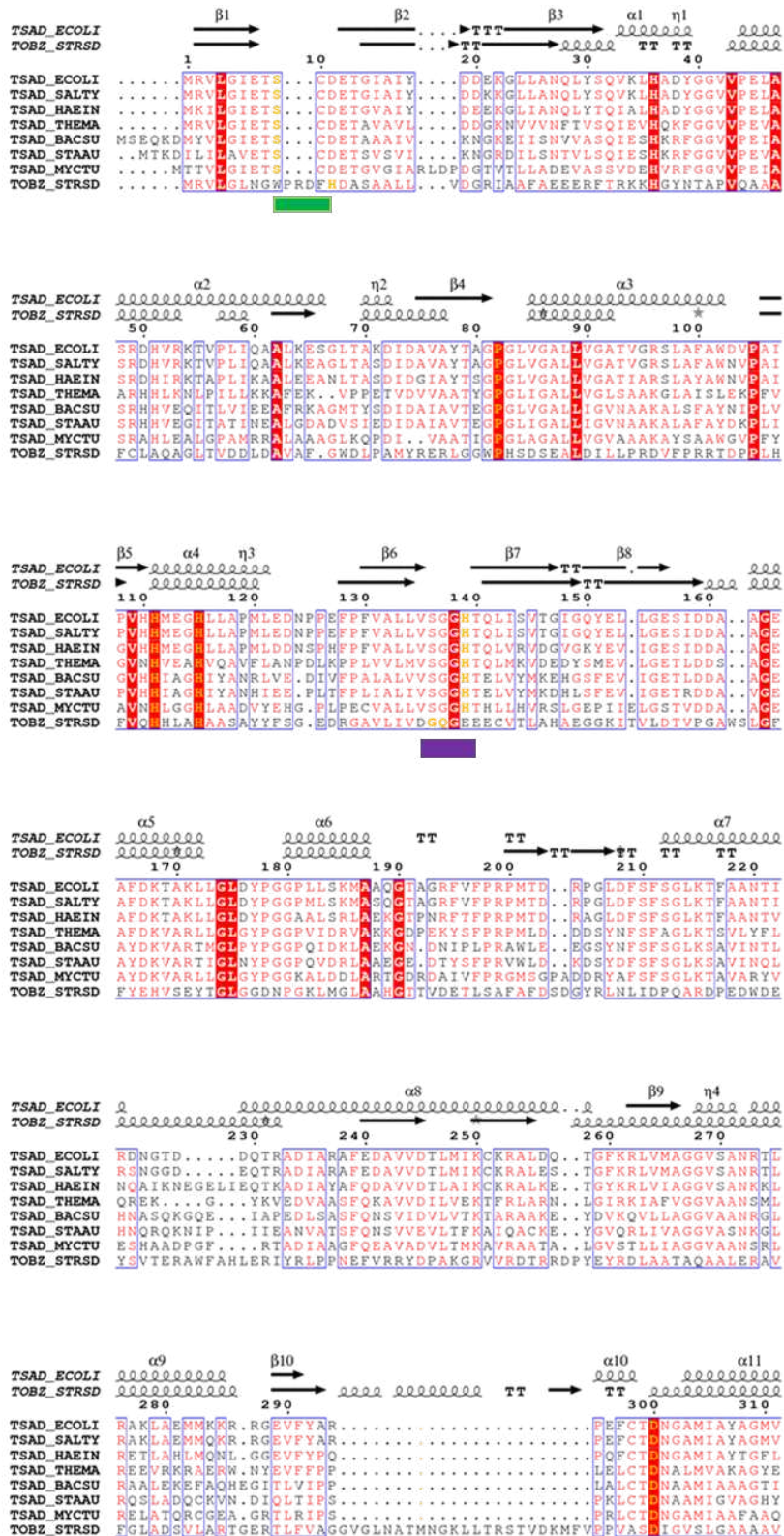


Figure 58 Alignement multiple des séquences des protéines Tsad de *Escherichia coli* (ECOLI), de *Salmonella typhirunium* (SALTY), de *Haemophilus influenzae* (HAEIN), de *Thermotoga maritima* (THEMA), de *Bacillus subtilis* (BACSU), de *Staphylococcus aureus* (STAAU), de *Mycobacterium tuberculosis* (MYCTU) et de la protéine TOBZ de *Streptoalloteichus tenebrarius* (STRSD). Les acides aminés conservés sont surlignés en rouge et ceux similaires sont écrits en rouge. Les acides aminés coloriés en jaunes sont discutés dans le texte.

Les structures cristallographiques de TsaBD des bactéries *Salmonella typhimurium* et *Escherichia coli*, ainsi que celle de TobZ de *Streptoalloteichus tenebrarius* ont été résolues en présence respectivement d'analogue non hydrolysable de l'ATP (ATP $\gamma$ S), d'ADP et de carbamoyle-AMP (Ca-AMP) liées aux sites actifs des protéines. La comparaison des sites actifs met en évidence que les acides aminés de TsaD impliqués dans la liaison des différents ligands et ceux du domaine Kae1 de TobZ impliqués dans la fixation du Ca-AMP sont très similaires (figure 57) et de plus, ces acides aminés s'alignent à des positions équivalentes sur l'alignement multiple de séquence illustré à la figure 58.

Les ions Zn<sup>2+</sup> et Fe<sup>2+</sup> sont coordonnés par les histidines (H<sub>111</sub> - H<sub>114</sub> et H<sub>115</sub> - H<sub>118</sub>) et les aspartates (D<sub>300</sub> - D<sub>338</sub>) conservés dans StTsaD et EcTsaD (figures 57 A et B). De façon similaire à la fixation de la molécule BK951, l'histidine 139 de TsaD partage une liaison hydrogène avec le phosphate  $\gamma$  de l'ATP $\gamma$ S.

Également, le peptide S<sub>136</sub> - G<sub>137</sub> - G<sub>138</sub>, conservé au sein de la famille TsaD/Kae1/Qri7, stabilise l'oxygène du carbamoyle de BK951, du phosphate  $\gamma$  de l'ATP $\gamma$ S et du phosphate  $\beta$  de l'ADP (ce peptide est localisé par une barre violette sur la figure 58). Chez TobZ, le peptide G<sub>138</sub> - Q<sub>139</sub> - G<sub>140</sub> (retrouvé à une position équivalente dans SmTsaD et EcTsaD) stabilise le groupement carbamoyle du Ca-AMP (figure 57 C).

En présence d'ATP $\gamma$ S et d'ADP les acides aminés D<sub>11</sub> et E<sub>12</sub>, conservés chez les procaryotes, coordonnent des molécules d'eau et un ion Mg<sup>2+</sup>, qui stabilisent l'oxygène du phosphate  $\alpha$  (figure 57 A et B). Cette liaison est très similaire à celle réalisée par une molécule d'eau, partageant une liaison hydrogène avec l'aspartate 15 de TobZ, sur l'oxygène du phosphate  $\alpha$  (figure 57 C) (ces acides aminés sont localisés par une barre verte sur la figure 58). Dans TobZ, les acides aminés H<sub>14</sub> - D<sub>15</sub>, localisés proche du groupement carbamoyle du Ca-AMP et de la tobramycine (accepteuse du carbamoyle), sont supposés participer à déprotonation du groupement 6' hydroxyle de la molécule tobramycine pour favoriser l'attaque nucléophile sur le Ca-AMP aboutissant au transfert du groupement carbamoyle sur la molécule de tobramycine formant le produit nebramycine (Parthier *et al.*, 2012).

La dyade H<sub>14</sub> - D<sub>15</sub> de TobZ et D<sub>11</sub> - E<sub>12</sub> de TsaD se superposent, suggérant une fonction commune. En revanche, nous n'avons pas pu observer de densité électronique dans le site actif de TsaD en présence de la molécule BK951 qui aurait pu correspondre à une molécule d'eau ou d'ion Mg<sup>2+</sup> en face du phosphate  $\alpha$ .

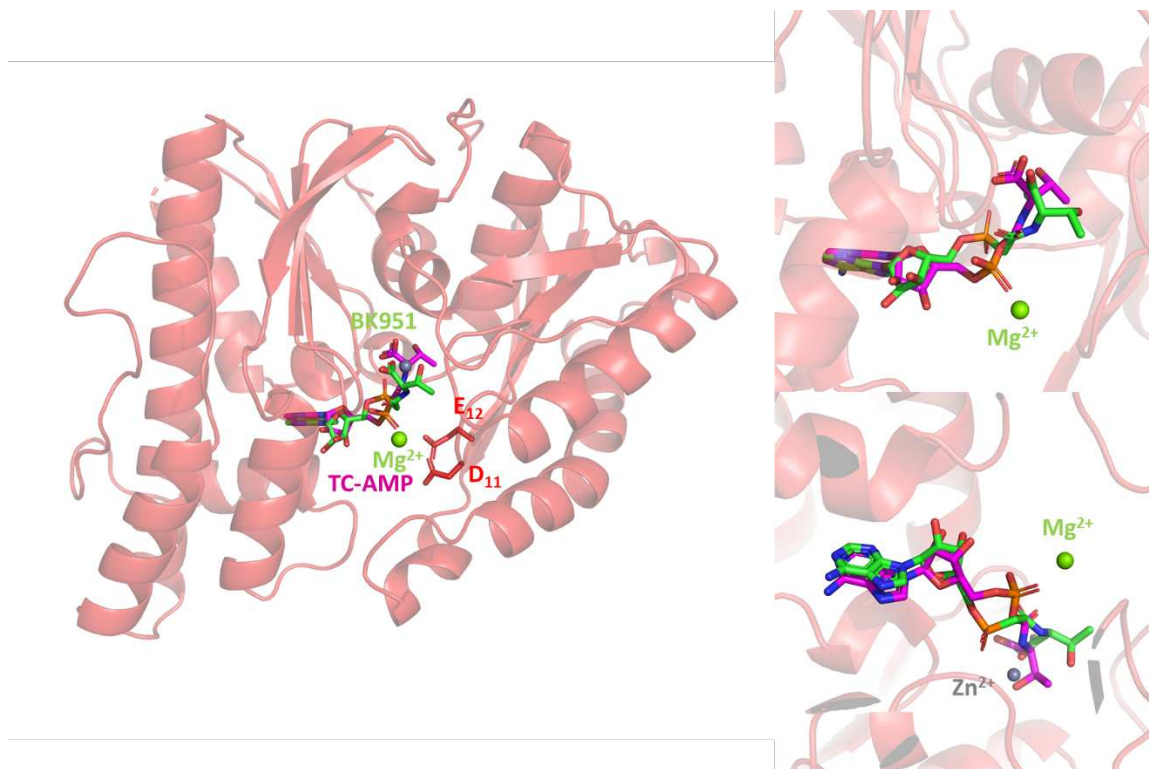


Figure 59 Superposition de l'intermédiaire TC-AMP (en magenta) (résolue dans la structure de Sua5 de *Sulfolobus tokodaii* code PDB 411B) sur l'analogue BK951 (en vert), dans TsaD de la bactérie *Escherichia coli*. Le métal  $Zn^{2+}$ , sphère grise, coordonne le groupement thréonyle de la molécule BK951. Le métal  $Mg^{2+}$ , sphère verte, coordonne l'oxygène du phosphate  $\alpha$  du TC-AMP dans Sua5. La superposition de l'intermédiaire TC-AMP sur la molécule BK951 met en évidence une rotation du phosphate  $\alpha$ .

La superposition de l'intermédiaire TC-AMP de la structure de Sua5 de *Sulfolobus tokodaii* (code PDB 4E1B) sur la molécule BK951 dans TsaD de *Escherichia coli* (figure 59) met en évidence une rotation du groupement phosphate  $\alpha$  de la molécule BK951 causée, probablement, par la substitution de l'oxygène en carbone dans la molécule BK951. Si le phosphate  $\alpha$  était orienté comme dans la molécule de TC-AMP, il pourrait être coordonné par des molécules d'eau et par un ion  $Mg^{2+}$ , eux même stabilisés par les résidus D<sub>11</sub> et E<sub>12</sub> et orienterait le ligand dans une conformation favorable à l'attaque nucléophile par le N<sup>6</sup> de l'A<sub>37</sub> de l'ARNt.

### II.3. Conclusion de l'article.

L'étude menée en collaboration avec l'équipe de Charles Lauhon a permis de caractériser pour la première fois les acides aminés impliqués dans la fixation d'un analogue stable de l'intermédiaire TC-AMP (BK951) au site actif de TsaD. La mutagenèse dirigée a permis d'identifier les acides aminés qui participent à la fixation de la molécule BK951 et ceux qui participent probablement à la catalyse de la modification t<sup>6</sup>A. La comparaison avec le site catalytique de l'enzyme TobZ, permet d'émettre des hypothèses quant au mécanisme catalytique de la modification t<sup>6</sup>A.

La boucle S<sub>9</sub>CDE<sub>12</sub> est très conservée chez les procaryotes, ce qui suggère qu'elle joue un rôle important dans la réaction de transfert t<sup>6</sup>A, éventuellement dans la déprotonation de l'adénosine 37 de l'ARNt substrat. Zhang et al ., ont mis en évidence que la substitution du glutamate 12 en alanine ne possédait plus d'activité t<sup>6</sup>A et ne fixait plus d'ion Mg<sup>2+</sup> (code PDB 4WQA) sans toutefois altérer la fixation du nucléotide ATP, suggérant un rôle catalytique pour l'ion Mg<sup>2+</sup> (Zhang *et al.*, 2015). Cependant, les acides aminés D<sub>11</sub> et E<sub>12</sub> de TsaD sont substitués par des acides aminés asparagine et lysine chez les eucaryotes et des acides aminés histidine et thréonine chez les archées. Le rôle de cette boucle reste donc à démontrer.

Également, l'ordre de fixation des différents substrats n'est pas connu. Comme vu dans le chapitre précédent, la présence de TsaE n'est pas compatible avec la fixation de l'ARNt substrat, le complexe TsaBDE pourrait être un état pré-actif compatible avec la fixation de l'intermédiaire TC-AMP, comme suggéré par la fixation de l'ATPγS dans le site actif de TsaD en conformation ouverte (*cf. chapitre I*). Une étude du complexe ternaire TsaBDE avec l'analogue BK951 pourrait être envisagée au sein du laboratoire.



## Chapitre III

### III. Effets de YrdC et de GON7 sur la modification t<sup>6</sup>A chez l'Homme.

#### III.1. Présentation de l'article « Defects in t<sup>6</sup>A tRNA modification due to GON7 and YRDC mutations leads to Galloway-Mowat ».

Les défauts de modifications des ARNt sont impliqués dans différentes pathologies, et ce champ d'investigations est en pleine expansion. Ces pathologies sont très souvent liées à des désordres neuro-développementaux chez l'Homme (Kapur et Ackerman, 2018; Ramos et Fu, 2019). Parmi ces pathologies, le syndrome de Galloway-Mowat (GAMOS) est une maladie neuro-rénale rare qui se caractérise par des syndromes néphrotiques stéroïdes résistants (SNRS), une microcéphalie, ainsi que des défauts neurologiques (Galloway et Mowat, 1968). Le GAMOS est donc cliniquement hétérogène et présente des origines génétiques diverses. En effet, de nombreuses mutations GAMOS ont été identifiées sur les gènes *WDR73* et *WDR4* (cf. § 4.10 page 80), mais également sur les gènes codant 4 protéines du complexe KEOPS humain (*OSGEP*, *TP53RK*, *TPRKB*, *LAGE3*) (Colin *et al.*, 2014; Edvardson *et al.*, 2017; Braun *et al.*, 2017, 2018; Wang *et al.*, 2018).

Il a été vu dans l'introduction que la première étape de la modification t<sup>6</sup>A chez les eucaryotes est assurée par la famille YrdC/Sua5 puis la seconde étape est réalisée par le complexe KEOPS (El Yacoubi *et al.*, 2009, 2011).

Récemment, l'orthologue de Gon7 de la levure ( $\gamma$ Gon7) a été identifié chez les métazoaires. Nommé dans un premier temps C14orf142, il a été renommé GON7.  $\gamma$ Gon7 est identifiée comme une IDP (intrinsically destructured protein) mais 50% de la protéine se structure lorsqu'elle est en interaction avec son partenaire Pcc1 (Zhang *et al.*, 2015). GON7 de l'Homme interagit avec LAGE3, formant avec OSGEP, TP53RK et TPRKB un pentamère (Wan *et al.*, 2017). Malgré le faible taux d'identité de séquence entre  $\gamma$ Gon7 et GON7 (17 % d'identité de séquence et 34 % de similitude de séquence), ces deux protéines partagent le même partenaire Pcc1/LAGE3 (Wan *et al.*, 2017; Drew *et al.*, 2017).

Dans cette étude, réalisée en collaboration avec l'équipe de Corine Antignac et Géraldine Mollet de l'institut IMAGINE, de nouvelles mutations sur les gènes codant YrdC et GON7 ont été identifiées, et leur impact sur la biosynthèse de la modification t<sup>6</sup>A conduisant au GAMOS a été étudié. De plus, la structure cristallographique du sous-complexe OSGEP-



LAGE3-GON7 à 1,9 Å de résolution est présentée (code PDB 6GWJ). La caractérisation structurale montre que GON7 acquiert sa structure tridimensionnelle en interaction avec LAGE3.

Le séquençage de l'exome entier de 11 patients présentant les caractéristiques du syndrome GAMOS, provenant de 5 familles différentes, a permis d'identifier des mutations sur les gènes *YRDC* et *GON7*. Le séquençage ainsi que les analyses génétiques et cellulaires ont été réalisés par l'équipe de Corine Antignac à l'institut IMAGINE.

Deux mutations hétérozygotes sur le gène *YRDC* au sein d'une famille ont été détectées. L'une est une mutation faux-sens ( $C_{251}>T$ ) qui conduit à la traduction de l'acide aminé valine à la position 84 à la place d'une alanine ( $A_{84}V$ ). La seconde est une délétion de 4 paires de base ( $_{721-724}del$ ) conduisant à un décalage de lecture et à la traduction d'une valine à la position 241 à la place d'une isoleucine ( $V_{241}I$ ). Enfin une troisième famille présente une mutation homozygote décalante de 3 paires de base ( $_{794-796}del$ ) conduisant à la délétion de l'acide aminé leucine à la position 265 ( $L_{265}del$ ). Afin de savoir où ces diverses mutations sur YrdC se localisent, un modèle *in silico* tridimensionnel a été réalisé à partir de la structure de son homologue Sua5 de *Sulfolobus tokadaii* (code PDB 4EIB, 20% d'identité de séquence entre les 2 protéines). L'intermédiaire TC-AMP et l'ion  $Mg^{2+}$  ont été placés par superposition des sites actifs des deux protéines. La mutation  $A_{84}V$  est localisée proche de l'ion  $Mg^{2+}$  qui coordonne le phosphate  $\alpha$  du TC-AMP, alors que les mutations  $V_{241}I$  et  $L_{265}del$  sont localisées à l'extrémité C-terminale de YrdC ([figure 1e de l'article, page 183](#)). Ces mutations n'altèrent ni l'expression de la protéine YrdC dans les cellules de podocytes, ni la stabilité de sa structure tridimensionnelle analysée par des spectres 1D  $H^1$ -HSQC avec toutefois une dégradation du mutant  $V_{241}I$  ([figure 3b de l'article, page 185](#)). Des études de complémentation chez la levure *Saccharomyces cerevisiae*, chez qui le gène *SUA5* a été invalidé et remplacé par les gènes sauvage et mutants *YRDC*, ont été réalisées. La croissance optimale des levures (fitness) et le taux de nucléotides modifiés  $t^6A$  sur les ARNt substrats sont affectés chez les mutants  $A_{84}V$  et  $L_{265}del$ , d'un facteur 1,3 et 1,75 respectivement. Les effets sur le fitness des levures sont plus importants pour le mutant  $V_{241}I$  et aucun nucléotide modifié n'a été détecté sur les ARNt substrats, confirmant la corrélation entre la synthèse de la modification  $t^6A$  et la croissance des levures ([figure 3 de l'article, page 185](#)). Dans les cellules humaines, tous les mutants de YrdC diminuent significativement le taux de nucléotides  $t^6A$  modifiés sur les ARNt substrats, ce qui a pour effet

d'altérer la prolifération et la synthèse des protéines dans les podocytes et d'augmenter l'apoptose de ces cellules.

Concernant GON7, 4 familles présentent une mutation homozygote non-sens (C<sub>21</sub>>A) qui conduit à l'introduction d'un codon stop à la place de la tyrosine 7. Une cinquième famille présente une mutation spontanée différente qui se caractérise par une duplication de nucléotide (C<sub>19</sub>dup) impliquant la tyrosine 7 et conduisant à un décalage de lecture du codon et un codon stop prématuré. Aucune complémentation dans la levure, chez qui le gène *GON7* a été invalidé, par la protéine humaine GON7 n'a pu être observée. Cette incapacité de GON7 à compléter *yGon7* chez la levure suggère une spécificité d'interaction de GON7 avec LAGE3 chez l'Homme et de *Gon7* avec *Pcc1* chez la levure. L'analyse de ces interfaces sera développée en partie supplémentaire.

Les deux mutations conduisent à l'absence de la protéine GON7 dans les cellules. Les mutants GON7 altèrent l'expression des protéines de 66 % dans les cellules ([figure 4c de l'article page 186](#)). Cependant, seule une légère diminution du taux de nucléotides modifiés ([figure 3d de l'article page 185](#)) et un faible impact sur le taux de prolifération et le taux d'apoptose dans les podocytes sont détectés chez ces mutants ([figure 4a-b de l'article page 186](#)). Pour expliquer le GAMOS des patients possédant les mutations de GON7, la stabilité de LAGE3, son partenaire direct a été étudiée dans des cellules HEK293T. En absence de GON7, LAGE3 est dégradée de 50 % en 6 heures ([figure 6b de l'article page 189](#)). En revanche, la présence de GON7 empêche la dégradation de la protéine LAGE3. Ce résultat montre que GON7 stabilise LAGE3, son partenaire direct dans le complexe KEOPS.

L'analyse par RMN 2D <sup>1</sup>H-<sup>15</sup>N SOFAST-HMQC de GON7 a révélé que cette protéine ne possède pas de structure secondaire en solution. Après addition de LAGE3 en solution, environ 35 pics du signal H<sup>1</sup>-N<sup>15</sup> se décalent ou s'éteignent, confirmant que GON7 interagit avec LAGE3 et acquiert ses structures secondaires. Lors de son interaction avec LAGE3, les signaux HSQC de 35 résidus restent localisés entre 8 et 8,5 ppm, suggérant qu'une partie de GON7 est toujours désordonnée malgré son interaction avec LAGE3 (*c.f.* [figure supplémentaire 6a, page 201](#)).

Les propriétés structurales en solution de GON7 ont été également étudiées par diffusion des rayons X aux petits angles couplée à une chromatographie d'exclusion stérique

(SEC-SAXS), ce qui a permis de confirmer l'absence de structure compacte/globulaire de la protéine lorsqu'elle est seule en solution. La partie linéaire de la courbe de Guinier ( $\ln I(q)$  versus  $q^2$ ) indique que GON7 est monodisperse en solution (c.f. figure supplémentaire 6d, page 201). De plus, la représentation selon Kratky confirme que GON7 est désordonnée mais le plateau d'intensité situé au voisinage de  $(qR_g)^2 I(q)/I(0)=1,5$  suggère la présence d'éléments de structure secondaires locales et/ou transitoires (c.f. figure 5a page 187). La fonction de Guinier et la distribution des distances Pr indiquent des rayons de giration  $R_g$  proches de 31,8 et 32,4 Å respectivement. La masse moléculaire dérivée de  $I_0/c$  de 14,7 kDa est compatible au poids moléculaire théorique de GON7 (12,59 kDa) d'une forme monomérique. La distance maximale mesurée s'étend jusqu'à 125 Å (c.f. figure supplémentaire 6b). L'ensemble de ces données confirme que l'extension maximale de GON7 est la conséquence d'un état déstructuré et non d'une oligomérisation.

	Rg Guinier (Å)	Rg Pr (Å)	$D_{max}$ Pr (Å)	Mw théorique (kDa)	Mw expérimentale (kDa)	Oligomérisation
GON7	31.8 ±0,4	32,4 ±0,2	125	12,59	14,7	1
LAGE3-GON7	35.6 ±0,5	38,1 ±0,8	155	29,05	56,5	2:2
OSGEP-LAGE3-GON7	30,6 ±0,1	31,4 ±0,1	115	62,88	58,5	1:1:1

Figure 60 Tableau récapitulatif des paramètres structuraux à basse résolution des objets étudiés par SEC-SAXS.

D'autre part, les sous-complexes LAGE3-GON7 (LG) et OSGEP-LAGE3-GON7 (OLG) ont également été étudiés par SEC-SAXS. Les représentations de Guinier ( $\ln I(q)$  versus  $q^2$ ) sont linéaires démontrant que les complexes LG et OLG sont monodisperses (cf. figure supplémentaire 6d, page 201). Leurs rayons de giration suggèrent que les complexes OLG et LG sont plus compacts que GON7 seule (figure 60). Les poids moléculaires mesurés à partir de  $I_0/c$  sont compatibles avec les stœchiométries 1 :1 :1 pour OLG et 2 :2 pour LG (figure 60). De plus les extensions maximales calculées à partir de la fonction de Patterson suggèrent une forme linéaire allongée des complexes et ainsi un enchainement linéaire des sous-unités (c.f. figure supplémentaire 6b, page 201). En présence de LAGE3 ou de LAGE3 et OSGEP, le Kratky-plot sans dimension présente des courbes en « cloche » avec un maximum situé en  $qR_g \sim 2$  et  $(qR_g)^2 I(q)/I(0)=1,1$ , confirmant que GON7 se structure au contact de LAGE3 avec toutefois des éléments flexibles dans les deux complexes (c.f. figure 5a de l'article, page 187). De plus, l'allure

des courbes, en présence ou en absence de OSGEP, sont similaires, ce qui montre que la sous-unité OSGEP ne joue pas de rôle dans l'acquisition du repliement tridimensionnel de GON7.

La structure cristallographique du sous-complexe OLG à 1,9 Å de résolution confirme toutes ces observations en solution. Les sous-unités forment un complexe linéaire. Dans ce complexe, 45 % de GON7 adopte une structure bien définie. Cependant aucune densité électronique n'a permis de modéliser les acides aminés de 51 à 106 de GON7, suggérant qu'ils restent flexibles. La structure de GON7 est composée de 2 brins  $\beta$  (des acides aminés 1 à 20), dont le brin  $\beta_1$  interagit avec le brin  $\beta_1$  de LAGE3, et en une hélice  $\alpha_1$  qui interagit avec l'hélice  $\alpha_2$  de LAGE3 ([figure supplémentaire 14d de l'article, page 209](#)). Les acides aminés 1 à 59 de LAGE3, prédits comme désordonnés, ne sont pas observés dans la structure de LAGE3. Le domaine N-terminal de LAGE3 et le domaine C-terminal de GON7 (les régions prédites flexibles du sous complexe OLG), ont été modélisés et leurs positions optimales ont été trouvées en ajustant la courbe calculée des modèles possibles avec la courbe expérimentale de SAXS. La courbe calculée du modèle final s'ajuste à la courbe expérimentale ( $\chi^2 = 0.33$ ) confirmant que ces régions sont flexibles en solution ([figure 5d de l'article, page 187](#)).

Le mode d'interaction de LG est identique à celui de Pcc1-Gon7 de la levure et les deux hétérodimères se superposent avec une déviation faible (RMSD de 1,4 Å ; [figure 5d de l'article, page 187](#)). Malgré une faible identité de séquence, la résolution de la structure tridimensionnelle de GON7 en complexe avec LAGE3 et OSGEP révèle donc une structure très similaire à celle qu'acquière Gon7 en complexe avec Pcc1 chez la levure. Cette forte similitude structurale confirme que ces deux protéines, majoritairement déstructurées seules en solution, des homologues structuraux en interaction avec leurs partenaires LAGE3/Pcc1, comme initialement prédit par l'équipe de Sicheri (Wan *et al.*, 2017). En revanche, la différence des séquences suggère une grande distance évolutive de ces gènes.



ARTICLE

<https://doi.org/10.1038/s41467-019-11951-x>

OPEN

# Defects in t<sup>6</sup>A tRNA modification due to *GON7* and *YRDC* mutations lead to Galloway-Mowat syndrome

Christelle Arrondel et al.<sup>#</sup>

N<sup>6</sup>-threonyl-carbamoylation of adenosine 37 of ANN-type tRNAs (t<sup>6</sup>A) is a universal modification essential for translational accuracy and efficiency. The t<sup>6</sup>A pathway uses two sequentially acting enzymes, YRDC and OSGEP, the latter being a subunit of the multiprotein KEOPS complex. We recently identified mutations in genes encoding four out of the five KEOPS subunits in children with Galloway-Mowat syndrome (GAMOS), a clinically heterogeneous autosomal recessive disease characterized by early-onset steroid-resistant nephrotic syndrome and microcephaly. Here we show that mutations in *YRDC* cause an extremely severe form of GAMOS whereas mutations in *GON7*, encoding the fifth KEOPS subunit, lead to a milder form of the disease. The crystal structure of the *GON7*/LAGE3/OSGEP subcomplex shows that the intrinsically disordered *GON7* protein becomes partially structured upon binding to LAGE3. The structure and cellular characterization of *GON7* suggest its involvement in the cellular stability and quaternary arrangement of the KEOPS complex.

---

Correspondence and requests for materials should be addressed to H.T. (email: [herman.van-tilbeurgh@i2bc.paris-saclay.fr](mailto:herman.van-tilbeurgh@i2bc.paris-saclay.fr)) or to G.M. (email: [geraldine.mollet@inserm.fr](mailto:geraldine.mollet@inserm.fr)). <sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

transfer RNAs (tRNA) are subject to multiple post-transcriptional modifications that are important for the stabilization of their ternary structure and the precision of the decoding process<sup>1</sup>. The majority of the complex modifications are concentrated in the anticodon region of the tRNAs and are crucial for accuracy of protein synthesis. The threonylcarbamoylation of the N<sup>6</sup> nitrogen of the adenosine at position 37 (t<sup>6</sup>A) of most ANN-accepting tRNAs represents one of the very few nucleotide modifications that exists in every domain of life<sup>2,3</sup>. The t<sup>6</sup>A biosynthesis pathway consists of two steps: firstly, the YRDC enzyme (Sua5 in yeast) synthesizes an unstable threonylcarbamoyl-AMP intermediate (TC-AMP) and in a second step, the KEOPS protein complex transfers the TC-moiety from TC-AMP onto the tRNA substrate<sup>4</sup>. Enzymes that synthesize TC-AMP exist in two versions depending on the organism: a short form which only has the YrdC domain (such as human YRDC) and a long form which has an extra Sua5 domain (yeast SUA5 for instance). The eukaryotic KEOPS complex contains five subunits GON7/LAGE3/OSGEP/TP53RK/TPRKB that are arranged linearly in that order<sup>5–8</sup>. OSGEP is the catalytic subunit that carries out the TC-transfer reaction and its orthologs are present in virtually all sequenced genomes<sup>9</sup>. The other subunits are essential for the t<sup>6</sup>A modification of tRNA, but their precise roles are as yet unknown, especially that of GON7, an intrinsically disordered protein (IDP), which was only recently identified in humans<sup>5,7</sup>. Fungal Gon7 was shown to be an IDP that adopts a well-defined structure covering 50% of its sequence upon complex formation with Pcc1 (LAGE3 homolog)<sup>8</sup>. In humans, GON7 was recently shown to be also structurally disordered in absence of the other KEOPS complex subunits. GON7 was proposed to be a very remote homolog of the yeast Gon7 protein although its structure upon complex formation remains unknown<sup>5,7</sup>.

tRNA modifications have been demonstrated to play a role in the development of the brain and nervous system, and an increasing number of defects in these modifications are now being linked to various human neurodevelopmental disorders<sup>10</sup>. We recently identified autosomal recessive mutations in genes encoding four of the five subunits of human KEOPS complex, namely *LAGE3*, *OSGEP*, *TP53RK*, and *TPRKB* in patients with Galloway-Mowat syndrome (GAMOS, OMIM#251300). GAMOS is a rare neuro-renal disorder characterized by the co-occurrence of steroid-resistant nephrotic syndrome (SRNS) with microcephaly and neurological impairment<sup>11</sup>. GAMOS is clinically heterogeneous, reflecting a genetic heterogeneity. Indeed, disease-causing mutations have been identified in eight genes to date: four in KEOPS genes and four in other unrelated genes, *WDR73*, *WDR4*, *NUP133*, and *NUP107* (refs. 12–17). SRNS is typically detected in the first months of life and most often rapidly progresses to end-stage renal disease (ESRD) within a few months; however, there are rarer cases with preserved renal function in adulthood<sup>18</sup>. Kidney lesions range from minimal change disease, to focal segmental glomerulosclerosis (FSGS) that might be of the collapsing type, or diffuse mesangial sclerosis (DMS). Cerebral imaging findings include cerebral and cerebellar atrophy, and gyration and/or myelination defects. These anomalies are associated with neurological deficits such as psychomotor impairment, hypotonia, seizures, and more rarely sensorineural blindness and deafness. Affected children may also present with facial and/or skeletal dysmorphic features. The prognosis of GAMOS is poor, and most affected children die before 6 years of age.

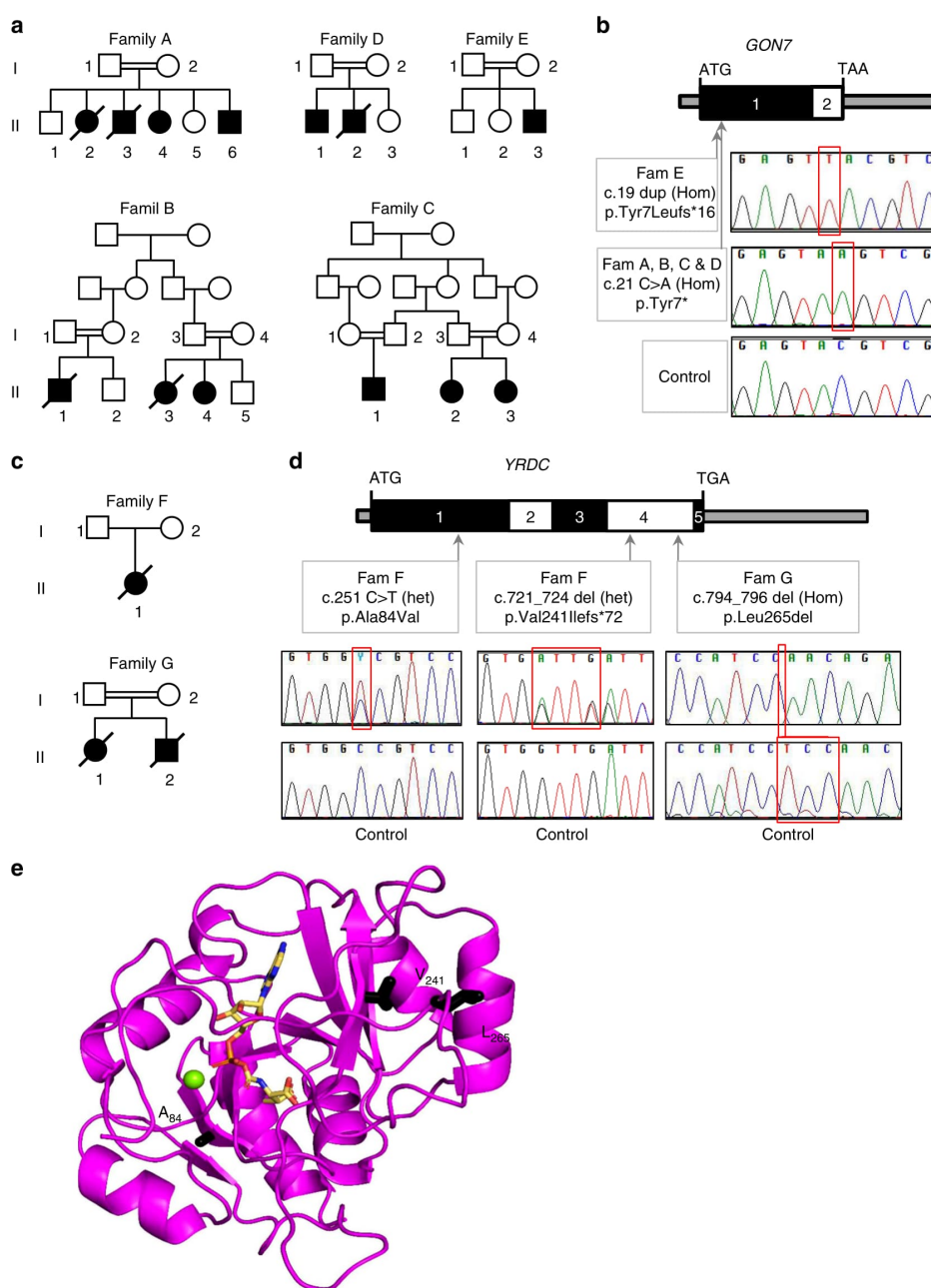
Here we present 14 GAMOS-affected individuals from seven families, with mutations in *GON7* (alias *C14orf142*) and *YRDC*, both genes encoding proteins involved in the biosynthesis of the t<sup>6</sup>A modification. These data, together with our previous work, show that mutations in genes encoding all the proteins involved in the two chemical steps of t<sup>6</sup>A lead to GAMOS. Furthermore,

we determine the crystal structure of the GON7/LAGE3/OSGEP KEOPS subcomplex showing that GON7 becomes structured upon binding to LAGE3. The structure also explains our observations that GON7 stabilizes the remainder of the KEOPS complex and directs its quaternary organization.

## Results

**Identification of *GON7* and *YRDC* mutations in GAMOS patients.** Through whole-exome sequencing in individuals with GAMOS, we identified mutations in the *GON7* gene in 11 affected individuals from 5 unrelated families and in the *YRDC* gene in 3 affected individuals from 2 unrelated families (Fig. 1a–d and Supplementary Table 1). Four of the families with *GON7* mutations (Families A to D) carried the same homozygous nonsense mutation (c.21 C>A, p.Tyr7\*) which causes a stop codon at position 7 of the protein. These families, all originating from the same region of Algeria, shared a common haplotype at the *GON7* locus indicating a founder effect (Supplementary Table 2). The affected individual of the fifth family (Family E) carried a different mutation involving the same residue at position 7 and leading to a frameshift (c.19dup, p.Tyr7Leufs\*16). Both *GON7* mutations are predicted to lead to the absence of protein expression and, as expected, no protein was detected in cells available from the affected individuals from family A, B, and C (Supplementary Fig. 1a, c). Two compound heterozygous *YRDC* mutations were identified in Family F: a missense mutation (c.251 C>T, p.Ala84Val) and a 4-base pair deletion leading to a frameshift (c.721\_724del, p.Val241Ilefs\*72). In Family G, we identified a homozygous in-frame deletion of Leucine 265 (c.794\_796 del, p.Leu265del). For both families, western blot and qPCR analysis on cell extracts from affected children showed the presence of *YRDC* transcripts and proteins (Supplementary Fig. 1b, d). To make a prediction of the effect of the *YRDC* mutations on the protein, we created a three-dimensional (3D) structural model of human *YRDC* using the structure of the *YRDC* domain of the archaeal Sua5 (PDB 4E1B, 20% sequence identity<sup>19</sup>) and mapped these mutations onto this in silico model (Fig. 1e). Structures of *YRDC* domains are very well conserved and sequence alignment shows that the human *YRDC* only has minor insertions/deletions compared to Sua5 (Supplementary Fig. 2). The replacement of Ala84, located in a hydrophobic region between a  $\beta$ -sheet and a connecting  $\alpha$ -helix, by the larger amino acid valine might perturb optimal packing and destabilize the structure of the protein. The *YRDC* Leu265del mutation affects a highly conserved amino acid and creates a deletion in a C-terminal peptide that hangs over the active site and could have a role in enzyme activity.

Early-onset proteinuria was observed in all affected children, with first detection ranging from between birth and 5 years. All but three children reached ESRD between 1.5 months and 6 years of age. All individuals carrying *YRDC* mutations presented with congenital or infantile SRNS detected from between birth and 4 months of age and died early, whereas most of the individuals carrying *GON7* mutations were alive at last follow-up, with either a functioning graft or with normal renal function despite a mild to heavy proteinuria (Supplementary Table 1). Kidney biopsies, when available, typically displayed FSGS (Families A, C, and E) or DMS (Families B, F, and G) (Fig. 2a–d). In addition to developmental delay, primary microcephaly was present in the two affected children of one family with *YRDC* mutations, whereas the affected child in the second *YRDC* family and all *GON7*-mutated individuals presented with post-natal microcephaly. Brain magnetic resonance imaging revealed a spectrum of cerebellar and cortical hypoplasia or atrophy with thin corpus callosum and ventricular dilation, myelination delay, and in one case a simplified gyral pattern (Individual G.II-2) (Fig. 2e–n,

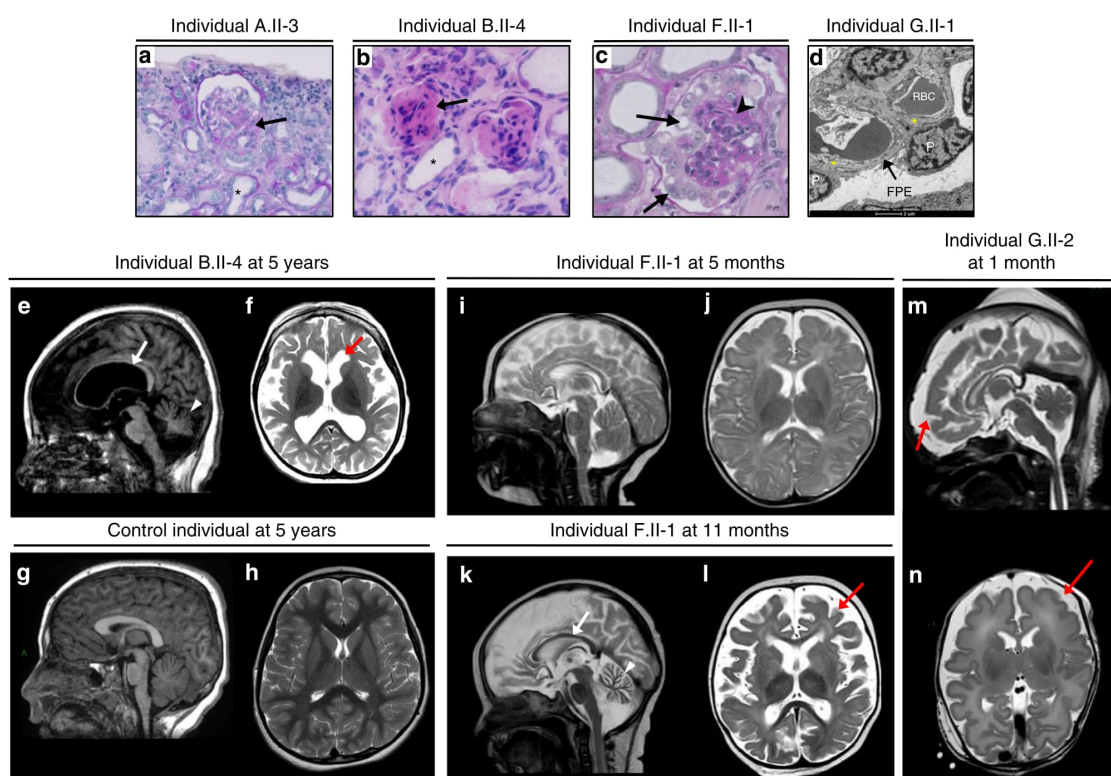


**Fig. 1** Identification of mutations in *GON7* and *YRDC* in patients with Galloway-Mowat syndrome. **a, c** Pedigrees of families with mutations in *GON7* (**a**) and in *YRDC* (**c**). Affected individuals are in black. **b, d** Organization of exons of human *GON7* and *YRDC* cDNAs. Positions of start and stop codons are indicated. Arrows indicate positions of the identified mutations. Lower panels show the sequencing traces for affected individuals with identified mutated nucleotide indicated with a red square (Hom: homozygous; het, heterozygous). **e** Representation of a 3D model of human *YRDC*, bound to the reaction product threonylcarbamoyl-adenylate, in sticks. The model was constructed using the crystal structure of *Sulfolobus tokodaii* Sua5 (PDB code 4E1B). The side chains of the three mutated residues are in black sticks. The green sphere represents an  $Mg^{2+}$  ion

Supplementary Fig. 3). Extra-renal features included facial dysmorphism, arachnodactyly, hiatal hernia with gastroesophageal reflux, congenital hypothyroidism (solely in the *YRDC* cases), and myoclonia. This clinical picture is highly reminiscent of that observed in *GAMOS*-affected individuals with

mutations in *LAGE3*, *OSGEP*, *TP53RK*, and *TPRKB*<sup>13</sup>. However, individuals with *GON7* mutations presented with milder neurological and renal manifestations, always with post-natal microcephaly and no gyration defects, later onset of proteinuria (median age 18 months vs. 1) and slower progression to ESRD



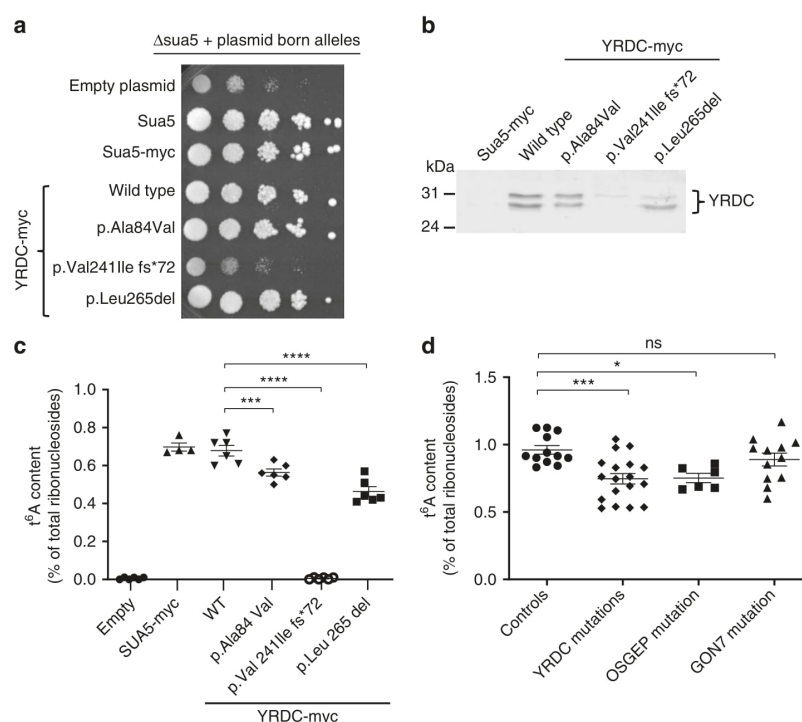


**Fig. 2** Kidney pathology analysis and neuroimaging. Light and transmission electron microscopy (TEM) of kidney sections of patients with GON7 (**a, b**) or YRDC mutations (**c, d**). **a** Individual A.II-3 displays a retracted glomerulus with a focal segmental glomerulosclerosis lesion at the vascular pole (black arrow) and tubular dilations (black star) (PAS;  $\times 200$  magnification). **b** Individual B.II-4 displays diffuse mesangial sclerosis with tiny, retracted and sclerosed glomeruli (black arrow) with dilated tubes surrounded by flat epithelial cells (black star) and interstitial fibrosis (H&E;  $\times 400$  magnification). **c** Individual F.II-1 displays a marked glomerular tuft collapsing (arrowhead) surrounded by a layer of enlarged and vacuolized podocytes (black arrows) (PAS stain;  $\times 400$  magnification, scale bar,  $10\ \mu\text{m}$ ). **d** TEM of individual G.II-1 shows diffuse foot process effacement (FPE; black arrow), a classical hallmark of nephrotic syndrome, along with abnormal folded and laminated segments (yellow stars), alternating with others with normal appearance. P podocyte, RBC red blood cell. Scale bar,  $2\ \mu\text{m}$ . Brain MRI of patients with GON7 (**e, f**) and YRDC mutations (**i-n**). **e, f** Brain MRI abnormalities in individual B.II-4 at 5 years. Sagittal T1-weighted image (**e**) shows important cortical subtentorial atrophy as well as corpus callosal (arrow) and cerebellar atrophy (arrowhead). The axial T2-weighted image (**f**) shows abnormal myelination and ventricular dilatation (red arrow). **g, h** Brain MRI of a 5-year old control showing sagittal T1 (**g**) and axial T2 (**h**) weighted images. **i-l** Brain MRI abnormalities in individual F.II-1 at 5 months (**i, j**) and 11 months (**k, l**). Sagittal T2-weighted image shows normal pattern at 5 months (**i**) evolving to a progressive major cerebellar (arrowhead) and cortical atrophy with a very thin corpus callosum (arrow) at 11 months (**k**). The axial T2-weighted image is normal at 5 months (**j**) but shows a very marked abnormality of myelination and cortical atrophy (red arrow) at 11 months (**l**). **m, n** Brain MRI abnormalities in individual G.II-2 at 1 month. Sagittal T2 (**m**) and axial T2 (**n**) weighted images show gyral anomalies with marked frontal gyral simplification (red arrow) and myelination delay

(median age 49 months in 8/11 children vs. 5 months in 3/3 children), and a longer survival compared to the YRDC cases.

**Impact of YRDC and GON7 mutations on  $t^6\text{A}$  biosynthesis.** To assess the pathogenicity of YRDC mutations, we first used a yeast heterologous expression and complementation assay as previously performed for OSGEP mutations identified in GAMOS individuals<sup>13</sup>. Indeed, the deletion of *SUA5*, the YRDC ortholog in *S. cerevisiae* leads to a very severe growth defect, similarly to the deletion of each of the genes encoding the five KEOPS subunits<sup>3,6,20,21</sup>. We therefore heterologously expressed the human YRDC cDNAs encoding wild-type (WT) and mutant proteins in the  $\Delta\text{sua5}$  strain to evaluate their ability to rescue the slow growth phenotype. Since the catalytic activity of YRDC does not require protein partners, the WT YRDC protein efficiently complemented the  $\Delta\text{sua5}$  growth defect (Fig. 3a). Although a

somewhat similar complementation was observed for the p. Ala84Val and p. Leu265del mutants, the p. Val241Ilefs\*72 mutant was notably unable to improve the poor growth of the  $\Delta\text{sua5}$  strain (Fig. 3a). All YRDC proteins were efficiently expressed in  $\Delta\text{sua5}$  strain, except the p. Val241Ilefs\*72 mutant that was barely detectable by western blot, suggesting that it is likely being degraded by an intracellular proteolytic machinery (Fig. 3b). Using mass spectrometry, we then analyzed the  $t^6\text{A}$  content of these transformed  $\Delta\text{sua5}$  yeast strains. As expected, since *Sua5* is the only enzyme in yeast that generates TC-AMP, the  $\Delta\text{sua5}$  mutant was unable to synthesize  $t^6\text{A}$ , whereas the WT YRDC expressing strain exhibited  $t^6\text{A}$  levels comparable with those measured for WT *Sua5* (Fig. 3c). The p. Ala84Val and p. Leu265del mutants showed a slight, but significant decrease in  $t^6\text{A}$  levels. In contrast, like in the  $\Delta\text{sua5}$  strain, no trace of  $t^6\text{A}$  modification could be detected in the p. Val241Ilefs\*72 mutant (Fig. 3c). In line with the results of the growth complementation assay, there was a



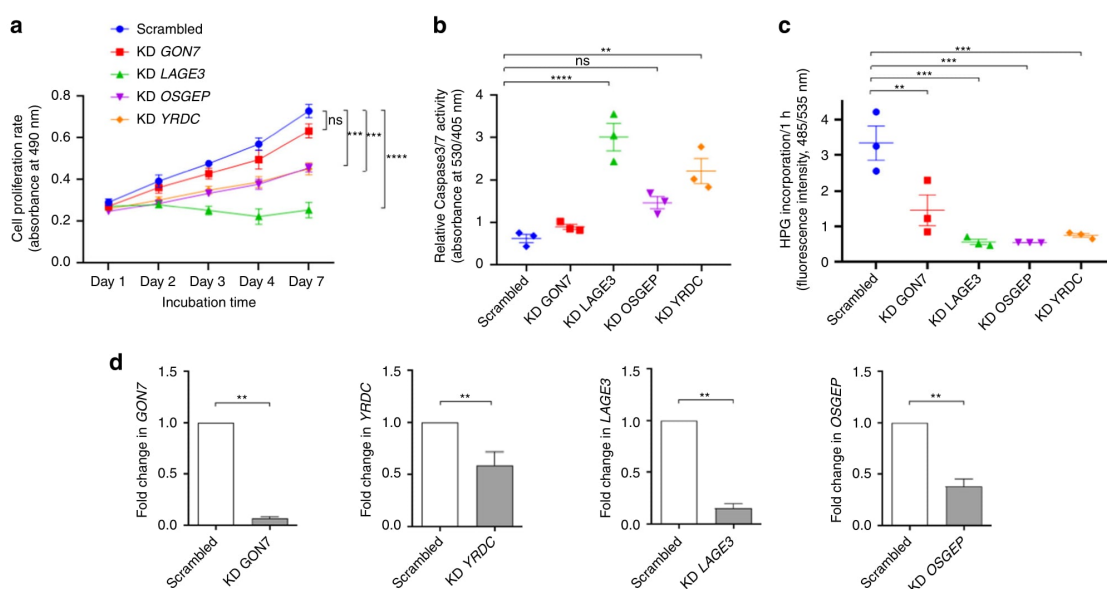
**Fig. 3** Effects of YRDC and GON7 mutations on t<sup>6</sup>A biosynthesis. **a** Evaluation of fitness of  $\Delta$ sua5 yeast strains expressing human YRDC variants (spots are from 10-fold serial dilutions of cell suspensions at OD<sub>600nm</sub> = 0.5, and three independent clones were evaluated) and **b** western blot analysis on total protein extracts from  $\Delta$ sua5 yeast cells expressing human YRDC variants using anti-hYRDC antibody. **c**, **d** Mass spectrometry (LC-MS/MS) quantification of t<sup>6</sup>A modification in total tRNAs extracted from  $\Delta$ sua5 yeast cells expressing human YRDC variants (**c**) (mean  $\pm$  s.e.m. of two independent LC-MS/MS experiments (technical replicates), each measuring samples from three independent yeast transformants; one-way ANOVA ( $F(5,28) = 269.4$ ,  $p < 0.0001$ ), Dunnett's multiple comparisons test,  $***p = 0.0008$ ,  $****p < 0.0001$ ) and from cultured primary skin fibroblasts from controls (two unaffected individuals) and affected individuals with either the p.Tyr7\* GON7 mutation (two individuals), or the YRDC mutations (three individuals) or with the p.Arg352Gln OSGEP mutation in the homozygous state (individual «CP» described in Braun et al.<sup>13</sup>) (**d**) (mean  $\pm$  s.e.m. of two independent LC-MS/MS experiments (technical replicates), each measuring samples from three independent cell culture experiments; one-way ANOVA ( $F(3,44) = 6.446$ ,  $p < 0.001$ ), Dunnett's multiple comparisons test, n.s. = 0.4894,  $*p = 0.0169$ ,  $***p = 0.008$ ). Source data are provided as a Source Data file

direct correlation between cell fitness and t<sup>6</sup>A content. This allowed us to classify YRDC mutations into hypomorphic (encoding p.Ala84Val and p.Leu265del) and amorphic (encoding p.Val241Ilefs\*72) alleles, as has been previously shown for OSGEP mutations<sup>13</sup>. A similar approach could not be applied for GON7 mutations since GON7 failed to complement the growth defect of the  $\Delta$ gon7 yeast strain (see ref. <sup>7</sup> and our data). We therefore measured the t<sup>6</sup>A content in fibroblasts from two individuals with the p.Tyr7\* GON7 mutation, three individuals with YRDC mutations and one individual with the p.Arg325Gln OSGEP mutation. The t<sup>6</sup>A levels were significantly decreased in both YRDC- and OSGEP-mutated fibroblasts, and to a lesser extent in GON7-mutated fibroblasts (Fig. 3d) confirming the impact of these mutations on t<sup>6</sup>A biosynthesis in affected individuals. In addition, we demonstrated that, similarly to individuals with OSGEP or TP53RK mutations<sup>13</sup>, telomere length was not affected in individuals with YRDC and GON7 mutations (Supplementary Fig. 4). This confirms that contrary to what has been demonstrated in yeast, human YRDC, and KEOPS complex are not involved in telomere maintenance in human cells<sup>22–24</sup>.

**In vitro characterization of WT and mutants YRDC.** To compare the stability and structure of the WT YRDC with those of the p.Ala84Val and p.Leu265del mutants, we expressed and

purified these proteins in an *E. coli* expression system (Supplementary Table 3). The three proteins could be purified but we noticed that both mutants were less stable and less soluble compared to the WT (Supplementary Methods). To probe the proper folding of the YRDC WT and mutants, we collect 1D <sup>1</sup>H-NMR spectra (Supplementary Fig. 5a). All the spectra displayed well-dispersed signals for amide protons as well as several signals at chemical shifts lower than 0.8 ppm that are typical of methyl groups in the hydrophobic core of proteins, suggesting the WT and mutants were well folded. To compare their enzymatic properties, we measured their TC-AMP synthesizing activities in vitro by quantifying the pyrophosphate reaction product. The p.Ala84Val and p.Leu265del mutants have lost about 75% and 30% of their activities respectively compared to WT (Supplementary Fig. 5b). The activities of these mutants are compatible with their hypomorphic nature, as suggested by the results of the yeast  $\Delta$ sua5 complementation experiments (Fig. 3a).

**Proliferation, apoptosis, and protein synthesis defects.** We have previously shown that transient gene expression knockdown (KD) of human KEOPS components OSGEP, TP53RK, and TPRKB leads to perturbations of various cellular processes including proliferation and apoptosis<sup>13</sup>. Similarly here, we transiently depleted the expression of GON7 and YRDC, as well as LAGE3 and OSGEP as

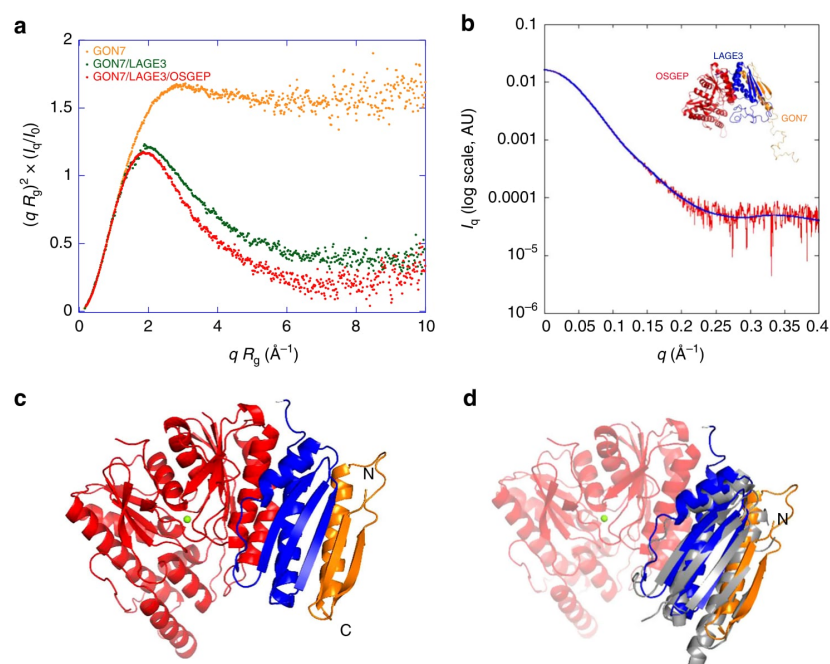


**Fig. 4** Proliferation, apoptosis, and protein synthesis defects upon *GON7* and *YRDC* knockdown. Transient knockdown (KD) of *GON7*, *LAGE3*, *YRDC*, and *OSGEP* was performed by lentiviral transduction of shRNA in immortalized human podocyte cell lines with a scrambled (non-targeting) shRNA as control. **a** Cell proliferation was assessed using a colorimetric MTT assay over 7 days, measuring absorbance at 490 nm at days 1, 2, 3, 4, and 7 (mean  $\pm$  s.e.m. of  $n = 5$  experiments, with each experiment performed in triplicate; two-way ANOVA ( $p < 0.0001$ ), Dunnett's multiple comparisons test, n.s. = 0.2031,  $***p < 0.0007$ ,  $****p < 0.0001$ ). **b** Cell apoptosis was evaluated by quantification of caspase 3/7 activation on the basis of fluorescence intensity (530/405 nm). Absolute values were normalized to DAPI fluorescence intensity as an internal control and compared to non-targeting shRNA-treated control cells (scrambled) (mean  $\pm$  s.e.m. of  $n = 3$  experiments with each experiment performed in triplicate; one-way ANOVA ( $F(4,10) = 21.42$ ,  $p < 0.0001$ ), Dunnett's multiple comparisons test, n.s. = 0.0556,  $*p = 0.0012$ ,  $****p < 0.0001$ ). **c** Protein biosynthesis rates were assessed on the basis of incorporation of HPG, an alkyne-containing methionine analog. After 2 h, alkyne-containing proteins were quantified on the basis of fluorescence intensity (485/535 nm). Absolute values were normalized to DAPI fluorescence intensity as an internal control and compared to control cells (mean  $\pm$  s.e.m. of  $n = 3$  experiments, with each experiment performed in triplicate, one-way ANOVA ( $F(4,10) = 16.36$ ,  $p = 0.0002$ ), Dunnett's multiple comparisons test,  $**p = 0.0035$ ,  $****p < 0.0003$ ). **d** Relative expression of *GON7*, *YRDC*, *LAGE3*, and *OSGEP* transcripts were normalized to that of HPRT in KD podocytes compared to non-targeting shRNA control treated cells (mean  $\pm$  s.e.m. of  $n = 5$  experiments, with each experiment being performed in triplicate; two-tailed Mann-Whitney test,  $**p < 0.05$ ). Source data are provided as a Source Data file

positive controls, in an immortalized human podocyte cell line<sup>25</sup>. We then demonstrated using a colorimetric cell proliferation assay that diminished expression of all four of these genes decreased cell proliferation, with the strongest decrease being observed in *LAGE3* KD podocytes (Fig. 4a). Despite efficient *GON7* KD, cells exhibited only a slight decrease in cell proliferation compared to cells treated with the control scrambled shRNA (Fig. 4a, d). The impairment of cell proliferation in *YRDC* and *OSGEP* KD cells was less marked than in *LAGE3* KD cells, which could be explained by a less efficient gene silencing (Fig. 4d). By measuring Caspase-3/7 activity, we next demonstrated that apoptosis was inversely related to proliferation with *LAGE3* KD podocytes displaying the highest rate of apoptosis (Fig. 4b). Since loss of t<sup>6</sup>A modification impacts global translation in yeast<sup>26</sup>, we also quantified the newly synthesized protein levels, which were decreased in all KD cells (Fig. 4c), even in *GON7* KD podocytes where proliferation and apoptosis rates were not drastically affected (Fig. 4a, b). Altogether, these results reinforce our previous findings for the other KEOPS subunits, *OSGEP*, *TP53RK*, and *TPRBK*, and confirm that mutations which alter t<sup>6</sup>A biosynthesis in human cells have an impact on cell survival through decreased proliferation and protein synthesis, ultimately leading to apoptosis.

**Structure of the human *GON7*/*LAGE3*/*OSGEP* subcomplex.** To better understand the role of human *GON7* and how its loss of

function could be connected with GAMOS, we set out to determine its structure and to establish how it interacts with the other KEOPS subunits. We had either crystal structures (*TPRBK*) or good quality 3D models (*LAGE3*, *OSGEP*, *TP53RK*) for all of the KEOPS subunits at our disposal, except for *GON7* (ref. 6). Based on very weak sequence similarity, it was proposed, that *GON7* is a remote homolog of yeast *Gon7* (ref. 7). We first investigated the structure of *GON7* in solution by collecting a 2D <sup>1</sup>H-<sup>15</sup>N Band-Selective Optimized Flip Angle Short Transient Heteronuclear Multiple-Quantum Correlation (SOFAST-HMQC) NMR spectrum of a <sup>15</sup>N-labeled *GON7* sample. The poor spectral dispersion in the <sup>1</sup>H dimension of the 2D correlation spectrum showed that *GON7* lacks well-defined structure, confirming the conclusions of Wan et al.<sup>7</sup>. Adding non-labeled *LAGE3* to the sample, caused the shift and/or disappearance for many crosspeaks, suggesting *GON7* interacts with *LAGE3* (Supplementary Fig. 6a). We further characterized the conformation of *GON7* in solution by small-angle X-ray scattering (SAXS) measurements. By representing the scattering data as a dimensionless Kratky plot ( $qR_g^2 \times I_q/I_0$  versus  $qR_g$ ), one can assess qualitative information on the degree of compactness of the scattering object<sup>27</sup>. The plateau observed for *GON7* is characteristic of a fully disordered protein, possibly with very short elements of secondary structure (Fig. 5a), confirming our NMR data (Supplementary Fig. 6a). We then purified the recombinant *GON7*/*LAGE3*/*OSGEP* complex and analyzed its behavior in solution by SAXS. Our SAXS data



**Fig. 5** Structure of the GON7/LAGE3/OSGEP complex. **a** Normalized Kratky plot of intensity scattering of GON7 (orange) and of the GON7/LAGE3 (green) and GON7/LAGE3/OSGEP complexes (red).  $q$ : scattering vector,  $R_g$  radius of gyration,  $I_q$ : scattering intensity,  $I_0$ : scattering intensity at zero angle. **b** Experimental x-ray scattering curve of the GON7/LAGE3/OSGEP complex (red). The blue curve represents the calculated scattering curve for the corresponding crystal structure of the complex. This yielded a good fit with the experimental data ( $\chi^2 = 0.33$ ). The inset shows the BUNCH model. **c** Representation of the crystal structure of the GON7/LAGE3/OSGEP complex: GON7 (gold), LAGE3 (blue), OSGEP (red). The N and C-termini of GON7 are labeled. The crystal lacked density for GON7 beyond residue 50. The active site of OSGEP is highlighted by the  $Mg^{2+}$  ion (green). **d** Superimposition of the yeast Gon7/Pcc1 complex (gray) onto GON7/LAGE3/OSGEP. GON7/LAGE3/OSGEP (same color code as in panel **c**)

established that the complex has a 1:1:1 stoichiometry in solution (Fig. 5b, Supplementary Table 4). In contrast with GON7, the dimensionless Kratky plot for GON7/LAGE3/OSGEP shows a bell-shaped curve with a maximum for  $qR_g \approx 2$  (Fig. 5a). This shape suggests that the complex is mainly compact, but that disordered regions are still present. In addition, the comparison of the distance distribution functions of GON7/LAGE3/OSGEP and GON7 shows that the latter alone is more extended than the complex (Supplementary Fig. 6b). In full agreement, the  $^{15}N$  SOFAST-HMQC NMR spectrum of the complex reveals that about 35 amino-acid residues of GON7 remains flexible and disordered in the complex whereas  $\sim 35$  crosspeaks vanished upon complex formation. These latter crosspeaks likely correspond to amino-acid residues engaged in the interaction with LAGE3 and thus experiencing extensive line-broadening due to the large molecular size of the complex. We therefore deduced that GON7 is becoming partially ordered upon complex formation with LAGE3/OSGEP and set out to determine its structure by X-ray crystallography. We obtained 1.9 Å resolution diffraction data of the GON7/LAGE3/OSGEP complex (Fig. 5c, d, Supplementary Table 5). The structure could be solved by molecular replacement using our 3D models of OSGEP and LAGE3 (refs. <sup>8,13,23</sup>). LAGE3 contains 60 residues at the N-terminus that are absent in the Pcc1 orthologs from yeast and archaea and that are predicted to be disordered. We did not observe any electron density for this N-terminal peptide, confirming this region indeed lacks a stable structure. We cannot exclude however that partial proteolysis removed this peptide during the long crystallization process. LAGE3 is at the center of the complex, binding on one side to OSGEP and on the other to GON7, which does not directly

interact with OSGEP (Fig. 5c). The structures of the LAGE3 and OSGEP subunits are very similar to their archaeal/fungal Pcc1 and Kae1 orthologs respectively. The two helices of LAGE3 associate with the N-terminal helices of OSGEP into a helical bundle. Only 45% of the GON7 sequence adopts a well-defined structure upon binding to LAGE3, confirming our SAXS- and NMR-based conclusions (Fig. 5a, Supplementary Fig. 6a–d). The N-terminal peptide of GON7 forms a  $\beta$ -hairpin between Met1 and Ser20 and the region between Gly25 and Pro50 forms a helix that lies parallel against the  $\beta$ -hairpin (Supplementary Fig. 7). Electron density for GON7 was absent for residues 21 to 24 and for the region beyond position 50. The C-terminal half of GON7 is highly enriched in acidic and sparse in hydrophobic amino acids and predicted to be unfolded. Despite their very weak sequence similarity (19% identity, 34% similarity), the structures of human GON7 and yeast Gon7 are almost identical (RMSD = 1.41 Å for 45 Ca positions; yeast Gon7 PDB entry: 4WXA) (Fig. 5d, Supplementary Fig. 7). The  $\beta$ -hairpin of GON7 aligns with the  $\beta$ -sheet of LAGE3 to form a continuous five stranded anti-parallel sheet. The helix of GON7 packs in an anti-parallel orientation against the C-terminal helix of LAGE3. The complex is stabilized mainly by the hydrophobic packing of side chains emanating from  $\beta 1$  and  $\alpha 1$  of GON7 and  $\alpha 2$  and  $\beta 1$  of LAGE3. The association mode between GON7 and LAGE3 is very similar to that of the yeast Pcc1/Gon7 complex, illustrated by their superposition (RMSD = 1.4 Å; Fig. 5d). Structure based sequence alignment between human GON7 and yeast Gon7 shows that only 6 out of 45 ordered residues (12%) are conserved (Supplementary Fig. 7). Compared to human GON7, yeast Gon7 is longer by about 20 residues that were disordered in its structure.

The experimental SAXS curve of the GON7/LAGE3/OSGEP complex in solution is in excellent agreement ( $\chi^2 = 0.33$ ) with the scattering curve calculated on the all-atom model built using the program BUNCH from the crystal structure (see Methods) (Fig. 5b). We therefore conclude that, although sharing very low sequence homology, human GON7 and yeast Gon7 are homologs that interact identically with their respective partners (LAGE3, Pcc1) in the human and yeast KEOPS complex.

**Role of human GON7 in KEOPS complex stability in vivo.** We further explored the deleterious cellular effects of the GAMOS-associated GON7 mutations. We first confirmed by mass spectrometry analysis that in a human podocyte cell line stably overexpressing either 2HA-tagged GON7 or V5-tagged LAGE3, the four additional KEOPS subunits significantly co-purified with GON7 or LAGE3, respectively, thus confirming that a five-subunit KEOPS complex is present in these renal glomerular cells (Supplementary Fig. 8a). We have previously demonstrated that the majority of GAMOS-associated mutations in genes encoding KEOPS complex components do not affect the intermolecular interactions between the LAGE3/OSGEP/TP53RK/TPRKB subunits<sup>13</sup>. In order to check whether mutations in LAGE3 affected GON7 binding, we co-expressed 2HA-tagged GON7 with WT or mutant V5-tagged LAGE3 in HEK293T cells. Our co-immunoprecipitation experiments demonstrated that the LAGE3 mutations found in GAMOS individuals do not prevent binding to GON7 (Supplementary Fig. 8b). Intriguingly, however, we noticed that co-expression of GON7 with LAGE3 in HEK293T cells led to an increased expression level of GON7, and to a lesser extent of LAGE3, suggesting that the interaction stabilizes both proteins (Fig. 6a). We therefore studied the stability of GON7 and LAGE3 in a time-course experiment using cycloheximide, an inhibitor of protein biosynthesis, in HEK293T cells transiently expressing either 2HA-GON7 or V5-LAGE3 alone or co-expressing both tagged-proteins. When expressed alone, GON7 and LAGE3 protein levels decreased, suggesting both proteins may be unstable in absence of their partner. This is particularly obvious for GON7 whose protein level decreased by half within an hour following cycloheximide addition (Fig. 6b, Supplementary Fig. 9). On the contrary, when co-expressed, we observed an increase of both GON7 and LAGE3 protein levels reflecting an increase in their stability. We wondered whether the absence of GON7 also impacts the stability of the whole KEOPS complex and indeed, we were able to demonstrate that the protein levels of the four KEOPS subunits were decreased in cells of individuals mutated for GON7, whereas they were not affected in cells of individuals with OSGEP or WDR73 mutations, the latter being also responsible for a specific subset of GAMOS cases not linked to a t<sup>6</sup>A biosynthesis defect<sup>15</sup> (Fig. 6c, Supplementary Fig. 10a). In addition, we demonstrated that this protein level decrease was not due to transcriptional regulation (Supplementary Fig. 10b). Altogether, these results suggest that the absence of GON7 affects KEOPS stability resulting in a decreased expression level of the four other subunits, which might impact t<sup>6</sup>A levels.

## Discussion

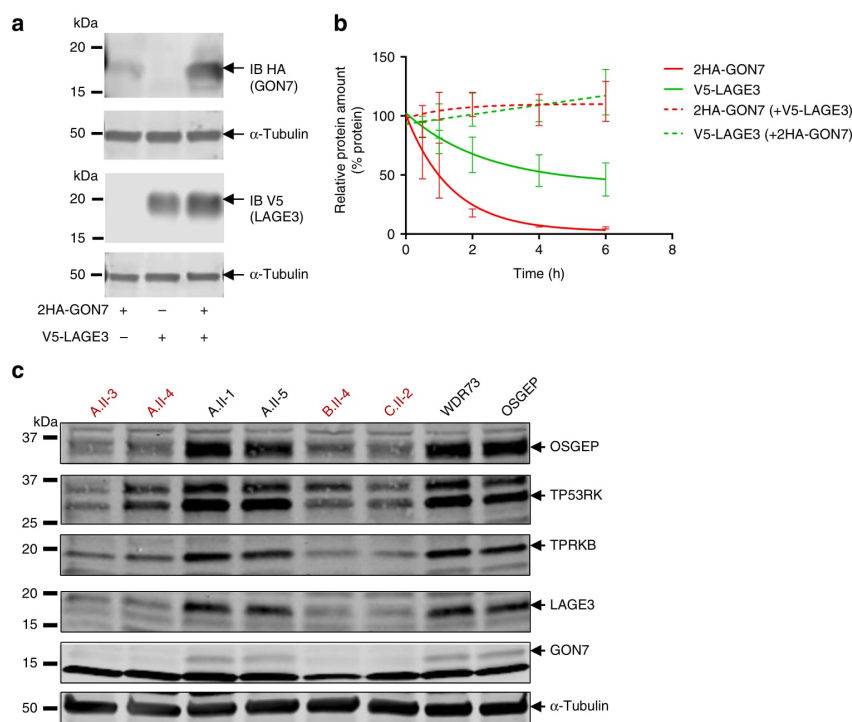
In this study, we identified mutations in two genes encoding proteins involved in t<sup>6</sup>A biosynthesis in GAMOS patients: YRDC encoding the enzyme that synthesizes the TC-AMP intermediate used by the KEOPS complex and GON7 encoding the fifth subunit of the KEOPS complex. Functional analysis of these specific mutations has revealed that they directly impact t<sup>6</sup>A modification (YRDC) and/or affect the stability of the KEOPS complex (GON7). These results complement our previous findings and

establish that mutations in all the genes involved in this pathway lead to GAMOS.

All individuals bearing GON7 or YRDC mutations present with the clinical features of GAMOS, similarly to the individuals previously reported to have mutations in the genes encoding the four other KEOPS subunits. In addition, we have expanded the GAMOS phenotype spectrum by describing congenital hypothyroidism to be associated with YRDC mutations. Although the two GON7 mutations encode truncated non-functional proteins, we noticed that they unexpectedly result in a less severe clinical outcome compared to that of individuals affected by mutations in other KEOPS subunit genes or in YRDC, for which biallelic null mutations were not found. This suggests that the absence of GON7 has less severe consequences for cell life compared to the other components of the t<sup>6</sup>A biosynthesis pathway, where four out of six are encoded by genes considered to be essential<sup>28</sup>. This less severe clinical outcome correlates with our data showing that GON7 loss of function and depletion in fibroblasts and podocytes, respectively, have globally a weaker effect on t<sup>6</sup>A levels, proliferation, apoptosis, and protein synthesis compared to that of other KEOPS subunits and YRDC mutations or depletion. However, although our data have confirmed that GON7 is a functional homolog of Gon7, the effect of their absence in human and yeast, respectively, is markedly different since in the absence of Gon7, the yeast KEOPS complex has only very low t<sup>6</sup>A activity and cell growth is dramatically affected<sup>3</sup>. Altogether, our data in humans suggest that GON7 is not as essential in humans as in yeast for t<sup>6</sup>A biosynthesis.

Our biochemical and structural data provide a molecular framework to understand the pathophysiological effects of the GAMOS-associated mutations. The structure of the GON7/LAGE3/OSGEP complex shows that GON7 is bound exclusively to the non-catalytic LAGE3 subunit, distant from the catalytic center of OSGEP. It has been shown in vitro that the intact human KEOPS complex has a 1:1:1:1:1 stoichiometry, in contrast with the complex lacking GON7 which has a 2:2:2:2 stoichiometry<sup>7</sup>. The latter stoichiometry has also been observed for the archaeal KEOPS complex, for which no fifth subunit similar to Gon7 has yet been discovered. The Pcc1 subunit constitutes the dimerization unit of archaeal KEOPS<sup>29</sup> and this is also very likely the case for the LAGE3 subunit of human KEOPS in absence of GON7 (ref. 7). In line with these results, our structure of the GON7/LAGE3/OSGEP complex shows that GON7 competes with LAGE3 for dimerization, explaining the different stoichiometries of the KEOPS complex observed in the absence or presence of GON7. Indeed, GON7 covers a large hydrophobic surface of LAGE3 (Supplementary Fig. 11), which is very likely occupied by another LAGE3 subunit in the context of a homomeric dimer, as observed in the structure of Pcc1 dimer<sup>29</sup>. The exposure of this hydrophobic surface due to the absence of GON7 in the GAMOS patients may affect the solubility and activity of the KEOPS complex, and indeed, our data from experiments on cell lines further indicate that GON7 contributes to the stability of the KEOPS complex and/or to the maintenance of the correct (catalytically active) quaternary structure as evidenced by the decrease in KEOPS subunits protein levels observed in GON7 patient cells. Taken together, our data demonstrated that GON7 impacts the stability of the KEOPS complex therefore having an effect on its enzymatic activity. This is in line with the in vitro data of Sicheri's group showing that in presence of GON7, KEOPS t<sup>6</sup>A activity is potentiated<sup>29</sup> and with our in vivo data showing that t<sup>6</sup>A levels in GON7-mutated patient fibroblasts are slightly decreased compared to YRDC- and OSGEP-mutated fibroblasts.

Although human GON7 and yeast Gon7 have low sequence identity, their structures and interactions with LAGE3 and Pcc1,



**Fig. 6** Role of GON7 on KEOPS complex stability. **a** Immunoblot analysis of HEK293T cell lysates expressing either 2HA-tagged GON7 or V5-tagged LAGE3 alone or co-expressing both proteins. Anti-HA and anti-V5 antibodies were used to assess GON7 and LAGE3 expression, respectively, with  $\alpha$ -tubulin used as loading control. **b** Representation of cycloheximide chase experiments by fitting a one-phase exponential decay curve to experimental data (one representative experiment is shown in Supplementary Fig. 9) (mean  $\pm$  s.e.m. of  $n = 3$  experiments). HEK293T cells were transfected with either 2HA-tagged GON7 or V5-tagged LAGE3 alone or with both proteins before being subjected to treatment with 100  $\mu$ g/ml cycloheximide for the indicated time points in order to assess rates of protein degradation followed by western blotting of the cell lysates for both proteins with anti-HA and anti-V5 antibodies, respectively. GON7 and LAGE3 protein levels were normalized to those of  $\alpha$ -tubulin at each time point. **c** Western blot analysis of protein expression level of the five KEOPS subunits in lymphoblastoid cell lines from two unaffected relatives (A.II-1 and A.II-5), four individuals with the GON7 mutation p.Tyr\*7, one individual with the OSGEP mutations p.Arg325Gln and p.Arg280His (individual «N2705» described in Braun et al.<sup>13</sup>), and one individual with GAMOS linked to WDR73 mutations (individual A.II-4 described in Colin et al.<sup>15</sup>). One representative western blot is shown (three independent experiments were performed).  $\alpha$ -Tubulin was used as a loading control. Source data are provided as a Source Data file

respectively, are nearly identical. It is therefore surprising that GON7 could not complement the yeast  $\Delta gon7$  deletion mutant<sup>7</sup>. Comparison of the GON7/LAGE3 and Gon7/Pcc1 complexes shows that the hydrophobic character of the residues at the interface is very well conserved (Supplementary Fig. 7). However quite a few amino-acid substitutions between GON7 and Gon7 might create steric clashes that weaken or disrupt the interaction with Pcc1, explaining the lack of complementation. Nevertheless, the exquisite superposition of GON7 and Gon7 qualifies them as orthologs and confirms that GON7 is the functional fifth subunit of the human KEOPS complex. Such discrepancies between the protein sequence and structure conservation between distant species might be relevant in other protein complex, with their characterization helping to identify new candidate genes for human monogenic disorders.

An increasing number of mutations are being identified in genes encoding tRNA-modifying enzymes that are linked to human neurodevelopmental disorders. Very recently, mutations in *WDR4*, initially described to cause a distinct form of microcephalic primordial dwarfism and brain malformations<sup>30,31</sup>, have been identified in individuals with GAMOS<sup>14</sup>. *WDR4* is a component of the METTL1/*WDR4* holoenzyme, an N<sup>7</sup>-methylguanosine (m<sup>7</sup>G) methyltransferase that is responsible for the highly

conserved m<sup>7</sup>G modification on a specific subset of tRNAs<sup>32,33</sup>. Interestingly, it has been shown that the absence of m<sup>7</sup>G tRNA modification leads to impaired cell proliferation, neural differentiation, and a decrease in global translation with a less efficient translation of mRNAs involved in cell division and brain development, consistent with the microcephaly and brain anomalies found in individuals with *WDR4* mutations<sup>34</sup>. Similarly, as a consequence of the decrease in t<sup>6</sup>A levels observed in affected individual's cells, perturbed protein translation could impact the translation of specific mRNA involved in kidney and brain development and/or podocyte/neuron maintenance. It is likely that the requirement for t<sup>6</sup>A-modified tRNAs levels is dependent of the cell-type and/or cell cycle as has been previously shown in *D. melanogaster* where highly proliferative cells of the wing imaginal discs are more affected by the absence of t<sup>6</sup>A modification than fully differentiated photoreceptors<sup>35</sup>. Neuronal progenitors that have high mitotic activity probably have higher demands for protein translation, making them more vulnerable to any perturbation in the tight regulation of tRNA modifications. Furthermore, another potential regulatory step to spatiotemporally modulate these tRNA modifications and thus protein translation is the tissue- and developmental stage-specific expression of the tRNA-modifying enzymes<sup>36,37</sup>. YRDC and

KEOPS subunits could be differentially expressed between specific cell-types in the brain (neural progenitors) and the kidney (podocytes) and/or during development/differentiation explaining the tissue involvement and the course of the disease as well as its clinical outcome. Further studies on neuronal/renal progenitors and neurons/podocytes differentiated from induced pluripotent stem cells obtained from individuals with mutations in YRDC and KEOPS subunits will probably provide further insights into the pathogenesis of GAMOS.

Together, our data strongly emphasize the importance and relevance of the t<sup>6</sup>A biosynthesis pathway in the pathogenesis of GAMOS. Further investigations are needed to fully characterize all the KEOPS mutations at the biochemical, structural, and enzyme activity levels to better understand their impact on KEOPS complex-dependent t<sup>6</sup>A biosynthesis activity and how they influence the clinical phenotypes. Genes encoding components of the t<sup>6</sup>A biosynthesis pathway have to be added to the growing list of translation-associated proteins whose loss of function are responsible for rare genetic disorders.

## Methods

**Patients and families.** Written informed consent was obtained from participants or their legal guardians, and the study was approved by the Comité de Protection des Personnes "Ile-De-France II." Genomic DNA samples were isolated from peripheral blood leukocytes using standard procedures.

**Whole-exome sequencing and mutation calling.** We performed whole-exome sequencing using Agilent SureSelect All Exon 51 Mb V5 capture-kit on a HiSeq2500 (Illumina) sequencer (paired-end reads: 2 × 100 bases). Sequences were aligned to the human Genome Reference Consortium Human Build 37 (GRCh37) genome assembly with the Lifescope suite from Life Technologies. Variant calling was made using the Genome Analysis Toolkit pipeline. Then, variants were annotated using a pipeline designed by the Paris Descartes University Bioinformatics platform. We assumed the causal variant: (i) segregates with the disease status, (ii) is novel or has a minor allele frequency <1/1000 in gnomAD, (iii) was not found in >10/2352 projects of our in-house database. Missense variant pathogenicity was evaluated using in silico prediction tools (PolyPhen2, SIFT and Mutation Taster). Sanger sequencing was used to validate the variant identified by exome sequencing and to perform segregation analysis in the families. Sequence were analyzed with the Sequencher software (Gene Codes, Ann Arbor, MI) and positions of mutations were numbered from the A of the ATG-translation initiation codon. For Family G, WES, and SNP-array were performed according to standard diagnostic procedures and WES quality criteria at the UMC Utrecht, the Netherlands. The patient–parent quartet WES with sibling-sharing analysis focused on the regions of homozygosity determined by SNP-array (parents are consanguineous in the eighth degree).

**Plasmids, cell culture, establishment of cell lines.** The following expression vectors were used in this publication: LentiORF pLEX-MCS (Open Biosystems), pESC-TRP with a c-myc tag (Agilent), and pLKO.1-TRC Cloning vector (# Plasmid 10878, Addgene). The LentiORF pLEX-MCS plasmid was modified by site-directed mutagenesis (QuickChange kit, Agilent) to insert one *NheI* restriction site, and either two copies of the Human influenza hemagglutinin (HA) tag or one copy of the V5 epitope tag allowing epitope-tagging at the N-terminal of the encoded protein. Human full-length *GON7*, *LAGE3*, and *YRDC* cDNA (NM\_032490.5, NM\_006014.4, and NM\_024640.4, respectively) were amplified by PCR from IMAGE cDNA clones (IMAGE 4796574, IMAGE 5485603, IMAGE 5211591, and IMAGE 6147134, respectively), and subcloned into the modified pLEX-MCS plasmid using either *SpeI* and *XhoI* (for *GON7* and *LAGE3*) or *NheI* and *XhoI* (for *YRDC*). Human *YRDC* cDNA was also subcloned into the *BamHI* and *Sall* sites of pESC-TRP. Site-directed mutagenesis (QuickChange kit, Agilent) was used to generate the mutations used in this study. An adapted cloning protocol was used to obtain the C-terminal extension found for the YRDC p.Val241Ilefs\*72 mutant. For gene silencing, the shRNA sequences described in Supplementary Table 6 were cloned into the lentiviral pLKO.1-TRC Cloning Vector using the *AgeI* and *EcoRI* restriction sites. This vector contains a cassette conferring puromycin resistance. All constructs were verified by Sanger sequencing.

The human immortalized podocyte cell line (AB8/13) provided by M. Saleem (University of Bristol, UK) was grown at 33 °C with 7% CO<sub>2</sub> in RPMI 1640 medium supplemented with 10% fetal bovine serum, insulin-transferrin-selenium, glutamine, and penicillin and streptomycin (all from Life Technologies), and human primary fibroblasts, obtained from patient skin biopsies, were cultured in OPTIMEM medium supplemented with 10% fetal bovine serum, sodium pyruvate, glutamine, fungizone, and penicillin and streptomycin (all from Life Technologies) at 37 °C with 7% CO<sub>2</sub>. Obtention and culture of lymphoblastoid cell lines are

detailed in Supplementary Methods. Human podocytes stably overexpressing 2HA-GON7 or V5-LAGE3, or transiently depleted for *GON7*, *LAGE3*, *OSGEP*, or *YRDC* were obtained by transduction with lentiviral particles and subsequent puromycin selection (2 μg/ml). HEK293T cells (ATCC CRL-3216) were transiently transfected using Lipofectamine<sup>®</sup> 2000 (ThermoFisher Scientific).

**Antibodies and chemical compounds.** The following antibodies were used in the study: mouse anti- $\alpha$ -tubulin (T5168, used at 1:1000), mouse anti-actin (A5316, used at 1:1000), mouse anti-HA (12CA5, at 1/1000), rabbit anti-GON7 (HPA 051832, used at 1:500), rabbit anti-LAGE3 (HPA 036122, used at 1/500), rabbit anti-TPRKB (HPA035712, used at 1:500), rabbit anti-OSGEP (HPA 039751, used at 1/1000), and mouse anti-GAPDH (MAB374, used at 1/2000) from Sigma-Aldrich; mouse anti-V5 (MCA1360, used at 1/1000) from Bio-Rad; rabbit anti-YRDC (PA5-56366, used at 1:500) from ThermoFisher Scientific; rabbit anti-LAGE3 (NBP2-32715, used at 1:1000) and mouse anti-OSGEP (NBP2-00823, used at 1:500) from Novus Biologicals; rabbit anti-TP53RK (API17010b, used at 1:500) from Abgent. Secondary antibodies for immunoblotting were sheep: anti-mouse and donkey anti-rabbit HRP-conjugated antibodies (GE Healthcare, UK), and IRDye 800CW Donkey anti-rabbit (926-32213) and IRDye 680RD Donkey anti-mouse (926-68072) antibodies (LI-COR). Cycloheximide (C7698), Nuclease P1 (N8630), phosphodiesterase I from snake venom (P3243), and alkaline phosphatase (P4252) were purchased from Sigma-Aldrich.

**Yeast culture and heterologous complementation assay.** Yeast cells were grown at 28 °C in standard rich medium YEPD (1% yeast extract, 2% peptone, 2% glucose) or minimal supplemented media (0.67% YNB, 2% carbon source). Cells were transformed using the lithium acetate method<sup>38</sup>. Media were supplemented with 2% agar for solid media. The *S. cerevisiae* W303 derived strain, *Asua5::KanMX* (YCplac33-SUA5)<sup>39</sup>, was used as the host for the complementation assay. For each pESC-TRP plasmid derivative to be tested, three independent clones were selected after transformation and grown on GLU-TRP media. Clones were then streaked onto GAL-TRP containing 0.1% 5-fluoroorotic acid (5-FOA) to counter-select the YCplac33-SUA5 plasmid (containing *URA3*). After two rounds of selection, clones were checked for their acquired Ura<sup>-</sup> phenotype, their plasmid content was confirmed by sequencing after plasmid rescue before being finally evaluated for fitness by a 10-fold serial dilution spotted onto GAL-TRP minimal supplemented media. Empty pESC-TRP, pESC-TRP-SUA5, and pESC-TRP-SUA5-*myc* were used as negative and positive controls, respectively.

**Quantitative real-time PCR.** Total mRNA from knocked down podocytes, primary skin fibroblasts, and LCLs was isolated using Qiagen Extraction RNeasy<sup>®</sup> Kit and treated with DNase I. One microgram total RNA was reverse-transcribed using Superscript II, according to the manufacturer's protocol (Life Technologies). The relative expression levels of the mRNA of interest were determined by real-time PCR using Power SYBR Green ROX Mix (ThermoFisher Scientific) with specific primers listed in Supplementary Table 7. Samples were run in triplicate and gene of interest expression was normalized to human hypoxanthine-guanine phosphoribosyl transferase (*Hgprt*). Data were analyzed using the 2<sup>- $\Delta\Delta C_t$</sup>  method.

**Quantification of t<sup>6</sup>A modification.** Yeast tRNAs were extracted and purified from actively growing cells (at OD<sub>600nm</sub> of approximately 3 × 10<sup>7</sup> cells/ml) with phenol induced cell permeabilization, LiCl-selective precipitation, and subsequent ion exchange-chromatography purification on an AXR-80 column (Nucleobond, Macherey-Nagel), according to the manufacturer's instructions. For human primary fibroblasts, the two-step protocol that was applied is detailed in Supplementary Methods. Ten micrograms of yeast tRNAs and approximately 1 μg of human fibroblast tRNAs were then enzymatically hydrolyzed into ribonucleosides with nuclease P1, phosphodiesterase, and alkaline phosphatase, dephosphorylated by filtration, and finally dried under vacuum according to the protocol of Thuring et al.<sup>40</sup> t<sup>6</sup>A ribonucleoside was analyzed using the quantitative LC/MS-MS protocol of Thuring et al.<sup>40</sup> Quantification of t<sup>6</sup>A was performed by integration of the peaks of interest and expressed relative to the total area of the peaks corresponding to the four canonical unmodified ribonucleosides assessed in the same sample for normalization. tRNA extracted from three independent samples were each measured twice (two technical replicates). Detailed information are provided in Supplementary Methods.

**Protein extraction and immunoblotting.** Proteins from KD podocytes, primary fibroblasts, and LCLs were extracted in lysis buffer containing 150 mM NaCl, 50 mM Tris-HCl pH 7, 0.5% Triton-X100 with Complete<sup>™</sup> protease inhibitors (Roche), as in Serrano-Perez et al.<sup>41</sup>. Fifty micrograms of proteins were loaded onto acrylamide gels and blotted onto nitrocellulose membranes (Amersham). The membranes were blocked in 1× Tris-buffered saline, 0.1% Tween 20 (TBST) with 5% bovine serum albumin or in Odyssey (LI-COR Bioscience) blocking buffer. Membranes were then incubated with the indicated primary antibodies, washed, and then incubated with either HRP-conjugated or LI-COR IRDye secondary antibodies. Signals were detected using ECL reagents (Amersham Biosciences) and acquired in a Fusion Fx7 darkroom (Vilber Lourmat) or acquired with Odyssey CLX near-infrared fluorescent imaging system (LI-COR Bioscience). Densitometry

quantification was performed either using Bio-1D software or using *Image studio lite* software (version 5.2). Uncropped and unprocessed blots are provided in the Source Data file.

**Immunoprecipitation and cycloheximide chase experiments.** For immunoprecipitation, HEK293T cells were transiently transfected with the adequate plasmids (2HA-tagged GON7, V5-tagged-LAGE3 wild-type (WT) and/or mutants) using calcium phosphate. Forty-eight hours post transfection, cells were lysed in 150 mM NaCl, 25 mM Tris-HCl pH 8, 0.5% Triton with protease inhibitors and HA-tagged GON7 was immunoprecipitated using the  $\mu$ MACS<sup>TM</sup> Epitope Tag Protein Isolation Kit (Miltenyi Biotec). Briefly, fresh lysates (1–1.5 mg of protein) were incubated either with mouse anti-V5 antibodies, followed by a 30-min incubation with magnetic beads coupled to protein A, or directly with magnetic beads coupled to an HA antibody. Immunoprecipitated proteins were isolated using  $\mu$ MACS<sup>®</sup> Separation Columns in a magnetic  $\mu$ MACS separator and subsequently eluted with 1 $\times$  Laemmli buffer. Lysates and immunoprecipitated samples were subjected to immunoblot<sup>41</sup>. To assess rates of protein degradation, HEK293T cells transiently expressing either 2HA-GON7 or V5-LAGE3 alone, or co-expressing both proteins were incubated with cycloheximide at a final concentration of 100  $\mu$ g/ml for the indicated time periods (0.5, 1, 2, 4, and 6 h). Total protein extracts and immunoblotting were performed as described above. Anti-HA and anti-V5 antibodies were used to reveal GON7 and LAGE3, respectively. Relative GON7 and LAGE3 protein amounts were normalized to those of  $\alpha$ -tubulin at each time point.

**Cell proliferation, apoptosis, and protein synthesis assays.** Cell proliferation, apoptosis level, and rates of protein synthesis were assessed in KD podocytes using the CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (MTT) (Promega), the Caspase-3/7 Green detection Reagent (C10423; ThermoFisher Scientific), and the Click-iT HPG Alexa Fluor 488 Protein Synthesis Assays (C10428; ThermoFisher Scientific), respectively, according to the manufacturer's instructions. Detailed information are provided in Supplementary Methods.

**Proteomic studies.** Human podocyte cell lines stably expressing either 2HA-GON7 or V5-LAGE3 were used to perform proteomic studies. 2HA-GON7 and V5-LAGE3 were immunoprecipitated as described in the section above. Eluates were processed according to Braun et al.<sup>13</sup>. Two groups (control IP versus IP HA or IP V5), each containing three biological replicates, were used for statistical analysis. Only proteins that were identified at least three times out of six were retained. A *t*-test was performed, and the data were represented in a volcano plot (FDR < 0.01, S0 = 2, 250 randomizations).

**Telomeric restriction fragment.** Measurement of the length of the terminal restriction fragments was performed by Southern blotting according to Touzot et al.<sup>42</sup>.

**Expression and purification of KEOPS subunits.** All structural work was done using the full-length proteins of GON7, LAGE3, and OSSEP.

For NMR experiments, two vectors were ordered from Genscript (Piscataway, USA) for the expression of either unlabeled his-tagged LAGE3 (vector "pET21a-LAGE3\_hisTEV\_op") or <sup>15</sup>N-labeled his-tagged GON7 (vector "pET24d-C14\_hisTEV\_op") whose sequences are shown in Supplementary Table 8. Expression and purification of LAGE3 and <sup>15</sup>N-GON7 and subcomplex LAGE3/<sup>15</sup>N-GON7 preparation are described in detail in Supplementary Methods.

For SAXS or crystallography experiments, preparation of unlabeled GON7 and GON7/LAGE3/OSSEP subcomplex, and co-expression and purification of the KEOPS subunits are detailed in Supplementary Methods and Supplementary Figs. 12 and 13. Fractions of the heterotrimeric GON7/LAGE3/OSSEP complex eluted from size exclusion chromatography (Supplementary Table 9) were then re-loaded onto NiIDA and washed with lysis buffer A supplemented with increasing concentrations of NaCl (0.2; 0.5; 1, and 2 M) in order to remove traces of contaminants. Bound proteins were eluted using three fractions of 2 ml of buffer A supplemented with 100, 200, and 400 mM imidazole and the three subunits complex was concentrated to 8.3 mg/ml for crystallization trials. A unique crystal was obtained using the sitting-drop vapor diffusion method after more than 6 months incubation at 4 °C. The successful condition was composed of 100 nl of protein solution and 100 nl of 30% PEG 4000, 0.1 M Tris-HCl pH 8.5 and 0.2 M magnesium chloride. The crystal was cryo-protected by quick-soaking in reservoir solution supplemented with 30% glycerol prior to flash freezing in liquid nitrogen.

**Modeling and crystal structure determination.** Modeling of YRDC: the Phyre2 and I-tasser webservers both proposed high confidence models for YRDC based on the StSua5 crystal structure despite weak sequence identity between the two species. A 3D model of YRDC was built using the MODELLER software<sup>43</sup>. X-ray diffraction data collection was carried out on beamline Proxima1 at the SOLEIL Synchrotron (Saint-Aubin, France) at 100 K. Data were processed, integrated, and scaled with the XDS program package<sup>44</sup>. The crystal belonged to space group P4<sub>3</sub>. The OSSEP and LAGE3 subunits were positioned by molecular replacement with the programs PHASER<sup>45</sup> and MOLREP, implemented in the CCP4 suite<sup>46</sup> using the structures of MjKae1 (PDB ID: 2VWB) and ScPcc1 (PDB ID: 4WX8) as search

models. Residual electron density showed clearly the presence of the GON7 subunit, which was constructed using the program BUCCANEER<sup>46</sup>. The initial structure was refined using the BUSTER program<sup>47</sup> and completed by interactive and manual model building using COOT<sup>48</sup>. The correctness of the assigned GON7 sequence was verified by omit mFo-DFc, 2mFo-DFc, Prime-and-switch electron density maps<sup>49</sup> (Supplementary Fig. 14). One copy of the GON7/LAGE3/OSSEP heterotrimer was present in the asymmetric unit. Data collection and refinement statistics are gathered in Supplementary Table 5. The coordinates have been deposited at the Protein Data Bank (code 6GWJ).

**Small-angle X-ray analysis.** SAXS experiments were carried out at the SOLEIL synchrotron SWING beamline (Saint-Aubin, France). The sample to detector (Aviex CCD) distance was set to 1500 mm, allowing reliable data collection over the momentum transfer range  $0.008 \text{ \AA}^{-1} < q < 0.5 \text{ \AA}^{-1}$  with  $q = 4\pi\sin\theta/\lambda$ , where  $2\theta$  is the scattering angle and  $\lambda$  is the wavelength of the X-rays ( $\lambda = 1.0 \text{ \AA}$ ). To isolate the various species in solution, SAXS data were collected on samples eluting from an online size exclusion high-performance liquid chromatography (SEHPLC/Bio-SEC3Agilent) column and directly connected to the SAXS measuring cell. sixty-five microliters of GON7/LAGE3/OSSEP and GON7 samples concentrated at 1.5 and 6.7 mg/l, respectively, were injected into the column pre-equilibrated with a buffer composed of 20 mM MES pH 6.5, 200 mM NaCl, and 5 mM 2-mercaptoethanol. Flow rate was 300  $\mu$ l/min, frame duration was 1.0 s, and the dead time between frames was 0.5 s. The protein concentration was estimated by UV absorption measurement at 280 and 295 nm using a spectrometer located immediately upstream of the SAXS measuring cell. A large number of frames were collected before the void volume and averaged to account for buffer scattering. SAXS data were normalized to the intensity of the incident beam and background (i.e. the elution buffer) subtracted using the program FoXTrot<sup>50</sup>, the Swing in-house software. The scattered intensities were displayed on an absolute scale using the scattering by water. Identical frames under the main elution peak were selected and averaged for further analysis. Radii of gyration, maximum particle dimensions, and molecular masses were determined using PrimusQT<sup>51</sup> (Supplementary Table 4). The BUNCH program<sup>52</sup> was then used to build atomic models of GON7/LAGE3/OSSEP starting from the crystal structure and by determining the optimal position of the missing regions as to fit the data. In a final step, we substituted the dummy residues of these flexible parts with all-atom descriptions using the programs PD2 and SCWRL4 (ref. 53). An ultimate adjustment was performed using the program CRYSOLO<sup>54</sup>. The modeling was repeated 10 times and the best model was deposited in SASBDB<sup>55</sup> with codes SASDFK8, SASDFM8, and SASDFL8 for GON7, GON7/LAGE3, and GON7/LAGE3/OSSEP, respectively.

**Statistical analyses.** GraphPad Prism 8.0 software was used for the graphical representation and statistical analysis of cell-based data. Results are presented as mean  $\pm$  s.e.m. of at least  $n = 3$  independent experiments. For statistical analysis, data sets comparing more than three conditions (to a control group) were analyzed with ANOVA followed by Dunnett's multiple comparisons test or by using Kruskal–Wallis test followed by a Dunn's multiple comparisons test. Data sets with only two conditions to compare were analyzed using an unpaired *t*-test or a Mann–Whitney test.  $P < 0.05$  was considered statistically significant. A standard confidence interval of 95% was applied in all analyses. Displayed in the figures are the mean values of all technical replicates for each of the independent experiments (displayed as single data points). Black lines indicate the mean values of all independent experiments.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The source data underlying Figs. 3c–d, 4 and 6 and Supplementary Figs. 1, 5b, 8b, 9 and 10 are provided as a Source Data file. Patients' consent was not obtained for public deposition of whole-exome sequencing data, but these are available from the corresponding author upon reasonable request. Other data generated during the current study are available from the corresponding authors upon reasonable request. Accession codes for deposited data: crystal structure of GON7/LAGE3/OSSEP (PDB ID: 6GWJ), [<https://www.rcsb.org/structure/6GWJ>]; SAXS model codes SASDFK8, SASDFM8, and SASDFL8 for GON7, GON7/LAGE3, and GON7/LAGE3/OSSEP, respectively.

Received: 18 March 2019 Accepted: 9 August 2019

Published online: 03 September 2019

## References

1. El Yacoubi, B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* **46**, 69–95 (2012).



2. El Yacoubi, B. et al. A role for the universal Kae1/Qri7/YgiD (COG0533) family in tRNA modification. *EMBO J.* **30**, 882–893 (2011).
3. Srinivasan, M. et al. The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *EMBO J.* **30**, 873–881 (2011).
4. Lauhon, C. T. Mechanism of N<sup>6</sup>-threonylcarbamoyladenine (t(6)A) biosynthesis: isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry* **51**, 8950–8963 (2012).
5. Costessi, A. et al. The human EKC/KEOPS complex is recruited to Cullin2 ubiquitin ligases by the human tumour antigen PRAME. *PLoS One* **7**, e42822 (2012).
6. Mao, D. Y. et al. Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine. *Mol. Cell* **32**, 259–275 (2008).
7. Wan, L. C. et al. Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acids Res.* **45**, 805–817 (2017).
8. Zhang, W. et al. 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 (2015).
9. Thiaville, P. C., Iwata-Reuyl, D. & de Crécy-Lagard, V. Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t(6)A), a universal modification of tRNA. *RNA Biol.* **11**, 1529–1539 (2014).
10. Ramos, J. & Fu, D. The emerging impact of tRNA modifications in the brain and nervous system. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**, 412–428 (2019).
11. Galloway, W. H. & Mowat, A. P. Congenital microcephaly with hiatus hernia and nephrotic syndrome in two sibs. *J. Med. Genet.* **5**, 319–321 (1968).
12. Braun, D. A. et al. Mutations in multiple components of the nuclear pore complex cause nephrotic syndrome. *J. Clin. Invest.* **128**, 4313–4328 (2018).
13. Braun, D. A. et al. Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat. Genet.* **49**, 1529–1538 (2017).
14. Braun, D. A. et al. Mutations in WDR4 as a new cause of Galloway-Mowat syndrome. *Am. J. Med. Genet. A* **176**, 2460–2465 (2018).
15. Colin, E. et al. Loss-of-function mutations in WDR73 are responsible for microcephaly and steroid-resistant nephrotic syndrome: Galloway-Mowat syndrome. *Am. J. Hum. Genet.* **95**, 637–648 (2014).
16. Fujita, A. et al. Homozygous splicing mutation in NUP133 causes Galloway-Mowat syndrome. *Ann. Neurol.* **84**, 814–828 (2018).
17. Rosti, R. O. et al. Homozygous mutation in NUP107 leads to microcephaly with steroid-resistant nephrotic condition similar to Galloway-Mowat syndrome. *J. Med. Genet.* **54**, 399–403 (2017).
18. Vodopituz, J. et al. WDR73 mutations cause infantile neurodegeneration and variable glomerular kidney disease. *Hum. Mutat.* **36**, 1021–1028 (2015).
19. Parthier, C. et al. The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem. Int. Ed. Engl.* **51**, 4046–4052 (2012).
20. El Yacoubi, B. et al. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Res.* **37**, 2894–2909 (2009).
21. Kisseleva-Romanova, E. et al. Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J.* **25**, 3576–3585 (2006).
22. Downey, M. et al. A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell* **124**, 1155–1168 (2006).
23. Hecker, A. et al. Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex. *EMBO J.* **27**, 2340–2351 (2008).
24. Liu, Y. Y. et al. Yeast KEOPS complex regulates telomere length independently of its t(6)A modification function. *J. Genet. Genomics* **45**, 247–257 (2018).
25. Saleem, M. A. et al. A conditionally immortalized human podocyte cell line demonstrating nephrin and podocin expression. *J. Am. Soc. Nephrol.* **13**, 630–638 (2002).
26. Thiaville, P. C. et al. Global translational impacts of the loss of the tRNA modification t(6)A in yeast. *Micro Cell.* **3**, 29–45 (2016).
27. Bizien, T. et al. A brief survey of state-of-the-art BioSAXS. *Protein Pept. Lett.* **23**, 217–231 (2016).
28. Blomen, V. A. et al. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).
29. Wan, L. C. et al. Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t6A biosynthesis. *Nucleic Acids Res.* **44**, 6971–6980 (2016).
30. Shaheen, R. et al. Mutation in WDR4 impairs tRNA m(7)G46 methylation and causes a distinct form of microcephalic primordial dwarfism. *Genome Biol.* **16**, 210 (2015).
31. Trimouille, A. et al. Further delineation of the phenotype caused by biallelic variants in the WDR4 gene. *Clin. Genet.* **93**, 374–377 (2018).
32. Alexandrov, A., Martzen, M. R. & Phizicky, E. M. Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* **8**, 1253–1266 (2002).
33. Leulliot, N. et al. Structure of the yeast tRNA m7G methylation complex. *Structure* **16**, 52–61 (2008).
34. Lin, S. et al. Mettl1/Wdr4-Mediated m(7)G tRNA methylation is required for normal mRNA translation and embryonic stem cell self-renewal and differentiation. *Mol. Cell* **71**, 244–255 e245 (2018).
35. Rojas-Benitez, D., Eggers, C. & Glavic, A. Modulation of the proteostasis machinery to overcome stress caused by diminished levels of t6A-modified tRNAs in *Drosophila*. *Biomolecules* **7**, <https://doi.org/10.3390/biom7010025> (2017).
36. de Crécy-Lagard, V. et al. Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* **47**, 2143–2159 (2019).
37. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
38. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
39. Pichard-Kostuch, A. et al. Structure-function analysis of Sua5 protein reveals novel functional motifs required for the biosynthesis of the universal t(6)A tRNA modification. *RNA* **24**, 926–938 (2018).
40. Thuring, K., Schmid, K., Keller, P. & Helm, M. Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods* **107**, 48–56 (2016).
41. Serrano-Perez, M. C. et al. Endoplasmic reticulum-retained podocin mutants are massively degraded by the proteasome. *J. Biol. Chem.* **293**, 4122–4133 (2018).
42. Touzot, F. et al. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. *Proc. Natl. Acad. Sci. USA* **107**, 10097–10102 (2010).
43. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318–326 (1995).
44. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
45. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
47. Brignone G. et al. BUSTER version 2.10.3 (Global Phasing Ltd, Cambridge, UK, 2017).
48. Amsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
49. Terwilliger, T. C. Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2144–2149 (2004).
50. David, G. & Perez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Cryst.* **42**, 892–900 (2009).
51. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
52. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250 (2005).
53. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778–795 (2009).
54. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLO—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).
55. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–D363 (2015).

### Acknowledgements

We are grateful to the patients and their families for their participation. We thank Dr A. Lahoche, Dr A. Bennour, Dr C. Loirat, Dr K. van Gassen, Dr M. Lilien, and Dr M. van der Heide-Jalving who helped in the diagnostic process and the follow-up of our patients. We thank G. Froment, D. Nègre, and C. Costa from the lentivector production facility/SFR BioSciences Gerland-Lyon Sud (UMS3444/US8). We acknowledge the use of the biosources of the Necker Imagine DNA biobank (BB-033-00065) and I. Rouvet from CBC BioTec at the CRB-HCL (Hospices Civils de Lyon). We thank Olivier Namy (I2BC, Gif-sur-Yvette, France) for his helpful advices and expertise during the revision process of this article. This work was supported by the Fondation pour le Recherche Médicale (project DEQ2015031682) (to C. Antignac), the European Union's Seventh Framework Programme (FP7/2012, grant 305608 EUrenOmics) (to C. Antignac), the Investments for the Future Program (grant ANR-10-IAHY-01) (to C. Antignac), ANR KeoGamo (ANR-18-CE11-0008-01) (to G. Mollet and H.v.T.). This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01) (to H.v.T.) and the Dutch Kidney Foundation (grant 15OP14) (to A.M.v.E.). S.M. is supported by a Ph.D. grant of the Fondation pour le Recherche Médicale (FRM). P.R. is a scientist from Centre National de la Recherche Scientifique (CNRS).

**Author contributions**

R.S., O.G., O.B., E.M., F.N., D.A.B., M.P., C.M., A.M.v.E., F.H., D.M., and C. Antignac performed exome sequencing, bioinformatics analysis of exome sequencing data, whole-exome evaluation, Sanger sequencing, and mutation analysis. R.S., O.B., M.C., S.D., R.N., M.-A.M., B.R., J.B., A.L., D.A.B., A.M.v.E., D.M., F.H., and C. Antignac recruited patients and collected detailed clinical information for the study. N.B. provided MRI from control individuals and critically interpreted MRI images from patients. S.C.-F., R.S., and A.M.v.E. provided and analyzed images of renal histology and electron microscopy. C. Arrondel, J.P., B.C., D.L., G. Martin, E.M., and F.N. performed design of expression vectors used in this study. C. Arrondel generated knockdown cell lines, performed *in vitro* studies (proliferation, apoptosis, and protein synthesis) in immortalized human podocytes, and performed purification of human tRNAs. C. Arrondel, J.P., G. Menara, L.B., G. Martin, E.M., and F.N. performed cell experiments (co-immunoprecipitation, cycloheximide chase, cell culture), qPCR, and western blot experiments. S.M. performed proteins expression and purifications, cristallogensis trials, diffraction data collection, 3D structure resolution, SAXS data collection, and analysis. B.C. performed OSGEP/LAGE3/GON7-his expression and purification, and nucleosides preparation from tRNA samples and YRDC WT and mutant enzymatic assay. D.L. performed yeast complementation studies, expression, and purification of yeast tRNAs. D.D. performed SAXS data collection and analysis. E.L. collected and analyzed NMR experiments. A.-C.B. and S.S. performed HPLC MS/MS <sup>16</sup>A modification analysis. G. Mollet, G. Martin, and I.C.G. performed proteomic studies in human podocyte cell lines. P.R. performed telomere restriction-fragment assays. C. Arrondel, S.M., J.P., G. Menara, B.C., D.L., D.D., O.G., O.B., L.B., G. Martin, E.M., F.N., E.L., A.-C.B., S.S., I.C.G., P.R., M.P., C.M., A.M.v.E., D.M., C. Antignac, H.v.T., and G. Mollet contributed to the interpretation of the data. G. Mollet, H.v.T. and C. Antignac conceived and coordinated the study, and wrote the manuscript with the input of B.C., D.L., S.M., and C. Arrondel. All authors critically analyzed and edited the manuscript.

**Additional information**


**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-11951-x>.

**Competing interests:** F.H. is a cofounder and SAB member of Goldfinch-Bio. The remaining authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Peer review information:** *Nature Communications* thanks Piotr Neumann, Kazuhito Tomizawa, and other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

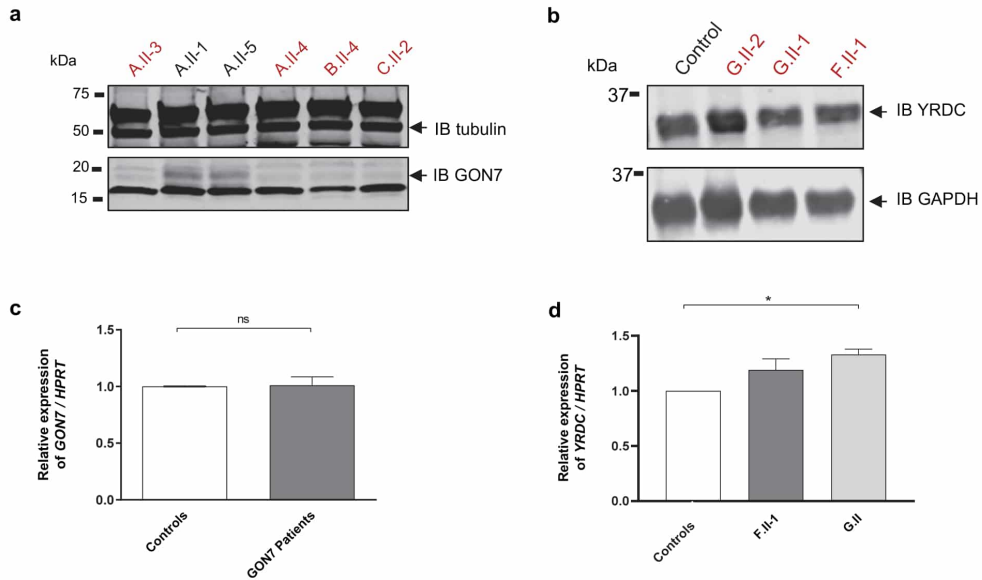
Christelle Arrondel<sup>1,2,4</sup>, Sophia Missouri<sup>2,24</sup>, Rozemarijn Snoek<sup>3,4</sup>, Julie Patat<sup>1</sup>, Giulia Menara<sup>1</sup>, Bruno Collinet<sup>2,5</sup>, Dominique Liger<sup>2</sup>, Dominique Durand<sup>2</sup>, Olivier Gribouval<sup>1</sup>, Olivia Boyer<sup>1,6</sup>, Laurine Buscara<sup>1</sup>, Gaëlle Martin<sup>1</sup>, Eduardo Machuca<sup>1</sup>, Fabien Nevo<sup>1</sup>, Ewen Lescop<sup>7</sup>, Daniela A. Braun<sup>8</sup>, Anne-Claire Boschat<sup>9</sup>, Sylvia Sanquer<sup>10,11</sup>, Ida Chiara Guerrero<sup>12</sup>, Patrick Revy<sup>13</sup>, Mélanie Parisot<sup>14</sup>, Cécile Masson<sup>15</sup>, Nathalie Boddart<sup>16</sup>, Marina Charbit<sup>6</sup>, Stéphane Decramer<sup>17</sup>, Robert Novo<sup>18</sup>, Marie-Alice Macher<sup>19</sup>, Bruno Ranchin<sup>20</sup>, Justine Bacchetta<sup>20</sup>, Audrey Laurent<sup>20</sup>, Sophie Collardeau-Frachon<sup>14</sup>, Albertien M. van Eerde<sup>3,4</sup>, Friedhelm Hildebrandt<sup>8</sup>, Daniella Magen<sup>22</sup>, Corinne Antignac<sup>1,2,3</sup>, Herman van Tilbeurgh<sup>2</sup> & Géraldine Mollet<sup>1</sup>

<sup>1</sup>Laboratory of Hereditary Kidney Diseases, INSERM UMR1163, Université de Paris, Imagine Institute, Paris, France. <sup>2</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France. <sup>3</sup>Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. <sup>4</sup>Center for Molecular Medicine, Utrecht University, Utrecht, The Netherlands. <sup>5</sup>Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, UMR7590 CNRS/Sorbonne-Université, UPMC, Paris, France. <sup>6</sup>Department of Pediatric Nephrology, AP-HP, Necker Hospital, Paris, France. <sup>7</sup>Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France. <sup>8</sup>Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. <sup>9</sup>Mass Spectrometry Facility, INSERM UMR1163, Imagine Institute, Paris, France. <sup>10</sup>Service de Biochimie métabolomique et protéomique, Hôpital Necker-Enfants Malades, Paris, France. <sup>11</sup>INSERM UMR-S1124, Université de Paris, Paris, France. <sup>12</sup>Proteomics Platform 3P5-Necker, Université de Paris—Structure Fédérative de Recherche Necker, Inserm US24/CNRS, UMS3633 Paris, France. <sup>13</sup>Inserm UMR1163, Laboratory of Genome Dynamics in the Immune System, Labellisé Ligue contre le Cancer, Université de Paris, Imagine Institute, Paris, France. <sup>14</sup>Genomics Core Facility, Structure Fédérative de Recherche Necker, INSERM U1163 and Inserm US24/CNRS UMS3633, Université de Paris, Paris, France. <sup>15</sup>Bioinformatics Platform, INSERM UMR1163, Université de Paris, Imagine Institute, Paris, France. <sup>16</sup>Department of Pediatric Radiology, and Imagine Institute, INSERM UMR 1163 and INSERM U1000, Université de Paris, Hôpital Necker-Enfants Malades, Paris, France. <sup>17</sup>Department of Pediatric Nephrology-Internal Medicine, Purpan Hospital, Toulouse, France. <sup>18</sup>Pediatric Nephrology Unit, University Hospital of Lille, Lille, France. <sup>19</sup>Department of Pediatric Nephrology, AP-HP, Robert Debre Hospital, Paris, France. <sup>20</sup>Service de Néphrologie, Rhumatologie et Dermatologie pédiatriques, Hospices Civils de Lyon, Hôpital Femme-Mère-Enfant, Centre de référence de maladies rénales rares, Université de Lyon, Bron, France. <sup>21</sup>Department of Pathology, Hospices Civils de Lyon-Hôpital Femme-Mère-Enfant, Claude Bernard Lyon 1 University, Bron, France. <sup>22</sup>Pediatric Nephrology Institute-Rambam Health Care Campus-Technion Faculty of Medicine, Haifa, Israel. <sup>23</sup>Department of Genetics, AP-HP, Necker Hospital, Paris, France. <sup>24</sup>These authors contributed equally: Christelle Arrondel, Sophia Missouri.

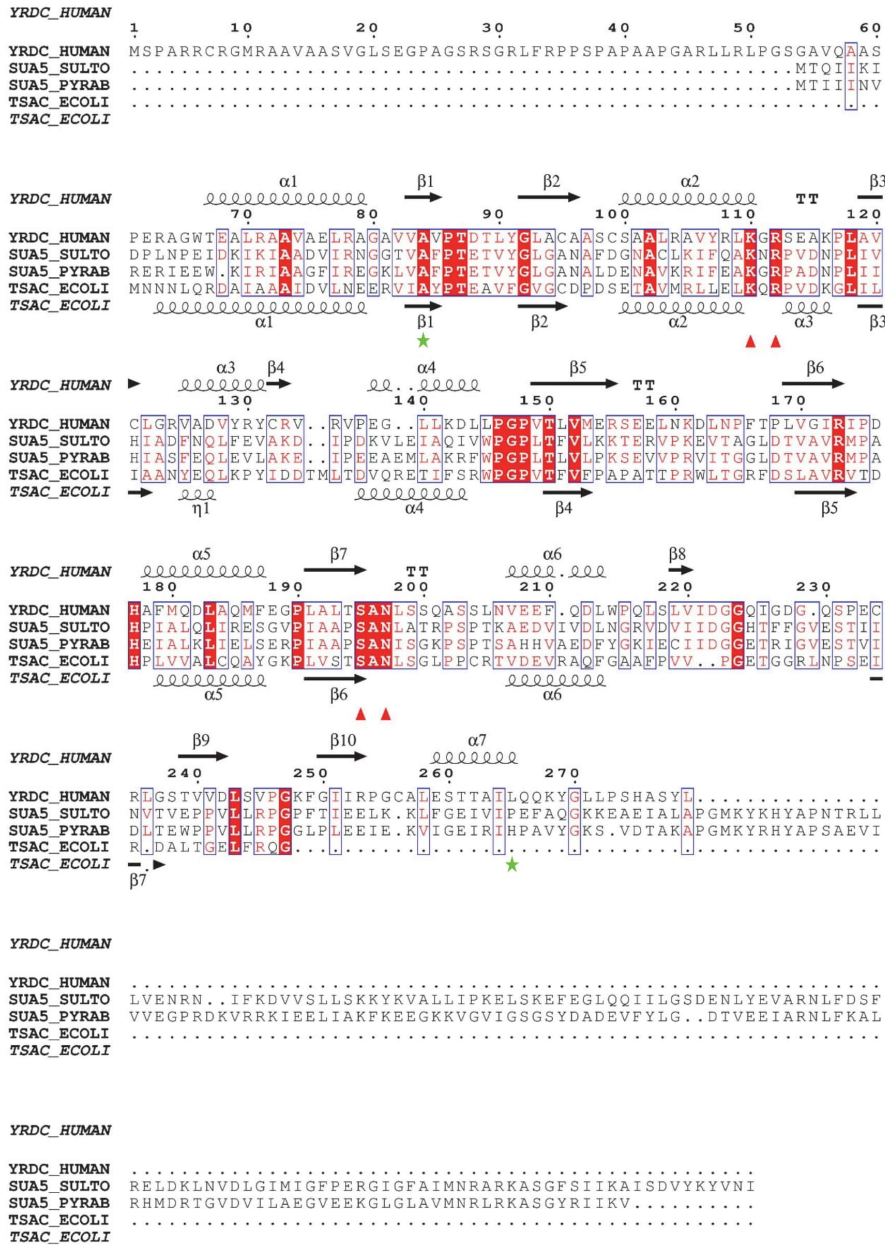
**Supplementary Information**

**Defects in t<sup>6</sup>A tRNA modification due to *GON7* and *YRDC* mutations lead to Galloway-Mowat syndrome**

Arrondel *et al.*

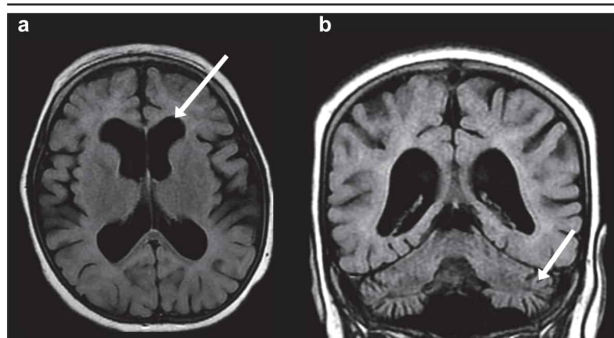


**Supplementary Figure 1 || Expression of GON7 and YRDC in cells from controls and affected individuals.** GON7 and YRDC protein expression levels were analyzed by western blot in lymphoblastoid cell lines (a) and in primary skin fibroblasts (b) from healthy children (in black) and affected individuals (in red) with the p.Tyr7\* GON7 mutation (a) or YRDC mutations (b). Tubulin and GAPDH were used as loading controls. (c-d) Relative expression of GON7 and YRDC transcripts normalized to HPRT and compared to controls (mean  $\pm$  s.e.m. of  $n=3$  experiments, unpaired t-test in (c) and Kruskal-Wallis test followed by Dunn's *post-hoc* test in (d); ns=0.2751, \* $p=0.00137$ ). Data correspond to the pool of 2 healthy individuals for the controls and 4 individuals with the p.Tyr7\*GON7 mutation for GON7 patients. For YRDC transcript expression, data of the 2 affected individuals from Family G have been pooled (G.II).

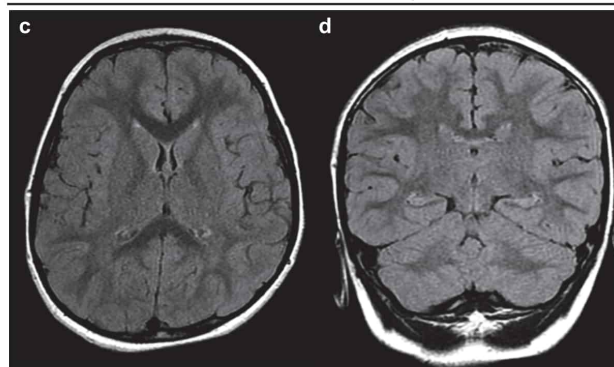


**Supplementary Figure 2 | Structure based sequence alignment of YRDC/Sua5 sequences:** human YRDC, Sua5 from *Sulfolobus tokadaii* and *Pyrococcus abyssi*, *Escherichia coli* YrdC (TsaC). Crystal structure for the latter three are known. Sua5 proteins have an extra domain that is not present in the YrdC members of the TC-synthesizing enzyme family. Secondary structure elements extracted from the YRDC and *E.coli* YrdC structures are at the top and bottom, respectively. Red triangles represent amino acid of the active site. Positions Ala84 and Leu265 are marked by a green star.

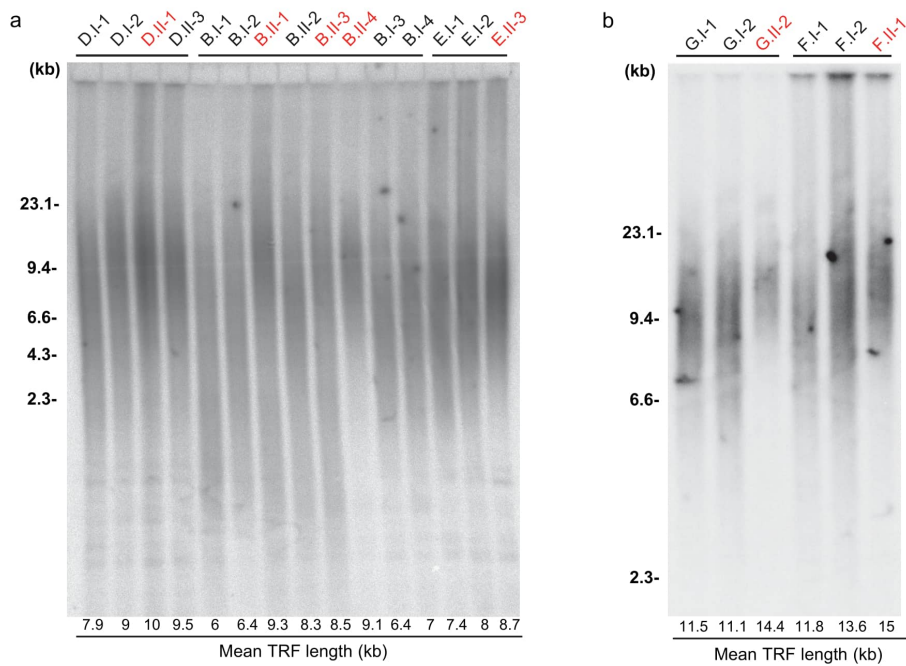
Individual B.II-4 at 5 years



Control individual at 5 years

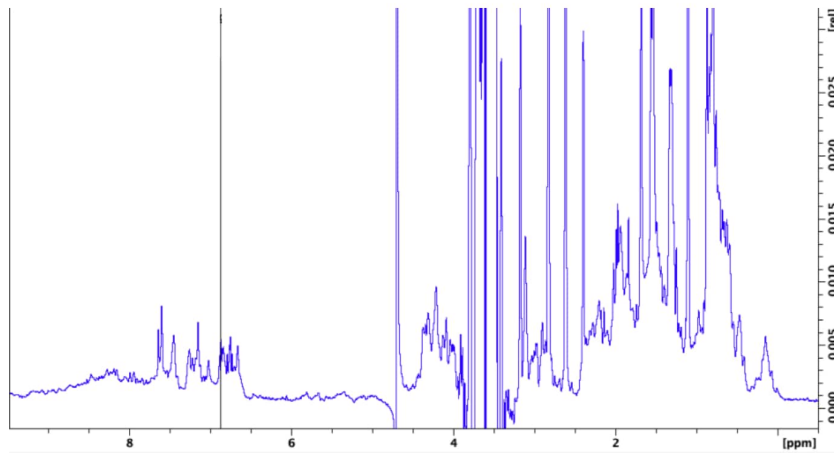


**Supplementary Figure 3 | Brain anomalies in a *GON7*-mutated individual.** Brain MRI of individual B.II-4 with *GON7* mutation at 5 years (**a-b**) compared to a control individual (**c-d**). The Flair (a) and coronal Flair (b) weighted images show ventricular dilatation and cerebellar atrophy (white arrows), respectively. (**c-d**) Brain MRI of a control individual at 5 years showing corresponding Flair (c) and coronal Flair (d) weighted images.

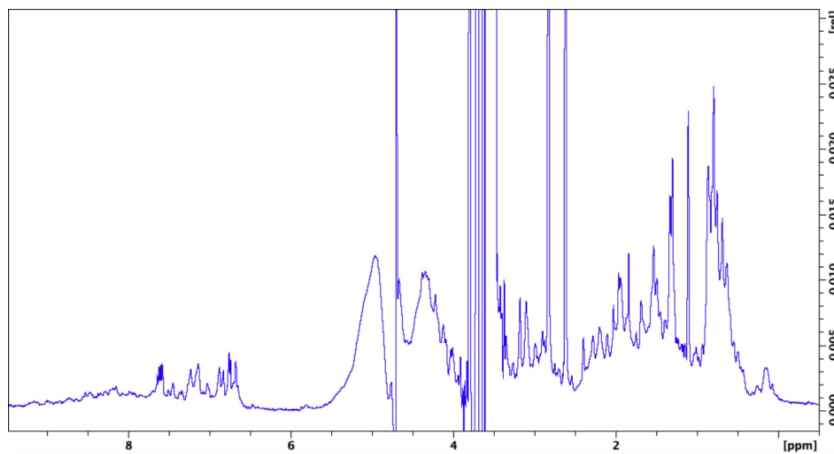


**Supplementary Figure 4 | Analysis of telomere length in individuals with mutations in *GON7* and *YRDC* and their parents.** Using the telomere restriction fragment (TRF) assay, the mean telomere lengths of whole blood cells from *GON7* (a) and *YRDC* (b) affected individuals and their healthy parents were estimated. This Southern blot shows a normal telomere length in affected children, with as expected, a telomere length higher than that of their parents. Genomic DNA was extracted from leukocytes.

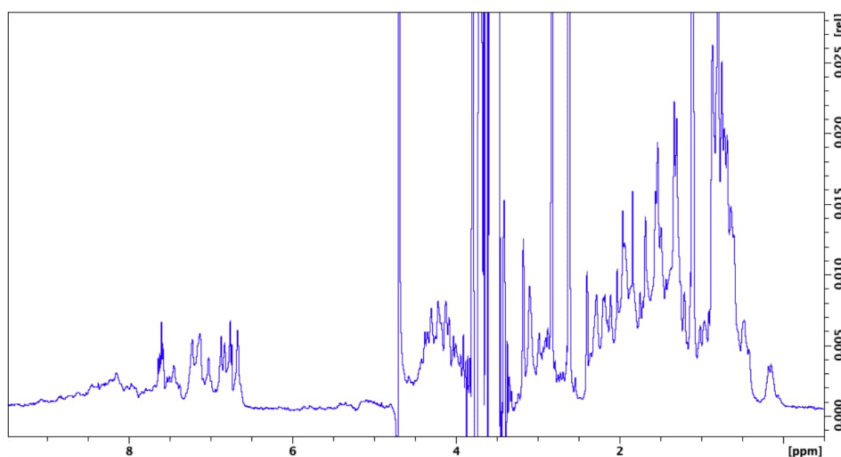
a



WT



p.Leu265del

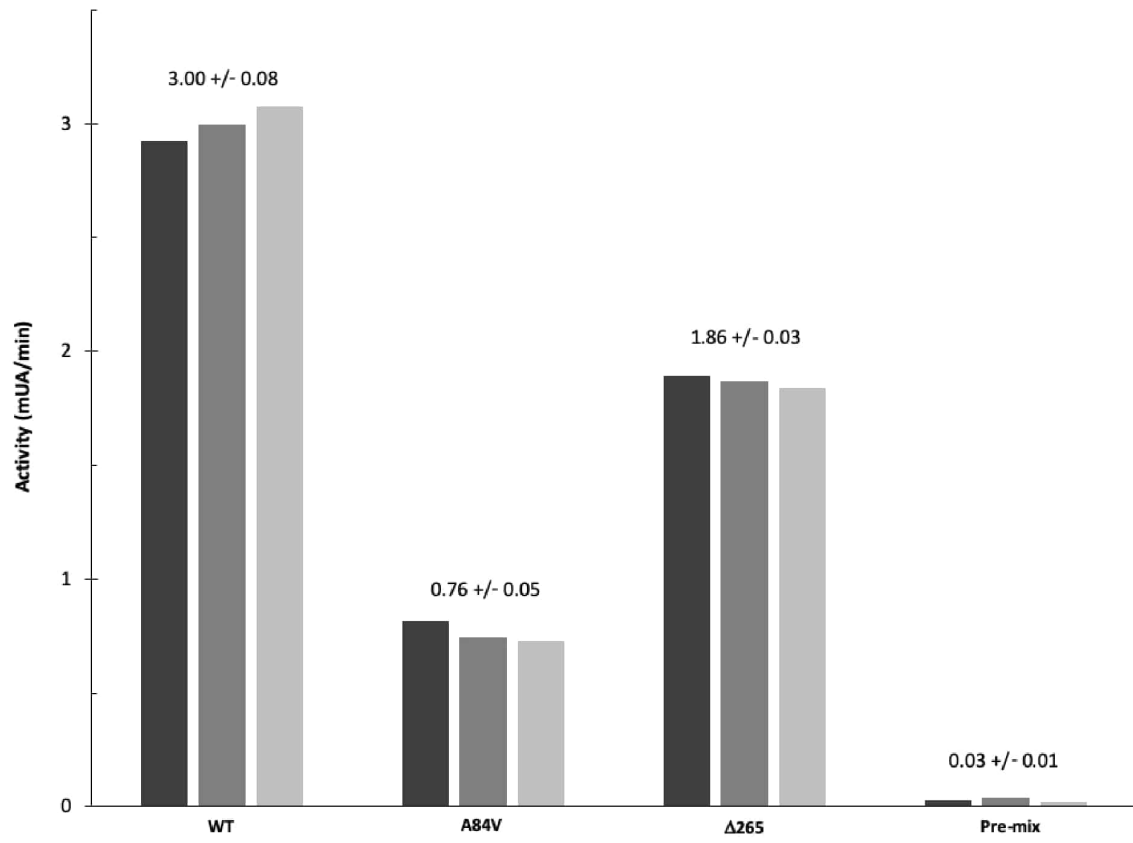


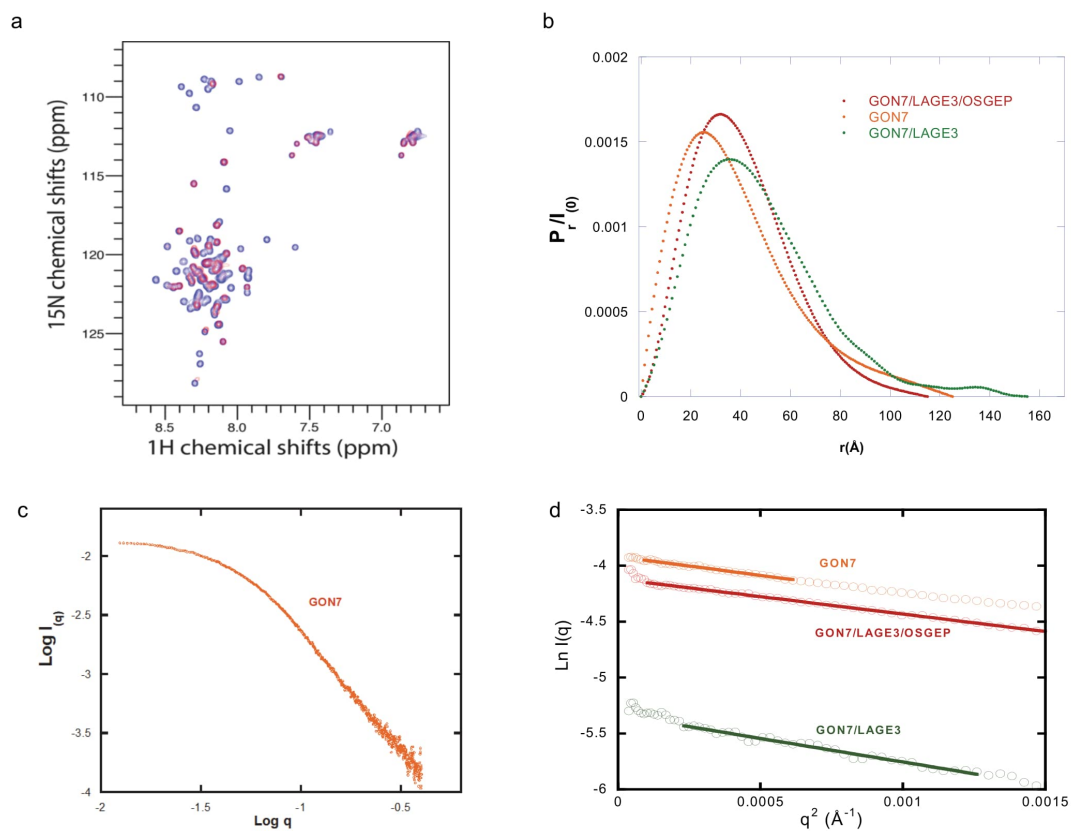
p.Ala84Val

**Supplementary Figure 5 | 1D  $^1\text{H}$  spectra and *in vitro* enzymatic activities of YRDC WT and mutants proteins.** (a) 1D  $^1\text{H}$  spectra. (b) *In vitro* enzymatic activities. Numbers above histograms indicate mean values in mUA/min from triplicate measurements ( $\pm$  standard deviation). Pre-mix: assay buffer containing all substrates and PPIase enzyme but without YRDC.

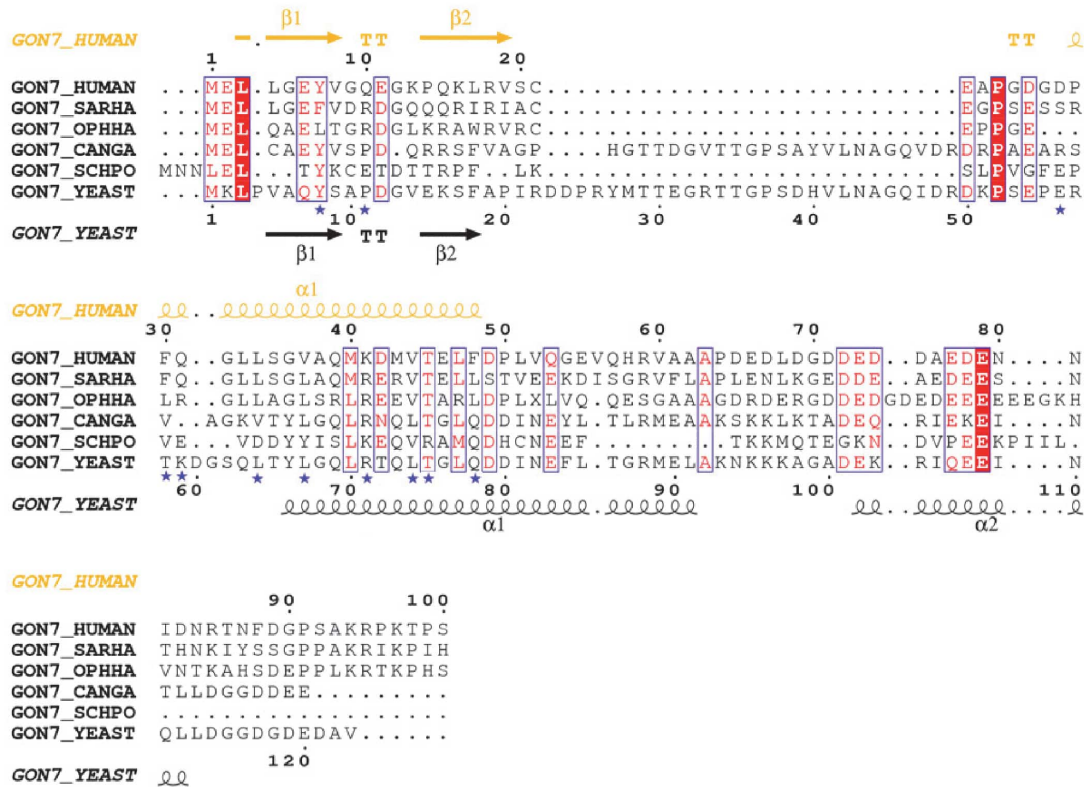


b

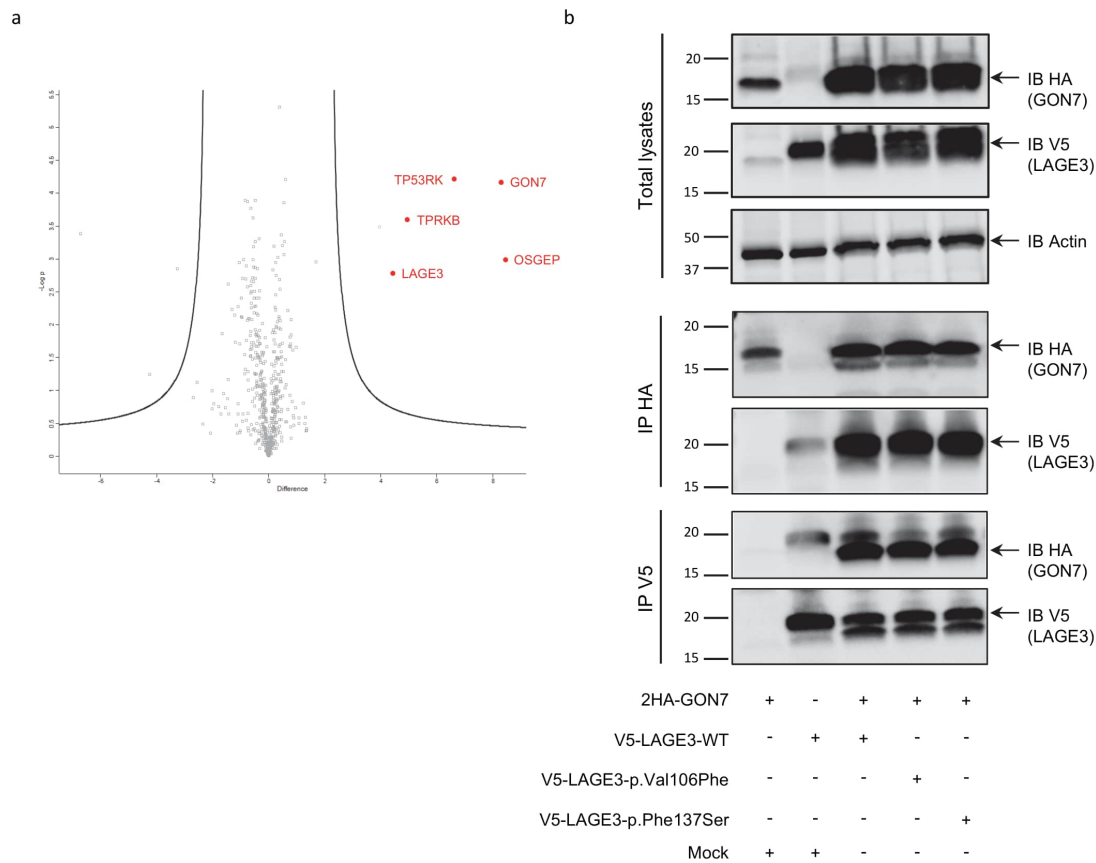




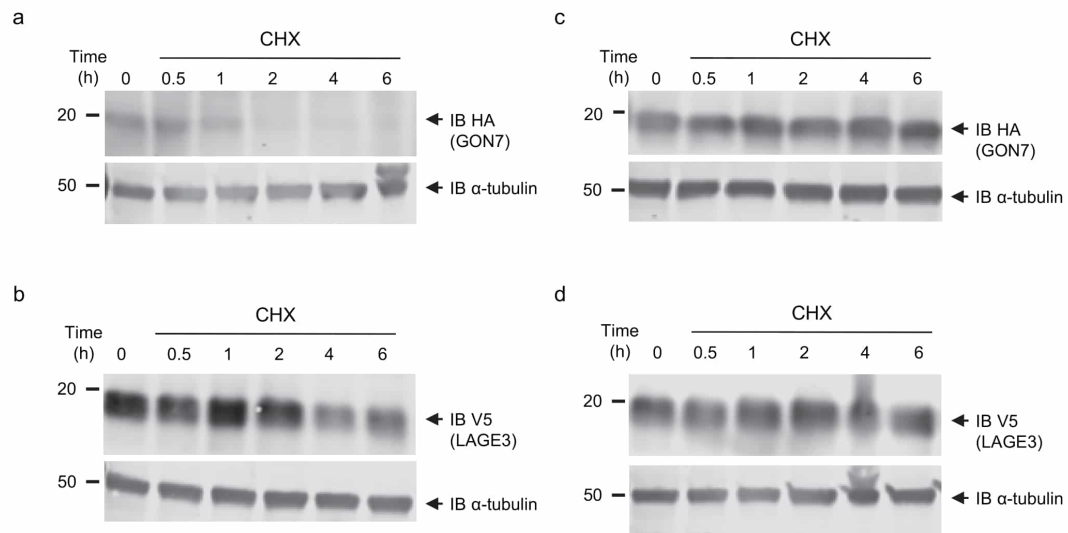
**Supplementary Figure 6 | Characterization in solution of GON7 and GON7 complexes.** (a) Overlap of 2D  $^{15}\text{N}$  SOFAST-HMQC spectra<sup>1</sup> collected on 200  $\mu\text{M}$   $^{15}\text{N}$ -labeled GON7 alone (blue spectrum) or 65  $\mu\text{M}$   $^{15}\text{N}$ -labeled GON7-unlabeled LAGE3 (red spectrum) in a buffer composed of 20mM MES pH 6.5 200mM NaCl 5mM mercapto-ethanol in 90%/10% H<sub>2</sub>O/D<sub>2</sub>O. The NMR spectra were recorded at 20 °C on a Bruker Avance III 800 MHz equipped with a TCI cryoprobe. The expression of  $^{15}\text{N}$ -labeled GON7 was achieved in the *E. coli* Gold strain in minimum media M68<sup>15</sup>N at 37°C during 4 hours. The expression of LAGE3 was achieved in the *E. coli* Gold strain in rich medium at 37°C during 3 hours. Cells were then harvested by centrifugation, suspended in 35 mL of buffer A. GON7-  $^{15}\text{N}$  and LAGE3 were purified independently as previously cited. 232 mM of LAGE3 were incubated with 156 mM of GON7- $^{15}\text{N}$  on ice. The subunits or binary complex were purified by size exclusion chromatography (Superdex 200 10-300). The purest fractions (analyzed by SDS PAGE) were pooled and concentrated up to 200  $\mu\text{M}$  and then analyzed by NMR. (b) Distance distribution ( $P_r$ ) curves for GON7 (orange) and for the GON7/LAGE3 (green) and GON7/LAGE3/OSGEP (red) sub-complexes. The normalized distance distribution was calculated by BioXTAS RAW software that uses the Bayesian method for determining the inverse Fourier transform (IFT) of a scattering profile<sup>2</sup>. (c) Log/log plot of the SAXS data for GON7. (d) Guinier plot of GON7 (orange), GON7/LAGE3 (green) and GON7/LAGE3/OSGEP (red).



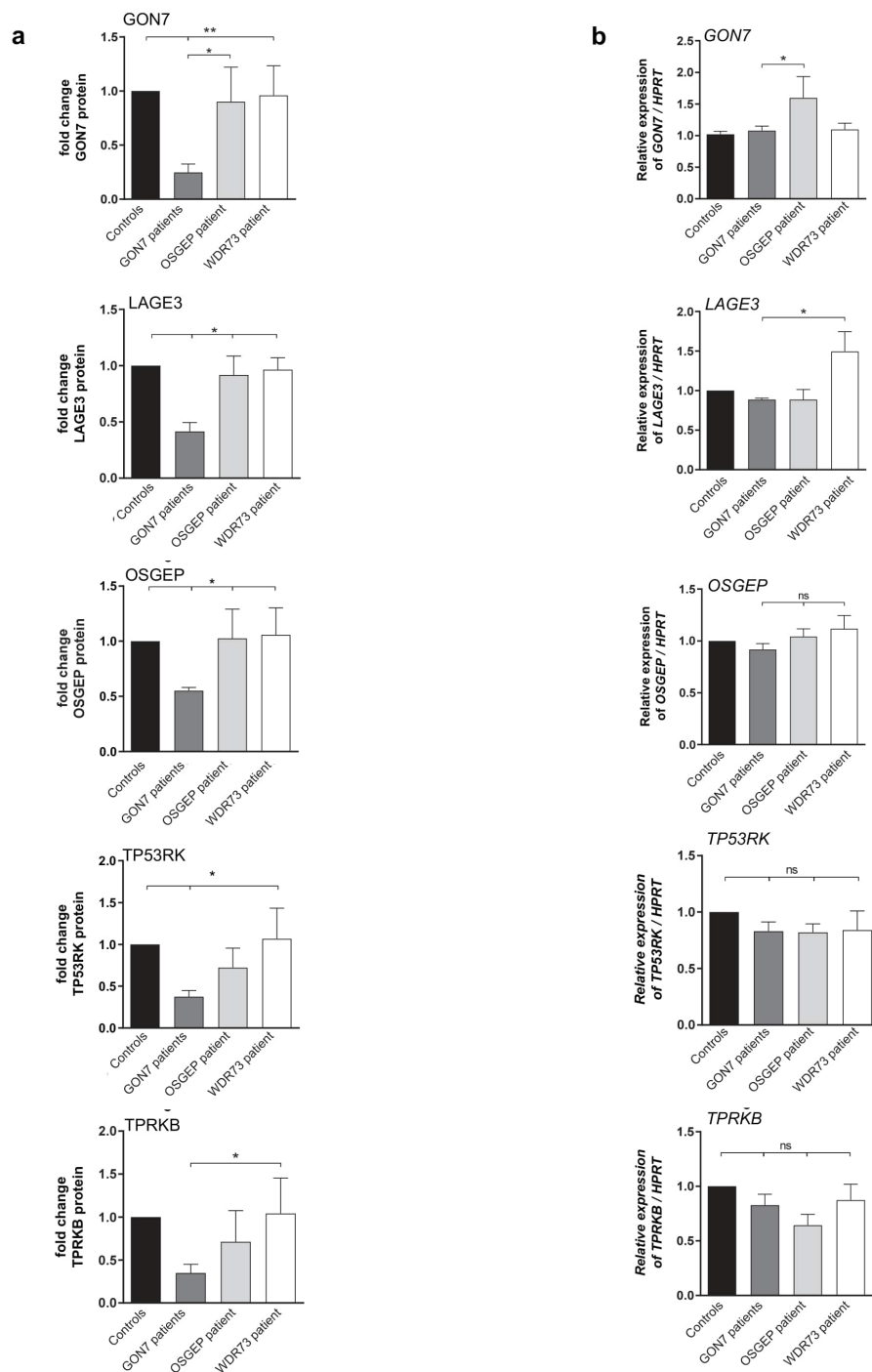
**Supplementary Figure 7 | Structure based sequence alignment between human GON7 and yeast Gon7.** Secondary structure elements were extracted from the crystal structures of GON7/LAGE3/OSGEP (top) and Gon7/Pcc1/Kae1 (bottom). Group 1 contains sequences from metazoans (*Homo sapiens*, *Sarcophilus harrisii* and *Ophiophagus hannah*) while group 2 contains sequences from fungi (*Candida glabrata*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*). Residues in contact with LAGE3 are marked by blue stars below the sequence.



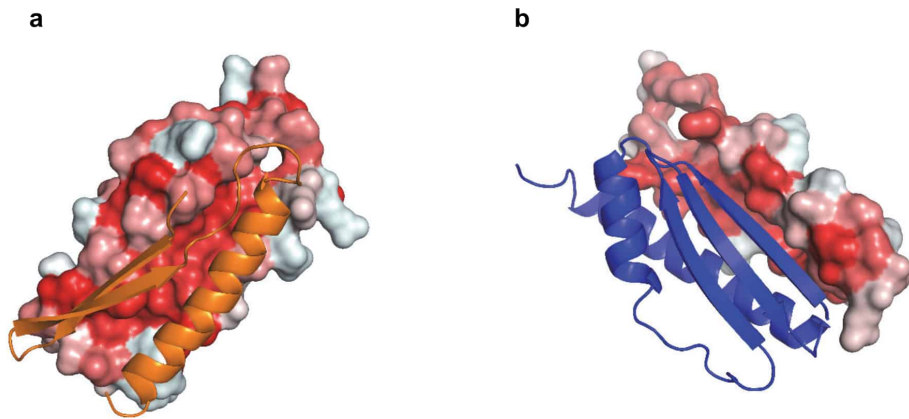
**Supplementary Figure 8 | GON7 interactions with KEOPS complex subunits in human podocytes and with LAGE3 mutants.** (a) Volcano plot representation of GON7-interacting proteins identified by mass spectrometry (nanoRSLC-Q Exactive PLUS MS) in 3 independent HA-immunoprecipitation experiments in a human podocyte cell line (empty HA-plasmid was used as control). X-axis reports the difference of the average of the logarithm of Label Free Quantification (LQF) intensities between the IP and its control; y-axis reports the negative logarithmic of t-test p value. Proteins showing a significant interaction are depicted outside of the black lines according to significance analysis (FRD=0.01, S0=2). Red dots indicate the four other human KEOPS subunits. (b) Lysates from HEK293T cells co-expressing HA-tagged GON7 and V5-tagged wild-type (WT) or mutant forms of LAGE3 were immunoprecipitated either with an anti-HA or anti-V5 antibody. Co-immunoprecipitated proteins were visualized by immunoblot (IB) using HA and V5 antibodies. Mock corresponds to empty vectors. Actin was used as a loading control.



**Supplementary Figure 9 | Cycloheximide chase experiments.** HEK293T expressing either 2HA-tagged GON7 (**a**) or V5-tagged LAGE3 (**b**) alone or co-expressing both proteins (**c** and **d**) and subjected to treatment with cycloheximide to observe protein degradation (100  $\mu$ M for 30min, 1 hour, 2 hrs, 4 hrs and 6 hrs) followed by western blotting of the cell lysates for both proteins with anti-HA (a and c) and anti-V5 (b and d) antibodies (experiments were performed 3 times).

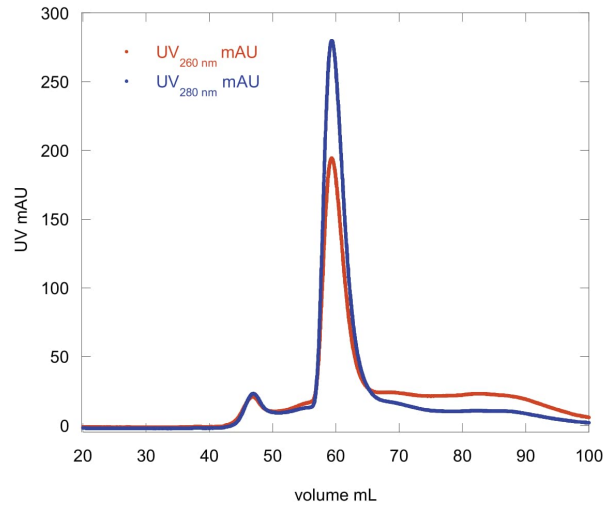


**Supplementary Figure 10 | KEOPS subunits protein and mRNA expression levels in cells from GON7 affected individuals.** Quantification of protein expression levels in (a) and transcript levels in (b) of the 5 KEOPS complex subunits. Proteins and mRNA were extracted from cell samples used in Figure 6d (mean  $\pm$  s.e.m of  $n=3$  experiments, with each experiment performed in triplicate; one-way ANOVA, Dunnett's *post-hoc* test).

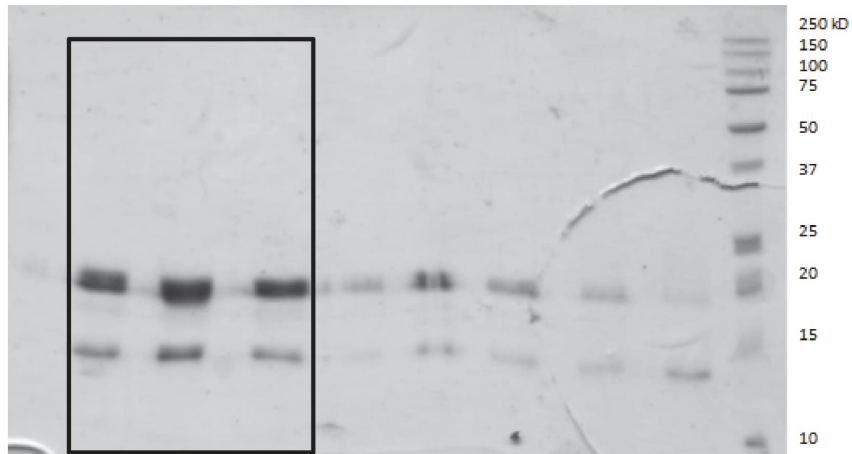


**Supplementary Figure 11 | Hydrophobic surfaces at the GON7/LAGE3 interface.** (a) Surface presentation of LAGE3, highlighting the hydrophobicity (increasing from white to red); GON7 presented in gold cartoon. (b) Surface presentation of GON7, highlighting the hydrophobicity (increasing from white to red); LAGE3 presented in blue cartoon.

a

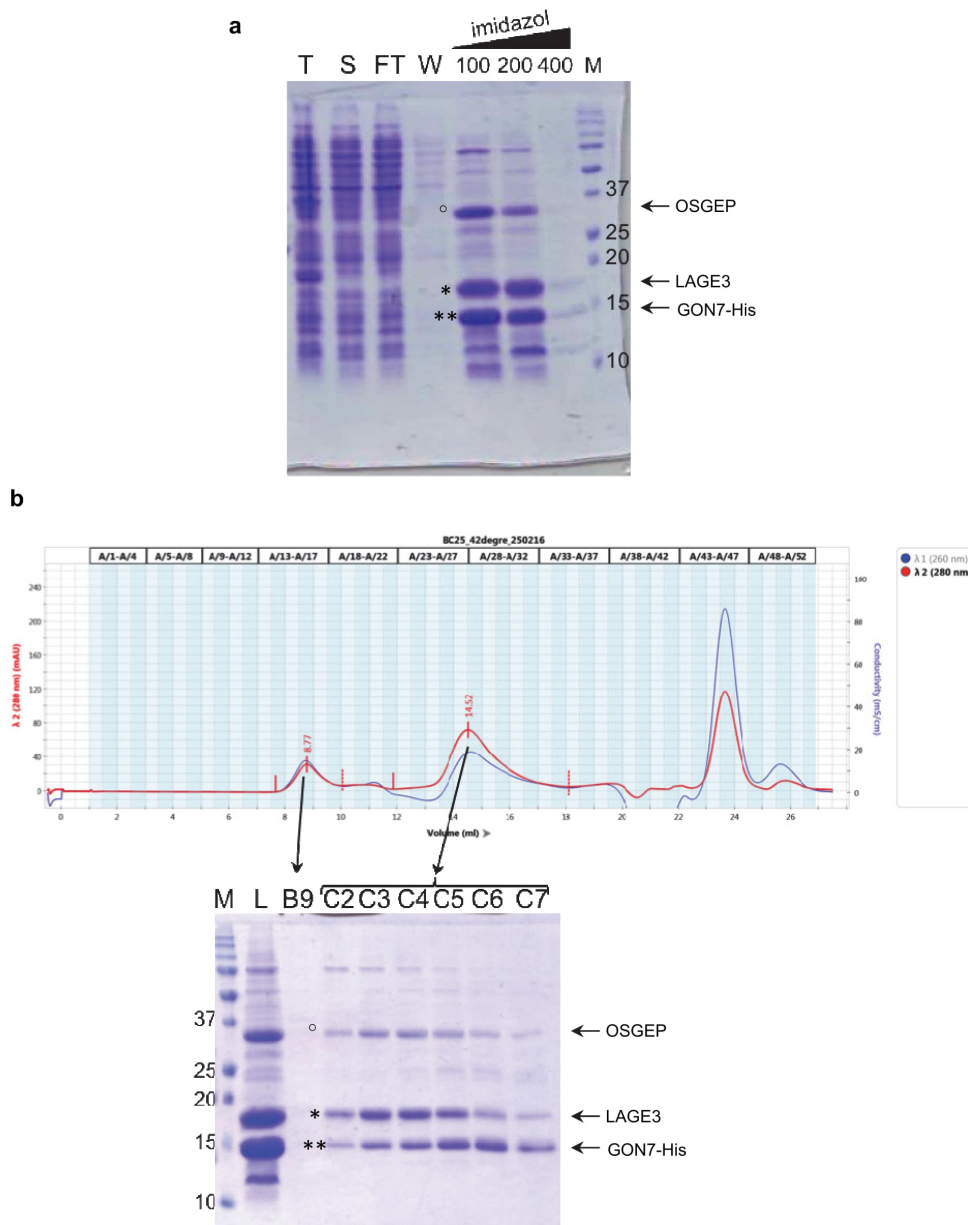


b

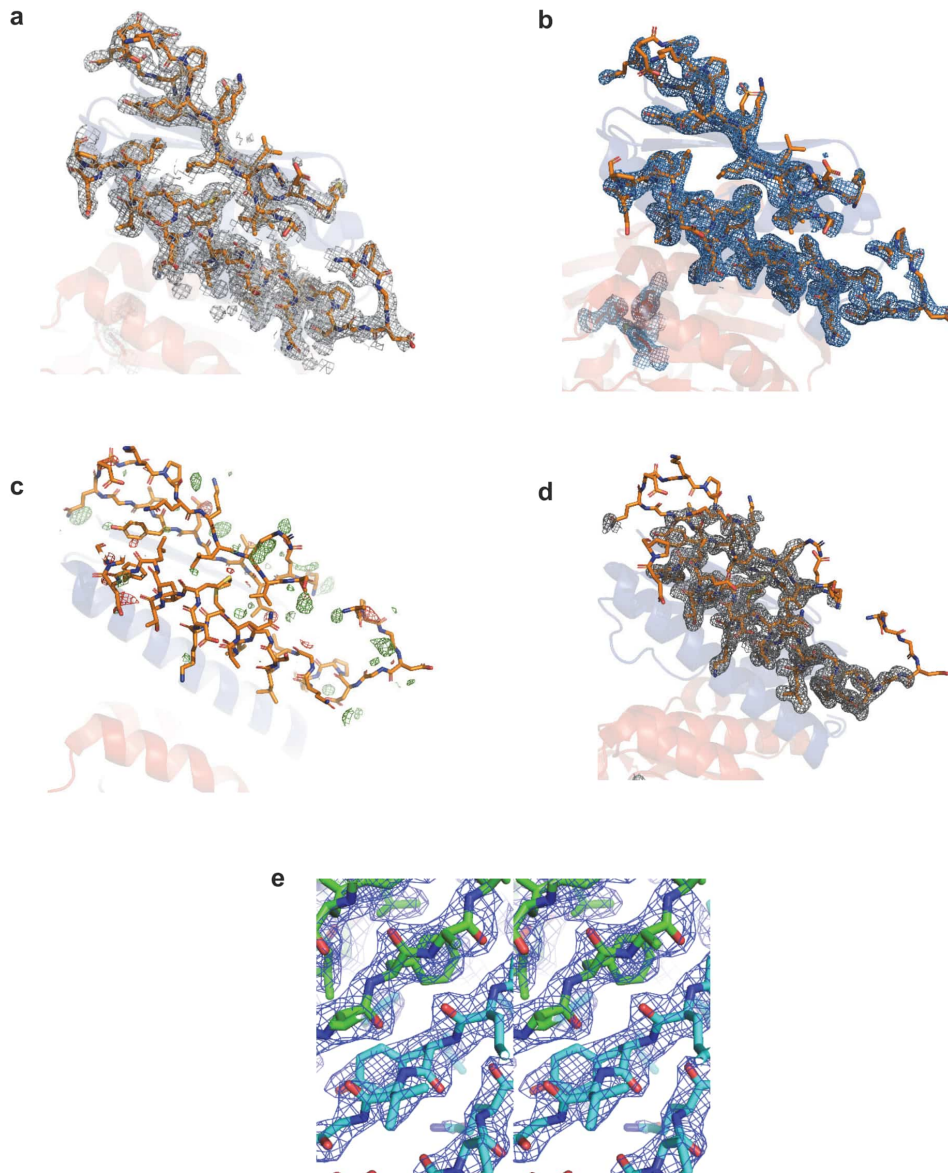


**Supplementary Figure 12 | Gel filtration analysis of GON7/LAGE3 complex. (a)** Gel filtration profile. **(b)** SDS-PAGE analysis of eluted fractions (black square: major peak fractions eluted at 59 mL)





**Supplementary Figure 13 | Co-purification of OSGEP/LAGE3/GON7-His subcomplex.**  
**(a)** SDS-PAGE analysis of  $\text{Ni}^{2+}$  chromatography fractions. T: Total whole cell extract; S: soluble cell extract; FT: Flowthrough fraction; W: Wash fraction; 100, 200, and 400: elution fraction with 100 mM, 200 mM and 400 mM imidazol, respectively. **(b)** Analytical S200 size-exclusion chromatography of the 100 mM and 200 mM imidazole fractions obtained in (a). Below chromatogram is the SDS-PAGE analysis of the 2 peaks eluted at 8.77 mL (B9) and 14.52 mL (fractions C2 to C7). Numbers: molecular weight markers in kDa. (°): OSGEP, (\*): LAGE3 and (\*\*): GON7-His.



**Supplementary Figure 14 | Calculated maps obtained from X-ray data.** (a) A composite-omit map calculated by omitting 5% at a time showed good agreement with the unrefined structure indicating the phases were not dominated by model bias. (b) the  $2mF_{obs} - DF_{calc}$  (contoured at  $1\sigma$ ) map covers protein residues 2-50 of GON7. (c)  $mF_{obs} - DF_{calc}$  difference map contoured at  $+3.0\sigma$  (green) and  $-3.0\sigma$  (red) for the same protein region. (d) Prime and switch map. (e)  $2Fo - Fc$  map of a few residues at the LAGE3 (green sticks) - GON7 (blue sticks) interface contoured at  $1\sigma$  (stereo view). The map (b) shows well-defined density around the vast majority of side chains and the difference map (c) shows no large discrepancies between the observed data ( $F_{obs}$ ) and the model ( $F_{calc}$ ).

Family Individual	Nucleotide change	Amino acid change	Exon (Zygosity, Segregation)	gnomAD (AC/Hom/AN)	Gender	Ethnic origin	Parental Consanguinity	Age of proteinuria (Age of ESRD) Age at Death	Extrarenal Manifestations	Biopsy (at age)	Therapy and response	Comments
<b>GON7 (5 families, NM_032490.4)</b>												
<b>Family A</b>												
II-3 (N1560)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	M	Algeria	Yes	2 mon (6 yrs) 33 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay, choreoathetotic encephalopathy. <u>Cranial imaging:</u> (12 yrs) moderate but progressive cerebellar atrophy, cortical atrophy, ventricular dilation, thin corpus callosum. <u>Facies:</u> dysmorphism: narrow forehead, large low-set ears, low nasal bridge, pinched nose, almond shaped eyes, hypotelorism, coarse hair. <u>Skeletal:</u> abnormal skull shape. <u>Other:</u> hiatal hernia with gastro-oesophageal reflux	FSGS (5 yrs)	SRNS	One affected sister died at 3 yrs of age. No available DNA.
II-4 (N1561)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	F	Algeria	Yes	4 mon (4 yrs) Alive with a functioning graft at 23 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> moderate cerebral and cerebellar atrophy, thin corpus callosum. <u>Facies:</u> narrow forehead, large low-set ears, low nasal bridge, pinched nose, almond shaped eyes, hypotelorism, coarse hair, hypertelorism. <u>Skeletal:</u> abnormal skull shape mammary. <u>Other:</u> hiatal hernia with gastro-oesophageal reflux	ND	No IS used	
II-6 (N1562)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	M	Algeria	Yes	5 mon (6 yrs) Alive with a functioning graft at 19 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> moderate cerebral and cerebellar atrophy, myelination delay. <u>Facies:</u> narrow forehead, large low-set ears, low nasal bridge, pinched nose, almond shaped eyes, hypotelorism. <u>Other:</u> hiatal hernia with gastro-oesophageal reflux	ND	No IS used	
<b>Family B</b>												
II-1 (N969)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	M	Algeria	No but same region	21 mon (32 mon) 15 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> (2 yrs): moderate cerebral and cerebellar atrophy, atrophic corpus callosum, myelination delay.	DMS	SRNS	
II-3 (N768)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	F	Algeria	No but same region	12 mon (33 mon) 3 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> (2 yrs): moderate cerebral and cerebellar atrophy, dilated lateral ventricles, myelination delay.	DMS	SRNS Ciclo-resistant NS	
II-4 (N1552)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	F	Algeria	No but same region	24 mon (30 mon) Alive with a functioning graft at 15 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> - 3 yrs: cerebral and cerebellar atrophy, myelination delay. - 5 yrs: stable cerebellar and cortical atrophy, ventricular dilation, thin corpus callosum, abnormal myelination.	DMS	No IS used	
<b>Family C</b>												
II-1 (N3290)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	M	Algeria	yes	12 mon (N/A) Alive with normal renal function and heavy Pu at 9 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> (6 mon): moderate cerebral atrophy.	ND	No IS used	
II-2 (N2516)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	1/0/249,098	F	Algeria	yes	12 mon (6 yrs) Alive with a functioning graft at 15 yrs	<u>Brain:</u> Secondary microcephaly, developmental delay. <u>Cranial imaging:</u> (6 yrs): severe cerebellar atrophy, cortical atrophy, ventricular dilation, thin corpus callosum. <u>Other:</u> sensorineural hearing loss	FSGS (18 mon)	SRNS Ciclo-resistant and mycophenolate mofetil-resistant NS	

II-3 (N3231)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	10/249,098	F	Algeria	yes	8 mon (N/A) Alive with normal renal function and mild Pu at 7.5 yrs	<u>Brain</u> : Secondary microcephaly, developmental delay. <u>Cranial imaging</u> : ND	ND	No IS used ACE inhibitors and ARBs	
<b>Family D</b> II-1 (N3780)	c.21 C>A	p.Tyr7*	Exon 1 (Hom), m(het), p(het)	10/249,098	M	Algeria	yes	5 yrs (N/A) Alive with normal renal function and heavy Pu at 7.5 yrs	<u>Brain</u> : Secondary microcephaly, developmental delay no visual contact. <u>Cranial imaging</u> : ND	ND	No IS used	An affected brother died when 5 years old, 3 mon after SRNS diagnosis and 2 mon after dialysis start
<b>Family E</b> II-3 (DM)	c.19dup	p.Tyr7Leufs*16	Exon 1(Hom); m(het), p(het);	0	M	Arab	yes	2 yrs (3 yrs) Alive with a functioning graft at 6 yrs	<u>Brain</u> : Microcephaly, neurodevelopmental delay <u>Cranial imaging</u> (16 mon): cerebellar hypoplasia, cortical atrophy, ventricular dilatation, thin corpus callosum, myelination delay. <u>Facies</u> : facial dysmorphism	FSGS	SRNS	health sibling 1 (het); health sibling 2 (WT);
<b>YRDC (2 families, NM_024640.3)</b>												
<b>Family F</b> II-1 (N4261)	c.251 C>T  c.721_724 del	p.Ala84Val §  p.Val241Ilefs*72	Exon 1 (het) #  Exon 4 (Het) #	0  1/0/251,484	F	European	No	2 mon (11 mon) 15 mon	<u>Brain</u> : hypotonia (4 mon), secondary microcephaly (6 mon), myoclonia, <u>Cranial imaging</u> : - 5 mon: normal - 11 mon: progressive major cerebellar and cortical atrophy and very marked myelination defects. <u>Facies</u> : facial dysmorphism. <u>Other</u> : congenital hypothyroidism	DMS (11 mon)	SRNS Ciclo-resistant	
<b>Family G</b> II-1 (NG_010_02)	c.794_796 del	p.Leu265del	Exon 4 (Hom), m(het), p(het)	0	M	European	Yes	1 mon (1.5 mon) 1.5 mon	<u>Brain</u> : primary microcephaly. <u>Cranial imaging</u> : ND <u>Skeletal</u> : arachnodactyly. <u>Facies</u> : facial dysmorphism. <u>Other</u> : congenital hypothyroidism.	DMS and FPE (1.5 mon)	No IS used	
II-2 (NG_010_01)	c.794_796 del	p.Leu265del	Exon 4 (Hom), m(het), p(het)	0	F	European	Yes	Birth (3 mon) 3 mon	<u>Brain</u> : primary microcephaly. <u>Cranial imaging</u> (11 mon): frontal simplified gyral pattern, left thalamic infarction, myelination delay. <u>Facies</u> : facial dysmorphism. <u>Skeletal</u> : arachnodactyly. <u>Other</u> : congenital hypothyroidism.	ND	No IS used	

**Supplementary Table 1. Mutations in GON7 and YRDC genes in 14 individuals of 7 families with Galloway-Mowat syndrome.**  
ESRD, end-stage renal disease; FSGS, focal segmental glomerulosclerosis; DMS, diffuse mesangial sclerosis; FPE, foot process effacement; SRNS, steroid resistant nephrotic syndrome; NS, nephrotic syndrome; IS, immunosuppression; ACE, angiotensin converting enzyme; ARBs, Angiotensin receptor blockers; Pu, proteinuria; yrs, years; mon, months; AC, allele count; AN, allele number; Hom, homozygous; het, heterozygous; WT, wild-type; p, paternal; m, maternal; N/A, not applicable; ND, not done. § PolyPhen2 (0.709), SIFT (0.03) and Mutation taster (disease causing: 1) scores. # Each mutation has been inherited from a different parent. In addition, none of the variants described in our study was present in the homozygous or compound heterozygous state in our in-house database (n>16,000 individuals both from European and Northern African ancestry).



plasmid name	sequence
>BC47_YRDC-WT	CCATGGGTATCAACATCAACATCAAGAACTGTACTTCCAGGGCCATATGCGGCGCGGTGGCGGCCAGCGTGGGCTGAGCGAAGTCCGGCCGGCAGCCGCAACGGTCTTATCCGCCCGCGA GCCCGCCCCGCGCCCGCGCGCCCGTCTTTACGCTGCCGGTAGCGGCGCGCTGCAAGCCCGGAGCCCGGAACGCGCGCTGGAACGAGGCGCTGCTGCCCGCTGGCCGAACTCCGCGCCGGG CCGTGGTGGCGTCCGCAACGATACGCTGTACGGCTGGCTGCGCGCGCAAGCTCTCGGCGCCCTGCGCGCGGTGTACCGCTGAAAGGTGCGACGGAAACCAACCGCTGGCCGTTGCTTAGGCGCG TGCCCCATGTCTACCGCTACTGCGGTGGCGGTGCCGAGGGCTTACTGAAAGACTTACTGCCGGTCCGGTGAACCTGGTGAATGGAAACGCAAGGAACTGAAACAAAGACTGAAACCGTTACGCG TTAGTGGGCAATCGTATCCGATCATGCTTTATGCAAGACTGGCGCAGATGTTTGAAGTCCGTTAGCCCTGACCAAGCCCAACTGAGCAAGCCAGCTCGTGAATGCGAAGAGTTCAGGAT CTGTGGCGCAGTTAAGCTGGTTATTGATGGTGGCAATTGGTATGGCCAGAGCCCGAATGTCGCTGGGCAAGCCGTGGTTGTTATCTGTGCCGGTAAATTTGGCATCATTGTCGGGCTGC GCCCTGAAAGCACCAAGCCATCTGCAACAGAAATACGGCTGTTACGAGCCATGCGAGCTACTGTAAGGATCC
>BC48_YRDC-A84V	CCATGGGTATCAACATCAACATCAAGAACTGTACTTCCAGGGCCATATGCGGCGCGGTGGCGGCCAGCGTGGGCTGAGCGAAGTCCGGCCGGCAGCCGCAACGGTCTTATCCGCCCGCGA GCCCGCCCCGCGCCCGCGCGCCCGTCTTTACGCTGCCGGTAGCGGCGCGCTGCAAGCCCGGAGCCCGGAACGCGCGCTGGAACGAGGCGCTGCTGCCCGCTGGCCGAACTCCGCGCCGGG CCGTGGTGGTGTCCGCAACGATACGCTGTACGGCTGGCTGCGCGCGCAAGCTCTCGGCGCCCTGCGCGCGGTGTACCGCTGAAAGGTGCGACGGAAACCAACCGCTGGCCGTTGCTTAGGCGCG TGCCCCATGTCTACCGCTACTGCGGTGGCGGTGCCGAGGGCTTACTGAAAGACTTACTGCCGGTCCGGTGAACCTGGTGAATGGAAACGCAAGGAACTGAAACAAAGACTGAAACCGTTACGCG TTAGTGGGCAATCGTATCCGATCATGCTTTATGCAAGACTGGCGCAGATGTTTGAAGTCCGTTAGCCCTGACCAAGCCCAACTGAGCAAGCCAGCTCGTGAATGCGAAGAGTTCAGGAT CTGTGGCGCAGTTAAGCTGGTTATTGATGGTGGCAATTGGTATGGCCAGAGCCCGAATGTCGCTGGGCAAGCCGTGGTTGTTATCTGTGCCGGTAAATTTGGCATCATTGTCGGGCTGC GCCCTGAAAGCACCAAGCCATCTGCAACAGAAATACGGCTGTTACGAGCCATGCGAGCTACTGTAAGGATCC
>BC49_YRDC-del165	CCATGGGTATCAACATCAACATCAAGAACTGTACTTCCAGGGCCATATGCGGCGCGGTGGCGGCCAGCGTGGGCTGAGCGAAGTCCGGCCGGCAGCCGCAACGGTCTTATCCGCCCGCGA GCCCGCCCCGCGCCCGCGCGCCCGTCTTTACGCTGCCGGTAGCGGCGCGCTGCAAGCCCGGAGCCCGGAACGCGCGCTGGAACGAGGCGCTGCTGCCCGCTGGCCGAACTCCGCGCCGGG CCGTGGTGGCGTCCGCAACGATACGCTGTACGGCTGGCTGCGCGCGCAAGCTCTCGGCGCCCTGCGCGCGGTGTACCGCTGAAAGGTGCGACGGAAACCAACCGCTGGCCGTTGCTTAGGCGCG TGCCCCATGTCTACCGCTACTGCGGTGGCGGTGCCGAGGGCTTACTGAAAGACTTACTGCCGGTCCGGTGAACCTGGTGAATGGAAACGCAAGGAACTGAAACAAAGACTGAAACCGTTACGCG TTAGTGGGCAATCGTATCCGATCATGCTTTATGCAAGACTGGCGCAGATGTTTGAAGTCCGTTAGCCCTGACCAAGCCCAACTGAGCAAGCCAGCTCGTGAATGCGAAGAGTTCAGGAT CTGTGGCGCAGTTAAGCTGGTTATTGATGGTGGCAATTGGTATGGCCAGAGCCCGAATGTCGCTGGGCAAGCCGTGGTTGTTATCTGTGCCGGTAAATTTGGCATCATTGTCGGGCTGC GCCCTGAAAGCACCAAGCCATCTGCAACAGAAATACGGCTGTTACGAGCCATGCGAGCTACTGTAAGGATCC

**Supplementary Table 3 | Sequences for expression of YRDC-WT and mutants for NMR-1D spectra and enzymatic activity assays.**

	GON7	GON7-LAGE3	GON7-LAGE3-OSGEP
<b>Data collection parameters</b>			
Instrument	SWING (SOLEIL)		
Detector	CCD-based AVIEX		
Beam geometry	0.8 mm x 0.15 mm		
Wavelength [Å]	1.0		
q-range [Å <sup>-1</sup> ]	0.007 < q < 0.50		
Absolute scaling	Comparison with scattering from pure H <sub>2</sub> O		
Exposure time [s]	2		
Temperature [K]	283		
SEC-SAXS column	Biosec 3 Agilent		
Loading concentration [mg/mL]	6.7	0.7 4.2	1.48
Injection volume [µL]	65		
Solvent ( solvent blanks taken from SEC flowthrough prior to elution of protein)	20 mM MES, 200 mM NaCl, 5 mM 2-mercaptoethanol		
Flow rate [mL.min <sup>-1</sup> ]	0.3		
Extinction coefficient [A <sub>280</sub> 0.1% (w/v)]	0.238	0.344	0.563
Partial specific volume [cm <sup>3</sup> .g <sup>-1</sup> ] <sup>a</sup>	0.7085	0.7155	0.7321
Average Concentration [g.L <sup>-1</sup> ] in combined data frames	~1.50	~-0.10	~-0.35
<b>Software used for SAXS data reduction and analysis</b>			
Foxtrot	Swing in-house software		
PRIMUS	ATSAS 2.8 suite, Franke et al. J. Appl. Cryst. 2017		
BUNCH	Petoukhov & Svergun, D.I. Biophys J.2005		
<b>Structural parameters</b>			
I(0) Guinier [cm <sup>-1</sup> ]	0.01988±0.00007	0.00483±0.00004	0.01625±0.00002
R <sub>g</sub> Guinier [Å]	31.8±0.4	35.6±0.5	30.6±0.1
qR <sub>g</sub> -range	0.30-0.79	0.54-1.26	0.31-1.29
I(0) p(r) [cm <sup>-1</sup> ]	0.01988±0.00005	0.00493±0.00006	0.01634±0.00003
R <sub>g</sub> p(r) [Å]	32.4±0.2	38.1±0.8	31.4±0.1
q-range [Å <sup>-1</sup> ]	0.0096 - 0.30	0.0152 - 0.40	0.01 - 0.40
D <sub>Max</sub> [Å]	125	155	115
<b>Molecular mass determination</b>			
Mol . L <sup>-1</sup> -1MM <sub>sequence</sub> [kDa] <sup>b</sup>	12.59	29.05	62.88
MM <sub>I(0)/c</sub> [kDa]	14.7	56.5	58.5
MM <sub>bayesian</sub> [kDa] <sup>c</sup>		65.45	59.5
Credibility interval		63.1-75.3	57.5-64.5
<b>BUNCH analysis</b>			
Bunch Model number			10
q-range [Å <sup>-1</sup> ]			0.01 - 0.40
χ <sup>2</sup> (Crysol)			0.3
SAS BDB ID	SASDFK8	SASDFM8	SASDFL8

<sup>a</sup> Calculated from the sequence using the program sednterp.

<sup>b</sup> The calculated masses were derived from the sequences.

<sup>c</sup> Molecular mass M obtained using the bayesian inference approach which combines four concentration independent MM estimators<sup>3</sup>. This mass determination does not depend on the value of the protein concentration c and provides a useful complement to the derivation of M from the I(0)/c value . Nevertheless this method is not suitable for strongly unstructured protein due to the plateau of I(q)q<sup>2</sup> and the I(0)/c method must be preferred in the case of GON7.

#### Supplementary Table 4 | SAXS results and data collection parameters.

<b>Data collection</b>	
Wavelength (Å)	0.978570
Space group	P4 <sub>3</sub>
Cell (Å, °)	a=105.12 b=105.12 c= 51.35 α=β=γ= 90
Resolution	46.14 - 1.95 (2.06 – 1.95)
R meas (%)	7.1 (118.8)
CC ½ (%)	99.9 (58.5)
I/σ(I)	15.97 ( 1.38)
Completeness (%)	99.93 (99.88)
<b>Refinement statistics</b>	
Unique reflections	41188 (4071)
Rwork / Rfree	0.1809 / 0.2106
Total atoms: protein / solvent	3845 / 223
Average B-factors (Å <sup>2</sup> )	54.83
RMSD bonds (Å) / RMSD angles (°)	0.014 / 1.62
Ramachadran favored (%)	97.38
Ramachadran allowed (%)	2.62
Ramachadran disallowed (%)	0

**Supplementary Table 5 | Crystallographic data collection and structure refinement statistics.**



Gene	Gene accession number	shRNA target sequence
<i>GON7</i>	NM_032490.5	CAAAGATAGTAAAGGCTTAGA
<i>LAGE3</i>	NM_006014.4	CTTTCTTGACCAGCTTCCCT
<i>OSGEP</i>	NM_017807.3	GGATTAACCTCCCAGGATATC
<i>YRDC</i>	NM_024640.4	CAAGACTTGGCTCAGATGTTT

**Supplementary Table 6 | shRNA targets for gene silencing**

Gene	Forward primer	Reverse primer
<i>GON7</i> qPCR	5'-CAGACGAGGACTTGGACGGTG-3'	5'-TTGGCCGTTTTGCAGATGGT-3'
<i>LAGE3</i> qPCR	5'-GGGTGGTTGGGAAGGATCTC-3'	5'-GATGACGGAAATTCGGAGCA-3'
<i>OSGEP</i> qPCR	5'-GTGATTGCATACTCGGAACA-3'	5'-TGGACTTGGGTCGTTAGAAA-3'
<i>YRDC</i> qPCR	5'-GGGCTCCTGAAAGACCTACT-3'	5'-GGAATCCGAATGCCTACAAG-3'
<i>TP53RK</i> qPCR	5'-CTGGAATATCTGCCCCAGTTG-3'	5'-CACTGAGCCTTCAATTTCTTCCAT-3'
<i>TPRKB</i> qPCR	5'-CATGGAAGGCACCATCGAT-3'	5'-CTGCCACAAGTATCTGAAATGGA-3'
<i>YRDC</i> Exon 1	5'-GTCGTGGTCCCTTTAAGCTG-3'	5'-AAGCCTGTCACCGGAAAC-3'
<i>YRDC</i> Exon 2 & 3	5'-GATTCCTTTTCTACGTCTGCC-3'	5'-CATCCCAGTGTACTATCAATGC-3'
<i>YRDC</i> Exon 4 & 5	5'-GCCTTTGATGACTGCGT-3'	5'-ATAGTATCCAGCACCAGGTCTT-3'
<i>GON7</i> Exon 1	5'-AGTGCCAGGAACCAATGACA-3'	5'-GGTCACTGCAGCACCCTCTC-3'
<i>GON7</i> Exon 2	5'-GCAGGTTTTATTAATCCCTA-3'	5'-CTCTTTGACAAAATTGTCCC-3'

**Supplementary Table 7 | sequence primers for qRT-PCR and genomic sequencing (exons)**

Vector name	Sequence
pET21a-LAGE3 <sub>his</sub> TEV <sub>op</sub> (cloned between <i>NdeI</i> and <i>EcoRI</i> )	CATATGCATCACCATCACCATCAGGAAACCTGTATTTTCAGGGCCGTGACGGG ACGCAGACCGAGGCGGTGGCCCTGACGGCGGGAGCCCGCGTGGCCACAGCTG CGCGGGGGGTGGACACCGCCGAGCTTCGGCCGGTGGCGCTCCGCCAGCCAC CGCCGGGTCCGGCCGTGACCGCCGCTCTGCGCCCGCGGGTCACTATGCGCC CGCACATCTTACCCTGAGCGTGGCAATCCCGACCCCATTTGAGGCGGAAATCGC CCACGTTCCCTGGCACAGATGCGAGCCGACCACGTTGGTTGGGAAGGAT CTGACCGTGGTGGCCGTATCCTGGTGGTGGTGGAAAGCTGAAGACTGTCGCC TGTGGCTATTTCCGTATCAACTTTCTTGCACGAGTTCCCTGGTGGTGGCCAC CATGCAGCGCTTGGTGGCCAGTTTCCCGCTAAGAATTC
pET24d-C14 <sub>his</sub> TEV <sub>op</sub> (cloned between <i>NcoI</i> and <i>XhoI</i> )	CCATGGGGCATCACCATCACCATCAGGAAACCTGTATTTTCAGGGCGAGCTGCT GGTGGTACGTCGGGAGGAAAGGAGCCGAGAGCTGGCTGTCTGTGAG GGCCGGGTGACGGGAGCCCATTCAGGGCCGTGGTGGCGTGGCCAGATGA AGGACATGGTAACGGAAATGTTTGCACCCACTGGTACAGGGTGAAGTGCAGCAGC CGTGGCGGGCCCTCCGGACGAGGACTTGGACGGTGGTGAAGATGATGCAGAA GATGAAAATAACATTTGTAACCGTACTAACTTCGATGGTCCATCTGCAAAACGCC CAAAAACCCGCTTAAAGCTCGAG
pET24a-BC25 <sub>hKEOPS</sub> <sub>opt</sub> (cloned between <i>NdeI</i> and <i>XhoI</i> )	CATATGCCGGCGGTGCTGGGTTTTGAAGGACGCGCAACAAAATTTGGCGTGGGG TGGTGGCGGATGGCAAAGTGTGGCGAACCCGCGTGCACCATATGTGACCGCCG GGCACCGGCTTTCTGGCGGTGATACCGCCGCGCATCACCGTGGGTTATCTCG GATCTGCTGAGGAAGCACTGACGGAAAGCGCTTAACAGCCAGGATATCGACT GCAATGCTACACAAAGCCCGGGCATGGGTGCCCGCTGGTTAGCGTGGCGGT TGTGGCCGTCAGTGGCCGAGCTGTGAATAAACCGCTGGTGGTGTGAACCC TGTATTTGGCCACATTTGAAATGGGCGCCGCTGATCACGGCGCCACCGCCGAGCC TGTATATGTGAGCGGGTAAACAGCGAGTATGCTTACAGCGAACATCGTTA TCGTATCTTTGGTGAACCATCGATAATGGCGTGGTAATGGCTGGATCGTTTT GGCGGTGTGCTGAAAATTTGAAAGCATCCGAGCCCGGGCTACACATTTGAACAGA TGGCAAAACCGGCAAAAATTTAGTTGAACTGCCGTACCCGTAAGAGGTATGGA TGTGACCTTTCCGGCATCTGAGCTTCAATGAAGATGTTGCCATCGCATGCTG GCCACGGGAGGTGTACCGGGAAGATGTTGTTTTCAGCTGCGAGGAAGCGTGT TTGCCATGCTGGTGGAGTACCGAAACCGCGGATGGCACATTTGGTCCAGGA AGCCCTGATTTGGGTGGCGTGGTGTAAATGTCCTTGCAGGAATGATGGCC ACCATGTCCAGGAAGTGGTGGCCGCTGTTTGGCACGATGAACGCTTCTGTA TTGACAATGGTGGATGATGCGCGCCGCGCTGGGAATGTTTCTGGCCGCTCA CGTACCCGCTTAAGTATTTGGCTTACCAGCGTATCGCACGGATGAAGTG GAGGTGACCTGGCGGATTAAGTCTAGGATCTCTGAATAATTTGTTAATTTA AGAAGGAGATAAACATGGCGGGCCGCGCCACACCGCGCCGATGGCGAA GAGCCGGCCCGGAAAGCGAGCGGCTGGCCGAGCCGTAACCGCAGCAGCCGCT TCTGAGCGGCTGGAACGTGGTGAACAGGGTGGCGAGCGCGGCTGTTCCGTGG CGCTTTCAGGGCCGCGGGCGGTGATCAAACCGCTTCCCGAAAGGCTACCGT CACCCGGCTGGAAGCGGCTTAGCCGCTGCGGTACGGTGCAGGAAGCCCGG CGCTGTACGCTGTCGCGCGCGGGCATTTAGCCCGCGTGTGTTTTTTTGGGA TTATGCCAGCACTGCTGTATATGGAAGAAATTTGAAGCTCGGTGACCGTTCGC GATATATTCAGAGCAGATGGAACGGAAGAAACCGCCGAGGCTGTGCAACT TAGCCAAAACCATTTGGTTCAGGTTTTAGCCCGATGCACGATGAAGACTGATTA TGGTATCTGACCAAGCAACATGCTGAAACCGCGCTGGAAAGATGAAC ATTTGTTAATGACTTTGGTCTGAGCTTCAATTTGGCCGCTGCCGGAAGATAAG GCGTGGACTATATGTTCTGAAAAGGCTTCTGAGCACCATCCGAAACCGGA AACGGTGTGGAAGCCTTCTGAAAAGCTACAGCAGCAGCTCGAAAAGCCCGT CCGGTCTGAAAATAATAGATGAAGTGGCCCTGCGCGCTGTAACCGCAGCATGG TTGGCTAACTAGAGGCTCATTGTTTAACTTTAAGAAGGAGATATAAATAGC AGCTGACCATCAGCTGGCCGTTTCCGGAATCCGCGTGAACCTTACTGTTAAT TAAAGATGTTAAAATGCGGGCGATCTCGCTGCAAAAGCGATGGAAGGCACATC GATGGTAGCTGATTAATCCGACCGTGAATTTGATCCGTTTCAGATTTGGTGG CGGCCAACAAAGCGGTTCACTTATACAACTGGTAAATGAAAACCCGACGCT GAGCCCGAAAATTTTCAACTTAAGCCGGAATAACAATATTAGCGAAGCCCTG AAAAAATTTGGTATCAGCGGAAATGACACCGATTTCTGATTTTACATTTGAAG AGGTGAAAAACAAATTAATCAGGAATACCTGATTAAGCCAAAGTGAAGGATCA GGTTTTCTGAAAATCTGCGGAAATCATGAACATTACGGAAGTGAAGAAAATTT TATAAATTAAGCTCGCAGGAAAGATTTGGCACCTGTAGATGCGATCATT GTCCATGACCAACCAAGATGCTGTAACTTAAGCTTATAAATTTGTTAAC TTAAGAAGGAGATATACCATGGCTGATGGGATCGGATCGAGCGGTGGCC GATGGCTGACGGCCGCTGGCCAGCTGCGCGCGCGCTGGATACCGCCG CGCCCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG CGCGAGCGCGCGCGCGCGCGCGGTAGCGCGCGCACATTTCACTTGGCGT CCGTTCCGACCCGCTGGAAGCGGAATCGCCAGGTTGCTGGCGCGGATG CGAACCGCACCGGTGGTGGTGAAGATCTGACCTGTGGCGCGTATCCT GGTGTGGCTGGAAGCGGAAGATTTGCGCTGTACGTATTTGGTCAACAAC TTCTGGACAGTAAAGCTGGTGGTGGCGCACATGACGCGCTTTGGTCCCGG TTAGCCGCTAAACTGCGCGCGCGCTAGCTTTGTTAACTTTAAGAAGGAGAT ATAGTTATGAACTGCTGGTGAATACGTTGGTGAAGGCAACCGCAGAAAC TGGCTGTGAGCTGTGAAGCGCGGCTGATGGCGACCCGTTCCAGGGCTGCTGAG CGCGTGGCCAGATGAAAGATATGTCAGCAATTTTCGATCCGCTGGTGGAG GGTGAAGTGCAGACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG ATGAAGATGATCGGAAGATAAATAACATTTGATAACCGTACCACTTCGATGG TCCGAGCCCAACCGCCGAAACCCCGTGGACCATCACCATCACCATTAACCTC GAG

**Supplementary Table 8 | Sequences for expression of LAGE3-his, GON7-his and OSSEP/LAGE3/GON7-his for NMR, SAXS and/or crystallogenes experiments.**

Accession	Description	Coverage	# Unique Peptides	# PSMs
Q9BXV9	EKC/KEOPS complex subunit GON7 [GON7_HUMAN]	91,51	10	212
Q9NPF4	Probable tRNA N6-adenosine threonylcarbamoyltransferase OS=Homo sapiens [OSGEP_HUMAN]	66,37	24	131
Q14657	EKC/KEOPS complex subunit LAGE3 OS=Homo sapiens [LAGE3_HUMAN]	74,13	11	54
Q9Y3C4	EKC/KEOPS complex subunit TPRKB OS=Homo sapiens [TPRKB_HUMAN]	21,03	3	4
Q96S44	TP53-regulating kinase OS=Homo sapiens [PRPK_HUMAN]	17,36	1	2

**Supplementary Table 9 | Identification of human KEOPS subunits by mass spectrometry in size-exclusion chromatography C3 fraction** (from Supplementary Figure 13b).

## Supplementary Methods

**Additional information for cell lines.** Lymphoblastoid cell lines (LCLs) were obtained by Epstein-Barr virus (EBV) transformation of B cells from peripheral blood (performed at the Necker Imagine DNA biobank, Paris, France) and were cultured in suspension at 37°C with 7% CO<sub>2</sub> in RPMI-1640 supplemented with 10% FBS, glutamine, ITS, and 50 IU/ml P/S (ThermoFisher Scientific).

**Additional information for tRNA extraction from human primary fibroblasts.** Briefly, small RNAs were extracted from 10.10<sup>6</sup> fibroblasts using Macherey Nagel's Isolation of small and large RNA kit, mixed with 2ml of R0 buffer and then loaded onto AXR-80 columns (Macherey Nagel's Genomic and total RNA purification kit). After binding to the column, tRNAs were selectively eluted using KCl gradient starting with 1ml of 450mM KCl followed by 2ml of 500mM KCl. The tRNA enriched fractions were pooled and precipitated with isopropanol. The pellet was washed two times with 70% Ethanol and resuspended in 50µl of water.

### Additional information for Mass spectrometry analysis of t<sup>6</sup>A.

*tRNA hydrolysis.* tRNA was unfolded at 95°C for 5 min and quickly incubated in a bath of water and ice for 5 min. Two units of nuclease P1 was mixed with the tRNA and incubated at 42°C for 2 hours. Then 0.01 unit of phosphodiesterase from snake venom was added and incubated for 1 hour at 37°C. Finally, the mixture was supplemented with 0.5 unit of alkaline phosphatase and incubated 1 hour at 37°C. To remove the enzymes, nucleosides were filtered onto 10kDa centrifugal units (Microcon) and filtrates were lyophilized and stored at -20°C.

*LC-MS/MS analysis of RNA nucleosides.* MS analysis was performed with a Hypersil Gold aQ column, 1.9 µm particle size, 2.1x100 mm (ThermoFisher Scientific) connected to an Acquity H-Class UPLC-XevoTQS instrument (Waters). The most intense MRM transitions used for integration and quantification were m/z 413>281 for t<sup>6</sup>A at 3 min retention time, 267>136 for adenosine at 2 min, 284>152 for guanosine at 2.3 min, 243>110 for uridine at 1.3 min and 244>112 for cytidine at 1 min.

**Additional information for Immunoprecipitation.** Briefly, 48 hrs after transient transfection, HEK293T cells were lysed in buffer composed of 50mM Tris-HCl, pH7, 500mM NaCl, 1% Triton. One mg of protein was incubated either with anti-V5 antibody for 1 hr, followed by a 30-min incubation with magnetic beads-coupled to Protein A/G or directly with magnetic beads coupled to a HA antibody for 30 minutes (Miltenyi Biotec). Immunoprecipitated proteins were isolated using µMACS<sup>®</sup> Separation Columns in a magnetic µMACS separator and subsequently eluted using 1X Laemmli buffer. Lysates and immunoprecipitated samples were subjected to immunoblot.

**Additional information for Cell-Proliferation Assay.** Podocytes were seeded in triplicate 6 days after lentiviral transduction into 96-well plates at a density of 6 000 cells per well and allowed to attach for 24 hrs. Cell proliferation was assessed with the CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (MTT) (Promega) according to the manufacturer's protocol after 1, 2, 3, 4, and 7 days of culture.

**Additional information for Apoptosis assay.** Levels of apoptosis in KD podocytes were assessed using Caspase 3/7 Green detection Reagent (C10423, ThermoFisher Scientific) according to the manufacturer's instructions. Cells were seeded in triplicates at a density of 12x10<sup>3</sup> cells per well into 96-well plates (655-892, SensoPlate 96 wells glass bottom microplates, Greiner Bio-One). After 48 hrs, cells were incubated with 2µM of the reagent in 1x PBS, 3% FBS for between 1 and 6 hrs. The Caspase-3 or 7 activity in apoptotic cells was measured with a spectrophotometer (Berthold) at an absorbance at 502nm and normalized to nuclear DAPI fluorescence intensity.

**Additional information for HPG Protein synthesis assay (1-homopropargylglycine based).** Rates of protein synthesis were assessed in KD podocytes using the Click-iT HPG Alexa Fluor 488 Protein Synthesis Assays (C10428, Thermo Fisher) according to manufacturer's instructions. Cells were seeded at a density of 15 x10<sup>3</sup> cells per well. After an incubation of 2 hrs with the reagent, intensity was measured with a microplate reader (LB 941 TriStar Berthold Technologies) and normalized to nuclear DAPI fluorescence intensity.

### Additional information for production and purification of human KEOPS subunits.

*For NMR experiments.* Expression of his-tagged LAGE3 was achieved in *E. coli* BL21 gold strain grown in rich 2xYT medium. After reaching the logarithmic phase, 0.5 mM of IPTG was added for 3 hours at 37°C,

then cells were harvested by centrifugation, resuspended in 35 mL of lysis buffer A (Tris-HCl 20 mM pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol) and stored at -20°C. Cells were lysed by sonication and centrifuged at 20 000g for 30 minutes. As LAGE3 is mainly insoluble, the pellet was recovered and washed with the same lysis buffer A supplemented with 0.5% Triton X100. After 30 minutes of centrifugation at 20 000 g, the pellet was resuspended in lysis buffer A + 8 M urea and centrifuged one more time at 20 000g. Supernatant containing solubilized, unfolded LAGE3 was then loaded onto NiNTA (Qiagen) affinity chromatography column and step-eluted in buffer A + 8 M urea containing increasing concentration of imidazole.

Expression of <sup>15</sup>N-GON7 was achieved in *E. coli* BL21 gold strain grown in M69 <sup>15</sup>N minimum medium and triggered by adding 0.5 mM of IPTG for 4 hours. Cells were collected by centrifugation and pellet was resuspended in 35 mL of lysis buffer A and stored at -20°C. Cells were lysed by sonication and centrifuged at 20 000 g for 30 minutes. The soluble supernatant containing <sup>15</sup>N-GON7 was purified by affinity chromatography on NiNTA. Purest fractions were pooled, concentrated and loaded on size exclusion chromatography column (Superdex 75 16/30) equilibrated in MES buffer 20 mM, pH 6.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol. 200 μM of pure <sup>15</sup>N-his-GON7 was used for recording <sup>15</sup>N-HSQC spectrum at 20°C on a Bruker Avance III 800 MHz.

Subcomplex LAGE3/<sup>15</sup>N-GON7 was prepared by refolding 387 μL of 600 μM of denatured his-LAGE3 with 200 μL 784 μM of <sup>15</sup>N-GON7 and loading the mixture on an analytical superdex S200 10/30 size exclusion chromatography column equilibrated in MES buffer 20 mM, pH 6.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol. Fraction containing the binary complex were pooled and concentrated to 64 μM for recording the <sup>15</sup>N-HSQC spectrum at 20°C on a Bruker Avance III 800 MHz.

To probe the folding of YRDC protein and the effect of the p.Ala84Val and p.Leu265del mutations, we collected 1D <sup>1</sup>H spectra of 70μM (for wild-type and p.Ala84Val) and 40 μM (p.Leu265del) proteins dissolved in 25mM Tris, pH 7.5, 200mM NaCl, 5mM beta-mercapto, and 95% H<sub>2</sub>O/ 5% D<sub>2</sub>O. Water suppression was achieved by the excitation sculpting method and the 1D spectra were collected at 298K on a 950 MHz Bruker Avance III HD spectrometer equipped with a cryoprobe.

*For SAXS or Crystallogensis experiments.* Unlabeled GON7 was prepared according to the above protocol except that *E. coli* BL21 GOLD cells were grown in rich 2xYT medium. The subcomplex LAGE3/GON7 was prepared by using two monocistronic vectors (pET21aLAGE3 and pET24d GON7). *E. coli* BL21 GOLD cells were cotransformed by adding 50 ng of each plasmid. The transformed cells were grown in rich 2YT medium during 3 hours at 37°C. Cells were then harvested by centrifugation, suspended in 40 mL of lysis buffer A (20 mM Tris buffer pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol) and stored at -20°C or directly used for purification. Cells were lysed on ice by three successive sonication cycles of 30 seconds using a Branson Sonifier and centrifuged at 20000 g for 30 minutes. Supernatant (i.e. soluble cell extract) was loaded on NiIDA silica resin (Macherey Nagel, Duren, Germany) and 40 mL of buffer A was applied as a washing step. Bound proteins were eluted using 3 fractions of 2 mL of buffer A supplemented with 100, 200 and 400 mM imidazole, respectively, with samples from each purification step being analyzed by SDS-PAGE. We loaded the pooled fractions onto an analytical S75 size exclusion chromatography column. LAGE3 and GON7 co-eluted as a single peak at 59 mL. Eluted fractions were analyzed by SDS PAGE (**Supplementary Fig.12**). The subcomplex OSGEP/LAGE3/GON7 was prepared by using a polycistronic vector (namely “pET24a-BC25\_hKEOPS\_opt”) designed to express the complete (i.e. 5 subunits) human KEOPS complex according to a previously published strategy<sup>4</sup>. Briefly, human KEOPS genes were codon-optimized for bacterial expression using EuGene software<sup>5</sup>, and subcloned into a pET24a vector between the NdeI and XhoI restriction sites and ordered from Genscript (Piscataway, USA). Sequence of the polycistronic gene is shown in **Supplementary Table 8**. The co-expression of the KEOPS subunits was achieved in the *E. coli* Rosetta pLysS strain at 37°C for 3 hrs. Cells were then harvested by centrifugation, suspended in 40 mL of lysis buffer A (20 mM Tris buffer pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol) and stored at -20°C or directly used for purification. Cells were lysed on ice by three successive sonication cycles of 30 seconds using a Branson Sonifier and centrifuged at 20000 g for 30 minutes. Supernatant (i.e. soluble cell extract) was loaded on NiIDA silica resin (Macherey Nagel, Duren, Germany) and 40 mL of buffer A was applied as a washing step. Bound proteins were eluted using 3 fractions of 2 mL of buffer A supplemented with 100, 200 and 400 mM imidazole, respectively, with samples from each purification step being analyzed by SDS-PAGE. Interestingly, we could only co-express and co-purify OSGEP, LAGE3 and GON7 on the NiIDA column. Indeed, SDS-PAGE analysis of the eluted fractions (**Supplementary Fig.13a**) showed the presence of OSGEP, LAGE3 and GON7, with barely detectable amounts of TP53RK and TPRKB (which is probably due to their low soluble expression levels in our polycistronic expression vector system). We loaded the pooled fractions onto an analytical S200 size exclusion chromatography column. As shown in **Supplementary Fig.13b**, OSGEP, LAGE3 and GON7

co-eluted as a single peak at 14.5 mL suggesting they are forming a complex. The identity of the three proteins in the eluted peak was confirmed by mass spectrometry (**Supplementary Table 9**).

#### **Additional information on sequence analysis**

Multiple sequence alignments with human YRDC, *Sulfolobus tokadaii* Sua5 and *Pyrococcus abyssi* Sua5, *Escherichia coli* YrdC and GON7 family were realised using blosum62 matrix<sup>6</sup>. The 3D model was generated using the protein fold recognition protocols of modeller software based on one – and three- dimensional sequence profiles<sup>7</sup>. Structure-based alignment of a subset was performed using the ESPRIPT platform<sup>8</sup>. All structural figures were generated with Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The hydrophobic surface creation was performed by PyMOL command to color protein molecules according to the Eisenberg hydrophobicity scale<sup>9</sup>. Structural comparisons was performed with the PDBeFold web server<sup>10</sup>.

#### **Additional information for production and purification of human WT YRDC and mutants**

DNA sequences of human YRDC WT, A84V and del265 have been purchased at Genscript (Piscataway NJ, USA) and subcloned in the pET21d vector between NcoI and BamHI, giving expression plasmids BC47\_YRDC-WT, BC48\_YRDC-A84V and BC49\_YRDC-del165 respectively (see enclosed sequences in **Supplementary Table 3**). Bacteria BL21 Gold have been transformed with 50 ng of plasmid and grown at 37°C in 800 mL of 2xYT rich medium supplemented with ampicillin. When bacteria started to reach exponential phase growth ( $OD_{600nm} = 0.4$ ), cultures have been shifted to 30°C for one hour and protein expression was triggered by adding 0.5 mM of IPTG. Cultures have been incubated for 3 more hours at 30°C and harvested by centrifugation at 6000 g for 20 minutes. Bacterial pellets were resuspended 35 mL of in Tris-HCl 25 mM, pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol.

Cell pellets have been lysed on ice by sonication using a Branson Sonifier (3 cycles of 30 seconds) and centrifuged at 20000 g for 30 minutes. Supernatant has been loaded on NiNTA resin (Qiagen) and purified by adding successive fraction of Tris-HCl 25 mM, pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol supplemented with 20 mM, 40 mM, 60 mM, 100 mM, 200 mM and 400 mM of imidazole. Protein fractions eluted at 200 and 400 mM of imidazole were pooled and further purified on S75 (16/60) superdex column (GE Healthcare) equilibrated with Tris-HCl 25 mM, pH 7.5 + 200 mM NaCl + 5 mM 2-mercaptoethanol. Purity was checked by SDS-PAGE and the purest fraction were pooled and concentrated by ultrafiltration. Finally, YRDC-WT could be concentrated up to 5 mg/mL, YRDC-A84V up to 3 mg/mL and YRDC-del265 up to 1 mg/mL.

#### **Additional information for the activity assay of YRDC WT and mutants *in vitro*.**

Activity of YRDC WT and mutants has been assayed in multi-well plate using the published malachite green method<sup>11</sup>. Pyrophosphatase (#I1643 from Sigma) was used as the coupling enzyme. The following solutions were prepared:

-Staining solution: 5 mL of Ammonium molybdate 4.2% (w/v) + 15 mL of malachite green 0.045 % (w/v) were mixed at room temperature for 30 minutes and filtered through 0.22  $\mu$ m.

-Stop solution: NaAcetate 0.25 M pH4

-Pre-mix solution (for 4 reactions): 60  $\mu$ L of ATP at 1.5 mM + 120  $\mu$ L of MgCl<sub>2</sub> at 1.5 mM + 30  $\mu$ L of NaHCO<sub>3</sub>, at 100 mM + 60  $\mu$ L of L-threonine at 1.5 mM + 2  $\mu$ L of 1U/ $\mu$ L Pyrophosphatase + 288  $\mu$ L of Tris-HCl 50 mM pH7.5. For the pre-mix solution, all of the mentioned stock solutions of substrates and Pyrophosphatase have been prepared in Tris-HCl 50 mM pH7.5.

-50  $\mu$ L of stop solution were dispensed in enough wells of a microtiter plate held on ice.

At t=0, 140  $\mu$ L of pre-mix was incubated with 10  $\mu$ L of 1.4 mg/mL of YRDC WT or mutants (or 10  $\mu$ L of Tris-HCl 50 mM pH7.5 for the pre-mix solution) and 20  $\mu$ L were immediately mixed in 50  $\mu$ L of stop solution. The remaining solution was incubated at 30°C and aliquots of 20  $\mu$ L were mixed with 50  $\mu$ L of stop solution at 10', 20', 30', 60' and 90'.

-60  $\mu$ L of staining solution were added in each wells of the microtiter plate which was incubated at room temperature. For standard curve determination, concentrations of inorganic phosphate (5 to 50  $\mu$ M) were assayed in parallel in the same microtiter plate

Absorbance at 620 nm was read after 15- and 30-minutes using a Tecan Infinite 200 PRO plate reader. Absorbance values were plotted versus time and each kinetic was fitted by linear regression using Excel software (Microsoft). Initial velocity (mAU/min) was calculated as the slope of each line. Each kinetic was recorded in triplicate.

### Additional references

1. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc.* **127**, 8014-8015 (2005)
2. Hopkins, J.B., Gillilan, R.E. & Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr.* **50**, 1545-1553 (2017)]
3. Hajizadeh NR, Franke D, Jeffries CM, Svergun DI. *Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data.* *Sci Rep.* 2018 8(1):7204
4. Collinet, B. et al. Strategies for the structural analysis of multi-protein complexes: lessons from the 3D-Repertoire project. *J Struct Biol.* **175**, 147-158 (2011).
5. Gaspar, P., Oliveira, J.L., Frommlet, J., Santos, M.A. & Moura, G. EuGene: maximizing synthetic gene design for heterologous expression. *Bioinformatics.* **28**, 2683-2684 (2012).
6. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881-10890 (1988).
7. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins.* **23**, 318-326 (1995).
8. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-324 (2014).
9. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol.* **179**, 125-142 (1984).
10. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr.* **60**, 2256-2268 (2004).
11. Sousa, E.H., Garay, P.A., Tinianow, J.N. & Gerber, N.C. Development of a spectrophotometric assay for cyclase activity. *Anal Biochem.* **348**, 57-63 (2006).

### III.2. Conclusion de l'article.

Cette étude, alliant des approches de génétique humaine à des approches biochimiques et structurales, a mis en évidence l'impact de mutations des gènes codant les protéines YrdC et GON7 sur le taux de modification t<sup>6</sup>A des ARNt et *in fine* sur des dérégulations cellulaires entraînant la pathologie GAMOS. Ainsi, nous avons montré que le défaut de la modification t<sup>6</sup>A des ARNt altère globalement la synthèse des protéines, conduisant à l'apoptose cellulaire.

Alors que le défaut de YrdC diminue significativement le taux d'ARNt modifié, ce qui a pour conséquence un phénotype GAMOS sévères, le défaut de GON7 entraîne des symptômes plus légers chez les patients GAMOS.

En absence de GON7, LAGE3 dimérise et les sous-unités du complexe KEOPS s'agencent en octamère de stœchiométrie 2 :2 :2 :2 (Wan *et al.*, 2017). En revanche, la présence de GON7, qui interagit avec la surface hydrophobe de LAGE3, induit la formation d'un pentamère de stœchiométrie 1 :1 :1 :1 :1 (Wan *et al.*, 2017). Chez la levure, il a été suggéré que yGon7 stabilise le complexe pentamérique (Zhang *et al.*, 2015). Cet agencement pentamérique impacte positivement l'activité d'un facteur 4 à 5 de la modification t<sup>6</sup>A (Wan *et al.*, 2017). Cependant, en absence de GON7, le complexe KEOPS *in vitro* de stœchiométrie 2 :2 :2 :2 assure tout de même une activité t<sup>6</sup>A (Wan *et al.*, 2017).

L'ensemble de ces données montre que GON7 contribue à la stabilisation de la forme stœchiométrique 1 :1 :1 :1 :1 du complexe KEOPS, qui est la forme la plus active d'un point de vue catalytique. Toutefois son absence n'altère pas drastiquement la modification t<sup>6</sup>A, expliquant le symptôme « adoucit » des patients GAMOS qui présentent des mutations sur GON7.

Nous avons évoqué en introduction générale, la disparité des phénotypes lors d'un défaut de la modification t<sup>6</sup>A à travers les espèces (*cf.* § 4.9. page 78). Ainsi, la modification t<sup>6</sup>A est essentielle à la survie de la majorité des bactéries. Quant aux champignons, l'altération de la modification t<sup>6</sup>A est tolérée sous des conditions standards non stressantes et conduit souvent à des erreurs de lecture des codons, altérant globalement la machinerie de la traduction.

Enfin, le défaut de la modification t<sup>6</sup>A modifie le développement des organismes pluricellulaires avec un impact d'intensité variable selon les organes. En effet, le défaut t<sup>6</sup>A altère globalement les podocytes et les cellules neuronales chez l'Homme et le



développement des disques imaginaux chez la *Drosophila melanogaster* (Rojas-Benítez *et al.*, 2017). L'exacerbation du phénotype au sein de ces cellules peut s'expliquer par une forte activité mitotique qui demande une machinerie de traduction des protéines hautement performante, rendant ces organes vulnérables en cours de développement à chaque perturbation.

Même si des fonctions alternatives ont été attribuées au complexe KEOPS, tel que l'homéostasie des télomères, pouvant expliquer les phénotypes pléiotropes dans les organismes (Downey *et al.*, 2006; Kisseleva-Romanova *et al.*, 2006; He *et al.*, 2019; Liu *et al.*, 2018), notre étude montre que le syndrome GAMOS est relié au défaut de la modification t<sup>6</sup>A causé par les mutations sur les gènes *YRDC* et *GON7*.

### III.3. Résultats supplémentaires.

#### III.3.1. Evolution des interfaces GON7/LAGE3 et Gon7/Pcc1.

Les protéines GON7/yGon7 sont considérées comme des IDPs (Zhang *et al.*, 2015; Wan *et al.*, 2017). Les IDPs sont des protéines qui évoluent très rapidement et il est très difficile de statuer sur leur origine évolutive (Brown *et al.*, 2011). De plus, les IDPs peuvent interagir avec des partenaires différents dans divers contextes biologiques et par conséquent acquérir des conformations différentes (Dyson et Wright, 2005). Même si les séquences peuvent diverger significativement pendant l'évolution, le « fold » des protéines est généralement plus préservé (Illergård *et al.*, 2009). Il est considéré que deux protéines qui possèdent un pli tridimensionnel similaire et plus de 30 % d'identité de séquence sont des homologues c'est-à-dire qu'elles partagent un ancêtre commun. Ici nous avons l'exemple des protéines GON7 et yGon7 qui présente une structure tridimensionnelle presque identique en interaction avec LAGE3 et Pcc1 et ayant une fonction commune alors qu'elles possèdent seulement 17 % d'identité de séquence.

De façon surprenante, au sein du laboratoire il a été observé que GON7 de l'Homme ne complémente pas la souche *Saccharomyces cerevisiae* lors de l'extinction du gène yGON7, observé aussi par l'équipe de F. Sicheri (Wan *et al.*, 2017), suggérant une spécificité des interfaces GON7/LAGE3 et yGon7/Pcc1.

Nous avons entrepris une analyse des séquences dans le but de comprendre cette incapacité de complémentation. 18 séquences de GON7, dont 6 séquences de métazoaires et 12 séquences de champignons, ont été alignées avec MAFFT (Kato *et al.*, 2019). L'alignement multiple des séquences a permis de révéler quelques acides aminés conservés et quelques acides aminés similaires au sein des GON7 (surlignés en rouge et jaune respectivement sur la [figure 62](#)). Nous avons aussi identifié les acides aminés impliqués dans l'interaction avec leur partenaire LAGE3/Pcc1 respectif (surlignés en cyan sur l'alignement en [figure 63](#)).

Les séquences des protéines de la famille LAGE3 et Pcc1 ont été également alignées avec MAFFT (Kato *et al.*, 2019). Nous avons ainsi identifié et localisé les acides aminés des protéines LAGE3 et Pcc1 impliqués dans l'interaction avec leurs partenaires GON7 et yGon7 respectifs ([figure 61](#)). Les acides aminés stabilisant les interfaces sont localisés à des positions équivalentes dans les deux structures GON7-LAGE3 et Gon7-Pcc1 ([figure 61](#)).

Les séquences Gon7 des champignons sont plus longues que celles des GON7 des métazoaires. A titre d'exemple GON7 de l'Homme possède 100 acides aminés alors que Gon7

de levure en possède 123. Gon7 de *Saccharomyces cerevisiae* (code PDB 4WXA) se replie en 2 brins  $\beta$  antiparallèles (acides aminés 1 à 19), suivis d'une région désordonnée (acides aminés 20 à 64) puis enfin de deux hélices  $\alpha$  consécutives. Il est important de noter que dans des séquences Gon7 de différents champignons, les acides aminés de cette région désordonnée sont conservés ou similaires chez les organismes (région encadrée en noir sur l'alignement en [figure 62](#)).

Cette région n'est pas conservée dans les séquences de GON7 des métazoaires et les deux hélices  $\alpha 1$  de GON7 et  $\gamma$ Gon7 s'alignent. Au cours de l'évolution, les séquences des métazoaires pourraient avoir perdu cette insertion ou bien les séquences des champignons auraient pu acquérir cette insertion en interaction avec d'autres partenaires.

La leucine 33 de GON7 est conservée dans toutes les séquences et est localisée au début des hélices  $\alpha 1$  (surlignée en rouge sur l'alignement, [figure 62 page 229](#)). De même, la tyrosine 7 est conservée et localisée sur les brins  $\beta 1$ , à l'exception de GON7 de *Sarcophilus harrisi* qui possède une phénylalanine à cette position (localisée par une étoile verte sur l'alignement en [figure 62](#)). L'encombrement stérique de ces deux résidus aromatiques est similaire. Un glutamate ou un aspartate sont localisés au début des brins  $\beta 2$  (E<sub>11</sub> pour la séquence GON7). Au bilan, il semblerait qu'au début de chaque élément de structure secondaire, les acides aminés sont conservés ou présentent des propriétés physico-chimiques similaires.

Aussi, nous avons étudié les acides aminés impliqués dans la stabilisation des hétérodimères. Par exemple, la K<sub>41</sub> de GON7 interagit avec l'aspartate D<sub>125</sub> de LAGE3. Ce dernier acide aminé est conservé dans Pcc1 et interagit avec la Q<sub>78</sub> qui présente un encombrement stérique similaire à la lysine. Ces acides aminés occupent des positions équivalentes dans les structures ([figure 61 A et B](#)). Mais les acides aminés qui stabilisent ces interfaces ne sont pas toujours conservés ou ne présentent pas toujours des propriétés physico-chimiques similaires.

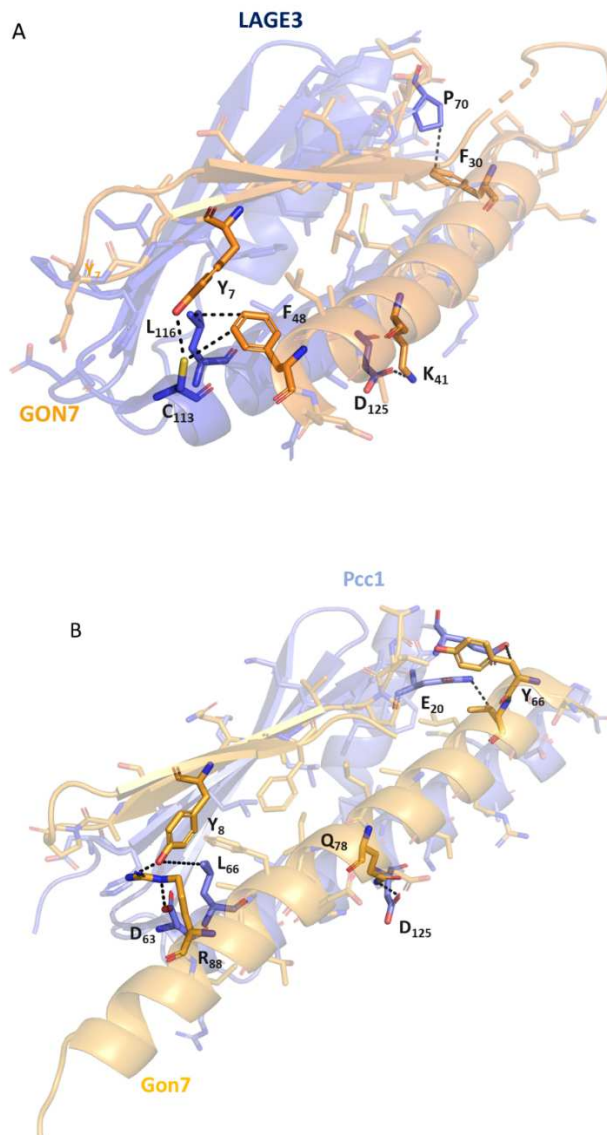


Figure 61 Structure quaternaire des complexes (A) GON7-LAGE3 de l'Homme (en orange et bleu respectivement) (code PDB 6GWJ) et (B) Gon7-Pcc1 de *Saccharomyces cerevisiae* (en orange et bleu respectivement) (code PDB 4WXA). Les acides aminés représentés en stick sont ceux qui sont impliqués dans l'interaction et sont surlignés sur les alignements multiples des séquences Figures x et y, respectivement en cyan et magenta pour GON7/Gon7 et LAGE3/Pcc1.

Par exemple, la tyrosine 7 et la phénylalanine 48 de GON7 forment des liaisons hydrogènes avec la cystéine 113 de LAGE3 (figure 61 A). La tyrosine 8 et l'arginine 88 de Gon7 de *Saccharomyces cerevisiae* forment des liaisons hydrogènes avec l'aspartate 63 de Pcc1 (figure 61 B). Ces acides aminés sont localisés à des positions équivalentes sur les structures (figure 61). Sur l'alignement de la famille de protéines LAGE3-Pcc1, les acides aminés C<sub>113</sub> et D<sub>63</sub> s'alignent (figure 63), présentant des encombrements stériques et des charges qui diffèrent. Ils interagissent avec la phénylalanine F<sub>48</sub> et l'arginine R<sub>88</sub> respectivement qui occupent des positions équivalentes dans GON7/Gon7 et qui ont également des propriétés

polaires différentes. Aussi, la phénylalanine F<sub>30</sub> de GON7 interagit avec la proline P<sub>70</sub> de LAGE3, alors qu'aux positions équivalentes dans la levure, la tyrosine Y<sub>66</sub> de Gon7 interagit avec le glutamate E<sub>20</sub> de Pcc1.

La nature de ces acides aminés diffère par leur propriété physico-chimique mais stabilisent l'interface à des positions équivalentes.

La co-évolution des protéines GON7, yGon7 et LAGE3, Pcc1 a donc permis de conserver le même mode d'interface d'un long feuillet de brins  $\beta$  anti-parallèles et de deux hélices  $\alpha$ . Néanmoins, la divergence des séquences entre GON7 et yGon7 permet d'expliquer que cette interface reste spécifique dans chaque espèce et que la protéine GON7 de l'Homme ne complémente pas la levure déléetée de son gène codant yGon7.

Pour tester cette pression évolutive, il serait intéressant d'essayer de compléter par LAGE3-GON7 dans la levure déléetée des gènes *PCC1-GON7*. Aussi, il serait intéressant de muter l'aspartate 63 par une cystéine et le glutamate 20 par une proline dans les séquences de Pcc1 de la levure et tester son interaction avec GON7 de l'Homme.

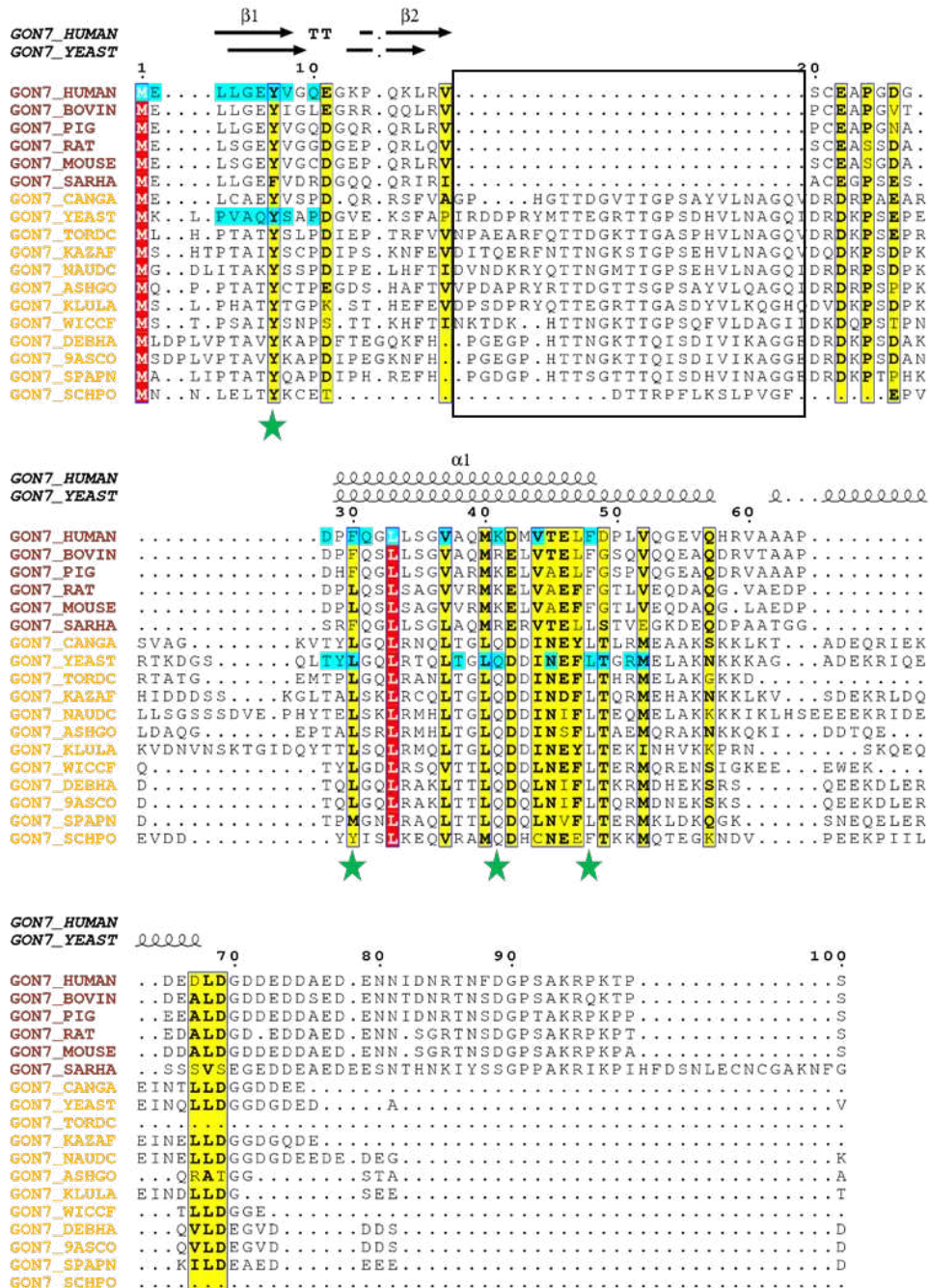


Figure 62 Alignement des sous-unités GON7 des métazoaires (en marron) et Gon7 des champignons (en orange). 18 séquences ont été alignées avec MAFFT (online version 7). La matrice utilisée est BLOSUM 40. Les séquences proviennent de UNIPROT (HUMAN : *Homo sapiens*, BOVIN : *Bos taurus*, PIG : *Sus scrofa*, RAT : *Rattus norvegicus*, MOUSE : *Mus musculus*, SARHA : *Sarcophilus harrisi*, CANGA : *Candida albicans*, YEAST : *Saccharomyces cerevisiae*, TORDC : *Torulaspota delbrueckii*, KAZAF : *Kazachstania africana*, NAUDC : *Naumovozyma dairenensis*, ASHGO : *Ashbya gossypii*, KLULA : *Kluyveromyces lactis*, WICCF : *Wickerhamomyces ciferrii*, DEBHA : *Debaryomyces hansenii*, 9ASCO : *Debaryomyces fabryi*, SPAPN : *Spathaspora passalidarum*, SCHPO : *Schizosaccharomyces pombe*.) L'alignement multiple de séquence a été ensuite traité par ESPRIPT. Les acides aminés surlignés en jaune sont similaires et en rouge sont strictement conservés. Les acides aminés surlignés en cyan sont impliqués dans l'interaction avec leurs partenaires respectifs LAGE3 ou Pcc1. La région des acides aminés 20 à 64 de yGon7, désordonnée, absente chez GON7, est encadrée en noire. Les étoiles vertes indiquent les acides aminés discutés dans le texte.

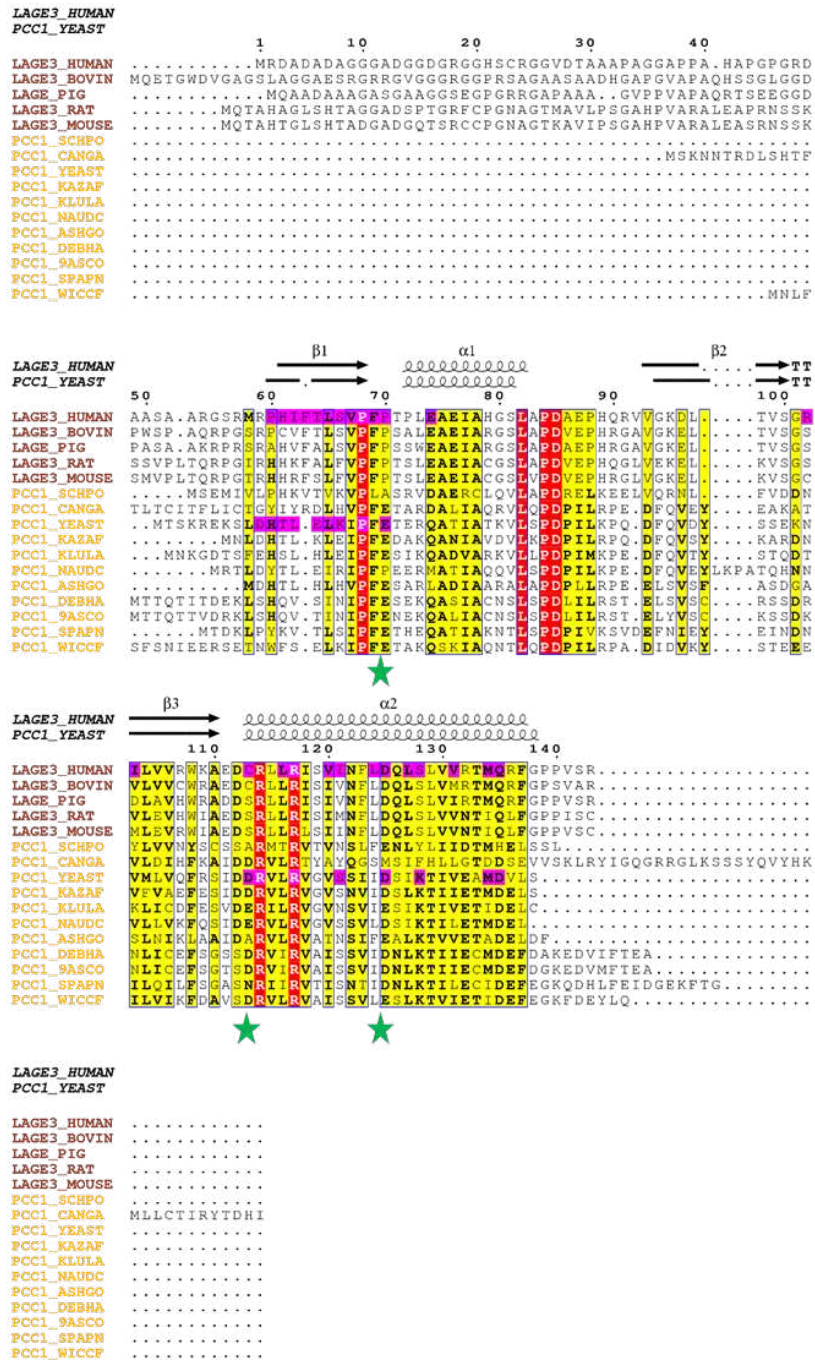


Figure 63 Alignement des sous-unités LAGE3 des métazoaires (en marron) et Pcc1 des levures (en orange). 16 séquences ont été alignées avec MAFFT (version 7). La matrice utilisée est BLOSUM68. Les séquences proviennent de UNIPROT (HUMAN : *Homo sapiens*, BOVIN : *Bos taurus*, PIG : *Sus scrofa*, RAT : *Rattus norvegicus*, MOUSE : *Mus musculus*, CANGA : *Candida albicans*, YEAST : *Saccharomyces cerevisiae*, KAZAF : *Kazachstania africana*, NAUDC : *Naumovozyma dairenensis*, ASHGO : *Ashbya gossypii*, KLULA : *Kluyveromyces lactis*, WICCF : *Wickerhamomyces ciferrii*, DEBHA : *Debaryomyces hansenii*, 9ASCO : *Debaryomyces fabryi*, SPAPN : *Spathaspora passalidarum*, SCHPO : *Schizosaccharomyces pombe*). L'alignement multiple de séquence a été ensuite traité par ESPRIT. Les acides aminés surlignés en jaune sont similaires et en rouge sont strictement conservés. Les acides aminés surlignés en magenta sont impliqués dans l'interaction avec leurs partenaires respectifs GON7 ou Gon7. Ceux surlignés en rouges sont strictement conservés et ceux en jaunes sont similaires. Les étoiles vertes indiquent les acides aminés discutés dans le texte.

III.3.2. Modélisation de la structure 3D du complexe entier KEOPS humain par diffusion des rayons X aux petits angles.

III.3.2.1. Matériels et Méthodes.

Production, expression et purification des protéines solubles.

Les gènes codant les protéines GON7, LAGE3, OSGEP, TP53RK, TPRKB ont été sous clonés entre les sites NdeI et XhoI dans un plasmide pET24a (nommé BC25) sous le contrôle du promoteur T7. Les protéines TP53RK et TPRKB ne s'expriment pas (*cf. article III, page 209*), nous avons donc construit des plasmides monocistroniques possédant les gènes codant TP53RK et TPRKB sous clonés respectivement dans des plasmides pET24d et pET21a (le détail des séquences est en annexe 3).

100 ng du plasmide BC25 ont été incubés dans 100  $\mu$ L de souches chimio compétentes BL21 gold DE3 (Novagen). 50 ng du plasmide pET24d-TP53RK et 50 ng du plasmide pET21a-TPRKB ont été co-incubés dans 100  $\mu$ L de souches chimio compétentes BL21 Rosetta (Novagen). L'expression des protéines a été réalisée dans un milieu 2YT riche supplémenté en antibiotiques appropriés et la production des protéines a été induite à l'IPTG (0,5  $\mu$ M). Après l'induction à l'IPTG, une centrifugation (4000 g, 20 minutes, 8°C) a été réalisée. Les culots bactériens sont resuspendus dans 35 mL de tampon 20 mM Tris-HCl pH 7,5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoéthanol. 70 mL de suspension bactérienne exprimant OLG et 30 mL de suspension bactérienne exprimant TP35RK et TPRKB sont mélangés puis conservés à -20°. Après décongélation, la lyse des cellules est réalisée par sonication à 0°C (Branson sonifier 250, 3 cycles de 1 min, 30% de puissance). Les extraits cellulaires totaux sont centrifugés 30 minutes à 20000 g et le surnageant, composés de la fraction protéique soluble dont les protéines recombinantes d'intérêt, est récupéré pour la suite de la purification (Quevillon-Cheruel *et al.*, 2004). Le surnageant est incubé avec la résine de silice Ni<sup>2+</sup>NTA (Qiagen). Ainsi les protéines du complexe KEOPS sont purifiées selon la méthode IMAC (Immobilized Metal Affinity Chromatography) et éluées par des sauts de concentrations croissantes d'imidazole (de 60 à 400 mM) qui rentrent en compétition avec l'affinité de l'étiquette 6 histidines avec les ions Ni<sup>2+</sup>. Les fractions d'éluion contenant les protéines d'intérêts sont réunies et concentrées par ultra-filtration sur membrane de porosité (seuil de coupure de 30 kDa (Centricon). Les protéines sont ensuite purifiées par chromatographie d'exclusion stérique



utilisant des colonnes Superdex 200 16-600 (GE-Healthcare) préalablement équilibrées dans un tampon 20 mM Tris-HCl pH 7,5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoéthanol. L'élution le long de la colonne se fait avec un débit de 1 ml.min<sup>-1</sup>. Des fractions de 1,5mL sont collectées en sortie de colonne. Un détecteur d'UV mesure l'absorbance à 280 nm et 260 nm en sortie de colonne, ce qui permet d'identifier les fractions où les protéines sont présentes et s'il y a présence de nucléotides. Enfin les fractions d'intérêt sont analysées en sortie de colonne par SDS-PAGE et celles contenant les protéines d'intérêts les plus pures sont réunies et concentrées par ultra-filtration sur membrane de porosité (seuil de coupure de 30 kDa (Centricon)). La concentration des protéines est estimée par lecture de l'absorbance à 280 nm. Pour cela, on utilise les valeurs théoriques du coefficient d'extinction molaire et du poids moléculaire (en faisant l'hypothèse que la DO à 280nm est uniquement due aux protéines d'intérêt). De plus, l'examen du spectre d'absorbance entre 240 et 340 nm permet de vérifier l'absence de contamination par des acides nucléiques.

#### SAXS : enregistrement, collection et traitement des données.

Les mesures de SAXS ont été réalisées sur la ligne SWING du synchrotron SOLEIL (Saint-Aubin). L'énergie du faisceau incident est fixée à 12,6 keV, soit  $\lambda = 1 \text{ \AA}$ , pour les mesures réalisées sur des protéines en solution. La distance entre l'échantillon et le détecteur (Eiger 4M Dectris) est fixée à 1,8 m. La gamme du vecteur de diffusion est ainsi de  $0,0085 < q < 0,5 \text{ \AA}^{-1}$ , où  $q = 4\pi\sin\theta/\lambda$  et  $2\theta$  étant l'angle de diffusion. L'échantillon est injecté dans une colonne de chromatographie d'exclusion de taille (GF-HPLC), directement connectée à la cellule de mesure, permettant ainsi de séparer les contributions provenant des différents objets éventuellement présents dans la solution injectée, et en particulier de s'affranchir de la contribution due aux agrégats. La cellule de mesure est un capillaire thermostaté à 15°C, de diamètre 1,5 mm.

Le complexe KEOPS a donc été injecté sur une colonne de chromatographie d'exclusion (SE- HPLC Bio Sec3, Agilent, 4.5x300 mm) équilibrée avec un tampon 20 mM Tris HCl pH 7,5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoéthanol dans le but de séparer le complexe d'éventuels agrégats ou des partenaires isolés. Le débit de l'élution était de 0,3  $\mu$ L/min. Pendant l'élution, les courbes de diffusion ont été enregistrées toutes les secondes avec un temps mort entre les courbes de 0,01s. Pour chaque courbe, la concentration des protéines est estimée à partir

de l'absorption à 280 nm par un spectrophotomètre UV localisé juste avant la cellule de mesure du SAXS. Les intensités diffusées sont mises en échelle absolue en normalisant par la diffusion de l'eau.

Un grand nombre de courbes de diffusion ont été collectées et moyennées avant l'élution du complexe étudié. Elles correspondent à la diffusion du tampon, qui est ainsi soustraite des courbes de diffusion de la solution de protéine, afin d'obtenir le signal de la protéine uniquement. La courbe finale représentative du complexe étudié est obtenue en comparant les différentes courbes mesurées au maximum du pic d'élution, en sélectionnant les courbes identiques et en les moyennant. Ces étapes ont été réalisées avec le programme Foxtrot développé sur la ligne SWING, et le programme US-SOMO (Brookes *et al.*, 2016). Le logiciel Primus a été utilisé pour déduire de la courbe  $I(q)$  la masse moléculaire, le rayon de giration, et la fonction de distribution des distances (Franke *et al.*, 2017).

Un modèle *in silico* du complexe KEOPS humain a été réalisé en utilisant le logiciel MODELLER (Sali *et al.*, 1995). Les positions des régions flexibles sont obtenues en ajustant la courbe calculée du modèle à la courbe expérimentale des intensités  $I(q)$  en utilisant le logiciel DADIMODO (Evrard *et al.*, 2011). Un dernier ajustement de la courbe calculée a été faite avec le programme CRY SOL de la suite ATSAS (Svergun *et al.*, 1995).

#### III.3.2.2. Résultats.

L'expression recombinante des protéines ainsi que leur purification puis la reconstitution du complexe ont été réalisés comme décrit dans la partie Matériel et Méthodes. L'analyse structurale du complexe en solution a été faite par SEC-SAXS sur la ligne Swing du synchrotron SOLEIL (St Aubin) (*cf. Matériel et Methodes*).

Le complexe KEOPS humain a été purifié par chromatographie d'affinité sur ions métalliques ( $Ni^{2+}$ ) immobilisés suivie d'une chromatographie d'exclusion stérique utilisant une colonne Superdex 200 16-600 (GE Healthcare). Le chromatogramme montre que des protéines sont éluées à un volume d'élution de 78 mL. Le volume d'élution est compatible avec la taille du complexe KEOPS de 112000 Da selon la calibration de la colonne préalablement effectué. Un pic minoritaire est élué vers 50 mL correspondant au volume mort de la colonne. L'absorbance à 260 nm montre que ce pic correspond probablement à l'élution d'acides nucléiques (*figure 64*). Un autre pic minoritaire est élué à un volume d'élution de 90

mL correspondant à des sous-complexes. Les fractions issues du pic majoritaire de la chromatographie d'exclusion ont été analysées par SDS-PAGE (figure 64 B) confirmant la présence des 5 sous-unités du complexe : GON7 (12 kDa), LAGE3 (17 kDa), OSGEP (35kDa), TP53RK (30 kDa) et TPRKB (21 kDa).

KEOPS forme donc un complexe suffisamment stable pour être élué par chromatographie d'exclusion stérique. Dans le but de caractériser la conformation et de confirmer l'état d'oligomérisation du complexe KEOPS (mesuré par MALLS Wan *et al.*, 2017), des expériences de SAXS couplées à une chromatographie d'exclusion ont été réalisées (SEC-SAXS) avec cet échantillon.

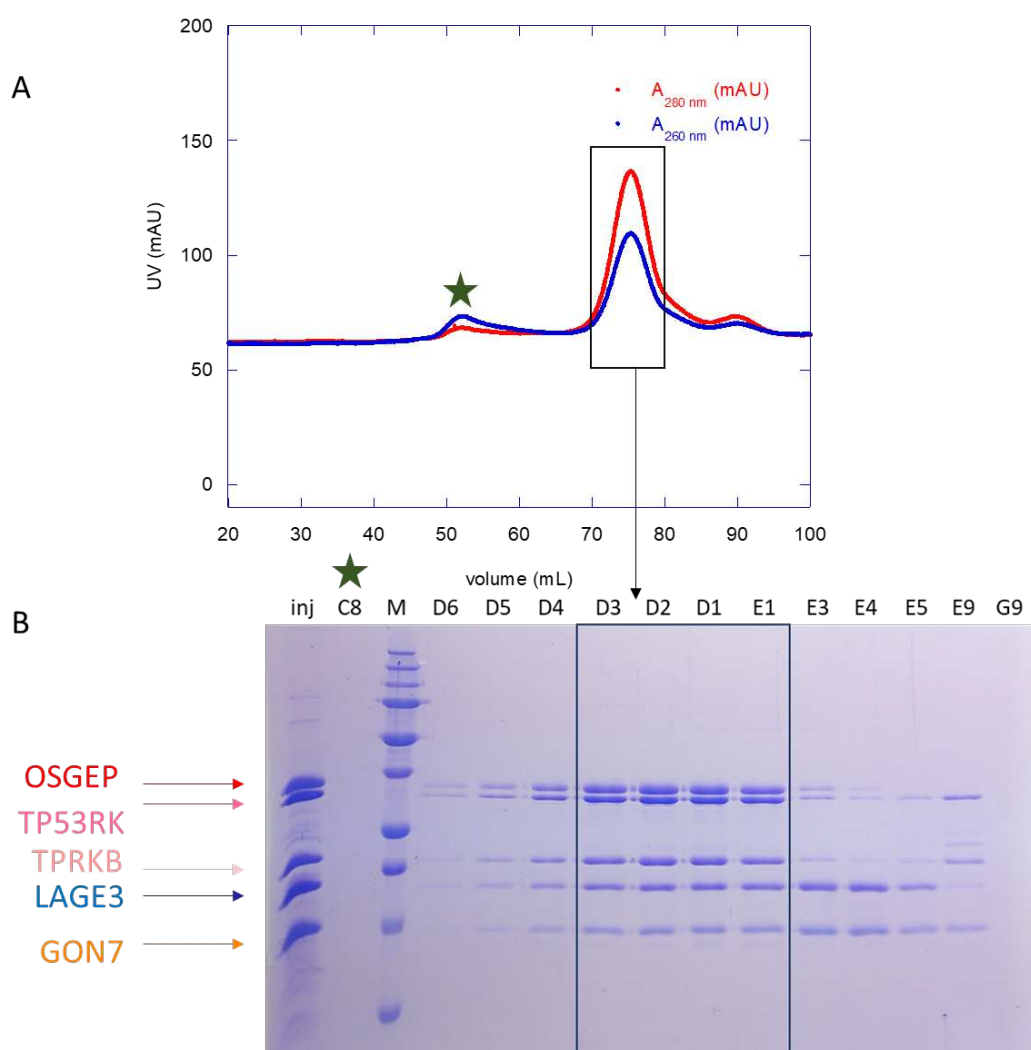


Figure 64 Purification du complexe hKEOPS. A. Le suivi de l'élué sur colonne S200 16-600 est représenté sur le chromatogramme. L'absorbance de l'UV à 260 et 280 nm est en bleu et rouge respectivement. B. Analyse SDS-Page des protéines du complexe KEOPS en sortie de colonne Superdex 200 16-600 (GE Healthcare). Un marqueur de taille a été chargé sur le gel, les bandes correspondent à 250 kDa, 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, 20 kDa, 15 kDa, 10 kDa. L'étoile verte correspond à la fraction du pic situé dans le volume mort de la colonne.

## Résultats.

Les analyses des expériences de SEC-SAXS ont permis d'obtenir de nombreuses informations structurales du complexe KEOPS humain à basse résolution en solution qui sont résumées dans le tableau suivant (figure 65).

<b>hKEOPS</b>	
<b>Data collection parameters</b>	
Instrument	SWING (SOLEIL)
Detector	CCD-based AVIEX
Beam geometry	0.8 mm x 0.15 mm
Wavelength [Å]	1.0
q-range [Å <sup>-1</sup> ]	0.007 < q < 0.50
Absolute scaling	Comparison with scattering from pure H <sub>2</sub> O
Exposure time [s]	2
Temperature [K]	283
SEC-SAXS column	Bioseac 3 Agilent
Loading concentration [mg/mL]	3,7
Injection volume [μL]	50
Flow rate [mL.min <sup>-1</sup> ]	0.3
<b>Structural parameters</b>	
I(0) Guinier [cm <sup>-1</sup> ]	8.19 10 <sup>-3</sup>
R <sub>g</sub> Guinier [Å]	38,68
qR <sub>g</sub> -range	0,51 - 1,3
I(0) p(r) [cm <sup>-1</sup> ]	8.20 10 <sup>-3</sup>
R <sub>g</sub> p(r) [Å]	38.77 ± 0.043
q-range [Å <sup>-1</sup> ]	0.0132 - 0.2120
D <sub>Max</sub> [Å]	134
<b>Molecular mass determination</b>	
Mol . L <sup>-1</sup> MM <sub>sequence</sub> [kDa] <sup>a</sup>	112,9
MM <sub>bayesian</sub> [kDa] <sup>b</sup> Credibility interval	109,075 99,200-111,25

**Figure 65** Paramètres d'acquisition des données SAXS et paramètres expérimentaux déduits de la courbe I(q) du complexe KEOPS humain. <sup>a</sup> La masse moléculaire est calculée à partir de la séquence en acide aminés. <sup>b</sup> La masse moléculaire expérimentale est obtenue en utilisant l'approche Bayésienne qui combine 4 méthodes différentes d'obtention de la masse à partir de la courbe I(q).

Les courbes de diffusion correspondant au maximum du pic d'éluion ont été moyennées. La courbe de Guinier (I<sub>q</sub> vs q<sup>2</sup>) linéaire montre que l'objet étudié en solution est monodisperse (figure 66 A). L'application de la loi de Guinier a permis de déterminer un rayon de giration de 38,7 Å. La masse moléculaire de KEOPS, déduite de la courbe I(q), est de 109 kDa, en accord avec la masse calculée sur la séquence (112,9 kDa). Les protéines du complexe KEOPS forment donc un complexe pentamérique de stœchiométrie 1 :1 :1 :1 :1.

La fonction de distribution des distances  $P(r)$  déduite de la courbe  $I(q)$  a permis une seconde estimation du rayon de giration ( $38.8 \text{ \AA}$ ) et une estimation de la dimension maximale de la particule ( $D_{\text{max}}$ ) de  $134 \text{ \AA}$ , ce qui en fait un objet plutôt allongé (figure 66 B).

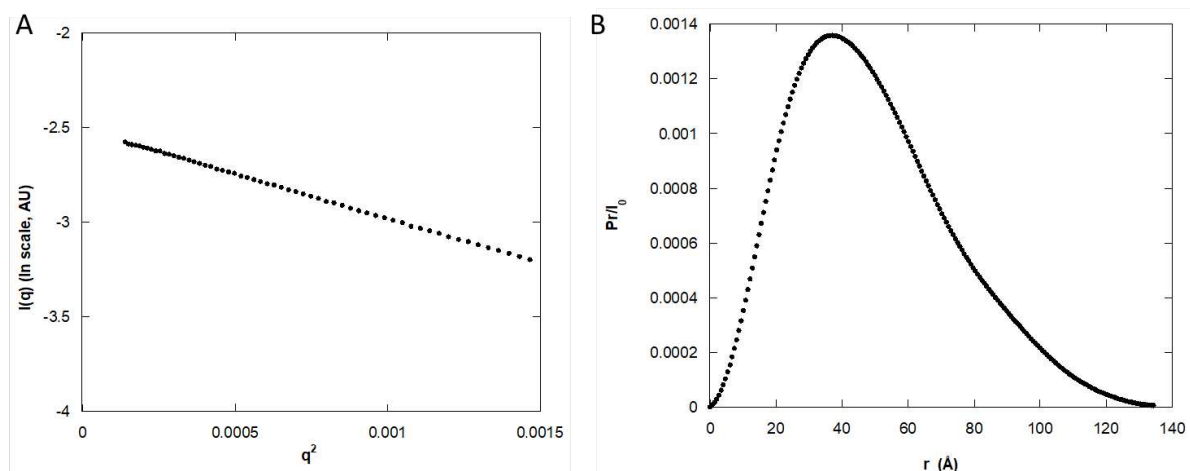


Figure 66 Résultats de SEC-SAXS. A. Représentation de Guinier. B. Fonction de Distribution des distances ( $Pr$ ) du complexe KEOPS humain, conduisant à un rayon de giration de  $38.8 \text{ \AA}$  et une extension maximale de  $134 \text{ \AA}$ .

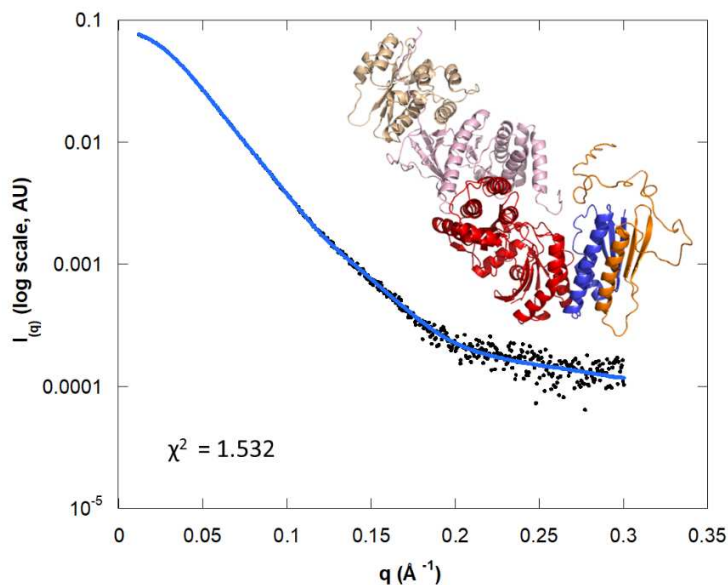


Figure 67 Ajustement du modèle *in silico* du complexe humain. La courbe expérimentale de diffusion des rayons X du complexe KEOPS humain est représentée en noire. La courbe bleue représente la courbe de diffusion calculée sur le modèle *in silico* du complexe KEOPS. La superposition des deux courbes s'ajuste parfaitement avec un  $\chi^2 = 1,532$ . L'encart en haut à droite représente le complexe KEOPS *in silico* ajusté par DADIMODO (en rouge ; OSGEP, en bleu ; LAGE3, en orange ; GON7, en rose ; TP53RK, en beige ; TPRKB).

Un modèle *in silico* de la structure tridimensionnelle du complexe KEOPS humain a été réalisé en utilisant Modeller (Sali *et al.*, 1995), en se basant sur les structures des sous-complexes ou des homologues connues (Bud32-Cgi121 (code PDB 4WW9), OSGEP-LAGE3-GON7 (code PDB

6GWJ)). Dans ce modèle, la région C-terminale de GON7 est manquante. Elle a été modélisée à partir de la courbe de SAXS en utilisant le programme DADIMODO (Evrard *et al.*, 2011) qui détermine la position des régions manquantes en ajustant la courbe calculée sur le modèle à la courbe expérimentale. Le modèle structural obtenu est en accord avec les résultats de SAXS comme en témoigne la valeur de  $\chi^2=1,532$  (figure 67). Il suggère que le complexe KEOPS humain est un complexe linéaire comme celui de la levure (Zhang *et al.*, 2015), et que l'extrémité C-terminale de GON7 est essentiellement désordonnée.

L'étude par SEC-SAXS nous a apporté de nombreuses informations structurales. Tout d'abord, la forme du complexe en solution est étendue avec une extension maximale de 138 Å. Aussi, l'ajustement du modèle *in silico* sur la courbe expérimentale  $I(q)$  confirme la flexibilité de l'extrémité C-terminale de GON7, confirmant que seuls 50 % de la protéine GON7 possède des structures secondaires et tertiaires comme observées dans la structure cristallographique du sous complexe OSGEP-LAGE-GON7 (code PDB 6GWJ). Les 5 sous-unités du complexe KEOPS forment un complexe stable de stœchiométrie 1 :1 :1 :1 :1 de la même façon que le complexe KEOPS de la levure. Wan *et al.*, ont mis en évidence par SEC-MALLS qu'en absence de GON7, les autres sous-unités du complexe KEOPS forment un objet de stœchiométrie 2 :2 :2 :2 en solution (Wan *et al.*, 2017). Nos données montrent sans ambiguïté que la présence de GON7 réorganise le complexe KEOPS humain en pentamère.

### III.3.3. Etude structurale du complexe KEOPS humain avec son substrat : l'ARNt<sup>lle</sup><sub>AUU</sub>.

#### III.3.3.1. Matériels et Méthodes.

##### Production et purification de l'ARNt<sup>lle</sup><sub>AUU</sub> par transcription *in vitro*.

Les ARNt ont été transcrits *in vitro* en insérant en 3' la séquence du ribozyme HDV (annexe 4). Les gènes codant l'ARNt<sup>lle</sup><sub>AUU</sub> suivi du ribozyme (ces deux séquences sont séparées par un site NdeI) ont été sous clonés entre les sites BglII et XhoI dans un plasmide pUC18 (nommé pUC18DNA<sup>ARNt</sup>).

100 ng du plasmide pUC18DNA<sup>ARNt</sup> ont été incubés dans 100 µL de la souche XL1 blue résistantes à la tétracycline (Novagen). Les bactéries transformées par choc thermique sont amplifiées dans un milieu riche 2YT supplémenté des marqueurs de sélection adéquate (ampicilline et tétracycline à 50 µg.mL<sup>-1</sup>) pendant 1 nuit à 37°C sous agitation (200 rpm). L'ADN

plasmidique est ensuite purifié à l'aide du kit commercial Endofree plasmid Maxi Kit (Qiagen). Une dernière étape de purification de l'ADN plasmidique est réalisée par précipitation à l'isopropanol 100% (Edelmann *et al.*, 2014). L'ADN plasmidique pUC18DNA<sup>ARNt</sup> est ensuite linéarisé par l'enzyme de restriction *XhoI* formant la matrice linéaire pUC18DNA<sup>ARNt</sup>. Cette matrice (0,02 µg.µL<sup>-1</sup>) est ensuite utilisée pour la réaction de transcription *in vitro* par la T7 polymérase (0,03 µg.µL<sup>-1</sup>) dans un tampon 40 mM Tris HCl pH 8, 1 mM spermidine, 5 mM DTT, 0,01 % triton 100 X, 2 mM de chaque NTPS (Thermofischer), 10 mM MgCl<sub>2</sub> pendant 4 heures à 37°C. Pour s'affranchir au maximum de la fraction de ribozyme HDV non clivé, la réaction est supplémentée de 5 mM final de MgCl<sub>2</sub> qui sert de cofacteur pour la réaction d'auto-clivage du ribozyme. La solution est incubée pendant 20 min à température ambiante puis 10 min à 65°C et enfin 30 min à température ambiante. La solution est centrifugée pendant 10 min à 7000g pour sédimenter le pyrophosphate P<sub>Pi</sub> (un des produits de la réaction de transcription). Le surnageant est concentré à l'aide d'une membrane de 5 kDa (centricon). Un volume de solution de réaction est mélangé à un volume d'urée à 7 M dans le but de dénaturer l'ARNt et le ribozyme. La fraction d'ARNt est purifiée sur gel d'acrylamide 10% et 7M urée. La migration est réalisée dans un tampon TBE 0,5x (44 mM Tris pH 8,3, 44 mM boric acid, 0.1 mM EDTA) à 300 V pendant 3 à 4h afin de séparer les bandes correspondantes à la fraction du ribozyme et à la fraction d'ARNt. Après découpage de la bande idoine, les ARNt purs sont récupérés par électro élution (35 min, 100V) dans un tampon TBE 0,5X et précipités avec 2,5 volumes d'éthanol 100 % froid supplémenté d'acétate de sodium (2.5 M pH 7,5) pendant une nuit à -20°C. La solution précipitée est finalement centrifugée (1h, 20000g, 4°C) et selon la taille du culot, les molécules d'ARNt sont suspendues dans 100 à 200 µL de tampon (20 mM HEPES, pH 7,5, 20mM MgCl<sub>2</sub>). La concentration finale d'ARNt est estimée par lecture de l'absorbance à 260 nm (Nanodrop).

#### Tests d'interaction préliminaires de l'ARNt<sup>lle</sup><sub>AUU</sub> avec le complexe KEOPS par technique de retard sur gel.

La réaction d'interaction entre le complexe KEOPS et l'ARNt<sup>lle</sup><sub>AUU</sub> a été réalisée dans un volume final de 40 µL contenant 0,18 pmol (45 nM) d'oligonucléotide (ARNt<sup>lle</sup><sub>AUU</sub>) radiomarqué au <sup>32</sup>P et de protéine à différentes concentrations, dans un tampon contenant 25 mM HEPES pH 7.5, 50 mM NaCl, 20 mM MgCl<sub>2</sub>, 5 mM β mercaptoéthanol. Après incubation pendant 30 min à

4°C ou à température ambiante, 2 µL de tampon de dépôt 6x (10 mM Tris-HCl pH 7,5, 20% glycerol, 1 mM EDTA, 0,1 mg/ml BSA, et 0,1% xylène cyanol) ont été ajoutés à la réaction, puis le mélange a été déposé sur un gel de polyacrylamide (37 :5 :1) 6%, 0,5 x TBE, 0,5 mM EDTA et les complexes nucléo-protéiques ont été séparés par électrophorèse dans un tampon TBE1X, sous une tension de 100 V (7.5 V/cm) à 4°C pendant 3 à 4 heures. Après migration, le gel a été séché puis la radioactivité a été révélée par exposition sur un écran PhosphorScreen (GE) et une lecture à 533 nm par un scanner Typhoon. L'intensité des bandes a été mesurée par le logiciel ImaJ.

#### Purification du complexe KEOPS- ARNt<sup>lle</sup><sub>AUU</sub> par chromatographie d'exclusion stérique.

L'ARNt<sup>lle</sup><sub>AUU</sub> pur a été incubé avec le complexe KEOPS humain à un ratio 1 : 1 pendant 1 heure à 4°C. La solution a été ensuite déposée sur chromatographie d'exclusion stérique (colonne Superdex 200 10-300 analytique) préalablement équilibrée dans un tampon 20 mM Tris HCl pH 7,5, 200 mM NaCl, 20 mM MgCl<sub>2</sub>. La colonne analytique permet de charger des petits volumes de 100 à 500 µL dans le but d'analyser la stabilité du complexe nucléoprotéique.

#### III.3.3.2. Résultats.

##### Purification de l'ARNt.

Après linéarisation du plasmide, l'ARNt est ensuite transcrit *in vitro*, et purifié selon un protocole décrit en Matériels et Méthodes. La révélation par ombrage UV du gel acrylamide

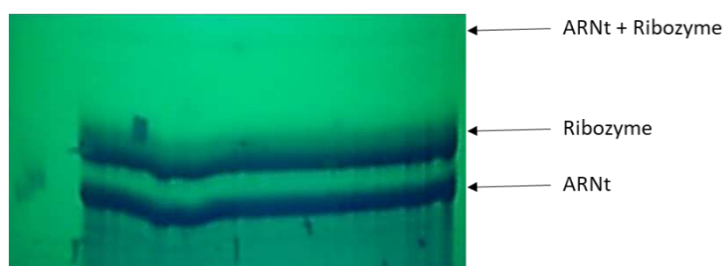


Figure 68 Analyse de la purification de l'ARNt<sup>lle</sup><sub>AUU</sub>. Révélation par ombrage UV ( $\lambda = 254$  nm) du gel 10% acrylamide 7M urée.

en condition dénaturante (Figure 68) permet d'observer deux bandes majoritaires correspondantes au ribozyme et à l'ARNt<sup>lle</sup>, ainsi qu'une bande peu intense en haut du gel correspondant à la fraction ARNt – ribozyme attestant une bonne efficacité de coupure du ribozyme. La bande de l'ARNt a été ensuite découpée et les molécules d'ARNt ont été extraites



par électro élution. Les molécules d'ARNt ont été précipitées à l'éthanol puis suspendues dans 100  $\mu$ L de tampon adéquat. Les ARNt sont concentrés à 12 mg/mL soit 495  $\mu$ M.

### Essai de tests d'interaction entre le complexe KEOPS et de l'ARNt.

Nous avons également étudié les propriétés de fixation *in vitro* du complexe KEOPS humain avec l'ARNt<sup>lle</sup><sub>AUU</sub> par la technique de retard sur gel. La figure 69 montre que la présence du complexe provoque un retard de migration des ARNt<sup>lle</sup><sub>AUU</sub> radiomarqués sur gel d'acrylamide en condition native, détectable dès 150 nM de protéine, et visible sous la forme d'une bande discrète retardées correspondant au complexe nucléoprotéiques (figure 69).

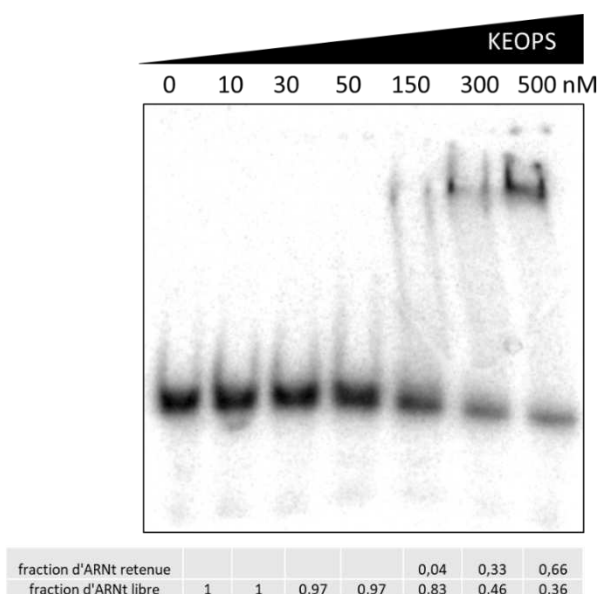
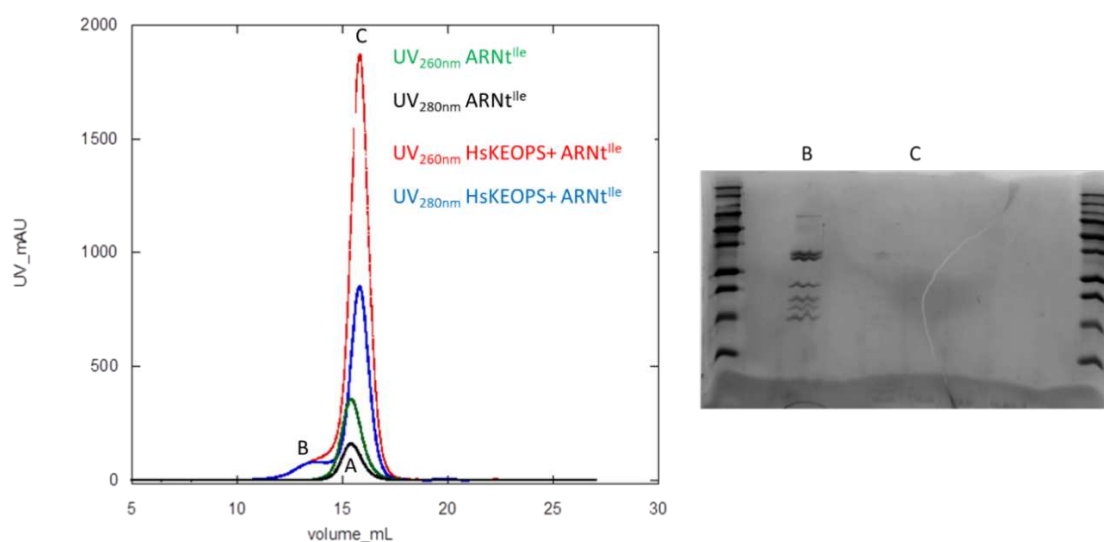


Figure 69 Propriétés de fixation à l'ARNt par le complexe KEOPS humain par retard sur gel. Les concentrations croissantes du complexe KEOPS sont écrites en haut du gel. Les fractions d'ARNt libre et retenue par le complexe KEOPS calculées à partir de l'aire des bandes par ImaJ sont notées dans le tableau en bas du gel.

L'intensité des bandes mesurées par ImaJ met en évidence que à une concentration 500nM de complexe KEOPS, 66% de l'ARNt<sup>lle</sup><sub>AUU</sub> substrat sont retardés. Malheureusement la gamme de concentration du complexe KEOPS ne permet pas d'observer un taux de 100% de l'ARNt<sup>lle</sup><sub>AUU</sub> substrat retardés. Nous pouvons tout de même estimer un ordre de grandeur de la constante de dissociation approximatif. Comme  $K_D = \frac{[Sl] \cdot [El]}{[ES]}$ , à concentration majoritaire en enzyme et à concentration limitante en substrat, on peut approximer que  $[El] = [E_T] - [ES]$  et  $[Sl] = [ES]$ . Ainsi  $K_D = \frac{ES \cdot (ET - ES)}{ES}$  et donc  $K_D = [E_T] - [ES]$ . A demi-saturation du substrat :  $[E_T] \gg [Sl]$  et  $[E_T] \gg [ES]$ , donc  $K_D = [E_T]$ . Comme 66 % de l'ARNt substrat sont retardés à

500 nM du complexe KEOPS, on peut approximer que dans ces conditions, le  $K_D$  est d'environ 380 nM ((500/0,66) x 0,5), ce qui a été confirmé plus tard dans le laboratoire par Dr. Sylvie Auxilien, qui a mesuré un  $K_D$  de 260 nM en absence de  $Mg^{2+}$  par anisotropie de fluorescence de l'ARNt possédant une sonde fluorescente (données non publiées).

La quantité d'ARNt purifié nous fait défaut et nous n'avons pas pu reconstituer le complexe nucléoprotéique avec un taux d'occupation du complexe KEOPS à 100 % par l'ARNt. Après une incubation de 1 heure à 4°C, le complexe KEOPS en présence de l'ARNt<sup>lle</sup><sub>AUU</sub> a été purifié par chromatographie d'exclusion stérique analytique Superdex 200 10/300 (GE Healthcare) à pH neutre et en présence de  $Mg^{2+}$  (voir [Matériel et Méthode](#)). Un échantillon d'ARNt<sup>lle</sup><sub>AUU</sub> a été également injecté seul dans les mêmes conditions expérimentales, afin de localiser son volume d'éluion sur cette colonne. Les chromatogrammes d'éluion et le gel SDS-PAGE correspondant aux fractions d'éluion du complexe sont illustrés à la [figure 70](#).



**Figure 70** A gauche est représenté les 2 chromatogrammes d'éluion superposés de l'ARNt<sup>lle</sup> seul (le suivi de l'absorbance à 260 nm et 280 nm est représenté en vert et noir respectivement) et du complexe KEOPS-ARNt<sup>lle</sup> (le suivi de l'absorbance à 260 nm et 280 nm est représenté en rouge et bleu respectivement). Les fractions correspondantes au pic B et C sont analysées sur gel SDS Page en condition dénaturante (à droite).

Un seul pic d'éluion est observé en sortie de colonne, grâce au suivi de l'absorbance à 260 et 280 nm, lors de la purification de l'ARNt<sup>lle</sup> (le pic A de la [figure 70](#)), alors que 2 pics sont visibles lors de la purification du complexe nucléoprotéique (les pics B et C). Le pic B présente une absorbance majoritaire à 280 nm et minoritaire à 260 nm, montrant que les fractions ne contiennent pas d'acides nucléiques. Les fractions du pic B sont analysées sur gel SDS-PAGE. Elles contiennent des protéines de taille correspondant à celles attendues pour les protéines du complexe KEOPS. Les fractions du pic C ne contiennent pas de protéine,

suggérant que l'absorbance mesurée provient de l'ARNt seul. Etant donné que les pics A et C sont élués approximativement au même volume d'environ 15 mL, cela suggère que le complexe nucléoprotéique n'est pas suffisamment stable pour être purifié par chromatographie d'exclusion stérique. En revanche les volumes de rétention des pics A et C sont trop faibles pour pouvoir correspondre à un ARNt unique qui a une masse moléculaire d'environ 28 kDa. Ceci suggère que l'objet élué présente une masse plus volumineuse qu'une molécule d'ARNt monomérique. Il est possible que des interactions attractives entre les molécules d'ARNt se soient produites. L'optimisation de la purification du complexe KEOPS avec son substrat ARNt<sup>lle</sup> sera poursuivie dans le laboratoire.

Malgré l'incapacité de purifier un complexe nucléoprotéique par chromatographie d'exclusion stérique, illustrant la faible interaction de l'ARNt avec le complexe protéique, nous avons décidé d'étudier le complexe KEOPS-ARNt<sup>lle</sup> par microscopie électronique en transmission en mélangeant extemporanément le complexe KEOPS humain et l'ARNt<sup>lle</sup> avec un ratio 1 : 1.

### III.3.3.3. Etudes préliminaires du complexe KEOPS humain en présence de l'ARNt<sup>lle</sup> par microscopie électronique en transmission.

Nous n'avons pas réussi à cristalliser le complexe KEOPS fixé à son substrat ARNt et nous avons donc décidé d'entamer une étude structurale par cryo-microscopie électronique combinée à l'analyse d'images de particules isolées. Le complexe nucléoprotéique présente une masse moléculaire théorique de 140 kDa et, bien que petit, se prête à l'analyse en cryo-microscopie électronique. Ces études ont été réalisées en collaboration avec Catherine Vénien-Bryan de l'IPMC (Sorbonne Université). La cryo-microscopie électronique en transmission (MET) s'est progressivement imposée comme une technique de choix pour la biologie structurale et la détermination de structures à haute résolution (Frank, 1995).

#### Propriétés du microscope électronique en transmission pour l'analyse des échantillons biologiques.

Le microscope électronique est constitué d'un canon à électrons (source des électrons), un système pour accélérer les électrons (avec des tensions d'accélération allant de 60 à 300 keV), un système pour focaliser le faisceau d'électrons (lentilles électromagnétiques)

un porte-échantillon avec un système de refroidissement à l'azote ainsi qu'un système d'enregistrement de l'image formée (figure 71).

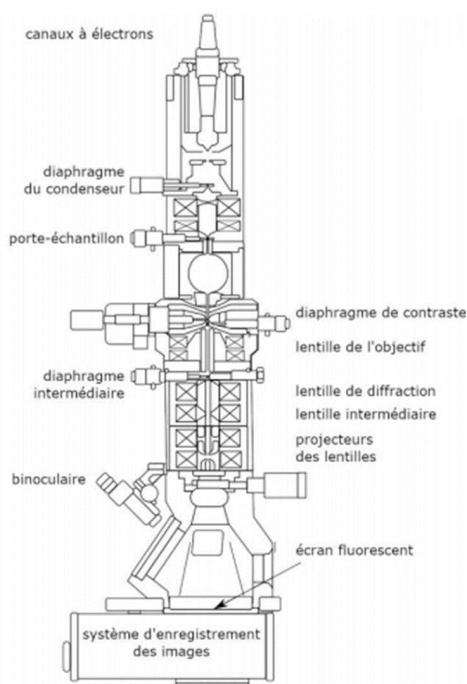


Figure 71 Représentation schématique d'un microscope électronique à transmission (MET).

La colonne dans laquelle se déplace les électrons est sous un vide très élevé afin que seule l'interaction entre les électrons et l'échantillon se produise. Trois défauts optiques existent dans le microscope (Bendory *et al.*, 2020)

Le premier défaut observé est l'aberration chromatique. Les électrons diffusent à différentes vitesses créant ainsi des faisceaux d'électrons à différentes longueurs d'ondes. Le développement technologique des microscopes permet de pallier

ce défaut et de produire des faisceaux d'électrons homogènes.

Le second défaut est l'aberration de sphéricité dû à des anomalies des lentilles. En effet les rayons des électrons traversant au centre de la lentille sont moins déviés que ceux qui sont éloignés et ainsi les électrons vont converger à des différentes distances focales. Pour corriger ces aberrations, une correction sera appliquée lors du calcul de la fonction de transfert du contraste (CTF) (Sigworth, 2016)

Enfin le troisième défaut est l'astigmatisme de l'image créé par la déviation de la convergence. Ce dernier peut se corriger par l'expérimentateur lors de la prise de vue mais aussi lors de l'étape de la correction de la CTF.

L'image de l'échantillon se forme grâce à l'interaction entre les électrons (chargés

négativement) et les atomes de l'échantillon (chargés positivement) (Frank, 2006). Cette interaction onde-particule est étudiée comme un ensemble d'ondes planes ou corpuscules. Deux types d'interaction ont lieu : la diffusion élastique ne montre pas de perte d'énergie lors de leur déviation des électrons et la diffusion inélastique (figure 72). Cette dernière se réalise lorsque l'électron du faisceau communique son énergie à un électron d'un atome de l'échantillon. Cette énergie peut éjecter l'électron de l'atome ou l'électron atteint un niveau d'énergie supérieur libérant un photon qui détruit l'échantillon. Cette perte de photon crée le bruit de fond de l'image. La diffusion élastique produit le signal pour former l'image mais crée un décalage de phase appelé contraste de phase.

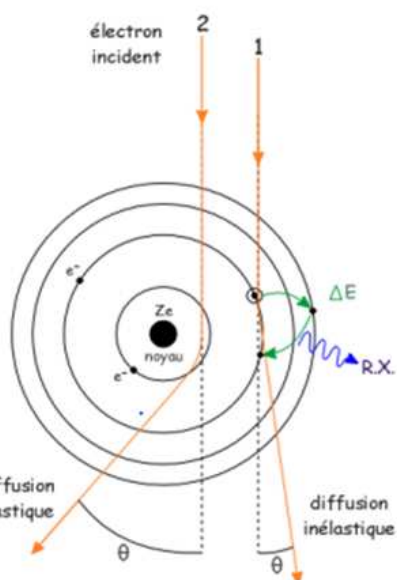


Figure 72 Représentation des phénomènes de diffusion élastique et inélastique d'un faisceau d'électron à travers les différentes couches d'électrons composant l'atome.

Les caméras numériques de dernières générations permettent de détecter directement les électrons contrairement aux caméras CCD sensibles aux photons et qui nécessitaient la conversion des électrons en photons, ce qui était une source du bruit formé sur les images (Bai *et al.*, 2015)

Comme dit précédemment, le calcul de la fonction de transfert du contraste et la prise en compte des corrections imputables aux différents défauts et aberrations du microscope électronique sont indispensables pour former l'image de l'échantillon (Heimowitz *et al.*, 2020).

La CTF est une méthode quantitative pour traduire la fonction d'onde de sortie et former une image finale. Lorsqu'une fonction d'onde d'électrons passe à travers une lentille, la fonction passe par une transformée de Fourier. La CTF est évaluée en déterminant la fréquence spatiale pour laquelle la fonction est nulle. Plus la fréquence spatiale correspondant à une CTF est élevée, plus l'image est de haute résolution.

Pour corriger les aberrations optiques du microscope, dont l'aberration sphérique et l'astigmatisme et pour permettre un plus grand contraste de phase, la transformée de Fourier a été multipliée par une CTF modélisée à plusieurs valeurs de défocalisation différentes (figure 73). Plus il y a d'images (avec des valeurs de défocalisation différentes) plus l'espace de Fourier est recouvert. Afin d'atteindre des données à haute résolution, une série de groupes d'images à des valeurs de défocalisations différentes est donc enregistrée (Velázquez-Muriel *et al.*, 2003).

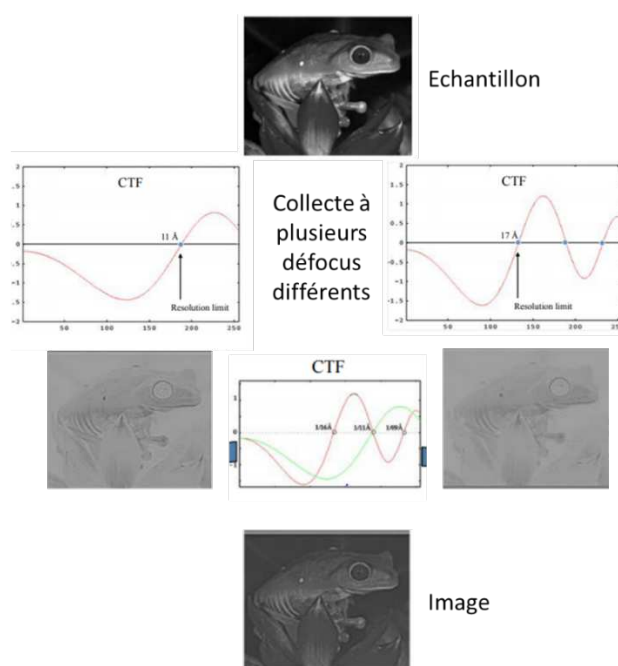


Figure 73 Effet de la CTF sur la résolution et le contraste des images. Pour observer l'effet de la CTF sur l'image, la transformée de Fourier a été multipliée par une CTF qui prend en compte deux valeurs de défocalisation différentes.

L'interaction onde – particule produit également de la chaleur et induit donc des mouvements à l'échantillon biologique qui se trouve à la température de l'azote liquide (-180°C). Les images collectées sont ainsi détériorées par ces mouvements qui introduisent énormément de bruit dans les images. Pour corriger cette limitation, les détecteurs les plus performants enregistrent plusieurs prises de vue dans le même champ : c'est-à-dire que chaque prise de vue enregistre une image "frame", il y a en a un certain nombre par seconde, et l'alignement de toutes ces images forment un film. Afin d'obtenir un film le plus net possible, les images sont sommées de la façon suivante : les deux dernières images sont alignées puis le résultat sera aligné avec l'image précédente jusqu'à la fin des séries d'images correspondant à cette prise de vue. L'alignement des images est un processus automatisé par le logiciel MotionCorr (Scheres, 2012; Li et al., 2013) implémenté dans les suites de programmes Relion 3.0.1 et Cryosparc (Scheres, 2015; Punjani et al., 2017). Lorsque les détecteurs ne peuvent pas collecter les images en mode film, ces dernières sont enregistrées en mode linéaire (Boutin et al., 2016).

#### Préparation des échantillons.

Pour l'observation, l'échantillon est déposé sur une grille (présentant un maillage fin) de 3mm de diamètre. Cette grille est généralement en Cuivre mais peut aussi être en Or. Au-dessus de cette grille, une fine couche de carbone est déposée. Cette couche peut être continue (pour la coloration négative) ou discontinu pour la cryo-MET, dans ce cas cette couche est formée de trous de diamètre variable et espacées régulièrement. C'est sur cette fine pellicule de carbone que l'échantillon se reposera ensuite pour pouvoir être examiné au microscope. Le carbone est généralement une substance hydrophobe. Pour rendre la surface hydrophile, les grilles sont soumises à une décharge lumineuse (plasma) formée par le passage d'un courant électrique au travers d'un gaz à basse pression « glow-discharge » (Vénien-Bryan *et al.*, 2017).

L'étude des objets biologiques nécessite des étapes de mises au point. Avant d'étudier les spécimens en cryo-EM, l'analyse des échantillons par coloration négative permet de rapidement observer les particules, de s'assurer de leurs homogénéités et de la faisabilité de l'étude en condition cryo.

La méthode de coloration négative (Brenner et Horne, 1959) est utilisée pour obtenir des images des macromolécules à très fort contraste. L'échantillon biologique est entouré

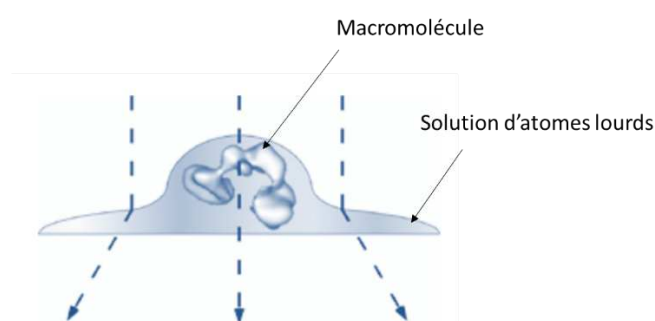


Figure 74 Présentation de la technique de coloration négative. Diffusion des faisceaux d'électrons à travers la solution d'atomes lourds qui entoure l'objet biologique.

d'une solution de sels de métaux lourds tels que les sels d'uranium sur une grille en cuivre recouverte d'une couche de carbone (Harris, 1991). Les atomes lourds sont utilisés parce qu'ils interagissent fortement avec le faisceau d'électrons et diffusent les électrons. Le degré de diffusion des faisceaux d'électrons est beaucoup plus

important dans les régions d'atomes lourds que dans celles où se trouve la protéine (figure 74). Les électrons hautement diffusés peuvent être stoppés par l'objectif du MET alors que la majorité des électrons traversant la protéine ou les zones où il n'y a pas ou peu de colorant sont récupérés sur le plan film. Ceci forme le contraste d'amplitude. Cette technique est rapide à mettre en œuvre mais induit des artefacts tels que parfois un aplatissement des structures qui s'élève à 10%, la résolution est limitée à 15-16 Å (Adrian *et al.*, 1984). De plus, le colorant ne pénètre pas à l'intérieur de l'échantillon, et ne recouvre que la surface de la protéine, comme un "moule" (figure 74). Il est réparti de façon inégale autour de la protéine. Le colorant étant très hydrophile, il se fixera davantage sur les surfaces les plus hydrophiles (Mannella et Frank, 1984), ainsi certaines régions peuvent présenter plus de contraste que d'autres. Cette étape de coloration négative est souvent utilisée pour caractériser l'échantillon biologique et permet de s'assurer de la qualité des particules, de la présence des protéines ainsi que des ligands de grande taille.

La méthode de la cryo-microscopie électronique (cryo-EM) (Adrian *et al.*, 1984; Dubochet *et al.*, 1988; Lepault et Dubochet, 1986) est plus délicate à mettre en œuvre que la méthode de coloration négative. En revanche, la cryo-EM préserve la structure native des macromolécules car elle élimine les artefacts dus à la coloration négative. Les informations sur les images en cryo sont dominées par le contraste de phase contrairement à la coloration négative. Les échantillons déposés sur les grilles sont plongés dans l'éthane liquide à -180°C à



très grande vitesse, ainsi le spécimen est maintenu dans un état hydraté congelé biologiquement intéressant. La vitesse rapide de congélation empêche l'eau de former des cristaux. Celle-ci se trouve sous forme solide amorphe, on dit que la glace est vitreuse et n'arrête pas les électrons contrairement aux cristaux de glace qui interagissent fortement avec les électrons et les arrêtent. Les grilles sont finalement conservées dans de l'azote liquide (-195°C) jusqu'à utilisation.

### III.3.3.3.1. Matériels et Méthodes.

Nous avons décidé d'étudier par cette technique le complexe KEOPS humain en interaction avec l'ARNt<sup>lle</sup><sub>AUU</sub>. Les séquences du complexe KEOPS humain et de l'ARNt<sup>lle</sup> se trouvent en annexe 3 et 4 (cf. page 327). Le complexe KEOPS a été purifié comme exposé au paragraphe III.3.2.1. (cf. page 233) et nous avons produit et purifié l'ARNt<sup>lle</sup> par transcription *in vitro* comme exposé au paragraphe III.3.3.1. (cf. pages 239).

#### Etude préliminaire du complexe KEOPS avec et sans l'ARNt en coloration négative.

Les différents complexes préparés ont été déposés sur des grilles préalablement ionisées (à 25 mA pendant 60 secondes), et incubés pendant 15 à 60 secondes. Les échantillons ont été ensuite colorés avec 2 % d'acétate d'uranyle. Les images ont été collectées sur un MET JEOL 2100 à 200 kV à un grossissement nominal de x30 000 avec une caméra CCD Gatan 2kx2k (correspondant à une taille de pixel de 3,57 Å au niveau de l'échantillon) par Catherine Vénien-Bryan à l'IPMC. Les paramètres de collecte sont donnés dans le tableau suivant (figure 75).

Paramètres	hKEOPS – ARNt <sup>lle</sup>	hKEOPS
JEOL FEG Voltage (kV)	200	200
Détecteur	Caméra CCD	Caméra CCD
Magnification réelle	39000	39000
Apix (Å/pix)	3.57	3.57
Cs (mm)	2	2
Défocus (µm)	0,8 - 4	0,8 - 4
Nombre d'images collectés	50	50
Concentration µM	168	203

**Figure 75 Paramètres de collecte pour l'analyse du complexe KEOPS et de son substrat ARNt<sup>lle</sup> par MET en coloration négative.**

50 images ont été collectées pour chaque complexe. Nous avons analysé les images avec la suite de programmes EMAN2 (Ludtke *et al.*, 1999; Tang *et al.*, 2007). 4000 et 6243 particules ont été sélectionnées ("picking") manuellement sur les images du complexe KEOPS et du complexe KEOPS-ARNt<sup>lle</sup> respectivement. Les particules ont été regroupées en classes 2D selon leur orientation et leur conformation par Relion 3.0.1. (Scheres, 2019). Les meilleures classes ont été sélectionnées dans le but de reconstruire un modèle tridimensionnel initial (classification 3D) (Scheres, 2012).

Etude préliminaire en cryo-EM de complexe KEOPS-ARNt<sup>lle</sup>, collecte sur un microscope Talos Artica à 200 kV.

Grace au consortium d'INSTRUCT ERIC (programme INEXT), nous avons eu l'opportunité de mettre au point les conditions de cryo-EM au Centre National de Biotechnologie de Madrid (CNB). Avant l'étape de vitrification, nous avons déchargé les grilles de type Cu R 0,6/1 (SPI supplies) pendant 60 secondes à 25 mA pour générer une surface hydrophile. 3 µL du complexe KEOPS préalablement incubé avec l'ARNt à un ratio 1: 1 à plusieurs concentrations 20, 10 et 5 µM ont été ensuite vitrifiés avec un Vitrobot (Thermofischer) à 4°C pendant 3 secondes.

Les images ont été collectées sur un microscope Talos Artica à 200 kV de la plateforme Cryo-EM du CNB (Madrid) à un grossissement nominal de x120 000 (correspondant à une taille de de 0,855 Å/pixel au niveau de l'échantillon) avec une caméra Falcon III Les paramètres de collecte sont récapitulés dans le tableau suivant (figure 76).

Paramètres	hKEOPS – ARNt <sup>lle</sup>
Talos Artica Voltage (kV)	200
Détecteur	Falcon III
Magnification	120000
Apix (Å/pix)	0,855
Cs (mm)	2,7
Défocus (µm)	-1 à -3
Nombre d'images collectées	400
Nombre de frames par images	0

A cause d'un problème technique sur le microscope, la collecte en mode "movie" n'était pas disponible. Ainsi la grille préparée avec le complexe KEOPS - ARNt<sup>lle</sup> à 5 µM a été sélectionnée pour acquérir 400 images en mode linéaire. Le traitement des données a été réalisé avec les suites de programmes EMAN2 et Relion 3.0.1 (Tang *et al.*, 2007;

Figure 76 Paramètres de collecte sur le microscope électronique Talos Artica pour l'analyse cryo-EM.

Scheres, 2012). 1760 particules ont été rigoureusement sélectionnées manuellement et ont été rangées dans 20 classes initiales différentes. Ces classes ont été ensuite utilisées pour l'apprentissage du logiciel automatique de "picking", nommé Cryolo (Wagner et Raunser, 2020). 41000 particules ont été ainsi sélectionnées et rangées dans 100 classes différentes. 35 classes ont été sélectionnées pour reconstruire un modèle initial 3D.

Etude en cryo-EM du complexe KEOPS-ARNt<sup>lle</sup> : collecte sur un microscope Titan Krios 300 kV.

De nouvelles grilles pour l'observation en condition cryo ont été préparées sur la plateforme de Cryo-microscopie électronique de Gif-sur-Yvette. Nous avons utilisé des grilles Quantifoil Ru Cu 400 de maille 1,2/1,3 (SPI supplies). 3 µL du complexe KEOPS-ARNt à 3,15 µM ont été déposés sur des grilles à l'aide du vitrobot (Thermofischer) avec les paramètres suivants utilisés : 60 secondes d'incubation et 8 secondes de "blotting" à 4°C.

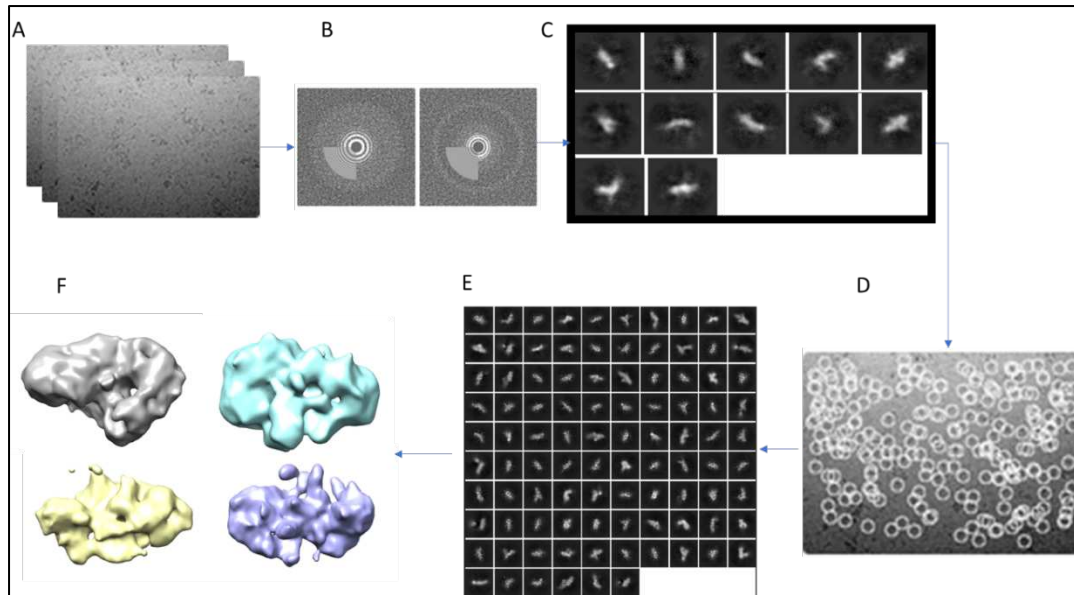
Les données ont été collectées sur un microscope Titan Krios (Thermofischer) à 300 kV de la plateforme Nanolmaging de l'institut Pasteur à un grossissement nominal de 130 000

Paramètres	hKEOPS – ARNt <sup>lle</sup>
Titan Krios Voltage (kV)	300
Détecteur	Caméra K3
Magnification	130000
Apix (Å/pix)	0,68
Cs (mm)	2,7
Défoc (µm)	-1 à -3
Nombre d'images collectées	6903
Nombre de frames par images	40
Dose électron e <sup>-</sup> /Å <sup>2</sup>	78

Figure 77 Paramètres de collecte sur le microscope électronique Titan Krios.

avec une caméra Gatan K3 (correspondant à une taille de 0,68 Å par pixel au niveau de l'échantillon). Les paramètres de collecte sont récapitulés dans le tableau suivant (figure 77). Le traitement des données a été réalisé avec les suites de programmes Relion 3.0.1. et cryosparc (Scheres, 2012; Punjani *et al.*, 2017).

6903 images dont chaque image possède 40 frames ont été collectées.



**Figure 78** Déroulement du traitement de données en cryo-EM. A. Alignement des 30 premiers frames. B. Analyse de la CTF global pour chaque image et sélection des images possédant une résolution comprise entre 3.5 et 5 Å. C. « Picking » manuel des particules sur 40 images. D. « Picking » automatique sur la totalité des images. E. Classification des particules en 100 classes. F. Reconstruction de 4 modèles 3D à partir des 96 classes sélectionnées.

Etant donné la forte dose utilisée ( $78 \text{ e}^- / \text{Å}^2$ ), les 6903 séries d'images ont été alignées avec les 30 premiers clichés, permettant de s'affranchir des 10 derniers pour lesquels l'échantillon se trouvait dégradé par l'application de la forte dose en électrons par MotionCorr implémenté dans Relion 3.0.1 (figure 78 A). En effet la dose accumulée sur ces derniers "frames" peuvent détruire l'échantillon, et détériorer la résolution. Une analyse de la CTF (fonction de transfert de contraste) par CTFind4 a été réalisée et seules les images comprises entre 3.5 et 5 Å de résolution ont été gardées soit 6054 images (figure 78 B). 2033 particules ont été sélectionnées manuellement sur 45 images différentes (la taille de la boîte utilisée est de 256 pixels et la taille de la boîte utilisée pour l'extraction des particules est de 300 pixels). 20 classes 2D ont été définies à partir des 2033 particules et 12 classes satisfaisantes ont été sélectionnées (figure 78 C) pour l'apprentissage du logiciel automatique de « picking », nommé Topaz, implémenté dans la suite cryosparc (figure 78 D) (Bepler *et al.*, 2019). 1 251 238 particules ont été isolées et rangées dans 100 classes 2D différentes (figure 78 E). 96 classes 2D ont été sélectionnées pour reconstruire 4 modèles initiaux 3D (figure 78 F).

### Modélisation *in silico* de la structure tridimensionnelle de l'ARNt<sup>lle</sup><sub>AUU</sub>.

Un modèle *in silico* de l'ARNt<sup>lle</sup><sub>AUU</sub> a été réalisé à partir de la structure tridimensionnelle de l'ARNt<sup>lle</sup> de *Oryctolagus cuniculus* (code PDB 4UJE chaîne A) par le serveur RNAcomposer (Antczak *et al.*, 2016). Ce modèle 3D *in silico* est réalisé dans le but de vérifier la présence de l'ARNt dans la structure de cryo-EM. De plus, si la présence de l'ARNt est confirmée ce modèle servira à affiner la structure de l'ARNt des données expérimentales de cryo-EM (Popenda *et al.*, 2012).

#### III.3.3.3.2. Résultats.

##### Coloration négative et construction d'un modèle préliminaire.

L'analyse de 50 images du complexe KEOPS et 50 images du complexe KEOPS-ARNt<sup>lle</sup> collectées sur un microscope JEOL à 200 kV en coloration négative ont permis d'analyser les objets à basse résolution et de calculer un modèle initial (figure 79). Les images collectées

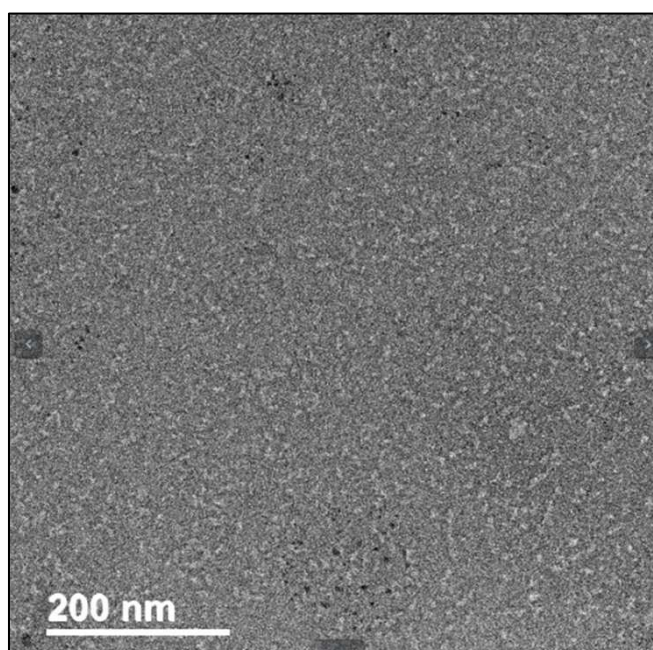


Figure 79 Image collectée du complexe KEOPS avec l'ARNt<sup>lle</sup> sur le microscope JEOL à 200 kV. La particule semble homogène et est suffisamment séparée des particules voisines pour l'analyse.

montrent que le complexe KEOPS avec l'ARNt est suffisamment homogène. Nous n'observons pas d'agrégats et les particules sont convenablement séparées ce qui est encourageant pour l'analyse en particules isolées. 4000 et 6243 particules pour les complexes KEOPS et KEOPS-ARNt<sup>lle</sup>, respectivement, ont été sélectionnées manuellement pour créer les classes 2D.

Les classes 2D montrent un objet linéaire mais nous n'avons pas pu observer de différences significatives entre le complexe KEOPS seul (figure 80 A) et en présence de l'ARNt<sup>lle</sup> (figure 80 B). Cependant, nous pouvons observer sur les classes 2D du complexe KEOPS seul (figure 80 A) que certaines orientations mettent en évidence un objet plus courbé comparé à l'objet observé sur les classes 2D du complexe KEOPS-ARNt<sup>lle</sup>. Cette différence suggère qu'en

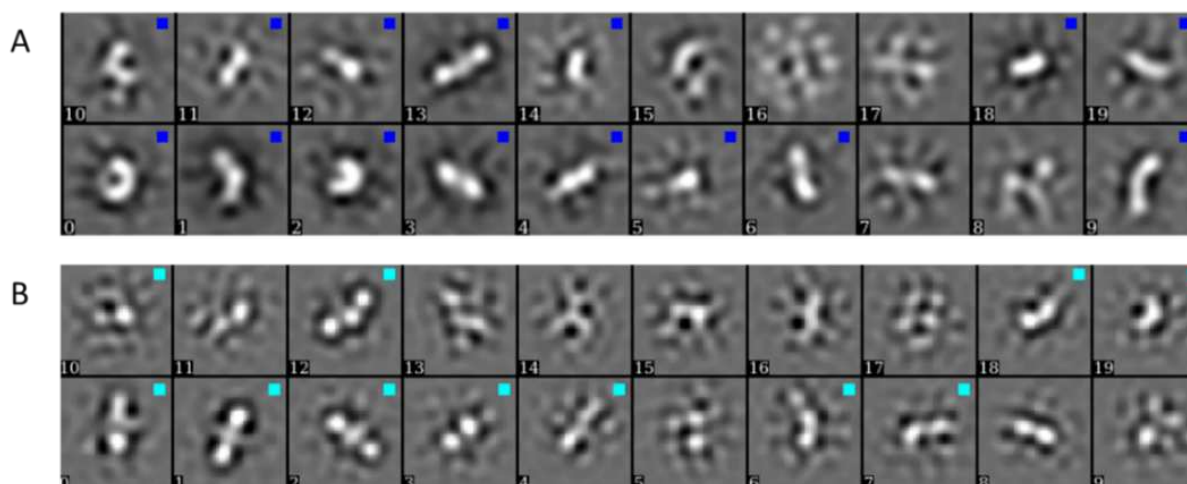
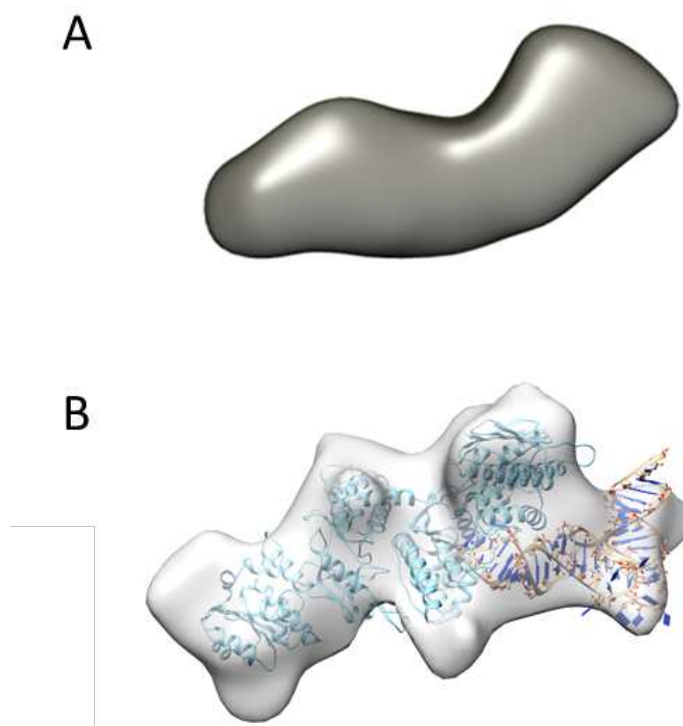


Figure 80 Classes 2D des particules collectées par MET en coloration négative (A) du complexe KEOPS humain et (B) du complexe KEOPS-ARNt<sup>lle</sup>.

présence de l'ARNt, un changement de conformation du complexe KEOPS s'est produit ou que nous n'avons pas sélectionné des particules dans cette orientation. Après analyse des classes 2D seules les classes de meilleure qualité ont été choisies (celles annotées avec un carré de couleur) pour reconstruire un modèle 3D initial.

Une structure tridimensionnelle à 20 Å de résolution a été obtenue pour les particules du complexe KEOPS, permettant de mettre en lumière plusieurs détails de la structure (figure 81 A). La reconstruction 3D révèle un objet linéaire compatible avec les modèles *in silico* du complexe KEOPS de l'Homme (cf. § III.3.3.2). L'objet s'étend sur 140 Å et présente un volume de l'ordre de 209000 Å<sup>3</sup> compatible avec l'extension maximale mesurée par SAXS qui est de 138 Å (cf. § III.3.2.2.). Concernant les particules du complexe KEOPS avec l'ARNt<sup>lle</sup>, une structure 3D à 16 Å de résolution a été obtenue (figure 81 B), qui montre un objet également linéaire et un volume de 275500 Å<sup>3</sup> potentiellement compatible avec le complexe KEOPS et l'ARNt<sup>lle</sup>. Ces valeurs sont approximatives car l'application du colorant en coloration négative augmentent significativement la masse de l'objet observé.



**Figure 81** Reconstruction 3D des particules sélectionnées par classification 2D (A) de l'objet KEOPS humain à 20 Å de résolution et (B) du complexe KEOPS-ARNt<sup>lle</sup> à 16 Å de résolution. La reconstruction a été réalisée par la méthode Bayesian implémentée dans Relion 3.0.1. Le modèle KEOPS humain et le modèle de l'ARNt<sup>lle</sup> ont été superposés dans la reconstruction 3D à 16 Å de résolution à l'aide de l'outil de visualisation chimera.

De plus, les modèles du complexe KEOPS et de l'ARNt<sup>lle</sup> ont été superposés manuellement au mieux sur le modèle 3D à 16 Å de résolution. Nous avons orienté la boucle anticodon de l'ARNt à proximité du site actif de la sous-unité catalytique OSGEP. Dans ce modèle 3D, le corps de l'ARNt s'oriente du côté des sous-unités LAGE3 et GON7 ([figure 81 B](#)).

Cette superposition met en évidence que le volume du modèle 3D est compatible avec le complexe KEOPS et une molécule d'ARNt<sup>lle</sup>. Cependant, nous arrivons à la limite résolutive de la coloration négative qui est de 15-16 Å du aux artefacts décrits ([cf. Matériels et Méthodes page 248](#)) et au grain du colorant. Ces résultats sont encourageants et justifient de poursuivre l'analyse structurale du complexe nucléoprotéique en cryo-microscopie électronique, afin d'obtenir des données à haute résolution, ce qui nous permettra de confirmer la position de l'ARNt dans ce modèle.

Optimisation des grilles pour la cryo-EM et construction d'un modèle initial du complexe KEOPS-ARNT<sup>lle</sup> à basse résolution.

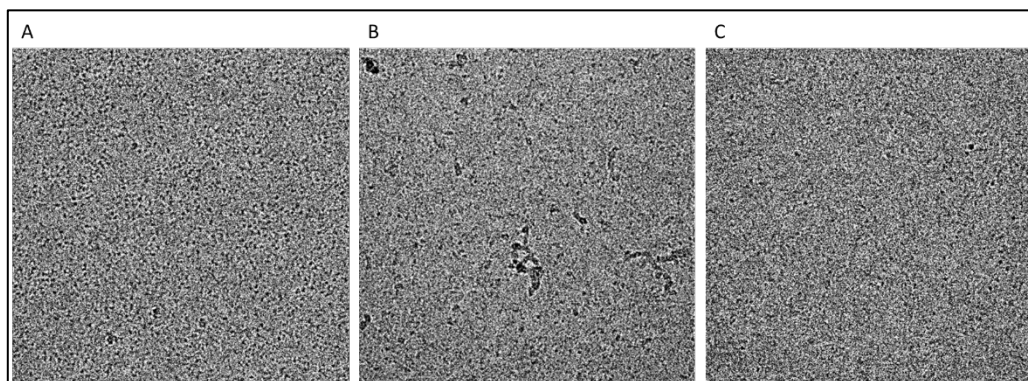


Figure 90 Images collectées sur le microscope Talos Artica à 200 kV des complexes KEOPS-ARNT<sup>lle</sup> à 20 (A), 10 (B) et 5 μM (C). Les particules sont en noirs et la répartition de l'objet est plus homogène sur l'image C.

Nous avons préparé différentes grilles à 3 concentrations différentes 20, 10 et 5 μM du complexe KEOPS - ARNT<sup>lle</sup>. Les images ont été analysées sur le microscope Talos Artica à 200 kV (CNB Madrid). Les différentes images du complexe à 20 μM et 10 μM montrent que les particules sont trop concentrées et ne sont pas suffisamment séparés (figure 90 A et B), alors que l'image montre que le complexe à 5 μM (figure 90 C) se réparti de façon homogène et que les particules sont suffisamment séparées pour l'analyse en particule isolée. Cette répartition homogène justifie la collecte de 400 images à cette concentration. Les images ont été traitées au laboratoire avec les suites de programmes EMAN2 et Relion 3.0.1 (cf. [Matériels et Méthodes](#)). A partir des particules sélectionnées avec le logiciel Cryolo, 100 classes 2D ont été réalisées pour ranger les particules selon leurs orientations et leurs conformations sur les images. Les 35 meilleures classes (figure 91) montrent un complexe linéaire.



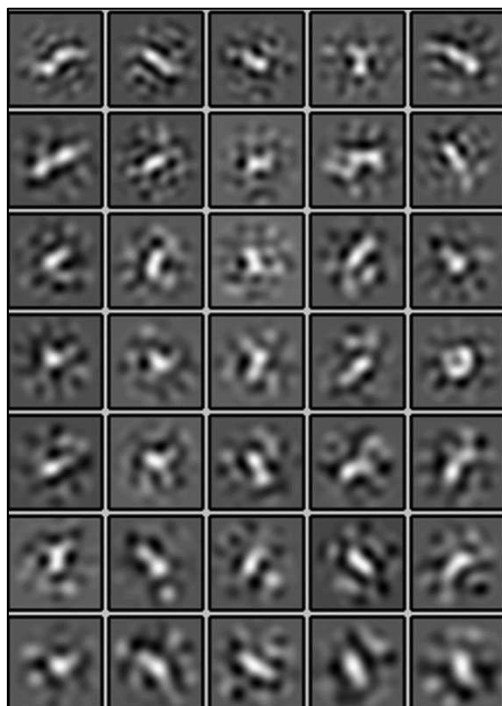


Figure 91 Classes 2D sélectionnées à partir de 19557 particules du complexe KEOPS -ARNt<sup>lle</sup> "pickées" sur les images à l'aide du logiciel CRYOLO.

Nous observons beaucoup de bruit sur les classes car les images ont été collectées en mode linéaire. Malgré le manque de détails, quelques classes montrent un objet en forme de croix ; ce pattern est proche du modèle en coloration négative du complexe KEOPS - ARNt<sup>lle</sup>. Un modèle 3D initial à partir de ces 35 classes a été construit (figure 92) et montre que l'objet est linéaire avec une densité électronique compatible à l'ARNt. Toutefois les classes 2D présentent énormément de bruit et la résolution du modèle est de 12 Å. Le modèle 3D initial montre que la position de l'ARNt serait peut-être différente de celle suggérée par le modèle initial en coloration négative.

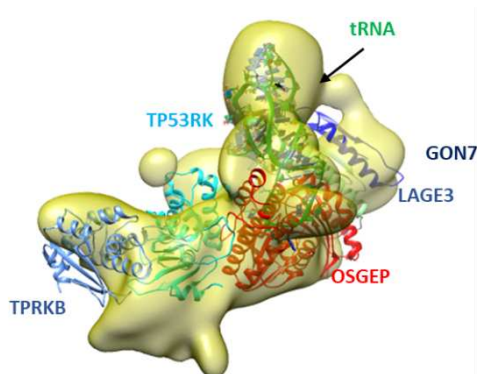


Figure 92 Modèle initial 3D à 12 Å de résolution du complexe KEOPS -ARNt<sup>lle</sup> par analyse cryo-EM.

Ces résultats nous ont encouragé à poursuivre l'étude en cryo-EM du complexe KEOPS-ARNT<sup>lle</sup> et de collecter les images en mode "movie" en utilisant un microscope plus performant.

Après avoir préparé les grilles avec le complexe KEOPS - ARNT sur la plateforme Gif-sur-Yvette, les images ont été enregistrées sur un microscope Titan Krios , 300 kV sur la plateforme de Nanomaging de l'institut Pasteur. Les images ont été traitées au laboratoire avec les suites de programmes Relion 3.0.1 et Cryosparc (*cf. Matériels et Méthodes*). A partir des particules sélectionnées avec le logiciel Topaz implémenté dans Cryosparc, 100 classes 2D ont été réalisées pour ranger les particules (environ un million) selon leurs orientations et leurs conformations sur les images. Certaines classes montrent beaucoup de signal et de détails alors que peu de bruit est observé. Selon certaines orientations on peut observer la projection de la densité électronique supplémentaire pouvant correspondre à l'ARNT au 1/3 de l'objet (*figure 93 B1, B8, C5, C6, C9, E8*). Certaines classes montrent beaucoup de finesse et sont de meilleures qualités que les classes 2D réalisées avec les jeux de données précédents. Par exemple, les classes A3, A6, A7, B1, C5, C6, C9, E5, E8, G1, G3, G5, G10, H9 et J4 montrent beaucoup de détails et on peut deviner possiblement certains éléments de structure secondaire (*figure 93*).

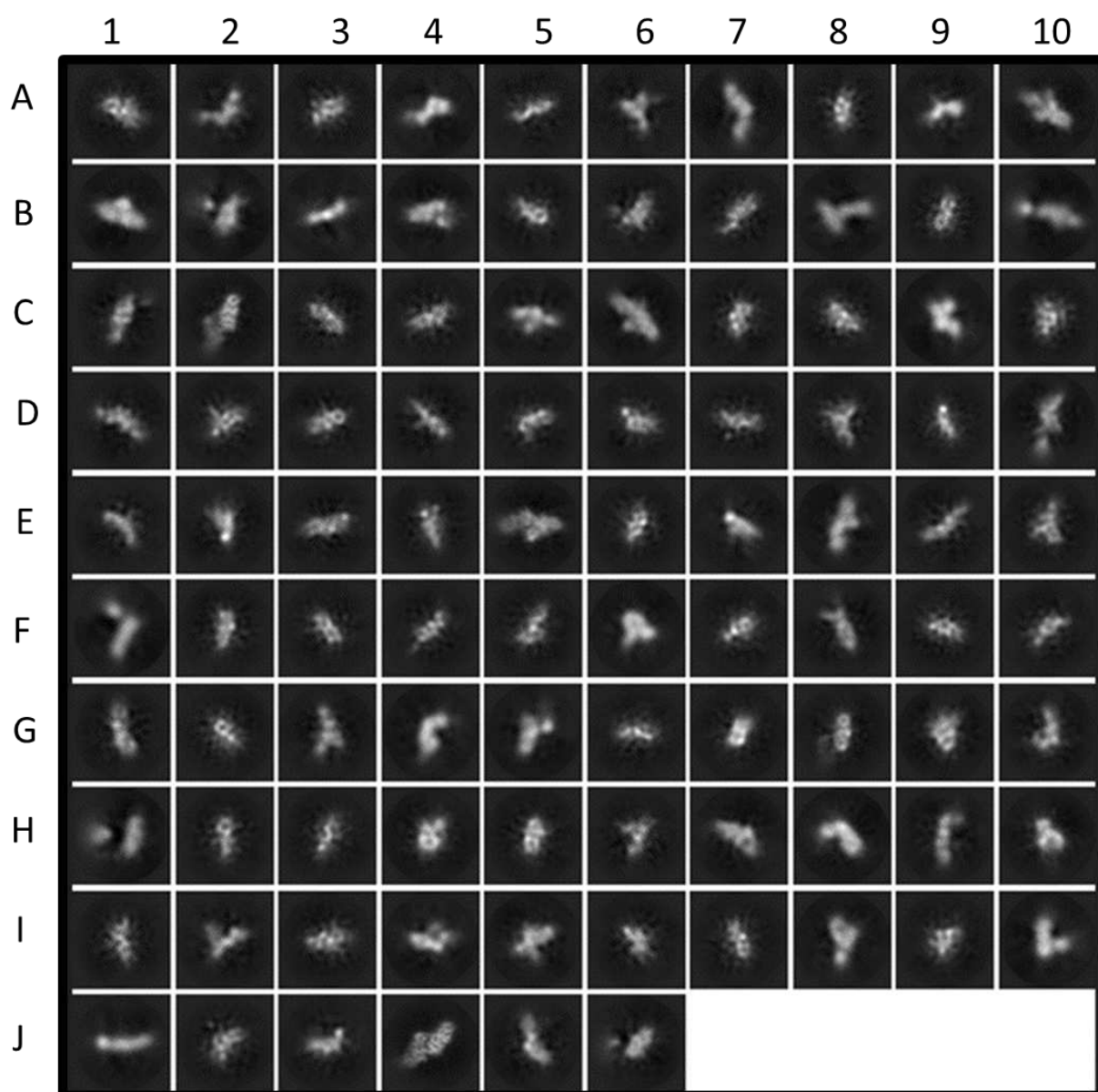


Figure 93 Classification 2D selon l'orientation des particules sur les images sélectionnées par Topaz. Les lettres et les chiffres sont ajoutés pour faciliter la discussion des classes 2D.

A partir de 96 classes 2D (1 251 238 particules) des étapes de reconstruction de 4 modèles initiaux en imposant une symétrie C1 ont été réalisées permettant de classer les différents objets, de s'affranchir au mieux des mauvaises particules mais aussi de détecter l'hétérogénéité de l'échantillon étudié avant d'affiner le modèle 3D (figure 94). Les 4 modèles initiaux A, B, C et D ont été reconstruits à partir de 423 528 particules, 266 898 particules, 295 642 particules et 265170 particules respectivement et atteignent une résolution de seulement 9 Å. Le modèle A présente le plus grand nombre de particules et semble être compatible avec

le complexe KEOPS-ARNt<sup>lle</sup>. Les modèles B, C et D, quant à eux, ne semblent pas compatibles avec une molécule d'ARNt complexée au complexe KEOPS. Au bilan, plus de 65 % des particules ne semblent pas complexées à l'ARNt, ce qui en fait une fraction importante et compatible au Kd estimé entre l'ARNt et le complexe KEOPS.

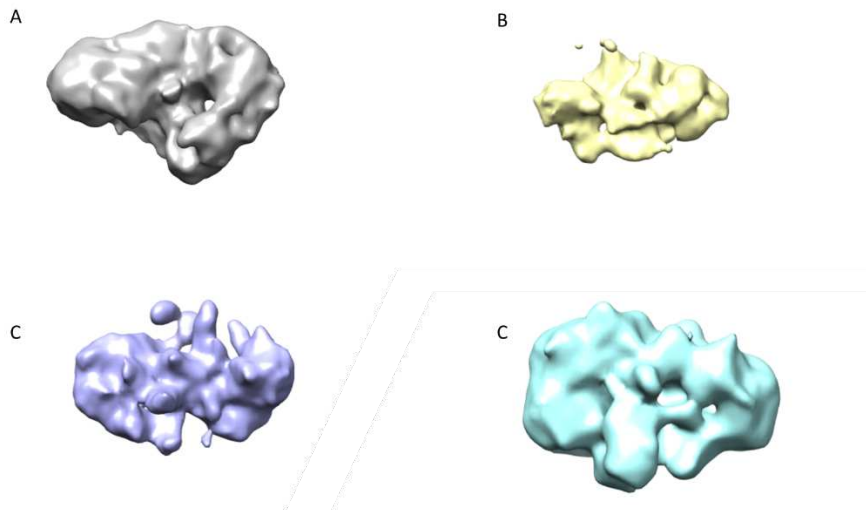


Figure 94 Reconstruction de 4 modèles initiaux à ~ 9 Å de résolution à partir des 96 classes 2D sélectionnées dans Cryosparc.

Il est, également, possible que le complexe KEOPS présente de la flexibilité intrinsèque augmentant l'hétérogénéité de l'échantillon. Malheureusement, pour le moment, nous ne sommes pas parvenus à un modèle initial à une résolution satisfaisante pour débiter les étapes d'affinement.

La résolution obtenue à 10 Å est décevante face à la qualité des classifications 2D. Même si nous travaillons sur un objet petit (140kDa) dépourvu de symétrie et peut-être flexible, nous avons identifié plusieurs pistes qui nous permettront d'obtenir une meilleure résolution :

i) Nous aimerions retraiter les données cryo-EM en sélectionnant rigoureusement les classes 2D.

ii) De nombreux paramètres utilisés lors de la reconstruction 3D devront être également réévalués.

iii) Nous pourrions également utiliser les plus belles classes 2D pour resélectionner les particules à l'aide du logiciel Topaz. Ce nouveau jeu de particules sélectionnées devra être retraité.

Dans un second temps, si les données collectées ne suffisent pas à obtenir une structure 3D à une résolution satisfaisante, la préparation du complexe KEOPS-ARNt<sup>lle</sup> devra être réévaluée. De nouvelles méthodes se développent ces dernières années pour améliorer la résolution des données en cryo-EM. Par exemple, des pontages chimiques peuvent être introduits pour stabiliser l'interaction entre les sous-unités des complexes oligomériques et/ou l'interaction avec le substrat. Également des pontages chimiques sont souvent réalisés entre les protéines de petite taille et un anticorps augmentant la taille de l'objet étudié et facilitant ainsi le « picking » des particules (Pal *et al.*, 2018). Ces différentes méthodes seront envisagées si le traitement des données n'aboutit pas. Aussi les essais de cristallisation du complexe KEOPS avec l'ARNt sont toujours en cours au laboratoire, et des tests de cristallisation avec un ARNt tronqué sont également envisagés.

## Chapitre IV

### IV. Identification de la cinquième sous-unité du complexe KEOPS chez les archées.

#### IV.1. Présentation de l'étude.

Nous avons vu dans le chapitre précédent (chapitre III) que l'identification de GON7 de l'Homme a permis d'élargir la famille des protéines GON7 aux métazoaires et aux champignons (Wan *et al.*, 2017). Malgré le faible taux d'identité de séquence entre GON7 et yGon7 (17 % d'identité de séquence), les deux protéines adoptent la même structure tridimensionnelle lorsqu'elles sont en complexe avec leurs partenaires (respectivement LAGE3 et Pcc1) (codes PDB 6GWJ et 4WXA). Chez l'Homme, nous avons mis en évidence que la présence de GON7 permet l'agencement du complexe KEOPS avec une stœchiométrie 1 :1 :1 :1 :1, comme chez la levure (Zhang *et al.*, 2015), alors que son absence permet une dimérisation du complexe KEOPS, avec une stœchiométrie 2 :2 :2 :2. Il a été vu que le taux de la modification t<sup>6</sup>A est augmenté sur les ARNt substrats en présence de GON7, suggérant que le complexe pentamérique est plus actif (Wan *et al.*, 2017, Arrondel *et al.*, 2019). A ce jour, aucun homologue de séquence ou de structure de Gon7 n'a été identifié chez les archées. Le complexe d'archée s'agence avec une stœchiométrie 2 :2 :2 :2. La surface hydrophobe de Pcc1 forme comme chez l'Homme ou la levure la plateforme de dimérisation, et l'octamère assure la synthèse de la modification t<sup>6</sup>A sur les ARNt substrats (Mao *et al.*, 2008; Perrochia *et al.*, 2013). Au regard de la présence des protéines GON7 et yGon7 chez deux groupes d'eucaryotes phylogénétiquement distants, de fortes présomptions quant à l'existence d'une cinquième sous-unité chez les archées sont émises.

Lors d'une analyse du contexte génomique des gènes de KEOPS de *Pyrococcus abyssi* GE5, réalisée dans le but de construire des souches dont les gènes codant les protéines du complexe KEOPS seront étiquetées au locus, ce fut par hasard que Marie-Claire Daugeron et Tamara Basta (de l'équipe Biologie Cellulaire des Archées I2BC) ont identifié deux gènes annotés *PCC1* dans ce génome. Les deux gènes sont localisés sur deux loci différents. Ce deuxième gène est renommé *PCC2* pour des soucis de clarté.

Dans le phylum *Thermococcales*, le gène *PCC1* est localisé dans un opéron encadré par les gènes codant une ribonucléoprotéine, impliquée dans la biogénèse des ribosomes, et le

facteur d'élongation de la traduction EF-2 (figure 95). Quant au gène *PCC2*, il est localisé dans un autre opéron, encadré par les gènes codant la protéine RecJ like (probablement une exonucléase) et une protéine ribosomale de la sous-unité 30S (figure 96). Dans le phylum *Thermococcales*, la localisation des gènes *PCC1* et *PCC2* est éloignée, suggérant un événement ancien de la duplication.

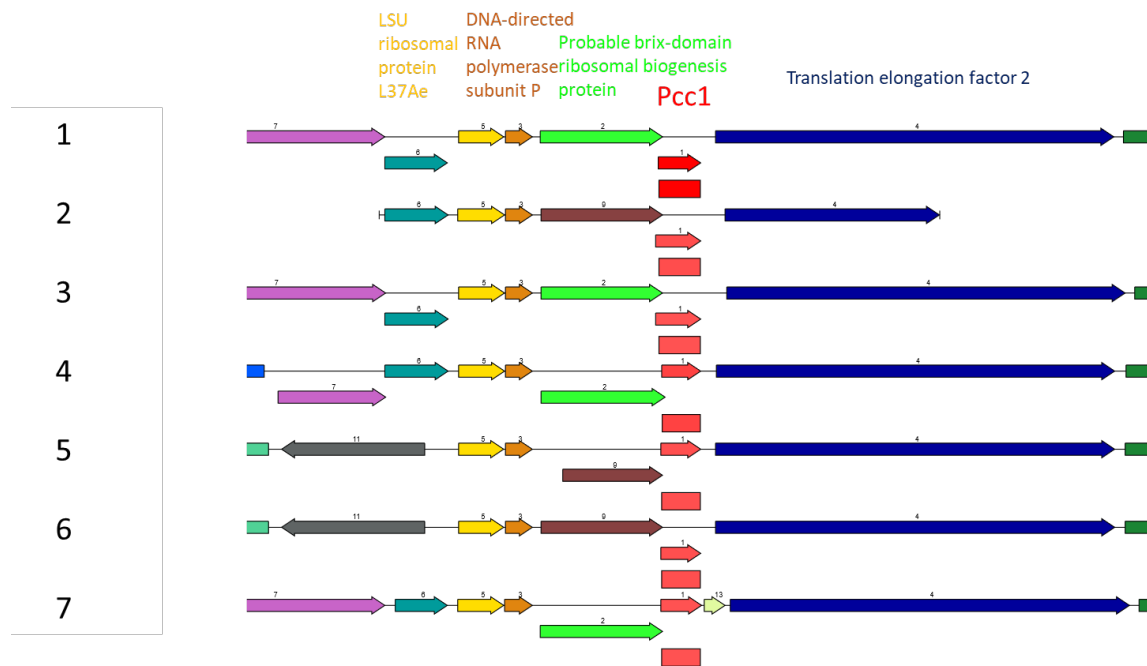


Figure 95 Contexte génomique de *Pcc1* dans les génomes (1) *Pyrococcus abyssi* GE5, (2) *Thermococci archaeon* strain B21\_G6, (3) *Pyrococcus* sp. NCB100, (4) *Pyrococcus* sp. NA2, (5) *Pyrococcus horikoshii* OT3, (6) *Pyrococcus horikoshii* strain UBA8834, (7) *Pyrococcus* sp. ST04. La syntenie a été réalisée à l'aide d'une recherche du gène *PCC1* (en rouge) localisé aux positions 280821-281054 sur le génome de *Pyrococcus abyssi* GE5 et comparé à 6 autres génomes dans la banque de données PATRIC (<https://www.patricbrc.org/>).

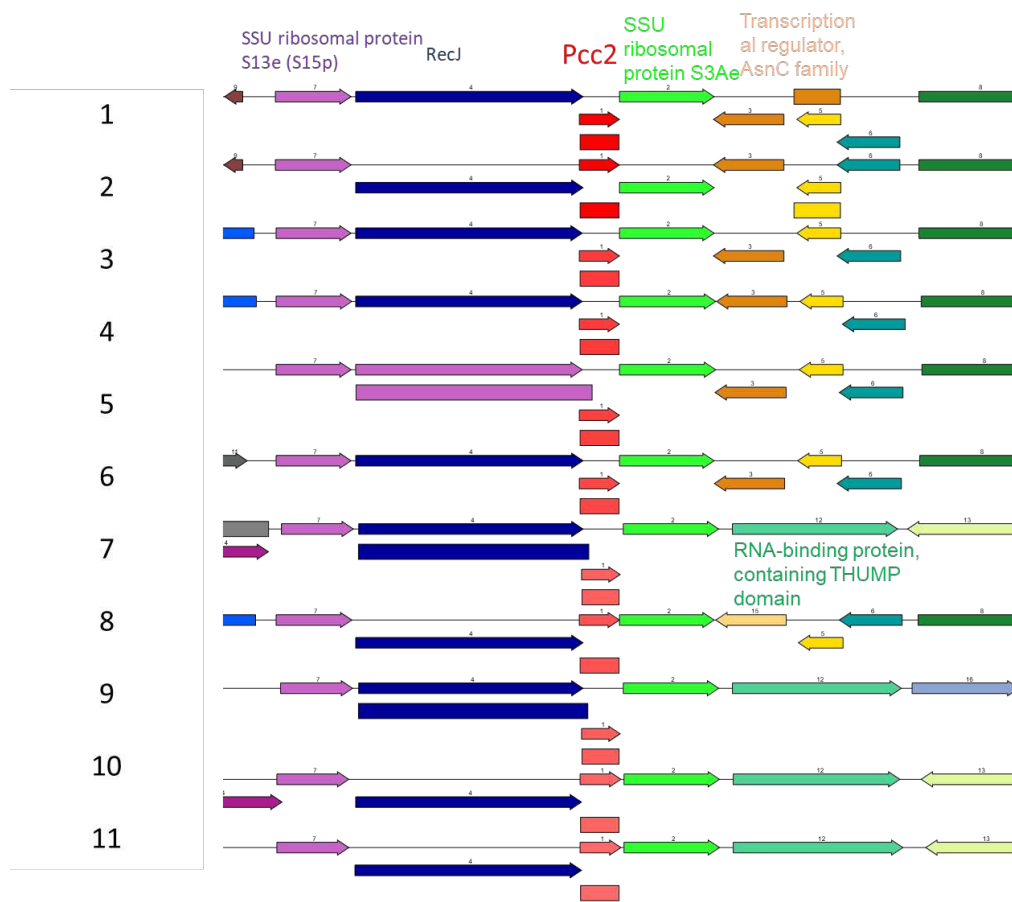


Figure 96 Contexte génomique de *PCC2* dans les génomes (1) *Pyrococcus furiosus* COM1, (2) *Pyrococcus furiosus* DSM3638, (3) *Pyrococcus horikoshii* OT3, (4) *Pyrococcus* sp. ST04, (5) *Pyrococcus* sp. NA2, (6) *Pyrococcus abyssi* GE5, (7) *Thermococcus barophilus* MP, (8) *Pyrococcus yayanosii* CH1, (9) *Thermococcus* sp. ES1, (10) *Thermococcus* sp.4557, (11) *Thermococcus celericrescens* strain DSM 17994. La synténie a été réalisée à l'aide d'une recherche du gène *PCC2* (en rouge) localisé aux positions 1509435-1509686 sur le génome de *Pyrococcus furiosus* COM1 et comparé à 10 autres génomes dans la banque de données PATRIC (<https://www.patricbrc.org/>).

Pour mieux comprendre la relation évolutive qui existe entre ces deux gènes, une comparaison des séquences a été réalisée par l'équipe de Biologie cellulaire des Archées. La majorité des génomes d'archées connus possèdent ces deux versions du gène, renommés *PCC1* et *PCC2*, à l'exception des génomes réduits chez qui le gène *PCC2* n'a pas été identifié comme les nanoarchées du clade DPAN (figure 97).



Tree scale: 1

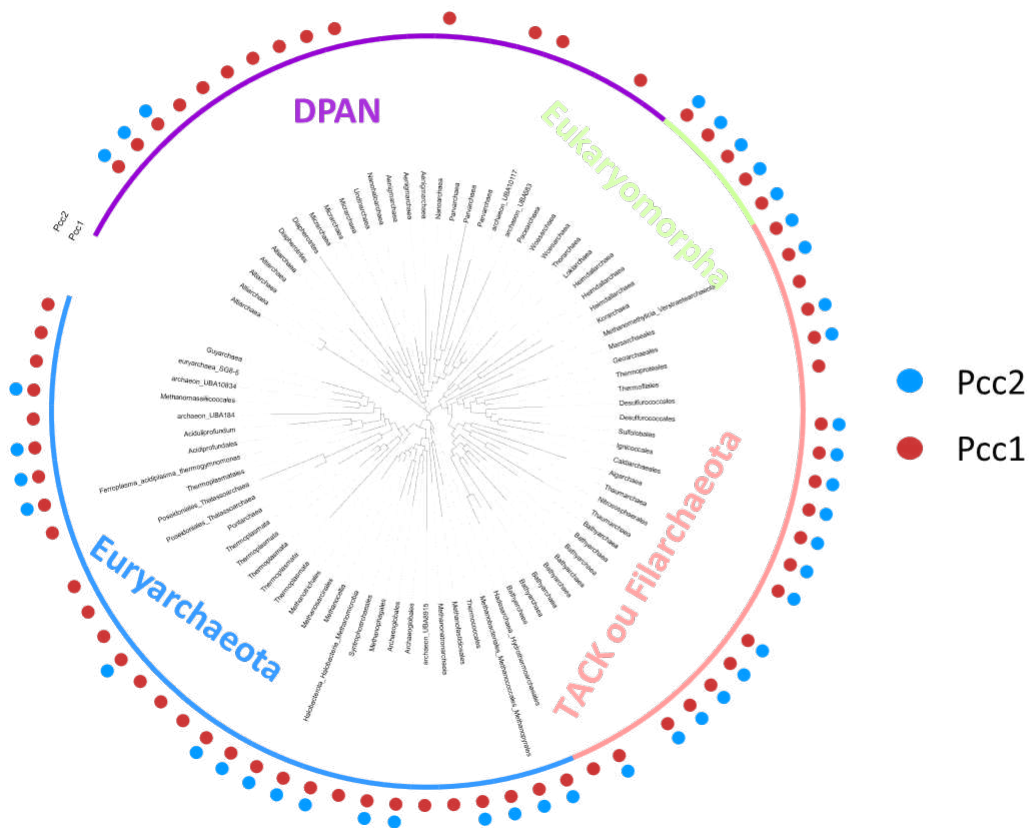


Figure 97 Distribution des gènes PCC1 et PCC2 au sein des Archées. Chaque phylum est représenté par un organisme référent dont le séquençage de son génome est entier.

L'arbre phylogénétique construit à partir d'un alignement des séquences de Pcc1 et Pcc2 présente deux groupes monophylétiques séparés par une longue branche. Une telle topologie peut correspondre à un évènement ancestral de duplication donnant lieu à deux gènes qui ont suivi une trajectoire évolutive distincte au sein des différentes lignées d'archées. Cette topologie et la présence simultanée des deux gènes dans la grande majorité des phylums d'Archées indique que Pcc1 et Pcc2 sont des paralogues (figure 1 de l'article en cours page 277).

A titre d'exemple, Pcc1 et Pcc2 de *Pyrococcus abyssi* partagent 20 % d'identité de séquence. Les séquences Pcc1 et Pcc2 connues chez les archées ont été alignées selon leur composition en acides aminés par un profil HMMER (Potter *et al.*, 2018), permettant de dresser un logo de la séquence pour ces deux protéines (figure 98). Les deux versions sont similaires selon 1 motif "signature" commun F/Y<sub>14</sub>-X-S/N/D/E<sub>16</sub>-X-K/D/E<sub>18</sub>-A20-X-X-X-

Y/F/A/L<sub>24</sub>-X-A/S<sub>26</sub>-L/I/V<sub>27</sub>-X-P/I/V<sub>29</sub>-E/D<sub>30</sub> sur les deux séquences logos. En revanche, elles diffèrent selon un motif "signature" caractéristique de chacune, composé de deux aspartates D<sub>70</sub>-D<sub>71</sub> dans les séquences Pcc2 et d'une asparagine et d'une sérine N<sub>67</sub>-S<sub>68</sub>-X-X-R/K<sub>71</sub>-W/L<sub>72</sub> dans les séquences Pcc1 (figure 98). Le motif commun et les deux motifs caractéristiques de Pcc1 et de Pcc2 sont localisés sur leurs structures secondaires respectives.

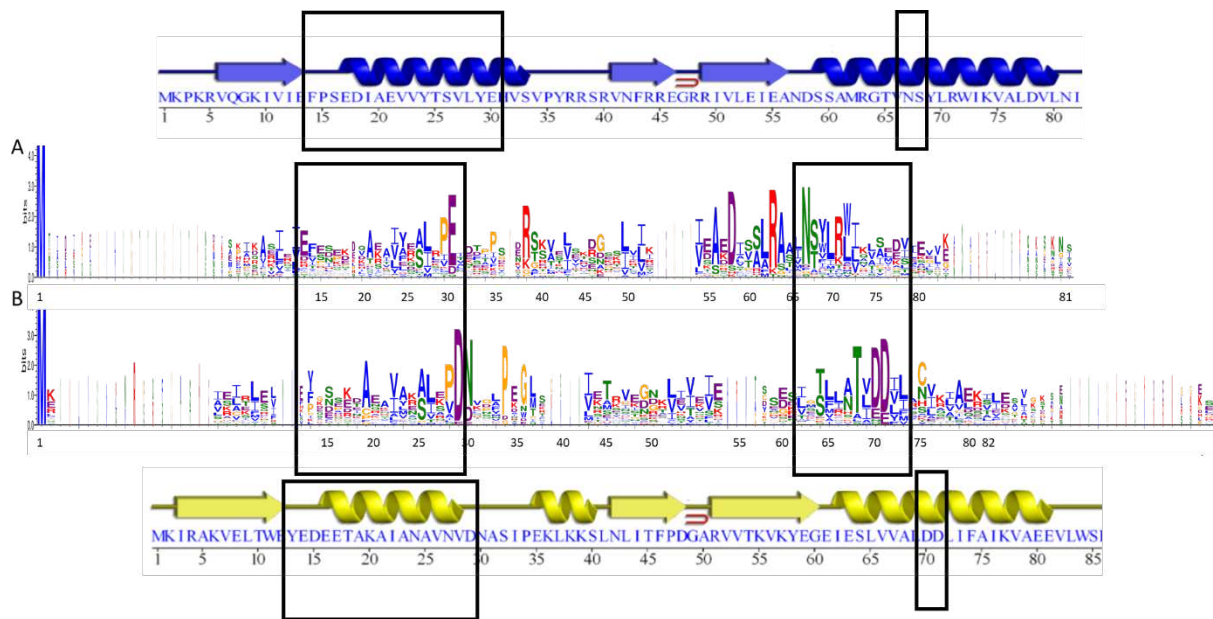


Figure 98 Les séquences logos des protéines (A) Pcc1 et (B) Pcc2 sont des représentations graphiques des séquences alignées des protéines de chaque famille. La taille relative de chaque lettre des acides aminés illustre leur fréquence à chaque position et leur degré de conservation à travers les séquences alignées. Les blocs de séquences signatures (partagé entre les deux versions pour celui localisé en N-terminal, et spécifique de chacune pour celui localisé en C-terminal), sont encadrés. Le domaine C-terminal de Pcc2 et de Pcc1 diffère fortement. D'après Schneider et Stephens, 1990.

Les prédictions structurales suggèrent que Pcc2 pourrait être un homologue de Pcc1 et comme il est connu que Pcc1 se dimérise (code PDB 3ENO), l'hypothèse que Pcc2 et Pcc1 forment un hétérodimère a été émise. Afin de tester cette hypothèse des expériences protéomiques et biochimiques ont été réalisées. De nombreuses questions ont également émergées. Forment-elles un complexe *in vivo* avec les protéines Kae1, Cgi121, Bud32 ? Quel est l'effet de Pcc2 sur la biosynthèse de la modification t<sup>6</sup>A *in vitro* et *in vivo* ? Quelle est la structure tridimensionnelle de cette protéine Pcc2 ? Pcc2 joue-t-elle le rôle de GON7 chez les archées ?

Pour répondre à ces questions nous avons entrepris une étude structurale de Pcc2 en complexe avec les autres sous-unités du complexe KEOPS de l'archée *Pyrococcus abyssi* (Pa). Tout d'abord, différents complexes et sous-complexes ont été purifiés par chromatographie d'affinité suivie d'une chromatographie d'exclusion stérique. Le complexe composé des protéines Kae1, Bud32, Cgi121, Pcc1 et Pcc2 (KBCP<sub>1</sub>P<sub>2</sub>) a été purifié ainsi que le sous-complexe KBCP<sub>1</sub>. Malgré le faible degré d'identité de séquence Pcc1 a pu être remplacé par Pcc2 pour former un complexe stable avec KBC *in vitro* (KBCP<sub>2</sub>) (dont les protocoles de purification sont décrits en annexe 5). Le sous-complexe Pcc2-Pcc1 a été également purifié. Enfin la protéine Pcc2 seule a été purifiée de la même manière. Selon l'algorithme de prédiction de désordre PONDR, la protéine Pcc2 est prédite comme une protéine ordonnée (figure 99), ce qui est encourageant pour sa cristallisation (Obradovic *et al.*, 2003).

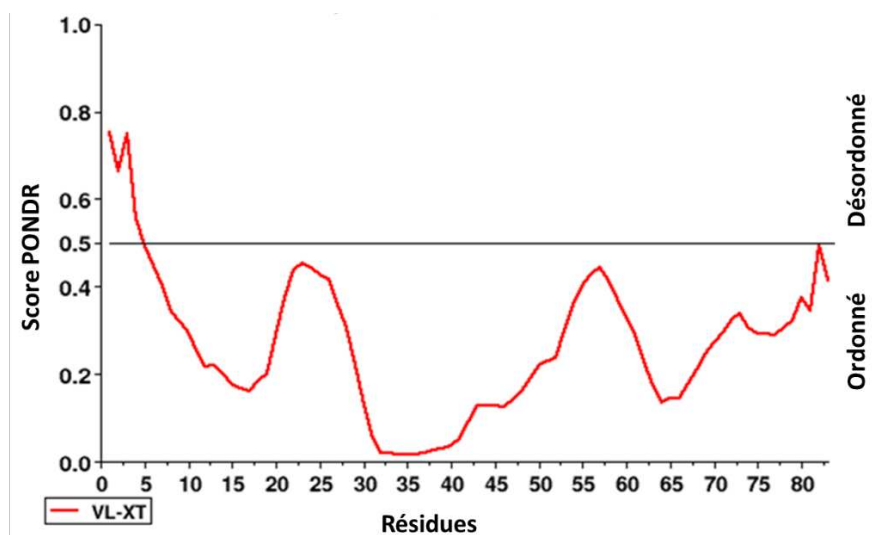


Figure 99 Prédiction de l'ordre et du désordre selon la séquence en acide aminé de Pcc2 de *Pyrococcus abyssi*. L'algorithme de PONDR utilise des fenêtres de 9 acides aminés qui coulisent le long de la séquence. L'algorithme mesure la charge nette et l'hydrophobie des acides aminés reflétant l'ordre et le désordre de la protéine. Un score compris entre 0 et 0,5 illustre une forte prédiction à acquérir des structures secondaires et un score entre 0,5 et 1 illustre une forte prédiction à être désordonné.

Nous avons obtenu des cristaux de PaPcc2 seule et en complexe avec PaPcc1, qui ont diffracté et qui ont permis de collecter des jeux de données dont l'analyse, décrite dans l'article en cours de rédaction, a abouti à la résolution des structures 3D à 3,2 et 1,8 Å de résolution (*cf. page 271*). PaPcc2 est organisée en 3 brins  $\beta$  antiparallèles et 3 hélices  $\alpha$  se positionnent sur l'une des faces du feuillet. 3 sous-unités de PaPcc2 sont retrouvées dans l'unité asymétrique. 2 monomères interagissent via leurs brins  $\beta_1$  et leurs hélices  $\alpha_3$ , formant un long feuillet de brins  $\beta$  antiparallèle stabilisé par des liaisons hydrogènes et des forces de

van der Waals (interface de type I) ([figure 3 de l'article en cours page 279](#)). La troisième sous-unité interagit via des liaisons hydrogènes et des forces de van der Waals entre les acides aminés des brins  $\beta_2$  et des hélices  $\alpha$  (interface de type II). Cette troisième sous-unité interagit avec une sous-unité de l'unité asymétrique voisine reproduisant l'interface de type I.

PaPcc1 est également organisée en 3 brins  $\beta$  antiparallèles et 2 hélices  $\alpha$  alors que GON7 est agencée en deux brins  $\beta$  et une hélice  $\alpha$ . Ainsi, PaPcc2 présente une structure tridimensionnelle très similaire à celle de PaPcc1 (RMSD = 1.48 Å, Z score = 7.2 sur 76 C $\alpha$ ) confirmant l'orthologie structurale partagée entre ces deux protéines.

La structure cristallographique de l'hétérodimère PaPcc1-Pcc2 nous a montré qu'elle est similaire à l'homodimère PfPcc1 (PDB code 3ENO). PaPcc1 et PaPcc2 interagissent via leurs brins  $\beta_1$  respectifs et leurs hélices  $\alpha_2$  terminales. En revanche, la superposition d'une sous-unité de PaPcc2 de l'homodimère sur une sous-unité PaPcc1 de l'hétérodimère montre un décalage de 10 Å de la seconde sous-unité PaPcc2 le long du brin  $\beta_1$  ce qui conduit à une interface légèrement différente ([figure 4 de l'article page 281](#)).

Aucun cristal des complexes KBCP<sub>1</sub>P<sub>2</sub>, KBCP<sub>1</sub> et KBCP<sub>2</sub> n'a pu être obtenu. Par contre, nous avons pu étudier ces complexes par diffusion des rayons X aux petits angles (SAXS), ce qui a donné des informations structurales en solution (forme de l'objet, rayon hydrodynamique, stœchiométrie, masse moléculaire) ([figure 100](#)). Ces analyses sont également décrites dans l'article en cours de rédaction. Les complexes KBCP<sub>1</sub> et KBCP<sub>2</sub> ont une masse moléculaire théorique de 91 et 92 kDa respectivement. La courbe des intensités I(q) obtenue par SAXS a permis de déterminer les masses moléculaires de KBCP<sub>1</sub> de l'ordre de 185 kDa et de KBCP<sub>2</sub> de l'ordre 180 kDa compatible avec une stœchiométrie 2 :2 :2 :2 en solution. Comme observé dans les structures cristallographiques des homodimères PfPcc1 (code PDB 3ENO) et PaPcc2 (la notre), il a été suggéré que Pcc1 et Pcc2 assurent la dimérisation des complexes KBCP1 et KBCP2 respectivement.

	Rg Guinier (Å)	Rg Pr (Å)	D <sub>max</sub> Pr (Å)	Mw théorique (kDa)	Mw expérimentale (kDa)	Oligomérisation
KBCP1P2	36,9	38,87 ±0,23	125	101,457	94,2	1:1:1:1:1
KBCP1	52,5 ±0,9	54,42 ±0,88	188	91,1	185,8	2:2:2:2:2
KBCP2	52,4±0,19	54,89 ±0,22	205	90,871	180,5	2:2:2:2:2

**Figure 100** Tableau récapitulatif des données structurales des différents complexes archées.

En revanche, le complexe KBCP<sub>1</sub>P<sub>2</sub> présente une masse moléculaire expérimentale de 95 kDa proche du poids moléculaire théorique de 101 kDa. KBCP<sub>1</sub>P<sub>2</sub> est donc compatible avec une stœchiométrie 1 :1 :1 :1 :1 suggérant que Pcc1 et Pcc2, ensemble, empêche la dimérisation du complexe. La fonction de distribution des distances montre que le complexe KBCP<sub>1</sub>P<sub>2</sub> possède une extension de 125 Å alors que les complexes KBCP<sub>1</sub> et KBCP<sub>2</sub> ont une extension maximale de 188 et 205 Å, respectivement, confirmant l'agencement linéaire du pentamère et des octamères. Des modèles *in silico* ont été réalisés à partir des structures cristallographiques connues (BC de la levure (code PDB 4WW9), KP<sub>1</sub> (code PDB 3ENC), et les structures présentées dans cette étude (P<sub>1</sub>P<sub>2</sub> et Pcc2 seule). Les courbes calculées Iq des modèles *in silico* ont été comparées aux courbes expérimentales et la position des sous-unités a été ajustée à l'aide du logiciel DADIMODO, confirmant l'agencement linéaire et leurs stœchiométries.

## ***Identification and characterization of the KEOPS fifth subunit in archaea.***

Daugeron Marie-Claire<sup>1</sup>, Missouri Sophia<sup>1</sup>, Da Cunha Violette<sup>1</sup>, Lazar Noureddine<sup>1</sup>, Collinet Bruno<sup>1,2</sup>, van Tilbeurgh Herman<sup>1</sup>, Basta Tamara<sup>1</sup>.

<sup>1</sup>Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France

<sup>2</sup> Institut de Minéralogie de Physique des Matériaux et de Cosmochimie (IMPMC), Sorbonne-Université, UMR7590 CNRS, MNHN, Paris, France.

### **Introduction**

tRNAs were discovered in 1958 and recognised as adaptor molecules allowing the genetic code to be translated into proteins (Hoagland *et al.*, 1958; Crick, 1958). Subsequent sequence analysis of these molecules revealed that they carried numerous post transcriptionally modified nucleosides (Holley *et al.*, 1965). Presently more than 100 different modified nucleosides are found in tRNA (Marck and Grosjean, 2002; Machnicka *et al.*, 2013; Boccaletto *et al.*, 2018). Much research effort went into the understanding of the mechanistic and functional aspects of these chemical modifications. Alterations in the chemistry of the nucleotides contained in the main body of the tRNA, are believed to have impact on the tertiary structure and hence the stability or to affect the interactions with their partners. Most of the more complex modifications are found at the level of the anticodon stem-loop of the tRNA molecules and often play critical functional roles such as the expanding or restricting the decoding properties of tRNA and maintaining translational fidelity (Jackman and Alfonzo, 2013; Agris *et al.*, 2007, 2017). The N<sup>6</sup>-threonylcarbamoyl adenosine (t<sup>6</sup>A) modification at position 37 of tRNAs decoding ANN codons, is one of the few universal modifications of tRNA (Chheda *et al.*, 1969; Schweizer *et al.*, 1969; Takemura *et al.*, 1969). The t<sup>6</sup>A modification facilitates correct anticodon–codon pairing by enhanced base stacking and by preventing intraloop base pairing of U33–A37, thereby helping the formation of the canonical U-turn structure and thus enhancing the translation fidelity (Grosjean *et al.*, 1995; Sundaram *et al.*, 2000; Murphy *et al.*, 2004; Vendeix *et al.*, 2012). A deficit of t<sup>6</sup>A formation leads to pleiotropic phenotypes. Absence of t<sup>6</sup>A is non-viable in bacteria, reduces severely fitness in yeast and

archaea, leads to developmental impairments in model multicellular organisms and to GAMOS in humans (Naor *et al.*, 2012; Thiaville *et al.*, 2014; Lin *et al.*, 2015; Kisseleva-Romanova *et al.*, 2006; Braun *et al.*, 2017; Arrondel *et al.*, 2019). Despite heterogeneous effects, phenotypes are globally caused by mis-regulation of translation causing reading errors and frameshifts (Thiaville *et al.*, 2014; de Crécy-Lagard and Jaroch, 2020; Ogura, 2020; Pollo-Oliveira *et al.*, 2020).

The formation of t<sup>6</sup>A in all organisms occurs through two distinct ATP-dependent reactions. First, the ubiquitous YrdC/Sua5 enzymes form an unstable intermediate threonyl-carbamoyl adenosine (TC-AMP) using L-threonine, bicarbonate and ATP (Lauhon, 2012). Subsequently, the universal Kae1/TsaD/Qri7 enzyme transfers the threonyl-carbamoyl moiety from TC-AMP onto A<sub>37</sub> of substrate tRNAs (Srinivasan *et al.*, 2011; El Yacoubi *et al.*, 2009, 2011; Perrochia *et al.*, 2013a). While Sua5/YrdC family members function independently, Kae1/TsaD/Qri7 always act within multi-subunit protein complexes whose composition varies among species. In mitochondria, Qri7 functions as a homodimer. In bacteria, TsaD forms a trimeric complex with two other proteins, TsaB and TsaE (Zhang *et al.*, 2015a; Luthra *et al.*, 2018; Missouri *et al.*, 2018; Luthra *et al.*, 2019). In eukaryotes and archaea, Kae1 is a part of an ancient macromolecular assembly named KEOPS (Kinase Endopeptidase and Others Proteins of Small size), further composed of Bud32, Cgi121, Pcc1. A fifth core subunit, named Gon7, was identified in fungi (Costessi *et al.*, 2012; Downey *et al.*, 2006; Kisseleva-Romanova *et al.*, 2006; Lopreiato *et al.*, 2004; Wan *et al.*, 2016; Zhang *et al.*, 2015b; Wan *et al.*, 2017). Bud32 is related to the small atypical RIO2 kinase proteins (Knüppel *et al.*, 2018). Its presence is required for t<sup>6</sup>A formation and deletion mutants are hardly viable in yeast, but its function is not yet understood (Perrochia *et al.*, 2013a). Cgi121 has no structural homologues and it is dispensable for t<sup>6</sup>A synthesis *in vitro*. It binds tightly to Bud32 and stimulates its autophosphorylation activity as well as t<sup>6</sup>A formation *in vitro*, suggesting an effector function (Mao *et al.*, 2008; Perrochia *et al.*, 2013). The two other subunits Pcc1 and Gon7 are small  $\beta$ -sheet proteins that are functionally important and necessary for *in vivo* t<sup>6</sup>A activity in yeast. Pcc1 consists of a three stranded anti-parallel  $\beta$  sheet packed into two  $\alpha$  helices on one face, and solvent exposed on the other. Pcc1 has a large hydrophobic surface that is covered either through homodimer formation (archaea) or through heterodimerization with Gon7 in yeast (Zhang *et al.*, 2015). Despite the low sequence similarity, human GON7 (hGON7) exhibits

similar biochemical characteristics to yGon7. Indeed, as yGon7, hGON7 is a small 11 kDa protein that, although is an intrinsically disordered protein alone in solution, adopts a well-defined structure for about 50% of its polypeptide chain upon interaction with LAGE3 (the human Pcc1 ortholog). yGon7 and hGON7 form an anti-parallel  $\beta$  sheet packed against 2  $\alpha$  helices (Zhang *et al.*, 2015b; Arrondel *et al.*, 2019). Upon complex formation, Pcc1/yGON7 and LAGE3/GON7 form a continuous six stranded  $\beta$ -sheet, completed by packing of helices such that the interaction mode is conserved. In presence of yGon7, the eukaryotic KEOPS complex has a 1:1:1:1:1 stoichiometry. In absence of hGON7, KEOPS dimerizes resulting in a hetero-octamer with 2:2:2:2 stoichiometry (Zhang *et al.*, 2015b; Wan *et al.*, 2017). It has been shown that the presence of hGON7 or yGon7 improves the amount of t6A modification *in vitro* and increases the cellular stability of the KEOPS complex *in vivo* (Wan *et al.*, 2017; Arrondel *et al.*, 2019).

Sequence similarity search for yGon7 orthologs outside Fungi was unfruitful. However, the recent discovery of hGON7 ortholog in Metazoans (Wan *et al.*, 2017) raised the possibility that a fifth KEOPS subunit could exist in other eukaryotic groups and in archaea.

We here present the identification of a putative core fifth subunit of archaeal KEOPS, that we named Pcc2. We show that Pcc2 protein of *Pyrococcus abyssi* forms a stable pentameric complex with the other four subunits *in vitro* or an octamer in absence of Pcc1. Furthermore, we determined the crystal structure of the Pcc2 and of the Pcc1-Pcc2 subcomplex, showing that Pcc2 is a structural homolog of Pcc1. Although Pcc2 has a different fold from yGon7 and hGON7, it associates in a very similar way with Pcc1 indicating that this association is functionally important requirement that has been conserved from Archaea to Eukaryotes.

## Materials and Methods

### Cloning procedures

The genes encoding PaKEOPS subunits PAB\_RS09585 (Kae1), PAB\_RS08575 (Bud32), PAB\_RS06930 (Cgi121) and PAB\_RS01470 (Pcc1) were amplified by polymerase chain reaction (PCR) using as DNA template, pKEOPS500 plasmid (Perrochia *et al.*, 2013a). The gene PAB\_RS00380 encoding Pcc2 was amplified by polymerase chain reaction (PCR) using genomic DNA of *P. abyssi* GE5 strain as template. Different combinations of bicistronic or polycistronic



sequences were cloned into Pet26b (Novagen) NdeI-XhoI either using T4 ligase (NEB) or by Gibson assembly (NEB) following the protocol supplied by the manufacturer. The sequences of used oligonucleotides are listed in supplementary information section.

pKP1His6 and pKP2His6 contain each a bicistronic sequence with Kae1 plus Pcc1 or Kae1 plus Pcc2 encoding genes, respectively. The genes encoding Pcc1 and Pcc2 are fused at their 3' end to a sequence encoding a hexa-histidine tag indicated as His6.

pKP2strep-P1his6 contains a bicistronic sequence with Kae1, Pcc2 and Pcc1 encoding genes. The gene encoding Pcc2 is fused at its 3' end to a sequence encoding a Strep-tag indicated as strep whereas the gene encoding Pcc1 is fused at 3' end to a sequence encoding a hexa-histidine tag.

pKBC and pKB contain Kae1, Bud32 and Cgi121 encoding genes or Bud32 and Cgi121 encoding genes respectively. These genes are not fused to an epitope tag encoding sequence.

#### Heterologous expression and purification of recombinant proteins

Recombinant proteins were expressed in Rosetta2 (DE3) pLysS *E. coli* strain (Novagen). The cells were transformed either with pKBC, pBC, pKP1-His6, pKP2-His6 or with pKP2strep-P1 his6 plasmids and were grown at 37°C in complete LB medium with kanamycin and chloramphenicol until the optical density (600 nm) reached of 0.6-0.7. Gene expression was induced by addition of IPTG (1mM), and incubation was continued at 37°C for 2h. Cells were collected by centrifugation (5000 g, 15 min) and resuspended in cold Lysis buffer (20 mM Tris-HCl (pH 8), 300 mM NaCl, 5mM  $\beta$ mercaptoethanol, 10 % Glycerol) supplemented with complete EDTA free protease inhibitors cocktail (Roche). Cells were lysed at 4°C with a one-shot cell disruptor (Branson sonifier 250, 3 x 1 min) and the lysate was clarified by centrifugation at 30 000 g for 30 min at 18°C. The clarified lysate was incubated 10 to 15 min at 65°C to precipitate *Escherichia coli* proteins. The precipitates were removed by centrifugation at 30 000g for 20 min at 18°C. The soluble fraction was adjusted to 10 mM imidazole and His-tagged proteins were purified by gravity flow affinity chromatography on an NiNTA resin column (Quiagen) at room temperature. Equilibration, washes and elution steps were done with lysis buffer containing respectively 10, 40 and 400 mM imidazole. Fractions of interest were pooled and loaded on HiTRAP Heparin affinity column (GE Healthcare) pre-equilibrated in lysis buffer. Elution was done with a linear NaCl gradient (from

150 mM to 1M). Fractions of interest were pooled, concentrated at 18°C and submitted to size exclusion chromatography on a HiLoadR 16/600 SuperdexR 200pg or a HiLoadR 16/600 SuperdexR 75pg (GE Healthcare).

To purify different complexes, cell pellets obtained with cells transformed with different plasmids were mixed in defined ratio and lysed together.

P1P2 and P2 sub-complexes used for SAXS analysis and crystallization were purified from pKBC transformed cells mixed with pK1-P2 transformed cells (ratio 2:1) and from pKBC transformed cells mixed with pK2 transformed cells (ratio 2:1), respectively. In these conditions, a significant excess of P1P2 and P2 complexes over the complete KEOPS complexes is obtained. The CBKP1P2, CBKP1 and CBKP2 complexes used for SAXS analysis and enzymatic assays were purified from pKBC transformed cells mixed with pBC transformed cells and either pKP1His6 transformed cells (2:1:1 ratio) or pKP2His6 transformed cells (2:1:1 ratio) or pKP2strep-P1His6 transformed cells (2:1:1 ratio). In these conditions, four or five subunit KEOPS complexes can be purified.

#### Phylogenetic analysis

For the phylogenetic analysis, we chose 27 couples of Pcc2 and Pcc1 paralogs from one representative of each phylum of Archaea. Protein alignment was done with MAFFT V7 with the amino acid matrix Blosum 30 and iterative refinement methods L-INS-I or G-INS-I (Kato *et al.*, 2019). The Trimming was performed with BMGE with a BLOSUM30 matrix and the -b 1 parameter (Criscuolo and Gribaldo, 2010). The Maximum likelihood trees were performed using IQ-TREE v1.7 software (<http://www.iqtree.org/>) (Nguyen *et al.*, 2015) with the best substitution model as suggested by the ModelFinder option -MFP. The branch robustness was estimated by the SH-aLRT approximate likelihood ratio test (Guindon *et al.*, 2010), the ultrafast bootstrap approximation (Hoang *et al.*, 2018) (10000 replicates for each), and with the transfer bootstrap expectation approaches (Lemoine *et al.*, 2018) based on 500 bootrees.

#### Crystallization of the PaPcc1-Pcc2 heterodimer

Fractions from purification by size exclusion chromatography containing the heterodimeric PaPcc1-Pcc2 complex were then pooled and concentrated to 7.9 mg.mL<sup>-1</sup> for crystallization trials. Crystals were obtained by mixing 100 nL of the protein solution with 100 nL of 15% PEG

6K and 5% glycerol using the sitting-drop vapor diffusion method. Crystals appeared after about 21 days of incubation at 18°C. The crystallization drop was equilibrated against a 70 µL reservoir solution of 15% PEG6K and 5% glycerol. Crystals were cryo-protected by quick-soaking in 20% PEG6K and 20% glycerol prior to flash freezing in liquid nitrogen.

#### Crystallization of the PaPcc2 monomer

Fractions containing pure monomeric PaPcc2 were concentrated to 22 mg.mL<sup>-1</sup> for crystallization trials. Crystals of Pa Pcc2 were obtained by mixing 100 nL of the protein solution with 100 nL of 30% PEG 3K 0.2 M lithium sulfate and 0.1 M Tris HCl pH 8.5 using the sitting drop vapor diffusion (crystals appeared after 17 days of incubation at 18°C). The crystallization drop was equilibrated against a 70 µL reservoir solution of same the precipitant. Crystals were cryo-protected by quick-soaking in 50% paraffin and 50% paratone prior to flash freezing in liquid nitrogen.

#### SAXS analysis

SAXS experiments were carried out at the SOLEIL synchrotron SWING beamline (Saint-Aubin, France). The sample to detector (Eiger 4M Dectris) distance was set to 1500 mm, allowing reliable data collection over the momentum transfer range  $0.005 \text{ \AA}^{-1} < q < 0.5 \text{ \AA}^{-1}$  with  $q = 4\pi \sin \theta / \lambda$ , where  $2\theta$  is the scattering angle and  $\lambda$  is the wavelength of the X-rays ( $\lambda = 1.0 \text{ \AA}$ ). To collect data on homogenous protein samples, SAXS data were collected at the exit of a size exclusion high-performance liquid chromatography (SEC HPLC Bio-SEC3 Agilent) column directly connected to the SAXS measuring cell. Sixty-five microliters of Kae1/Bud32/Cgi121/Pcc1/Pcc2 (KBCP<sub>1</sub>P<sub>2</sub>), KBCP<sub>1</sub>, and KBCP<sub>2</sub> samples concentrated at 2.55, 0.72 and 3 mg/ml, respectively, were injected into the column preequilibrated with a buffer composed of 20 mM Tris pH 8.5, 300 mM NaCl, and 5 mM 2-mercaptoethanol and 5% glycerol. Flow rate was set at 300 µl/min, frame duration was 1.0 s, and the dead time between frames was 0.01 s. The protein concentration was estimated by UV absorption measurement at 280 and 295 nm using a spectrometer located immediately upstream of the SAXS measuring cell. A large number of frames were collected before the void volume and averaged to account for buffer scattering. SAXS data were normalized to the intensity of the incident beam and background (i.e. the elution buffer) subtracted using the program FoxTrot50, the Swing in-house software. The scattered intensities were displayed on an absolute scale using the

scattering by water. Identical frames under the main elution peak were selected and averaged for further analysis. Radii of gyration, maximum particle dimensions, and molecular masses were determined using PrimusQT (Franke *et al.*, 2017).

#### Modelling of the complete KEOPS complex

A composite KEOPS complex was modelled using a combination of structures of archaea KEOPS subcomplexes and of the present PaPcc1-Pcc2 structure by the Modeller software (Sali *et al.*, 1995). The Dadimodo and FoXsDock program (Evrard *et al.*, 2016) were then used to adjust/complete the models of KBCP1P2, KBCP1 and KBCP2 respectively as to optimally fit the experimental curve I(q). An ultimate adjustment was performed using the program CRY SOL (Svergun *et al.*, 1995).

#### Comparative sequence and structure analysis

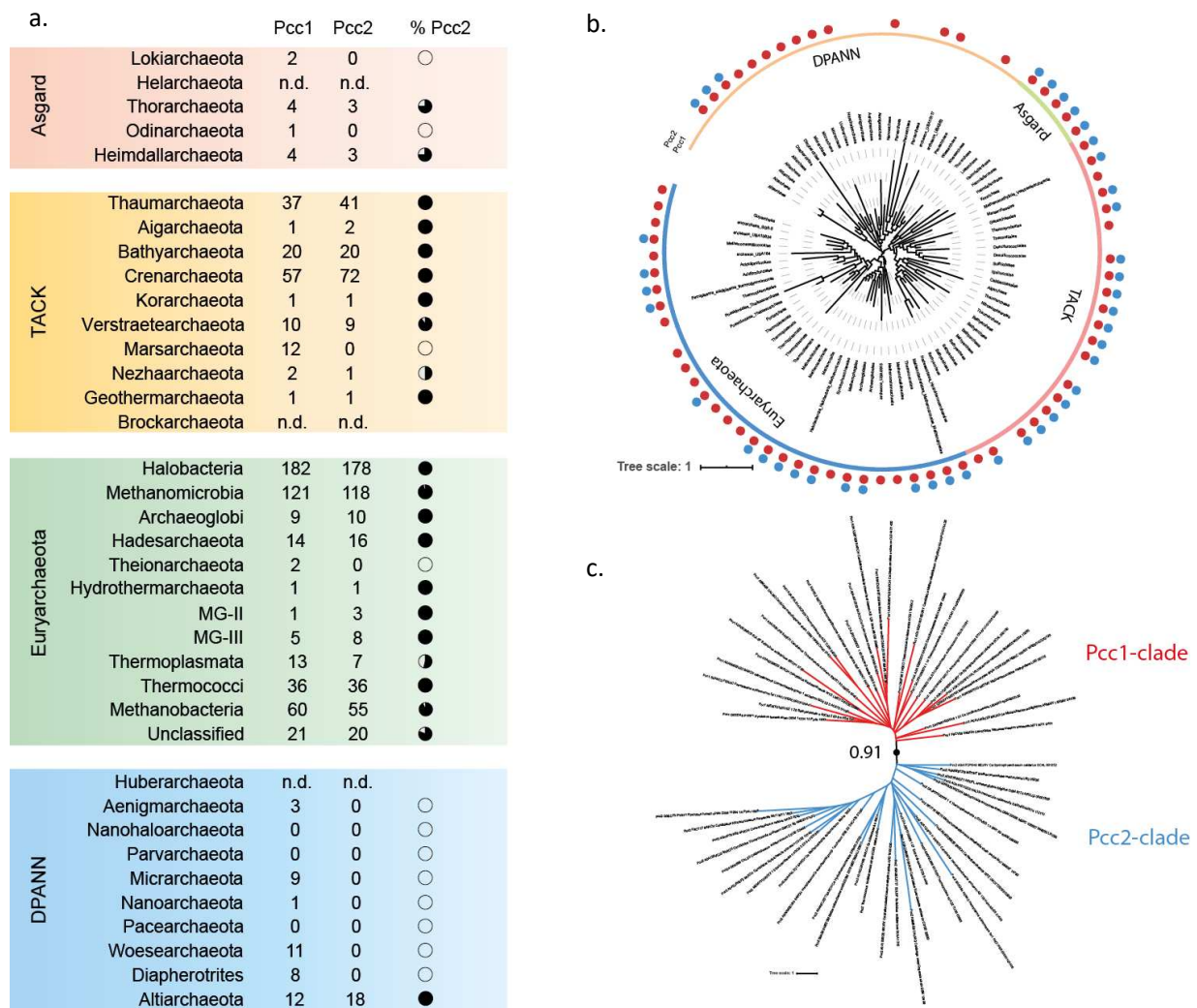
Multiple sequence alignment of the archaea Pcc2 family members was realized using the blosum62 matrix6 with MAFFT (online version 7) (Kato *et al.*, 2019). Structure-based alignment of a subset of these members was performed using the ESPRIPT web-server (Robert and Gouet, 2014). All figures of structures were generated with Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Structure comparisons was performed with the PDBeFold web server.

## Results

#### Identification of a Pcc1 paralog in archaea

The recent discovery of the fifth core subunit of KEOPS in metazoa raised the possibility that a corresponding subunit could exist also in other eukaryotes and in archaea. In search for such a subunit we noticed in the archaea *Thermococcales* the existence of two genes annotated as *Pcc1* and encoded on two distinct loci on the chromosome ([supplementary figure 1](#)). These genes encode two proteins of similar size (82 amino acids for Pcc1, 85 amino acids for Pcc1-like) and exhibit modest but significant pairwise sequence identity (22 % identity, 55 % similarity for *Thermococcus kodakarensis* orthologs). Using HMM-based sequence similarity searches, we detected Pcc1-like proteins in 11% of DPANN, 38% of Asgard, 83% of TACK and 88% of Euryarchaeaota analysed genomes. Except for several DPANN phyla, all the analyzed

archaeal genomes contained both Pcc1 and Pcc1-like encoding genes (figure 1 a. and b.). The alignment of representative Pcc1 and Pcc1-like sequences from 27 archaeal phyla allowed the identification of conserved motifs (supplementary figure 2). Of note, the motif N<sub>67</sub>S/T<sub>68</sub> in Pcc1 sequences is replaced by D/E<sub>70</sub>D/E<sub>71</sub> in Pcc1-like sequences (*T. kodakarensis* numbering) indicating that this might be a distinguishing feature between the two proteins. The phylogenetic tree constructed from this alignment showed bipartite topology with one monophyletic clade corresponding to Pcc1 orthologs and the other to Pcc1-like orthologs (figure 1 c.). Collectively, these data indicated that Pcc1-like and Pcc1 proteins are paralogs present in the vast majority of Archaea. To underscore the evolutionary connection with Pcc1 we named its paralog Pcc2.



**Figure 1: Identification of Pcc1 paralogs among archaea domain. a.** prevalence of Pcc1 and Pcc2 paralogs among archaeal phyla. Number of found Pcc1 or Pcc2 paralogs is indicated in the corresponding columns. Full black circle corresponds to 100 % and white circle to 0% of genomes containing both Pcc1 and Pcc2 encoding genes. n.d. means not determined. **b.** distribution of Pcc1 (red circles) and Pcc2 (blue circles) mapped onto archaeal phylogenetic tree. The position of the four superphyla is indicated. **c.** maximum likelihood phylogenetic tree made from alignment of 27 Pcc1 or Pcc2 sequences representative of the archaeal diversity. The statistical support for the branch separating Pcc1 and Pcc2 clades was evaluated from 500 boot trees using transfer bootstrap expectation. This means that the branch separating the two clades occurred in 455 bootstrap trees out of 500.

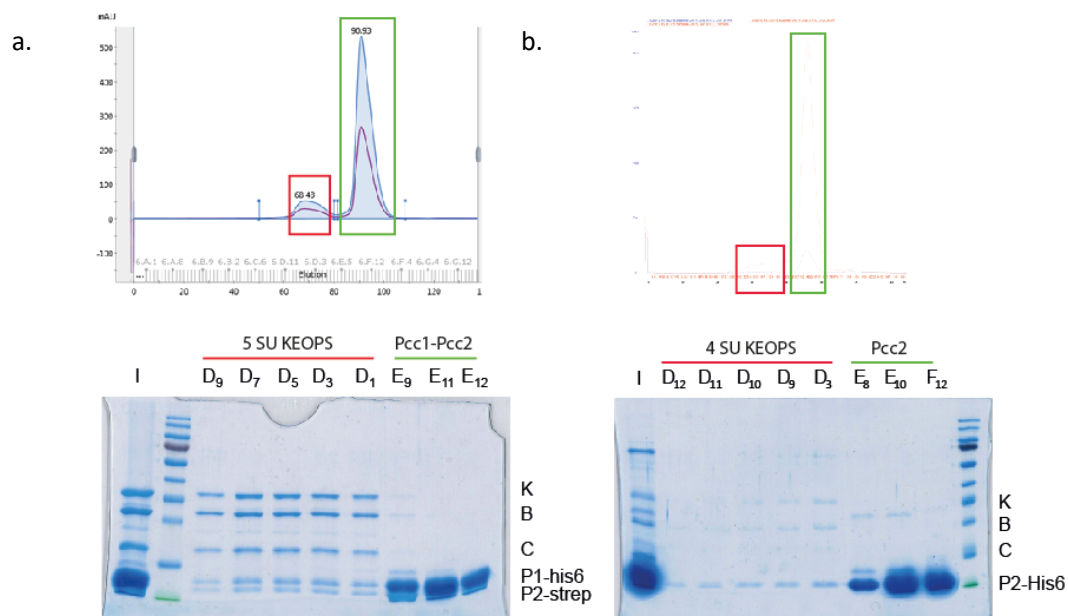
#### PaPcc2 protein co-purifies with archaeal KEOPS subunits

To investigate if Pcc2 could interact with KEOPS subunits we expressed the four known subunits of KEOPS from *Pyrococcus abyssi* (PaKEOPS) among which Pcc1 carried hexahistidine tag together with *Pyrococcus abyssi* Pcc2 ortholog (PaPcc2) and performed Ni-NTA affinity chromatography followed by heparin and size exclusion chromatography. With this procedure

we purified a complex containing Kae1, Bud32, Cgi121, Pcc1 and a fifth band with apparent molecular mass corresponding to that of Pcc2 (Figure 2). The apparent stoichiometry of the purified complex was 1:1:1:1:1 as judged from SDS-PAGE analysis. We also purified an excess of Pcc1-Pcc2 complex which was used for crystallization trials (see below).

We next tested whether Pcc2 can replace Pcc1 in a four-subunit complex. Similar purification scheme was employed with Pcc2 being the sole His-tagged protein. The purified complex contained four bands corresponding to Kae1, Bud32, Cgi121 and Pcc2 with apparent stoichiometry of 1:1:1:1. An excess of Pcc2 was also purified and used for crystallization trials (see below) (Figure 2).

These experiments demonstrated that *in vitro* Pcc2 can associate with KEOPS subunits either as a fifth subunit or by replacing Pcc1 to form four membered complex.

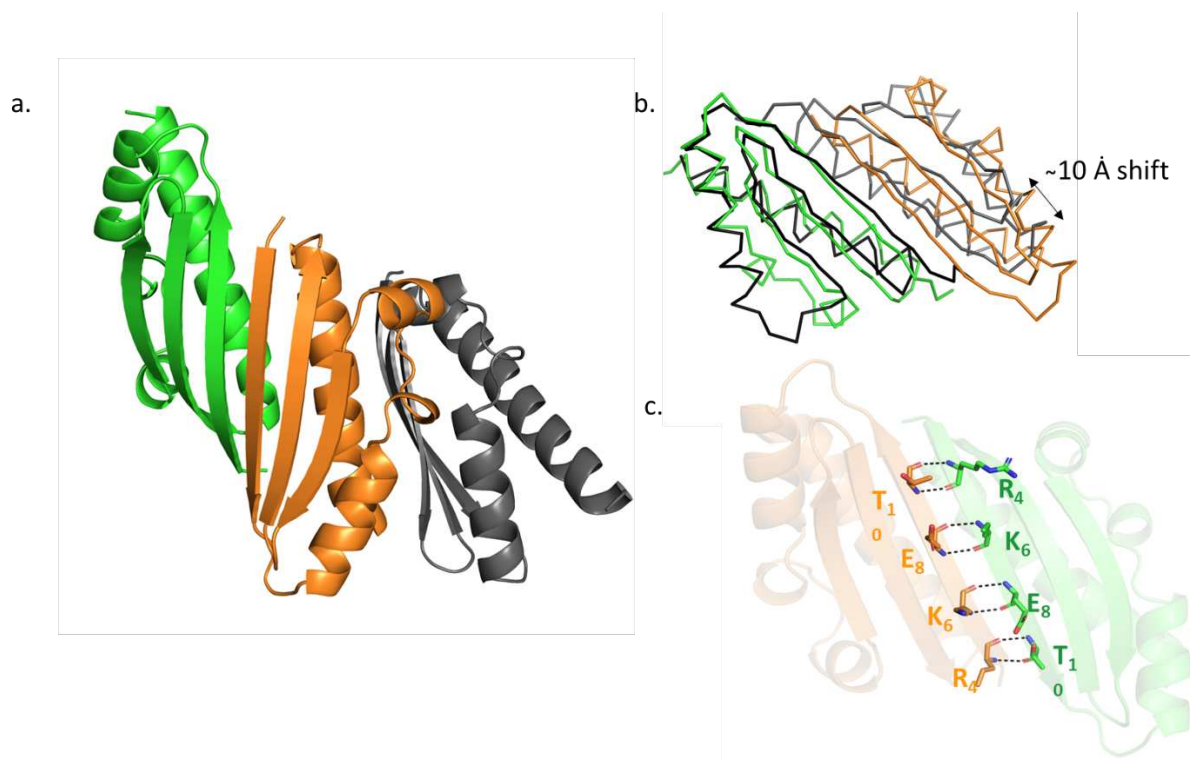


**Figure 2: Purification of five and four subunit Pcc2-containing KEOPS complexes.** The size exclusion chromatography analysis of five (a.) or four (b.) subunit KEOPS complex. The corresponding SDS-PAGE analysis of the two boxed peaks is shown below each chromatogram. I corresponds to the protein sample injected on the column. The fractions analyzed are indicated above each gel. K stands for Kae1, B for Bud32, C for Cgi121, P1 for Pcc1 and P2 for Pcc2.

### The Pcc2 crystal structure

The structure of PaPcc2 was solved by molecular replacement at a resolution of 1.84 Å and contains three PaPcc2 copies in the asymmetric unit. The structures of the 3 copies in the asymmetric unit is nearly identical (RMSD around 1.3 Å). Two copies form a tight dimer

whereas the third copy is more loosely associated with the two others. An identical dimer is created by the third copy through crystal symmetry contacts (Figure 3.a). PaPcc2 forms a three stranded anti-parallel  $\beta$ -sheet with three  $\alpha$  helices packed against one face of the sheet ( $\beta\alpha\alpha'\beta\beta\alpha$  topology) (figure 3.a). PaPcc2 and Pcc1 share the same fold and superpose very well with a RMSD of 1.48 Å (Z score of 7.2 for 76 C $\alpha$  positions) (figure 3.b).



**Figure 3:** a. Structure and packing of the three copies of PaPcc2 in the asymmetric unit. The dimer that resembles the PaPcc1-Pcc2 heterodimer is represented by the subunits in green and orange, the third subunit is in grey. b. Superimposition of the PaPcc2 homodimer (pink and orange) with one subunit of the PfPcc1 homodimer (each subunit is colored in black and grey). c. Packing of PaPcc2 homodimer into the asymmetric unit. Interface type I sticks residues are involved into the packing stabilization.

The main structural difference between Pcc1 and Pcc2 resides in the connection between strands  $\beta$  1 and  $\beta$ 2 that consists of a single  $\alpha$  helix for Pcc1 and 2 perpendicularly oriented  $\alpha$  helices for PaPcc2. The Pcc2 dimer is very similar to that of the Pcc1 homodimer and also to the PaPcc1-Pcc2 heterodimer, described below. The two  $\beta$ 1 strands associate to form a continuous six stranded anti-parallel  $\beta$ -sheet. This Pcc2 homodimer is stabilized by 9 hydrogen bonds between  $\beta$ <sub>1</sub> strand (interface area 815.6 Å<sup>2</sup>: interface type I) (figure 3.c). The second dimer interface in the asymmetric unit is formed by packing of the  $\alpha$ <sub>1</sub> and  $\alpha$ <sub>2</sub> helices against the  $\beta$ -sheet (interface type II). This last subunit contacting the  $\beta$ <sub>1</sub>-strand as the

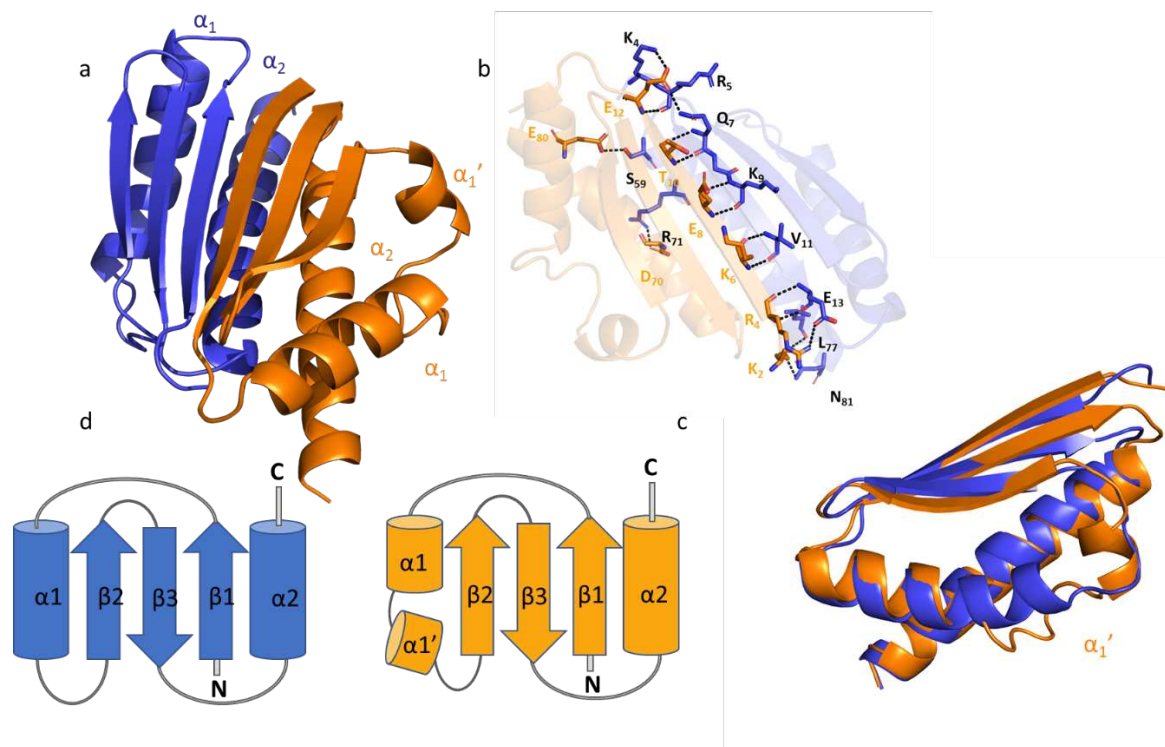


interface type I of a neighboring Pcc2 heterotrimer in the crystal (not shown) suggesting that this interface is probably purely crystallographic. The superimposition of PaPcc2 (type I) with the PfPcc1 homodimer highlights a transverse sliding of second subunit along its  $\beta$ 1 strand axis (figure 3.b).

These data revealed that PaPcc2 structurally closely resembles PaPcc1 despite the low level of sequence identity suggesting, after the gene duplication event, the existence of strong positive selection to conserve the overall structure.

#### Structure of the Pcc1-Pcc2 heterodimer

The structure PaPcc1-Pcc2 was solved at 3.18 Å resolution. The asymmetric unit of the crystal contains one copy of the PaPcc1-Pcc2 heterodimer (figure 4.a). The  $\beta$ 1-strands of PaPcc1 and PaPcc2 align to create a six stranded anti-parallel  $\beta$ -sheet. The packing of the two C-terminal helices of Pcc1 and Pcc2 provides the main hydrophobic interface of the heterodimer (interface area of 1154.5 Å<sup>2</sup> (figure 4.b). PaPcc2 provides a similar fold that PaPcc1, but PaPcc2 has an helix insertion (figure 4 c. and d.). Moreover the PaPcc1-Pcc2 dimer structure is very similar to that of the Pcc1 homodimer (figure 4.b) (PDB ID 3ENC) but is somewhat different from the Pcc2 homodimer (RMSD = 2.4 Å, Z-score = 7.37, 144 positions of C $\alpha$  aligned) (figure 6.b). However, the structure of one subunit PaPcc2 as present in its homodimer and in the PaPcc2-Pcc1 heterodimer are nearly the same (RMSD 0.98 Å) suggesting that binding of Pcc2 to Pcc1 does not induce a significant conformational change in Pcc2 structure (supplementary figure 3).



**Figure 4: Structure of PaPcc1-Pcc2 heterodimer. a.** Cartoon presentation of PaPcc1-Pcc2 (in blue and orange respectively). **b.** Packing of PaPcc1Pcc2 heterodimer into the asymmetric unit. sticks residues are involved into the packing stabilization **c.** Superimposition of PaPcc2 (in orange) onto PaPcc1 (in blue). **d.** Topology diagrams of PaPcc1 and PaPcc2 respectively.

#### Analysis of KEOPS complexes in solution.

To learn more about the oligomeric status of the Pcc1 and Pcc2 complexes in solution, we performed SAXS experiments. Three complexes were analyzed: KBCP<sub>1</sub>P<sub>2</sub>, KBCP<sub>1</sub> and KBCP<sub>2</sub>. Samples of these were loaded onto a Bio-SEC3 Agilent size exclusion chromatography mounted before the SAXS cell. Guinier plot shows a linear curve testifying of monodispersy of the complexes ([supplementary figure 7 a.](#)). Our SAXS data established that the KBCP<sub>1</sub>P<sub>2</sub> complex has a Mw of 90 kDa approximatively in agreement with the theoretical Mw of 101 kDa and most compatible with a heteropentamer of 1 :1 :1 :1 :1 stoichiometry in accordance with size exclusion chromatography data ([Table 2](#)). Based on a combination of crystal structures (KP<sub>1</sub> (PDB code 3ENC), BC (PDB code 4WW9) and the present P<sub>1</sub>P<sub>2</sub>), we constructed a 3D-model of the KBCP<sub>1</sub>P<sub>2</sub> pentamer using the software MODELLER. The calculated scattering curve of this pentamer model fits very well the experimental SAXS curve, yielding a  $\chi^2$  value of 0.619 ([supplementary figure 4](#)). The Patterson function allowed us to calculate the maximal extension of the object, that is around 125 Å ([supplementary figure 7 b.](#)). Next, we showed

by SAXS that KBCP<sub>2</sub> complex is stable and has a Mw of approximately 180 kDa, not compatible with the theoretical Mw of 91 kDa. This strongly suggests that KBCP<sub>2</sub> forms an octamer in solution with (Table 2) 2 :2 :2 :2 stoichiometry. We created a model for this complex, using a combination of previous BKP<sub>1</sub> and the present Pcc2 homodimer structures. We replaced P<sub>1</sub> by the closely related P<sub>2</sub> in the KBP<sub>1</sub> complex and imposed P<sub>2</sub> homodimer (PDB code 7A66) as the central dimerization unit to construct the octamer. This model fitted the experimental SAXS curve satisfactorily with a  $\chi^2$  value of 1.64 (supplementary figure 5).

Archaeal KEOPS complex was previously shown to form octamers via Pcc1 dimerization (Mao *et al.*, 2008; Perrochia, Guetta, *et al.*, 2013). To get structural information about this octamer we collected SAXS data and established that the KBCP<sub>1</sub> complex has a Mw of approximately 186 kDa, confirming KBCP<sub>1</sub> has 2 :2 :2 :2 stoichiometry in solution (Table 2). We then made a model of octameric KBCP<sub>1</sub>, using the structure of the Pcc1 as the dimerization unit (PDB ID 3ENC). The initial model created a few steric clashes (Wan *et al.*, 2016) which could be released using the FoXsDock software. The adjusted KBCP<sub>1</sub> octamer model fitted very well to the experimental SAXS curve with a  $\chi^2$  value of 1.39 (supplementary figure 6). The KBCP<sub>1</sub> octamer adopts a linear shape, in line with a maximal extension around 188 Å. The KP<sub>1</sub> interface in the octamer is slightly twisted compared to that in the KP<sub>1</sub> structure (PDB code 3ENO) but we modelled it as the interface of OSGEP-LAGE3 (PDB code 6GWJ) (supplementary figure 8). The proximity of the Kae1 subunits in the octamer raises the question about the existence of possible steric hindrance for tRNA binding. However, it has been shown that there is no difference in the *in vitro* t<sup>6</sup>A activities of the tetrameric and octameric KEOPS complexes (Wan *et al.*, 2016).

Together, the SAXS data indicated that PaPcc2 when associated to four-subunit PaKEOPS complex prevents its dimerization like the action of GON7 within the human KEOPS complex.

## Discussion

The t<sup>6</sup>A modification pathway is present in virtually all living cells. In eukaryotic organisms the threonyl-carbamoyl transferase activity is executed by the universal Kae1/TsaD enzyme within the KEOPS protein complex composed of five subunits. The sequences of four subunits (Kae1, Bud32, Cgi121, Pcc1) , are very well conserved among organisms, while that of the fifth subunit Gon7 diverges considerably between fungi and metazoans (Wan *et al.*, 2017). Apart

for Kae1, the functional role and associated mechanism of the other four units is still not understood. Deletion of Gon7, Bud32 and Pcc1 abolishes t<sup>6</sup>A activity and is lethal for prokaryotic cells while it results in growth and developmental retardation in single and multicellular eukaryotes (Thiaville *et al.*, 2016). Recently, mutations in the five KEOPS subunits in human were linked to a very debilitating genetic disease with low life expectancy (Braun *et al.*, 2017; Arrondel *et al.*, 2019).

In this study we have identified a paralog of KEOPS subunit Pcc1 which we named Pcc2. Both paralogs are widely conserved in Archaea suggesting that the gene duplication resulting in Pcc1 and Pcc2 is an ancient event that occurred before the diversification of Archaea. The high prevalence of both paralogs in Archaea indicate that Pcc1 and Pcc2 have important and distinct functions in Archaea.

We solved crystal structures of Pcc1-Pcc2 heterodimer and Pcc2-Pcc2 homodimer. Pcc1 has structural similarities with the well documented KH domain (RMSD 1.95 Å on 71 C $\alpha$  between PaPcc1 and KH1 domain of poly C binding protein PDB ID 2AXY), which functions as a versatile RNA, ssDNA, RNA hairpin binder (Grishin, 2001; Valverde *et al.*, 2008; Nicastro *et al.*, 2015) in line with its capacity to bind tRNA (Perrochia *et al.*, 2013b). Pcc2 is also related to the KH domain family, but the typical GxxG motif is absent between  $\alpha$ 1 and  $\alpha$ 1' helices. The closest structural homolog of Pcc2 is the KH3 domain of hnRNP K (RMSD 2.764 Å on 64 C $\alpha$ , PDB code 1J5K). Structure of hnRNP K - ssDNA complex shows that ssDNA contacts the protein in a narrow hydrophobic groove formed by four positively charged residues (K-K/R-K-R) and capped by hydrophobic residues (Braddock *et al.*, 2002). PaPcc2 has three positively charged residues located in  $\alpha$ 1' helix (K<sub>36</sub>, K<sub>38</sub>, K<sub>39</sub>) surrounded by hydrophobic residues (S<sub>32</sub> and S<sub>40</sub>). We were unable to model the side chain of K<sub>39</sub>, suggesting that this residue is flexible in contact with solvent or in absence of a ligand. In analogy to hnRNP K, Pcc2 could therefore harbor the capacity to bind nucleic acids. These speculative observations should be tested by isothermal titration calorimetry.

The superposition of the Pcc1-Pcc2 heterodimer with Pcc2-Pcc2 or Pcc1-Pcc1 homodimers revealed that the contact with Pcc1 does not induce a significant conformational change in Pcc2 ([supplementary figure 8](#)). However, we observed a significant shift alongside the heterodimer contact interface (about 10 Å) as compared with homodimers ([figure 3 b.](#))

impacting the maximum dimension of KBCP2 octamer ([supplementary figure 7 b.](#)). It is tempting to speculate that this shift explains, at least in part why, unlike KBCP1 or KBCP2, the KBCP1P2 does not form a hetero-octamer *in vitro*. The SAXS data collected for different KEOPS complexes and their corresponding models may provide an additional explanation ([supplementary figure SAXS data](#)). Indeed, the interface between the Kae1 and Pcc1 (PDB code 3ENO) is stabilised by a four-helix bundle created between  $\alpha 1$  and  $\alpha 2$  helices of Pcc1 and  $\alpha 2$  and  $\alpha 1$  helices of Kae1. In the model of the equivalent interface in the KBCP2 hetero-octamer the  $\alpha 2$  helix of PaPcc2 is perpendicular to the  $\alpha 2$  helix of Pcc1 ([supplementary figure 8](#)). Such conformation results, in our model structure, in smaller contact surface with Kae1 and consequently, less stable interaction with Kae1. Coincidentally, the major difference between Pcc1 and Pcc2 orthologs resides in the sequence of  $\alpha 2$  helices ([supplementary figure sequence alignment and logo](#)) which contains several conserved residues specific either for Pcc1 or for Pcc2. The conserved Asn<sub>67</sub> of Pcc1 interacts with Pcc2-specific Asp/Glu<sub>70</sub> in the Pcc1-Pcc2 heterodimer interface while the conserved Thr/Ser<sub>67</sub> contributes together with conserved Arg<sub>71</sub>Trp<sub>72</sub> to the stabilization of the Kae1-Pcc1 interface. The corresponding Pcc2 residue is Asp/Glu<sub>71</sub> which cannot establish an equivalent interaction. Based on these data we propose that the arrangement Kae1-Pcc1-Pcc2 is the most stable conformation for the archaeal KEOPS complex.

In spite of absence of sequence or structural similarity between Pcc2 and Gon7 the interface PaPcc1-PaPcc2 and hGon7-LAGE3 is remarkably similar ([figure 5](#)). The superposition of these crystallographic structures (RMSD = 1.64 Å, Z-score = 8, 106 positions of C $\alpha$  aligned) revealed that both heterodimers are stabilised through a 6-strand  $\beta$  or 5-strand  $\beta$  antiparallel sheets and two helices. This suggest that this interface with the KEOPS fifth subunit is conserved from archaea to eukaryotes.

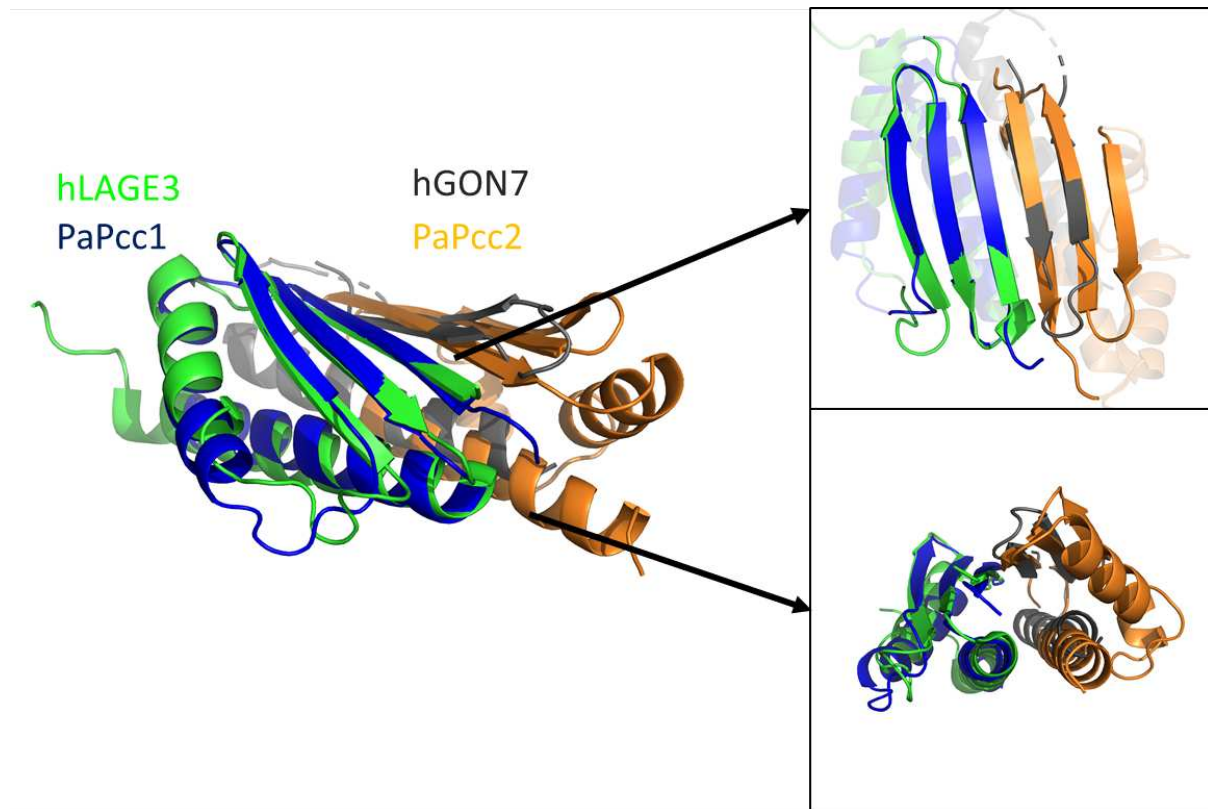


Figure 5: Structural comparison of Pcc1-Pcc2 and hLAGE3-hGon7 heterodimers. Superimposition of the crystallographic PaPcc1-PaPcc2 structures (in blue and orange, respectively) on hLAGE3-hGON7 (in green and grey, respectively) is shown on the left. The top right insert is a zoom on the 6-strand  $\beta$  and 5-strand  $\beta$  antiparallel sheets stabilising the two interfaces, respectively. The bottom right insert is a zoom on the helices that also contribute to the building of the interfaces.

Archaeal KEOPS complex catalyses the formation of  $t^6A$  *in vitro* in absence of the fifth subunit (Perrochia *et al.* 2013). What is the oligomeric state and composition of archaeal KEOPS *in vivo* and the function of Pcc2 in the  $t^6A$  synthesis remains to be investigated. In analogy to Gon7 proteins we can speculate that Pcc2 may have a regulatory role rather than direct function in the  $t^6A$  synthesis.

The HMM sequence searches for Pcc2 orthologs in public databases did not retrieve any significant hits outside Archaea indicating that if Pcc2 related proteins exist in eukaryotes they are highly divergent. This would not be surprising given that the  $\gamma$ Gon7 and hGon7, both found in opisthokonts, are highly divergent. Nevertheless, our study provides evidence that the interface between Pcc1 orthologs and the archaeal or eukaryotic fifth subunits is highly conserved even across the long evolutionary distances. We can therefore predict that fifth

KEOPS subunits different from Gon7 will be found in eukaryotes outside of Fungi and Metazoans.

	<b>PaPcc2</b>	<b>PaPcc1Pcc2</b>
PDB code	<b>7A66</b>	<b>7A67</b>
Number of molecules in asymmetric unit	3	2
Space group	C 1 2 1	P 32 2 1
Cell parameters (Å, °)	76.80 81.31 66.48 90.00 121.59 90.00	69.898 69.898 92.751 90.00 90.00 120.00
Resolution (Å)	50-1.84 (1.96-1.84)	50.00-3.18 (3.36-3.18)
X-ray wavelength	0.983988	0.980066
No. of total reflections	205991 (32315)	87363 (11857)
No. of unique reflections	29812 (4675)	4659 (678)
Completeness (%)	99.5 (96.9)	98.4 (91)
$I/\sigma$	18.24 (1.08)	10.54 (1.54)
$CC_{1/2}$	99.9 (65.5)	99.7 (85.9)
$R_{\text{cryst}}$	0.204 ( )	0.185
$R_{\text{free}}$	0.229 (0.3270)	0.250
RMSD bond lengths (Å)	0.010	0.010
RMSD bond angles (°)	1.12	1.25
Average $B$ (Å <sup>2</sup> )	45.0	60.0
Clashscore <sup>a</sup>	2.07	10.5
MolProbity score <sup>a</sup>	0.98	3.09
Ramachandran plot (%)		
Favored	98.37	89.38
Allowed	1.63	9.37
Outliers	0.00	1.25

**Table 1: Data collection and refinement statistics. Values for the highest resolution shell are in parentheses.  $CC_{1/2}$  = percentage of correlation between intensities from random half-dataset. <sup>a</sup> Calculated with MolProbity (MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution).**



	KBCP1P2	KBCP1	KBCP2
<b>Data collection parameters</b>			
Instrument	SWING (SOLEIL) CCD-based AVIEX 0.8 mm x 0.15 mm 1.0 $0.007 < q < 0.50$ Comparison with scattering from pure H <sub>2</sub> O 2 283		
Detector			
Beam geometry			
Wavelength [Å]			
q-range [Å <sup>-1</sup> ]			
Absolute scaling			
Exposure time [s]			
Temperature [K]			
SEC-SAXS column	Biosec 3 Agilent		
Loading concentration [mg/mL]	2.55	0.72	3
Injection volume [μL]	50		
Flow rate [mL.min <sup>-1</sup> ]	0.3		
<b>Structural parameters</b>			
I(0) Guinier [cm <sup>-1</sup> ]	<b>8.75 10<sup>-3</sup></b>	<b>7.46 10<sup>-3</sup> ± 8.44 10<sup>-5</sup></b>	<b>4.77 10<sup>-2</sup> ± 1.13 10<sup>-4</sup></b>
R <sub>g</sub> Guinier [Å]	<b>36.9</b>	<b>52.5 ± 0.873</b>	<b>52.4 ± 0.19</b>
qR <sub>g</sub> -range	<b>0.45-1.3</b>	<b>0.479-1.29</b>	<b>0.47-1.27</b>
I(0) p(r) [cm <sup>-1</sup> ]	<b>8.90 10<sup>-3</sup></b>	<b>7.54 10<sup>-3</sup> ± 0.90-4</b>	<b>4.83 10<sup>-2</sup> ± 1.22 10<sup>-5</sup></b>
R <sub>g</sub> p(r) [Å]	<b>38.87 ± 0.23</b>	<b>54.42 ± 0.88</b>	<b>54.89 ± 0.22</b>
q-range [Å <sup>-1</sup> ]	<b>0.0091-0.2639</b>	<b>0.091 – 0.197</b>	<b>0.0078 – 0.1856</b>
D <sub>Max</sub> [Å]	<b>125</b>	<b>188</b>	<b>205</b>
<b>Molecular mass determination</b>			
Mol . L <sup>-1</sup> MM <sub>sequence</sub> [kDa] <sup>b</sup>	<b>101.457</b>	<b>91.146</b>	<b>90.871</b>
MM <sub>bayesian</sub> [kDa] <sup>c</sup> Credibility interval	<b>~94.225</b> <b>85.6-100.3</b>	<b>~185.775</b> <b>162.65-194.95</b>	<b>~ 180.5</b> <b>151.45-194.95</b>
Oligomerization Stoichiometry	<b>pentamer</b> <b>1:1:1:1:1</b>	<b>Octamer</b> <b>2:2:2:2</b>	<b>Octamer</b> <b>2:2:2:2</b>

**Table 2: SAXS results and data collection parameters.**

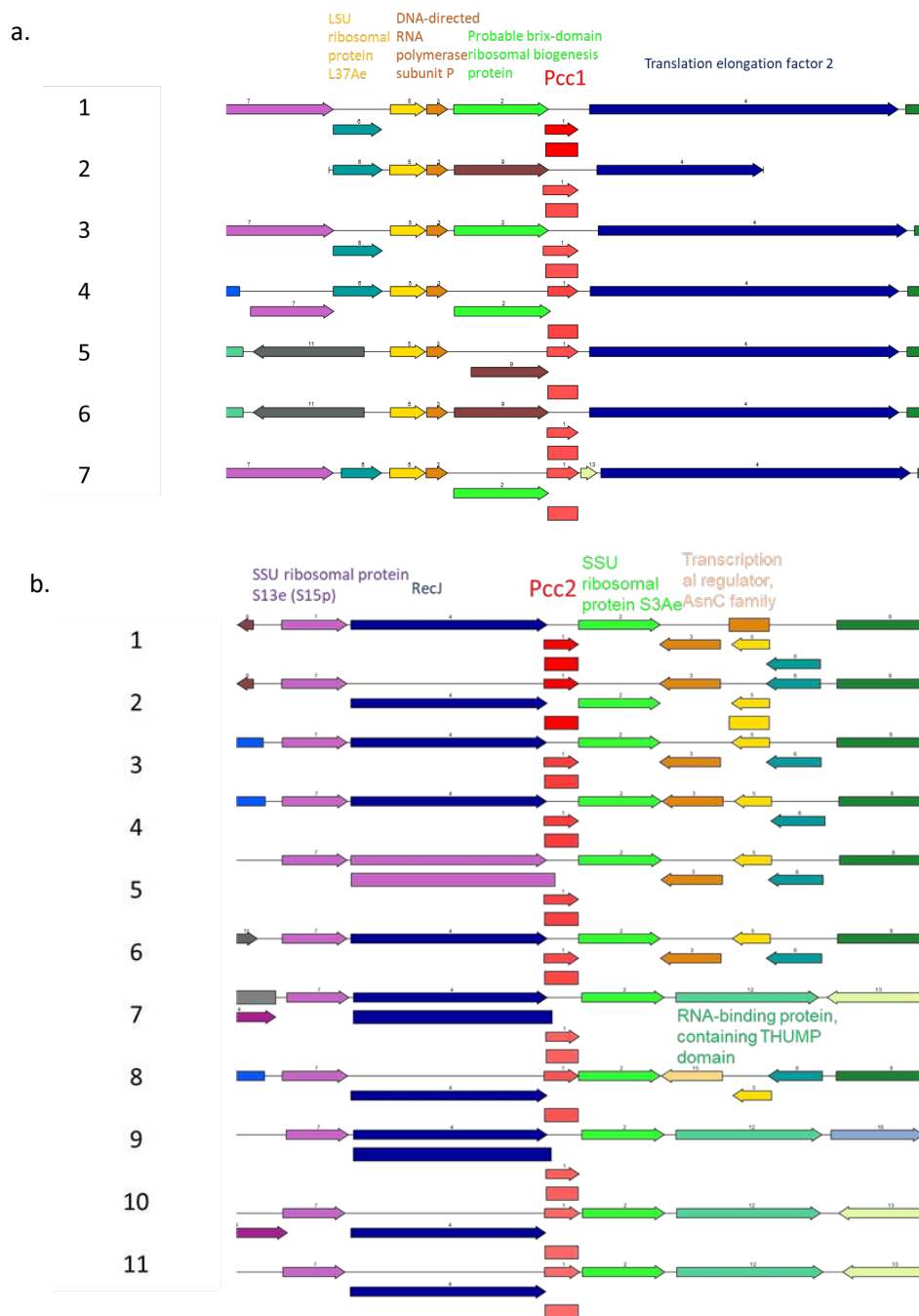
- Agris, P. F. *et al.* (2017) 'The Importance of Being Modified', in *The Enzymes*. Elsevier, pp. 1–50. doi: 10.1016/bs.enz.2017.03.005.
- Agris, P. F., Vendeix, F. A. P. and Graham, W. D. (2007) 'tRNA's Wobble Decoding of the Genome: 40 Years of Modification', *Journal of Molecular Biology*, 366(1), pp. 1–13. doi: 10.1016/j.jmb.2006.11.046.
- Arrondel, C. *et al.* (2019) 'Defects in t6A tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome', *Nature Communications*, 10. doi: 10.1038/s41467-019-11951-x.
- Boccaletto, P. *et al.* (2018) 'MODOMICS: a database of RNA modification pathways. 2017 update', *Nucleic Acids Research*, 46(Database issue), pp. D303–D307. doi: 10.1093/nar/gkx1030.
- Braddock, D. T. *et al.* (2002) 'Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA', *The EMBO Journal*, 21(13), pp. 3476–3485. doi: 10.1093/emboj/cdf352.
- Braun, D. A. *et al.* (2017) 'Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly', *Nature Genetics*, 49(10), pp. 1529–1538. doi: 10.1038/ng.3933.
- Chheda, G. B. *et al.* (1969) 'Aminoacyl nucleosides. VI. Isolation and preliminary characterization of threonyladenine derivatives from transfer ribonucleic acid', *Biochemistry*, 8(8), pp. 3278–3282. doi: 10.1021/bi00836a022.
- Costessi, A. *et al.* (2012) 'The Human EKC/KEOPS Complex Is Recruited to Cullin2 Ubiquitin Ligases by the Human Tumour Antigen PRAME', *PLoS ONE*. Edited by L. Tora, 7(8), p. e42822. doi: 10.1371/journal.pone.0042822.
- de Crécy-Lagard, V. and Jaroch, M. (2020) 'Functions of Bacterial tRNA Modifications: From Ubiquity to Diversity', *Trends in Microbiology*, p. S0966842X2030189X. doi: 10.1016/j.tim.2020.06.010.
- Crick, F. H. (1958) 'On protein synthesis', *Symposia of the Society for Experimental Biology*, 12, pp. 138–163.
- 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 evolutionary biology*, 10, p. 210. doi: 10.1186/1471-2148-10-210.
- Downey, M. *et al.* (2006) 'A Genome-Wide Screen Identifies the Evolutionarily Conserved KEOPS Complex as a Telomere Regulator', *Cell*, 124(6), pp. 1155–1168. doi: 10.1016/j.cell.2005.12.044.
- El Yacoubi, B. *et al.* (2009) 'The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA', *Nucleic Acids Research*, 37(9), pp. 2894–2909. doi: 10.1093/nar/gkp152.
- El Yacoubi, B. *et al.* (2011) 'A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification', *The EMBO Journal*, 30(5), pp. 882–893. doi: 10.1038/emboj.2010.363.
- Evrard, G. *et al.* (2011) 'DADIMODO : a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints', *Journal of Applied Crystallography*, 44(6), pp. 1264–1271. doi: 10.1107/S0021889811035758.
- Franke, D. *et al.* (2017) 'ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions', *Journal of Applied Crystallography*, 50(Pt 4), pp. 1212–1225. doi: 10.1107/S1600576717007786.

- Grishin, N. V. (2001) 'KH domain: one motif, two folds', *Nucleic Acids Research*, 29(3), pp. 638–643. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC30387/> (Accessed: 13 November 2018).
- Grosjean, H., Sprinzl, M. and Steinberg, S. (1995) 'Posttranscriptionally modified nucleosides in transfer RNA: Their locations and frequencies', *Biochimie*, 77(1), pp. 139–141. doi: 10.1016/0300-9084(96)88117-X.
- Guindon, S. *et al.* (2010) 'New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0', *Systematic Biology*, 59(3), pp. 307–321. doi: 10.1093/sysbio/syq010.
- Hoagland, M. B. *et al.* (1958) 'A Soluble Ribonucleic Acid Intermediate in Protein Synthesis', *Journal of Biological Chemistry*, 231(1), pp. 241–257. Available at: <http://www.jbc.org/content/231/1/241> (Accessed: 19 March 2020).
- Hoang, D. T. *et al.* (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35(2), pp. 518–522. doi: 10.1093/molbev/msx281.
- Holley, R. W. *et al.* (1965) 'Structure of a Ribonucleic Acid', *Science*, 147(3664), pp. 1462–1465. doi: 10.1126/science.147.3664.1462.
- Jackman, J. E. and Alfonzo, J. D. (2013) 'Transfer RNA modifications: Nature's combinatorial chemistry playground', *Wiley interdisciplinary reviews. RNA*, 4(1), pp. 35–48. doi: 10.1002/wrna.1144.
- Katoh, K., Rozewicki, J. and Yamada, K. D. (2019) 'MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization', *Briefings in Bioinformatics*, 20(4), pp. 1160–1166. doi: 10.1093/bib/bbx108.
- Kisseleva-Romanova, E. *et al.* (2006) 'Yeast homolog of a cancer-testis antigen defines a new transcription complex', *The EMBO Journal*, 25(15), pp. 3576–3585. doi: 10.1038/sj.emboj.7601235.
- Knüppel, R. *et al.* (2018) 'Insights into the evolutionary conserved regulation of Rio ATPase activity', *Nucleic Acids Research*, 46(3), pp. 1441–1456. doi: 10.1093/nar/gkx1236.
- Lauhon, C. T. (2012) 'Mechanism of N<sup>6</sup>-Threonylcarbamoyladenonsine (t<sup>6</sup>A) Biosynthesis: Isolation and Characterization of the Intermediate Threonylcarbamoyl-AMP', *Biochemistry*, 51(44), pp. 8950–8963. doi: 10.1021/bi301233d.
- Lemoine, F. *et al.* (2018) 'Renewing Felsenstein's phylogenetic bootstrap in the era of big data', *Nature*, 556(7702), pp. 452–456. doi: 10.1038/s41586-018-0043-0.
- Lin, C.-J. *et al.* (2015) 'An extensive allelic series of *Drosophila* kae1 mutants reveals diverse and tissue-specific requirements for t<sup>6</sup>A biogenesis', *RNA*, 21(12), pp. 2103–2118. doi: 10.1261/rna.053934.115.
- Lopreiato, R. *et al.* (2004) 'Analysis of the interaction between piD261/Bud32, an evolutionarily conserved protein kinase of *Saccharomyces cerevisiae*, and the Grx4 glutaredoxin', *Biochemical Journal*, 377(2), pp. 395–405. doi: 10.1042/bj20030638.
- Luthra, A. *et al.* (2018) 'Structure and mechanism of a bacterial t<sup>6</sup>A biosynthesis system', *Nucleic Acids Research*, 46(3), pp. 1395–1411. doi: 10.1093/nar/gkx1300.
- Luthra, A. *et al.* (2019) 'Conformational communication mediates the reset step in t<sup>6</sup>A biosynthesis', *Nucleic Acids Research*, 47(12), pp. 6551–6567. doi: 10.1093/nar/gkz439.

- Machnicka *et al.* (2013) 'MODOMICS: a database of RNA modification pathways—2013 update', pp. D262–D267. Available at: doi:10.1093/nar/gks1007.
- Mao, D. Y. L. *et al.* (2008) 'Atomic Structure of the KEOPS Complex: An Ancient Protein Kinase-Containing Molecular Machine', *Molecular Cell*, 32(2), pp. 259–275. doi: 10.1016/j.molcel.2008.10.002.
- Marck, C. and Grosjean, H. (2002) 'tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features', *RNA*, 8(10), pp. 1189–1232. doi: 10.1017/S1355838202022021.
- Missoury, S. *et al.* (2018) 'The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t6A tRNA-modification', *Nucleic Acids Research*, 46(11), pp. 5850–5860. doi: 10.1093/nar/gky323.
- Murphy, F. V. *et al.* (2004) 'The role of modifications in codon discrimination by tRNALysUUU', *Nature Structural & Molecular Biology*, 11(12), pp. 1186–1191. doi: 10.1038/nsmb861.
- Naor, A. *et al.* (2012) 'A Genetic Investigation of the KEOPS Complex in Halophilic Archaea', *PLoS ONE*. Edited by A. J. Lustig, 7(8), p. e43013. doi: 10.1371/journal.pone.0043013.
- Nguyen, L.-T. *et al.* (2015) 'IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Molecular Biology and Evolution*, 32(1), pp. 268–274. doi: 10.1093/molbev/msu300.
- Nicastro, G., Taylor, I. A. and Ramos, A. (2015) 'KH–RNA interactions: back in the groove', *Current Opinion in Structural Biology*, 30, pp. 63–70. doi: 10.1016/j.sbi.2015.01.002.
- Ogura, M. (2020) 'Glucose-Mediated Protein Arginine Phosphorylation/Dephosphorylation Regulates ylxR Encoding Nucleoid-Associated Protein and Cell Growth in *Bacillus subtilis*', *Frontiers in Microbiology*, 11, p. 590828. doi: 10.3389/fmicb.2020.590828.
- Perrochia, L., Guetta, D., *et al.* (2013) 'Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification', *Nucleic Acids Research*, 41(20), pp. 9484–9499. doi: 10.1093/nar/gkt720.
- Perrochia, L., Crozat, E., *et al.* (2013) 'In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya', *Nucleic Acids Research*, 41(3), pp. 1953–1964. doi: 10.1093/nar/gks1287.
- 'Pollo-Oliveira *et al.* - 2020 - Loss of Elongator- and KEOPS-Dependent tRNA Modifi.pdf' (no date).
- Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42(W1), pp. W320–W324. doi: 10.1093/nar/gku316.
- Sali, A. *et al.* (1995) 'Evaluation of comparative protein modeling by MODELLER', *Proteins*, 23(3), pp. 318–326. doi: 10.1002/prot.340230306.
- Schneidman-Duhovny, D. *et al.* (2016) 'FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles', *Nucleic Acids Research*, 44(W1), pp. W424–W429. doi: 10.1093/nar/gkw389.
- Schweizer, M. P. *et al.* (1969) 'Aminoacyl nucleosides. VII. N-(purin-6-ylcarbonyl)threonine. A new component of transfer ribonucleic acid', *Biochemistry*, 8(8), pp. 3283–3289. doi: 10.1021/bi00836a023.

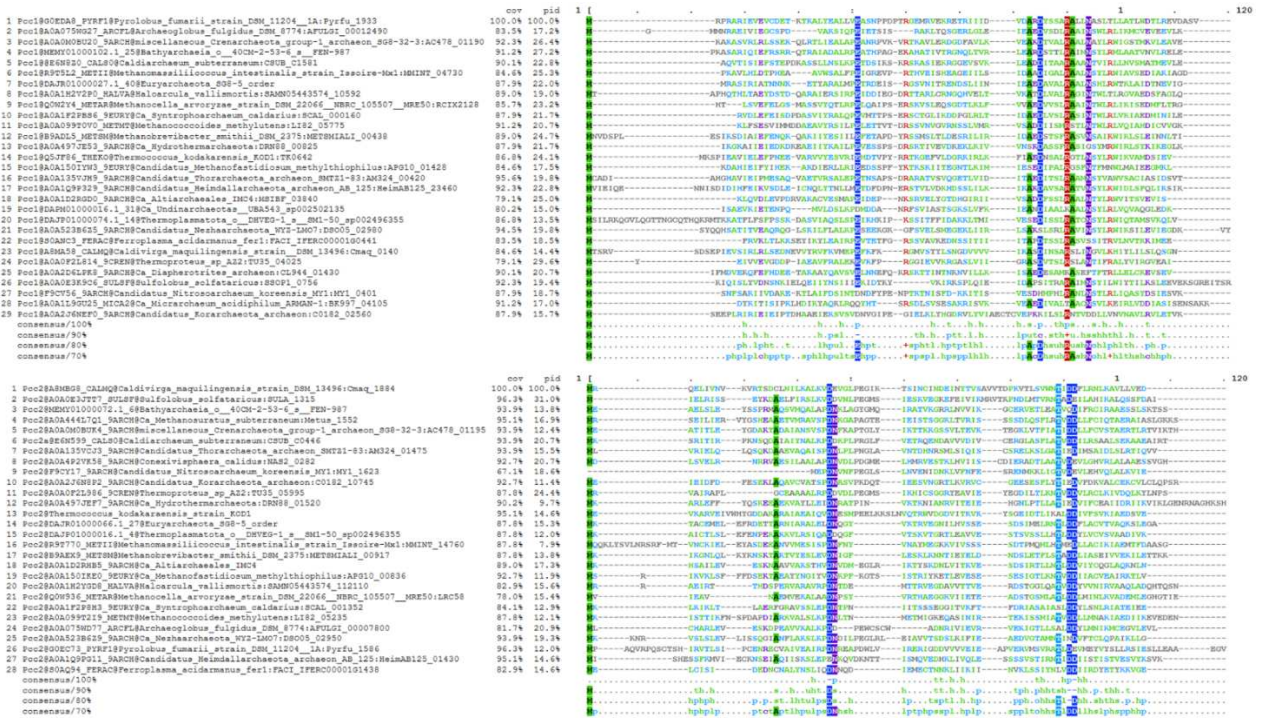
- Srinivasan, M. *et al.* (2011) 'The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A', *The EMBO Journal*, 30(5), pp. 873–881. doi: 10.1038/emboj.2010.343.
- Sundaram, M., Durant, P. C. and Davis, D. R. (2000) 'Hypermodified nucleosides in the anticodon of tRNA<sup>Lys</sup> stabilize a canonical U-turn structure', *Biochemistry*, 39(41), pp. 12575–12584. doi: 10.1021/bi0014655.
- Svergun, D., Barberato, C. and Koch, M. H. J. (1995) 'CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates', *Journal of Applied Crystallography*, 28(6), pp. 768–773. doi: 10.1107/S0021889895007047.
- Takemura, S., Murakami, M. and Miyazaki, M. (1969) 'Nucleotide Sequence of Isoleucine Transfer RNA from *Torulopsis utilis*', *The Journal of Biochemistry*, 65(3), pp. 489–491. doi: 10.1093/oxfordjournals.jbchem.a129040.
- Thiaville, P. C. *et al.* (2014) 'Cross Kingdom Functional Conservation of the Core Universally Conserved Threonylcarbamoyladenine tRNA Synthesis Enzymes', *Eukaryotic Cell*, 13(9), pp. 1222–1231. doi: 10.1128/EC.00147-14.
- Thiaville, P. C. *et al.* (2016) 'Global translational impacts of the loss of the tRNA modification t6A in yeast', *Microbial Cell (Graz, Austria)*, 3(1), pp. 29–45. doi: 10.15698/mic2016.01.473.
- Valverde, R., Edwards, L. and Regan, L. (2008) 'Structure and function of KH domains', *The FEBS Journal*, 275(11), pp. 2712–2726. doi: 10.1111/j.1742-4658.2008.06411.x.
- Vendeix, F. A. P. *et al.* (2012) 'Human tRNA<sup>Lys</sup>3UUU Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto–Enol Tautomerism', *Journal of Molecular Biology*, 416(4), pp. 467–485. doi: 10.1016/j.jmb.2011.12.048.
- Wan, L. C. K. *et al.* (2016) 'Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t6A biosynthesis', *Nucleic Acids Research*, 44(14), pp. 6971–6980. doi: 10.1093/nar/gkw542.
- Wan, L. C. K. *et al.* (2017) 'Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7', *Nucleic Acids Research*, 45(2), pp. 805–817. doi: 10.1093/nar/gkw1181.
- Zhang, W., Collinet, B., Graille, M., *et al.* (2015) 'Crystal structures of the Gon7/Pcc1 and Bud32/Cgi121 complexes provide a model for the complete yeast KEOPS complex', *Nucleic Acids Research*, 43(6), pp. 3358–3372. doi: 10.1093/nar/gkv155.
- Zhang, W., Collinet, B., Perrochia, L., *et al.* (2015) 'The ATP-mediated formation of the YgjD–YeaZ–YjeE complex is required for the biosynthesis of tRNA t6A in *Escherichia coli*', *Nucleic Acids Research*, 43(3), pp. 1804–1817. doi: 10.1093/nar/gku1397.

## Supplementary information



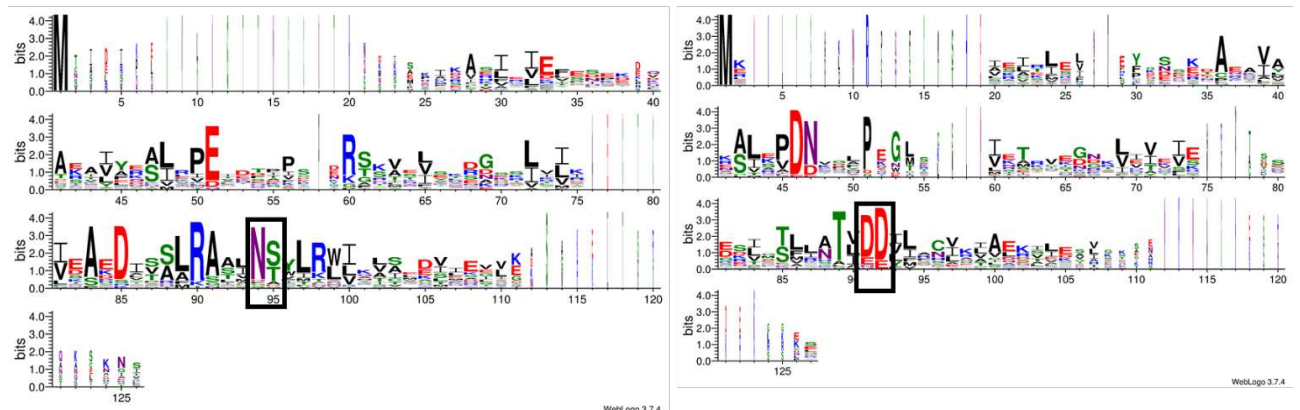
Supplementary figure 1: a. Genomic context of *PCC1* into (1) *Pyrococcus abyssi* GE5, (2) *Thermococci archaeon strain B21\_G6*, (3) *Pyrococcus sp. NCB100*, (4) *Pyrococcus sp. NA2*, (5) *Pyrococcus horikoshii* OT3, (6) *Pyrococcus horikoshii* strain UBA8834, (7) *Pyrococcus sp. ST04*. Synténie was performed from research of *PCC1* (in red) localised from 280821 to 281054 on *Pyrococcus abyssi* GE5 genome and compared to other genomes into database PATRIC (<https://www.patricbrc.org/>). b. Genomic context of *PCC2* into (1) *Pyrococcus furiosus* COM1, (2) *Pyrococcus furiosus* DSM3638, (3) *Pyrococcus horikoshii* OT3, (4) *Pyrococcus sp. ST04*, (5) *Pyrococcus sp. NA2*, (6) *Pyrococcus abyssi* GE5, (7) *Thermococcus barophilus* MP, (8) *Pyrococcus yayanosii* CH1, (9) *Thermococcus sp. ES1*, (10) *Thermococcus sp.4557*, (11) *Thermococcus celericrescens* strain DSM 17994. Synténie was performed as *PCC1*.

a.

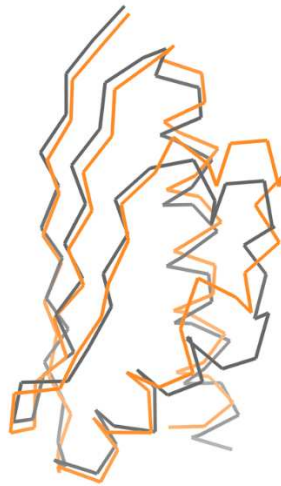


b. Pcc1

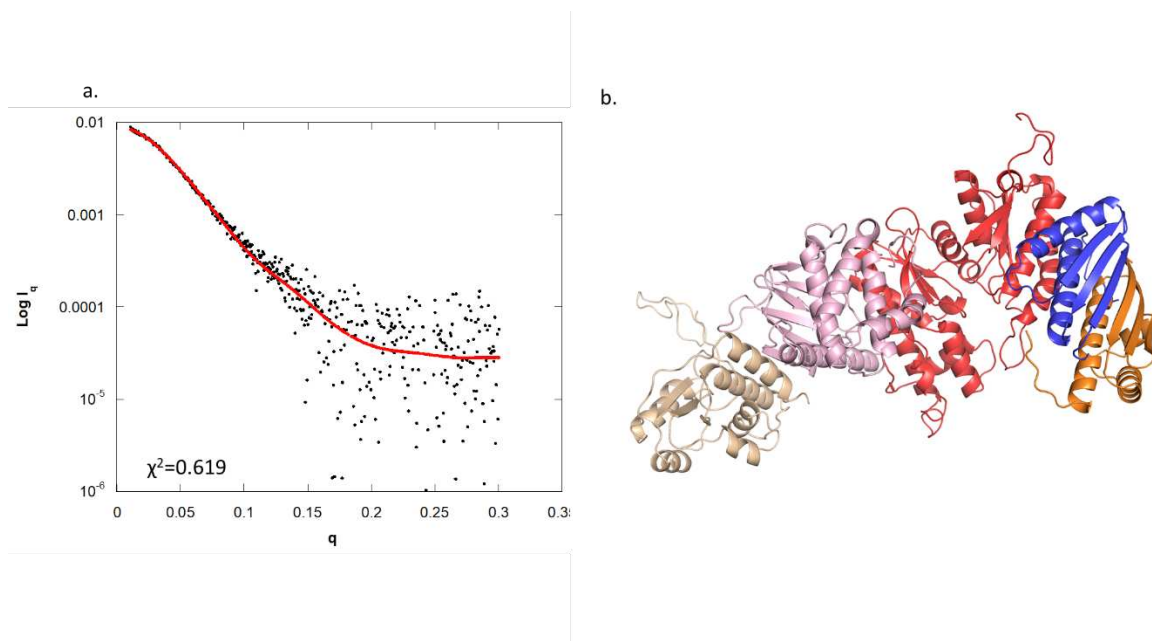
Pcc1-like



Supplementary figure 2: Alignment of Pcc1 and Pcc1-like sequences. a. alignment of Pcc1 (top) and Pcc1-like (below) sequences. The alignment was done using MAFFT v7. MView program was used to visualize the alignment. The residues conserved in 70 % of sequences or more are highlighted and the residues with similar physico-chemical properties are indicated in the same color. The percent of pairwise identity with respect to the top sequence is indicated next to the sequence name. b. sequence logo obtained from the alignment using WebLogo 3 (see materials and methods). The position in the sequence is indicated on the X-axes, the conservation level for each position is indicated in bits on the Y axes. The size of each letter is proportional to the frequency of their occurrence in the alignment. The motifs N<sub>67</sub>S/T<sub>68</sub> and D/E<sub>70</sub>D/E<sub>71</sub>T specific for Pcc1 or Pcc1-like sequences, respectively, are boxed.

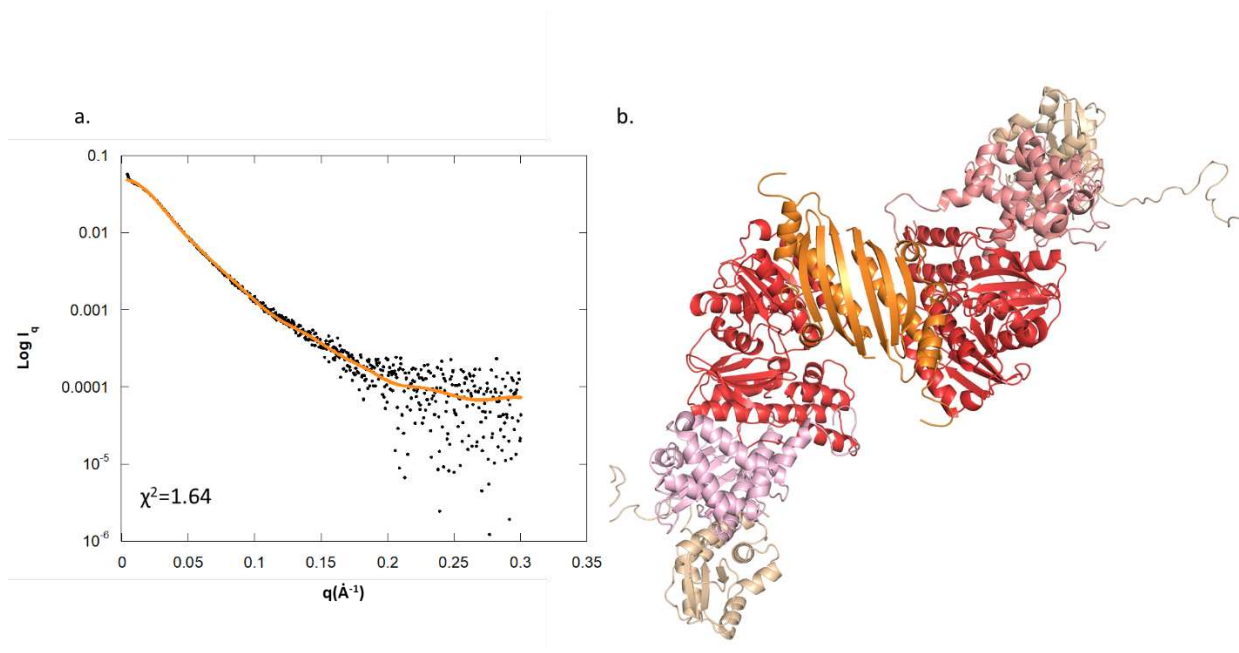


Supplementary figure 3: Superimposition of Pcc2 alone (in orange) and Pcc2 in complex with Pcc1 (in grey).

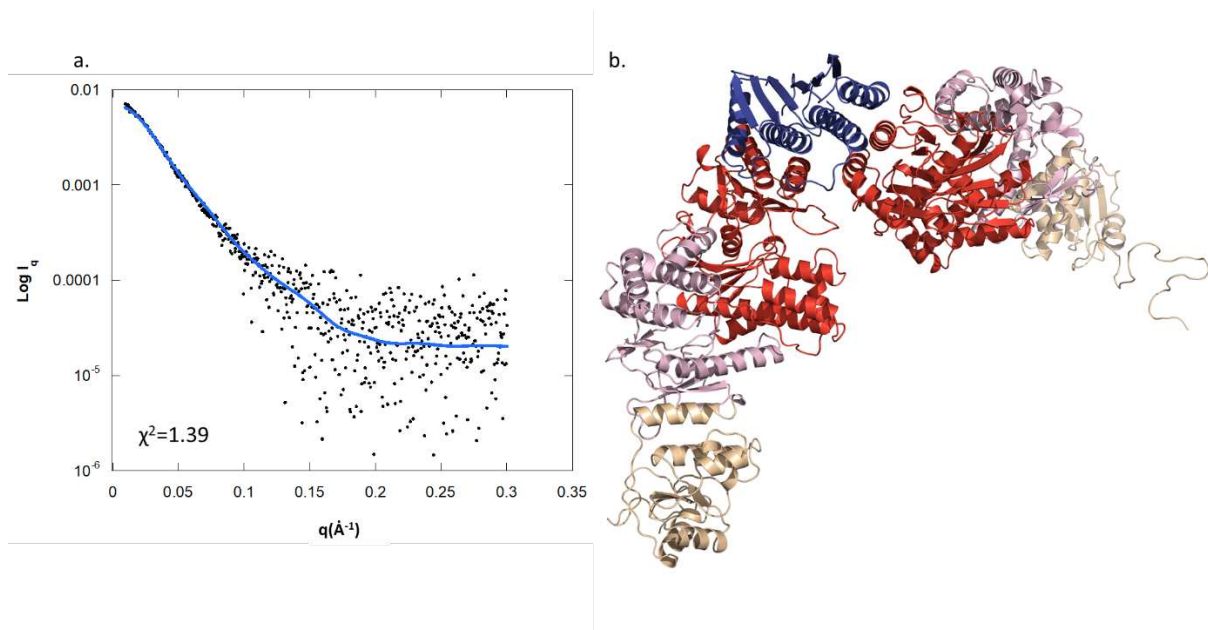


Supplementary figure 4: a. Saxs data (black dots) of the KBCP1P2 complex. The red curve represents a theoretical scattering curve of a model of the pentameric KBCP1P2 complex, represented in (b). Cgi1321 (Peach), Bud32 (Rose), Kae1 (red), Pcc1 (blue) and Pcc2 (orange).

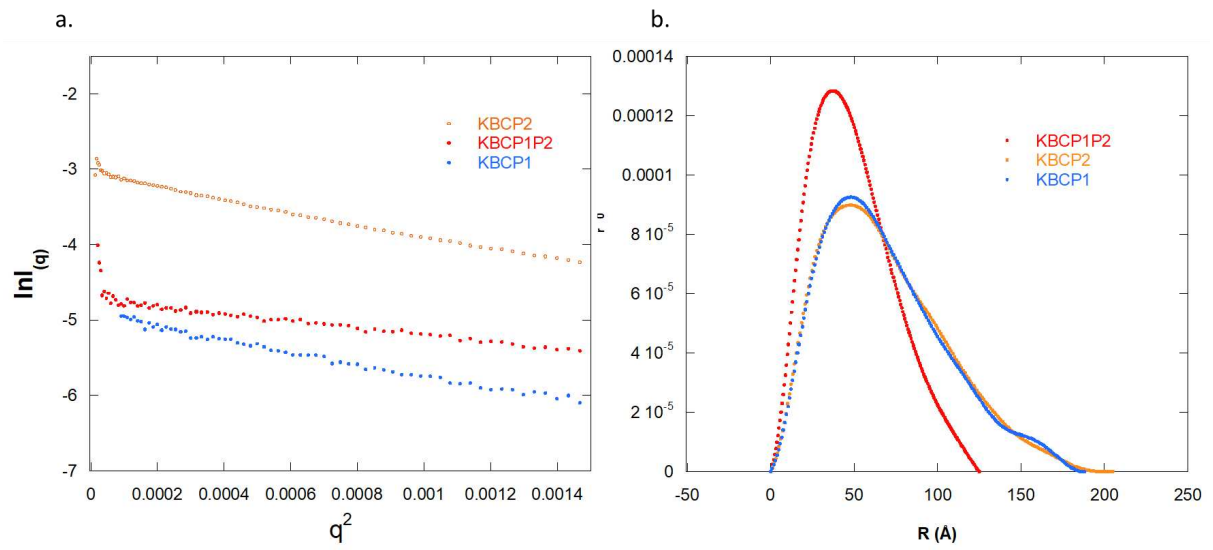




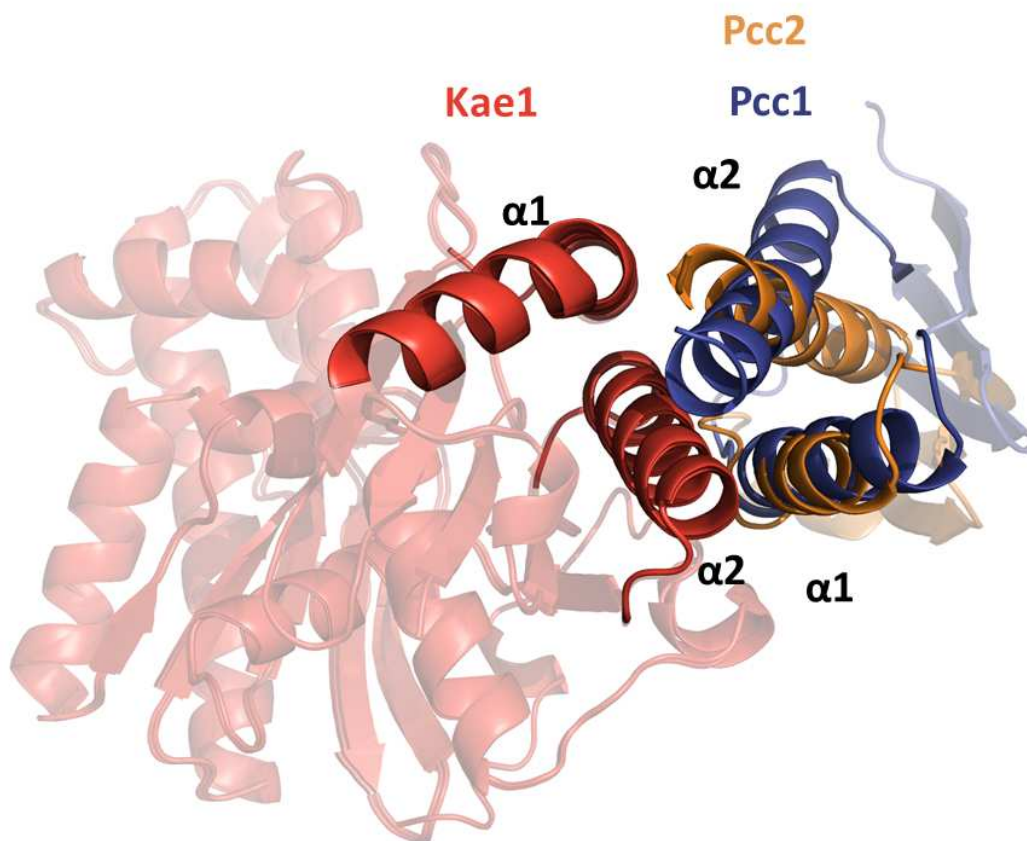
Supplementary figure 5: a. Saxes data (black dots) of the KBCP2 complex. The orange curve represents a theoretical scattering curve of a model of an octameric KBCP2 complex, represented in (b). Cgi132 (Peach), Bud32 (Rose), Kae1 (red) and Pcc2 (orange). Pcc2 forms the dimerization unit of the octamer.



Supplementary figure 6: a. Experimental saxes data (black dots) of the KBCP1 complex. The blue curve represents a theoretical scattering curve of a model of an octameric KBCP1P2 complex, represented in (b). Cgi132 (Peach), Bud32 (Rose), Kae1 (red) and Pcc1 (blue). Pcc1 forms the dimerization unit of the octamer.



Supplementary figure 7: a. Guinier plot of KBCP2 in orange, KBCP1P2 in red and KBCP1 in blue b. Distance distribution ( $P(r)$ ) curves for KBCP2 (orange) and for the KBCP1 (blue) and KBCP1P2 (red) complexes. The normalized distance distribution was calculated by PrimusQT implemented in Atsas software.



Supplementary figure 8: Comparison of interaction mode between Kae1 and Pcc1 or Pcc2.



## IV.2. Conclusion.

L'identification de PaPcc2 et la résolution de sa structure tridimensionnelle a mis en évidence et confirmé plusieurs points :

- Malgré seulement 20 % d'identité de séquence, PaPcc1 et PaPcc2 possèdent des structures 3D très similaires confirmant que PaPcc2 est le paralogue de PaPcc1.
- *In vitro*, le complexe KBCP<sub>1</sub> se dimérise via la sous-unité PaPcc1 et forme un octamère de stœchiométrie 2 :2 :2 :2. La structure 3D de l'homodimère PfPcc1 (code PDB 3ENO) montre que l'oligomérisation permet d'enfouir et de stabiliser la surface hydrophobe de chaque monomère PfPcc1.
- *In vitro*, la présence de Pcc2 empêche la dimérisation du complexe KBCP<sub>1</sub> permettant un agencement de stœchiométrie 1 :1 :1 :1 :1. La courbe de diffusion des intensités SAXS de KBCP<sub>1</sub>P<sub>2</sub> mesure sans ambiguïté que ce complexe est un pentamère.
- *De façon surprenante, in vitro*, le complexe KBCP<sub>2</sub> se dimérise aussi et forme un octamère de stœchiométrie 2 :2 :2 :2. Comme la structure cristallographique de PaPcc2 montre un homodimère dans l'unité asymétrique semblable à celui de PfPcc1, l'hypothèse que PaPcc2 forme la plateforme de dimérisation a été émise et un modèle 3D *in silico* de KBCP<sub>2</sub> dimérique a été réalisé. Ce modèle s'ajuste de façon satisfaisante à la courbe des intensités expérimentales de SAXS confirmant cet agencement.

Perrochia *et al.*, a mis en évidence que le complexe KBCP<sub>1</sub> interagit *in vitro* avec un ARNt substrat ( $K_d \sim 100$  à  $500$  nM) et catalyse la modification t<sup>6</sup>A (Perrochia *et al.*, 2013). Cependant aucune donnée, *ex vivo*, n'est disponible et permettrait de trancher quant à l'activité t<sup>6</sup>A supportée par une stœchiométrie 2 :2 :2 :2 ou 1 :1 :1 :1.

*In vitro*, Pcc2 réorganise l'oligomère en empêchant la dimérisation du complexe KEOPS et en stabilisant le complexe en pentamère. Des expériences biochimiques quant à l'effet de Pcc2 sur l'activité de la modification t<sup>6</sup>A sur les ARNt substrats sont prévues et permettront d'affiner le rôle de Pcc2 au sein du complexe KEOPS.

L'effet de Pcc2 sur l'organisation et la stœchiométrie du complexe est identique à celui de GON7/Gon7 sur le complexe KEOPS dans les champignons et les métazoaires. De plus, l'interface des hétérodimères PaPcc1Pcc2 et LAGE3-GON7 (code PDB 6GWJ) sont proches

(RMSD = 1,64 Å, Z score = 8 sur 106 positions C $\alpha$  alignées) ([figure 5 de l'article en cours page 285](#)). Ces similitudes montrent une conservation du mode d'interaction des différents hétérodimères. Il serait intéressant de faire des essais de complémentation de PaPcc1-Pcc2 dans la levure chez qui les gènes *PCC1* et *yGON7* ont été invalidés préalablement. Également, des tests *ex vivo* ou *in cellulo* pourraient être envisagés pour tester quant à l'existence du pentamère chez les archées.

## CONCLUSION GENERALE ET PERSPECTIVES.

### **Rappels sur la modification t<sup>6</sup>A.**

L'absence de la modification t<sup>6</sup>A est reliée à des phénotypes pléiotropes au sein des organismes, allant de la non-viabilité des bactéries à des défauts de développement de certains organes chez les métazoaires, dont le GAMOS observé chez l'Homme (Braun *et al.*, 2017; Lin *et al.*, 2010; Lin *et al.*, 2018; Thiaville *et al.*, 2014; Thiaville *et al.*, 2015). Chez la levure, le défaut t<sup>6</sup>A conduit à une croissance très altérée des clones mais une croissance minimale est toutefois observée en condition permissive. Ces phénotypes diffèrent selon les organismes et sont souvent la conséquence d'une mauvaise homéostasie de la synthèse des protéines, conduisant à leur agrégation ou à leur mauvais repliement (Thiaville *et al.*, 2016).

La famille d'enzyme ubiquitaire Kae1/OSGEP/TsaD/Qri7 nécessite des partenaires divers selon les espèces, formant des complexes oligomériques différents. Les fonctions des partenaires ne sont pas toujours bien comprises, ce qui nécessite leurs caractérisations afin de mieux comprendre leurs implications dans la modification t<sup>6</sup>A. Selon leurs contributions, un défaut de ces sous-unités pourrait expliquer aussi les variations des phénotypes dans les organismes.

Les travaux conduits pendant ma thèse sont majoritairement des études structurales des différents complexes oligomériques au sein des trois domaines du vivant.

## Les t<sup>6</sup>A synthèses au sein des procaryotes.

### Réarrangements structuraux de TsaD par TsaE chez les procaryotes.

A la suite des travaux de W. Zhang dans l'équipe, qui avait résolu la structure du sous complexe TsaBD d'*Escherichia coli* et avait remarqué que TsaE était essentielle pour l'activité t<sup>6</sup>A (Zhang *et al.*, 2015), nous avons poursuivi la caractérisation du complexe ternaire TsaBDE de *Thermotoga maritima*. La structure cristallographique du complexe de TsaBDE a permis de clarifier de nombreuses informations biochimiques (figure 101).

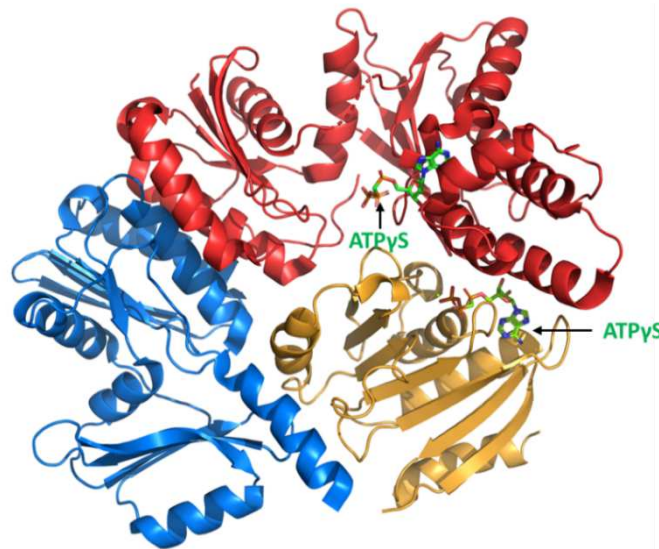


Figure 101 Structure quaternaire de TsaBDE de *Thermotoga maritima*. TsaE est en orange, TsaD en rouge et TsaB en bleu. Les deux nucléotides ATPγS sont représentés en sticks verts.

TsaE interagit avec TsaBD mais majoritairement avec TsaD (figure 101). La présence de l'ATP est nécessaire pour l'interaction entre TsaE et TsaBD (Zhang *et al.*, 2015). La structure de TsaBDE a effectivement montré qu'un analogue de l'ATP (ATPγS) est présent à l'interface de TsaD et TsaE (figure 102). En plus de son rôle de stabilisateur, le site de fixation du nucléotide partagé entre TsaD et TsaE explique l'augmentation de l'activité ATPase de TsaE observée en présence de TsaBD (Zhang *et al.*, 2015). Les acides aminés de TsaD impliqués dans la fixation du nucléotide sont très conservés dans les protéines TsaD (figure 102), en effet la lysine 166 est conservée dans les trois domaines du vivant alors que la lysine 213 est conservée uniquement chez les procaryotes. Ainsi, les protéines TsaD dans les procaryotes semblent avoir évoluées pour assurer la fixation du nucléotide avec TsaE et participer à l'hydrolyse de l'ATP.

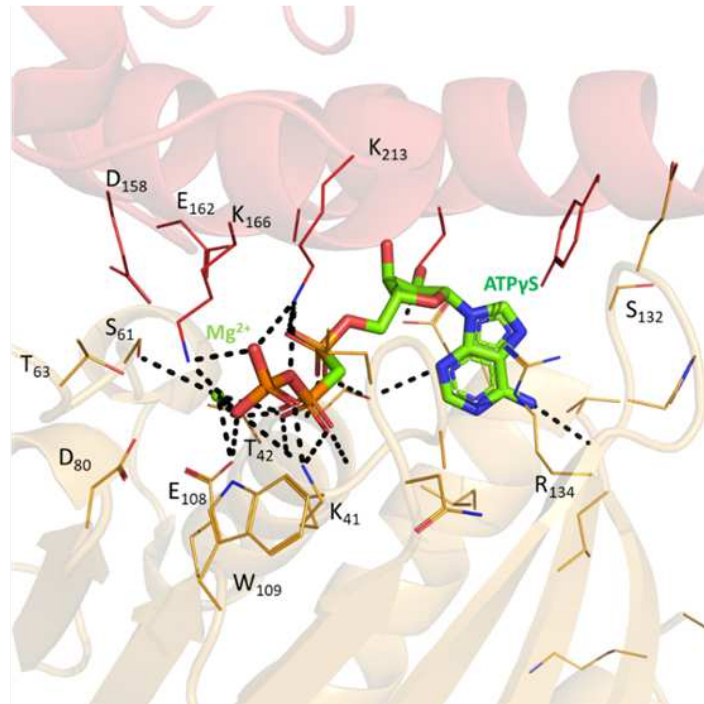


Figure 102 Site de fixation de l'ATP $\gamma$ S à l'interface de TsaD et TsaE. L'ATP $\gamma$ S est représenté en stick vert et est stabilisé par les acides aminés de TsaE et TsaD. Un ion Mg $^{2+}$  (ajouté dans la solution de cristallisation) est localisé à proximité et est coordonné par les chaînes latérales des acides aminés de TsaE et TsaD et également par les phosphates  $\beta$  et  $\gamma$  du nucléotide.

Il a été mis en évidence par l'équipe de M. Swarjko que le complexe ternaire TsaBDE est incompatible avec la fixation d'un ARNt (Luthra *et al.*, 2018). Notre structure quaternaire TmTsaBDE confirme pourquoi la fixation d'une molécule d'ARNt est incompatible avec la présence de TsaE (figure 101), car celle-ci occupe la surface de TsaBD majoritairement électropositive et l'entrée du site actif de TsaD.



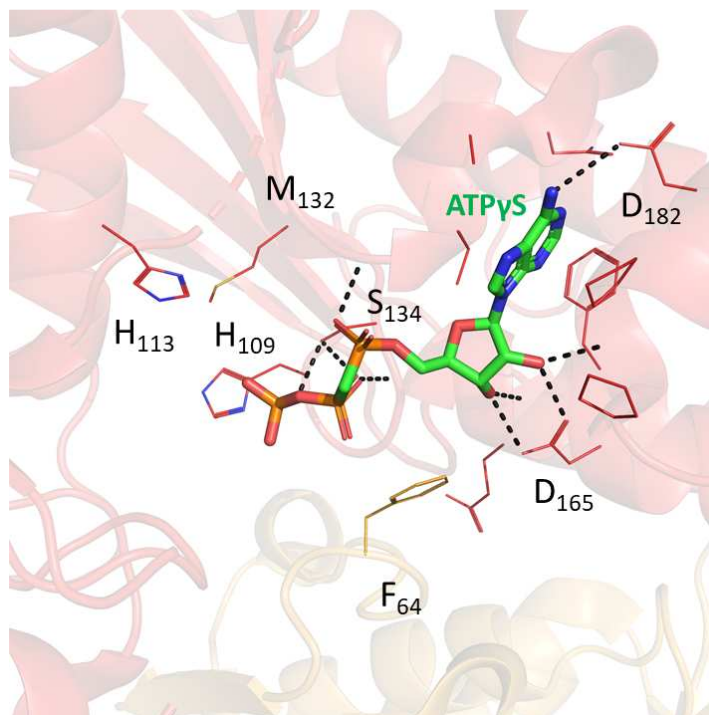


Figure 103 Site de fixation de l'ATP $\gamma$ S dans le site actif de TsaD. L'ATP $\gamma$ S est représenté en stick vert et est stabilisé par les acides aminés du domaine C-terminal de TsaD. La phénylalanine 64 de TsaE en orange se positionne à proximité de la boucle S<sub>134</sub>GGH<sub>137</sub> de TsaD qui participe à la stabilisation du nucléotide.

L'interaction avec TsaE induit aussi des changements drastiques de conformation de la sous-unité catalytique TsaD. D'abord, les domaines N et C-terminaux s'écartent l'un de l'autre provoquant une ouverture du site actif de TsaD. Malgré cette ouverture du site actif, TsaD est toujours capable de fixer une molécule d'ATP $\gamma$ S (figure 103) en maintenant les interactions de l'adénosine et du ribose avec les chaînes latérales des acides aminés du domaine C-terminal de TsaD, les mêmes qui sont observées dans des complexes TsaBD-nucléotides en absence de TsaE. Par contre les interactions de l'ATP $\gamma$ S avec les acides aminés du domaine N-terminal de TsaD ont été rompues par la présence de TsaE. En effet, les acides aminés H<sub>109</sub> et H<sub>111</sub> de TsaD sont désordonnés et ne fixent plus de métal. Ces résultats sont différents de ceux obtenus par l'équipe de M. Swarjo. Dans leur structure de TmTsaBDE, un métal est toujours présent malgré l'ouverture du site actif de TsaD (Luthra *et al.*, 2019), mais il n'est ni coordonné par les mêmes histidines ni par le phosphate  $\gamma$  du nucléotide comme dans EcTsaBD (code PDB 4YDU) (Zhang *et al.*, 2015). On peut donc se poser la question de savoir si on est en présence de structures qui représenteraient deux étapes différentes le long du chemin réactionnel.

La présence du nucléotide ATP $\gamma$ S dans la conformation ouverte de TsaD suggère que l'intermédiaire TC-AMP pourrait se fixer sur le complexe TsaBDE avant que l'ARNt substrat

n'arrive dans le site actif de TsaD. L'étude menée en collaboration avec Charles Lauhon (Université du Wisconsin) a permis d'observer le mode de fixation d'un analogue non hydrolysable de l'intermédiaire TC-AMP (la molécule BK951). Dans ce complexe, TsaD est en conformation fermée et un ion  $Zn^{2+}$  est coordonné par les histidines 109, 111 et l'aspartate 300 mais également par le groupement carboxylate du groupement thréonine de la molécule BK951. Il en a été conclu que ce métal jouerait un rôle dans la fixation de l'intermédiaire plutôt que dans la catalyse.

Des données cristallographiques ont montré qu'un second métal ( $Mg^{2+}$ ) se fixe dans le site actif de EcTsaD en conformation fermée (code PDB 4YDU) (Zhang *et al.*, 2015). Il est coordonné, d'une part, par les phosphates  $\alpha$  et  $\beta$  du nucléotide et d'autre part, par le glutamate 12 ( $E_{12}$ ), totalement conservé chez les procaryotes mais pas dans les orthologues eucaryotes ou d'archaea. Le mutant  $E_{12}A$  ne présente plus aucune activité  $t^6A$ , montrant qu'il est indispensable pour la catalyse (Zhang *et al.*, 2015). En revanche, nous n'avons pas observé d'ion  $Mg^{2+}$  dans notre structure de EcTsaBD (co-cristallisé avec la molécule BK951) (code PDB 6S84). L'une des explications possibles de l'absence de ce métal est la rotation du groupement phosphate  $\alpha$  causée, probablement, par la substitution d'un oxygène en carbone dans la molécule BK951. Le rôle du glutamate et la participation d'un éventuel second métal dans le mécanisme réactionnel n'ont donc pas été bien établis.

L'étude de mutagenèse dirigée, réalisée par l'équipe de C. Lauhon (Chapitre II), a permis en plus du glutamate 12 et des histidines 109 et 111, d'identifier des acides aminés supplémentaires qui sont indispensables pour la catalyse de la modification  $t^6A$ . La sérine 9 et la cystéine 10 qui contactent la partie thréonyle de la molécule BK951, sont des acides aminés conservés chez les procaryotes et localisés sur la même boucle que le  $E_{12}$ . Leur mutagenèse affecte également la catalyse, montrant que cette boucle joue un rôle important.

### **L'activité ATPase de TsaE est nécessaire pour la catalyse de la modification $t^6A$ .**

Il a été montré *in vitro*, dans le laboratoire, que TsaE était nécessaire pour la modification  $t^6A$  (Zhang *et al.*, 2015). En revanche, notre structure montre que l'interaction de TsaE avec TsaBD empêche la fixation de l'ARNt. Ces deux résultats qui semblent contradictoires dans un premier temps ont été partiellement conciliés par des expériences biochimiques menées dans l'équipe de M. Swarj. L'ensemble de ces données montre que TsaBD seul n'est capable

d'effecteur qu'un seul cycle de catalyse. En revanche, l'interaction avec TsaE, couplée à l'hydrolyse de l'ATP, maintient la machine catalytique de TsaD (Luthra *et al.*, 2018). Ceci a été montré grâce à l'étude du mutant E<sub>108</sub>A de TsaE, (E<sub>108</sub> interagit avec le nucléotide ATPγS) qui ne présente plus d'activité ATPase et n'interagit plus avec TsaBD (Zhang *et al.*, 2015). En présence de TsaBD, ce mutant ne permet qu'une seule modification t<sup>6</sup>A ("single turnover"), alors qu'en présence de la protéine sauvage, TsaBD effectue plusieurs cycles de modification t<sup>6</sup>A ("multiple turnover") (Luthra *et al.*, 2018).

Nos observations et celles de l'équipe Swarjoo permettent de dresser un schéma théorique de la modification t<sup>6</sup>A (figure 104). La fixation de TsaE provoque une ouverture et une perturbation partielle du site actif de TsaD. Cette conformation semble capable de fixer le substrat TC-AMP. L'hydrolyse de l'ATP à l'interface de TsaD et TsaE conduirait à la dissociation de TsaE. L'ARNt devrait alors interagir avec TsaBD et occuper un site de fixation, similaire à

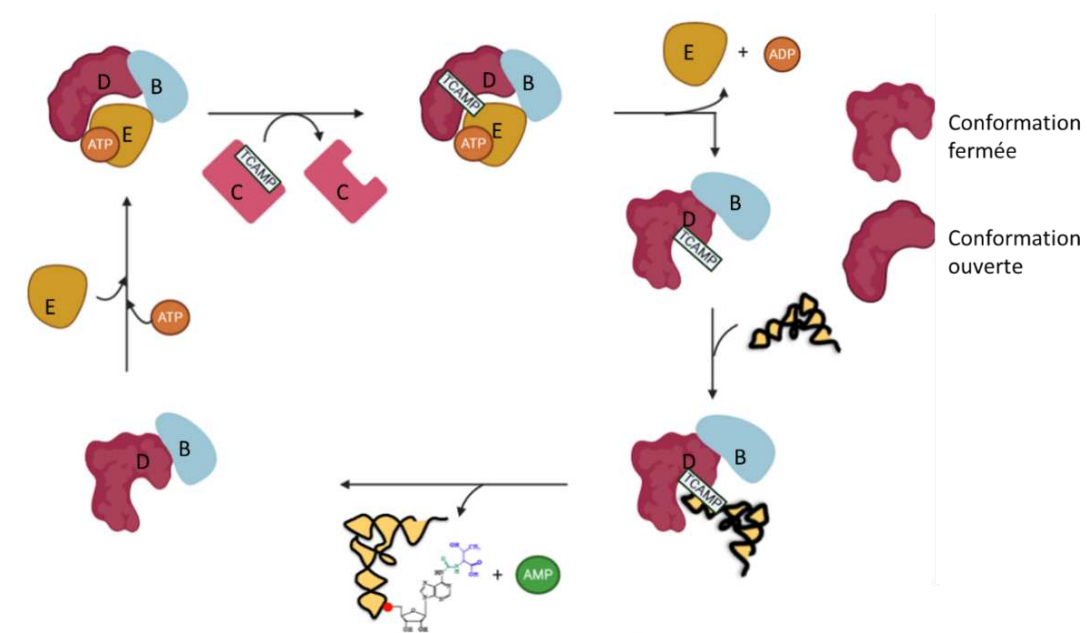


Figure 104 Schéma réactionnel hypothétique de la modification t<sup>6</sup>A chez les procaryotes.

celui de TsaE, pour subir la modification t<sup>6</sup>A. Ces derniers points sont encore très spéculatifs et leur confirmation nécessitera d'avoir des informations sur la structure du complexe TsaBD avec l'ARNt. Il se pourrait aussi que la conformation de TsaD soit inactive après dissociation de TsaE mais que la fixation de l'ARNt induise une forme fermée et active.

Stabiliser la forme inactive de TsaD pendant une partie du cycle réactionnel pourrait se justifier afin d'éviter l'hydrolyse de l'intermédiaire TC-AMP qui a été déjà montré comme étant

extrêmement labile et qu'il ne peut se maintenir qu'en se fixant dans un site actif, comme cela a été montré dans le cas de TsaC (Lauhon, 2012). La conformation ouverte de TsaD permettrait peut-être de fixer et de stabiliser le TC-AMP avant que l'ARNt n'arrive dans le site actif. Ce mécanisme putatif n'est certainement pas la seule possibilité du cycle réactionnel puisque d'autres homologues (QRI7, Kae1, OSGEP) n'ont pas besoin de TsaE.

### **Inhibition de TsaE : cible thérapeutique potentielle ?**

Lors de la modification t<sup>6</sup>A, la majorité des procaryotes nécessitent TsaE qui, en induisant des changements conformationnels, promeut le renouvellement de la sous-unité catalytique TsaD. Il a été montré que TsaE est présente dans la majorité des génomes des bactéries, à l'exception de certains organismes tels que des *Mollicutes* qui possèdent des génomes réduits (Deutsch *et al.*, 2012). De plus il n'a pas été identifié, à ce jour, d'homologues chez les eucaryotes. Ainsi, le développement de molécules inhibitrices qui ciblent la fonction de TsaE pourrait potentiellement empêcher la modification t<sup>6</sup>A sur les ARNt, conduisant à la mort des bactéries.

Le site de fixation de l'ATP qui stabilise l'interaction TsaD et TsaE est atypique et pourrait présenter une cavité intéressante pour le développement d'inhibiteurs. Un inhibiteur qui empêcherait l'interaction entre TsaD et TsaE pourrait diminuer le niveau de modification t<sup>6</sup>A ce qui aurait pour conséquence la mort cellulaire.

Au contraire, un composé qui renforcerait l'interaction entre TsaD et TsaE et qui empêcherait la dissociation de TsaE, pourrait aussi empêcher la fixation de l'ARNt. Des molécules qui préviennent ou renforcent l'interaction entre TsaD et TsaE pourraient ainsi avoir un potentiel d'antibiotique.

Bien que la molécule BK951 soit déjà un bon inhibiteur, il serait difficile de le faire évoluer vers une molécule thérapeutique. En effet, les structures d'OSGEP humain et de TsaD sont très proches et il est probable que la molécule BK951 inhibe également l'homologue humain (des expériences sont prévues au laboratoire). Son utilisation causerait donc sans doute des problèmes d'effets secondaires. De plus, le BK951 est hydrophile et aura sans doute des difficultés pour traverser la membrane bactérienne.

## L'activité protéine kinase de TsaE.

Il a été montré que TsaE possède également une activité kinase (cf. § 4.5.1.2) et pourrait phosphoryler TsaD (Nguyen *et al.*, 2017). Dans notre structure, la sérine 60 coordonne le  $Mg^{2+}$  et la thréonine 62 est localisée proche du phosphate  $\gamma$  de l'ATP (du motif SPT) (figure 103 page 304). La conservation de ce motif classe TsaE dans la famille des kinases du type Ser/Thr Hanks (Nguyen *et al.*, 2017). Ces kinases s'organisent en deux lobes formant la poche du site actif (Janczarek *et al.*, 2018). Le domaine N-terminal est impliqué dans la liaison et dans l'orientation d'une molécule d'ATP, alors que le domaine C-terminal est responsable de la liaison au substrat protéique et du transfert du groupement phosphate (Endicott *et al.*, 2012). Le domaine N-terminal des kinases contient la boucle P, la boucle de fixation du  $Mg^{2+}$  et la boucle catalytique est retrouvée entre les domaines N et C-terminaux (Janczarek *et al.*, 2018). Dans toutes les kinases, le substrat est orienté afin que l'hydroxyle soit dirigé vers l'aspartate catalytique (ici le  $D_{166}$ ). Dans les Ser/Thr kinases, une lysine située deux résidus après ( $K_{168}$ ) interagit avec le phosphate  $\gamma$  en stabilisant la charge négative portée par l'oxygène du phosphate  $\gamma$  (figure 105) (Endicott *et al.*, 2012).

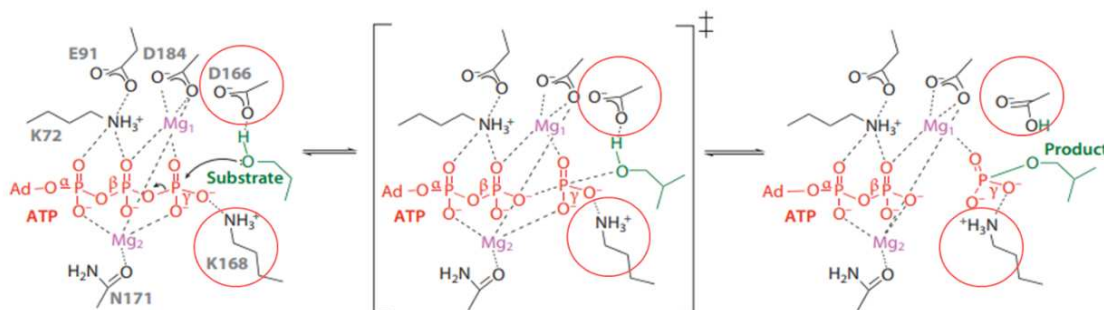


Figure 105 Diagramme schématique du mécanisme catalytique de la protéine kinase A (PKA). La molécule d'ATP est en rouge. La sérine substrat qui est phosphorylée est en vert. Les acides aminés catalytiques  $D_{166}$  et  $K_{168}$  sont indiqués par un cercle rouge. D'après Endicott *et al.*, 2012.

TsaE est une ATPase qui présente une faible activité ATPase seule et ne présente pas de domaine C-terminal. La poly-L-lysine (PLL) a déjà démontré qu'elle stimulait fortement l'activité de plusieurs protéines kinases (Absalon *et al.*, 2009). Nguyen a montré que l'activité d'autophosphorylation de TsaE est augmentée par la PLL (Nguyen *et al.*, 2017). Les hélices du domaine C-terminal de TsaD qui participent à la fixation du nucléotide engagent les chaînes latérales des lysines  $K_{166}$  et  $K_{213}$  pour former des liaisons hydrogènes avec les phosphates  $\beta/\gamma$

et un aspartate D<sub>158</sub> est localisé très proche des hydroxyles de la sérine 60 et de la thréonine 62 du motif SPT ([figure 102 page 303](#)). Cette proximité rappelle le site actif des kinases.

Il faudrait tout d'abord tester l'activité de phosphorylation de TsaE dans le contexte de la modification t<sup>6</sup>A. Le développement des peptides inhibiteurs des kinases est un champ d'investigation en essor. Il serait donc intéressant de tester l'activité inhibitrice de ces peptides sur TsaE si celle-ci possède une activité kinase dans le contexte de la modification t<sup>6</sup>A.

## Les t<sup>6</sup>A synthèses au sein des eucaryotes et des archées.

### **Contribution de GON7 chez l'Homme : formation d'un complexe à 5 sous-unités.**

yGon7 constitue la cinquième sous-unité du complexe KEOPS de *Saccharomyces cerevisiae*. Cette petite protéine de 123 acides aminés (code UNIPROT P46984) est intrinsèquement déstructurée seule en solution et seulement 60% de sa séquence se structure lorsqu'elle entre en interaction avec Pcc1 (Zhang *et al.*, 2015). Avec les autres sous-unités du complexe KEOPS, yGON7 forme un complexe pentamérique en solution (Zhang *et al.*, 2015).

Avant le début de ma thèse, l'homologue de yGon7 chez l'Homme, C14, avait été identifiée et ensuite renommé GON7 (Wan *et al.*, 2017). Tout comme chez la levure, GON7 semble former avec les autres partenaires du complexe KEOPS un complexe à 5 sous unités (figure 106) et sa présence augmente de 70% l'activité t<sup>6</sup>A du complexe (Wan *et al.*, 2017). Cependant, GON7 et yGon7 ne présentent que 17 % d'identité de séquence et aucune donnée structurale n'étant disponible, il n'était ni possible de statuer quant à l'orthologie de ces deux protéines, ni quant à la façon dont GON7 interagit avec les partenaires du complexe KEOPS.

Nos études structurales, présentées dans le chapitre III, ont permis de mettre en évidence que GON7 interagit avec LAGE3 et que malgré une faible similitude de séquence entre GON7 et yGon7, ces deux sous-unités, désordonnées seules en solution, se structurent en un feuillet de 2 brins  $\beta$  antiparallèles et en une longue hélice  $\alpha$  en interaction avec leurs partenaires respectifs LAGE3 et Pcc1. Leurs interfaces forment un feuillet de 5 brins  $\beta$  antiparallèles et les hélices  $\alpha$  de GON7 et yGon7 interagissent avec l'hélice C-terminale de LAGE3 et Pcc1. Malgré des structures identiques, GON7 ne peut pas compléter la levure déléetée de yGon7, ce qui suggère une évolution des interfaces au sien des organismes. Pour confirmer ces observations, il serait intéressant d'essayer de compléter la levure, déléetée de Pcc1 et yGon7, par LAGE3 et GON7.

Des mutations sur GON7 et YRDC conduisant au syndrome GAMOS chez l'Homme ont été étudiées en collaboration avec l'équipe de Corine Antignac (article du chapitre III). Les données obtenues *ex vivo* et *in vitro* ont permis de mettre en évidence que GON7 stabilise le complexe KEOPS et améliore la catalyse de la modification t<sup>6</sup>A par rapport au complexe à 4

sous unités, mais son absence est moins délétère que les mutations de la sous-unité catalytique YrdC.

Nous avons également montré que les 5 sous-unités forment un pentamère en solution et que GON7 empêche la formation d'un octamère (figure 106). La signification fonctionnelle de ce changement drastique de la structure quaternaire du complexe KEOPS n'est pas très claire mais l'ensemble des données montre que le complexe à 4 sous unités peut assurer une modification t<sup>6</sup>A minimale pour la survie des organismes eucaryotes. En revanche, la présence de GON7 implique une modification t<sup>6</sup>A optimale pour le bon fonctionnement de l'organisme.

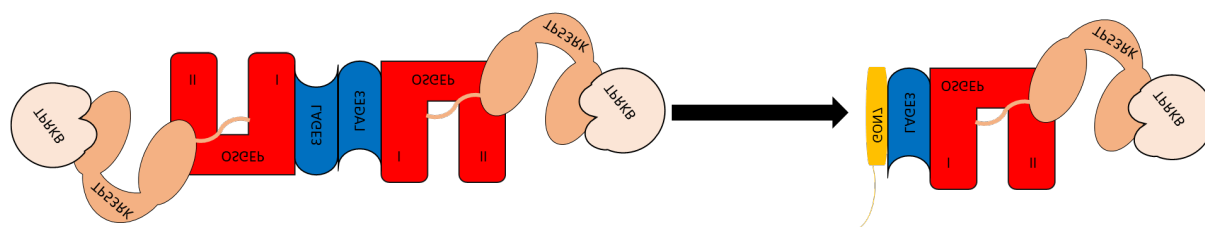


Figure 106 Schématisation de l'effet de GON7 sur l'agencement du complexe KEOPS chez les métazoaires.

### Identification d'une cinquième sous unité du complexe KEOPS archée.

Après l'identification de la cinquième sous unité GON7/Gon7 du complexe KEOPS dans les métazoaires et les champignons, de fortes présomptions concernant l'existence d'une cinquième sous unité chez les archées ont émergé. Utilisant des techniques d'analyse des génomes, Tamara Basta et Marie-Claire Daugeron de l'équipe Biologie cellulaire des Archées (I2BC, Gif sur Yvette) ont identifié un gène dans *Pyrococcus abyssi* codant une protéine PaPcc2 qui présente 20 % d'identité de séquence avec PaPcc1. Ce gène de fonction inconnue est conservé dans la majorité des génomes connus à l'exception des génomes réduits tels que les organismes du clade DPAN. Deux hypothèses concernant la fonction de ce gène ont été proposées : soit Pcc2 est un paralogue de Pcc1 et les deux protéines peuvent s'échanger selon les conditions cellulaires, soit Pcc2 est la cinquième sous-unité de KEOPS chez les archées. La structure de PaPcc2 est similaire à celle de PaPcc1 mais différente des protéines GON7/Gon7.



Il a été clairement montré *in vitro* que Pcc2 seule dimérise et forme des hétérodimères en présence de Pcc1.

Selon nos données de SAXS, en solution le complexe KBCP<sub>1</sub>P<sub>2</sub> forme un pentamère et KBCP<sub>1</sub> dimérise et forme un octamère. De façon surprenante le complexe KBCP<sub>2</sub> a pu être purifié, montrant que PaPcc2 peut remplacer PaPcc1. L'analyse de ce complexe a montré que tout comme KBCP<sub>1</sub>, KBCP<sub>2</sub> dimérise et forme un octamère en solution (figure 107).

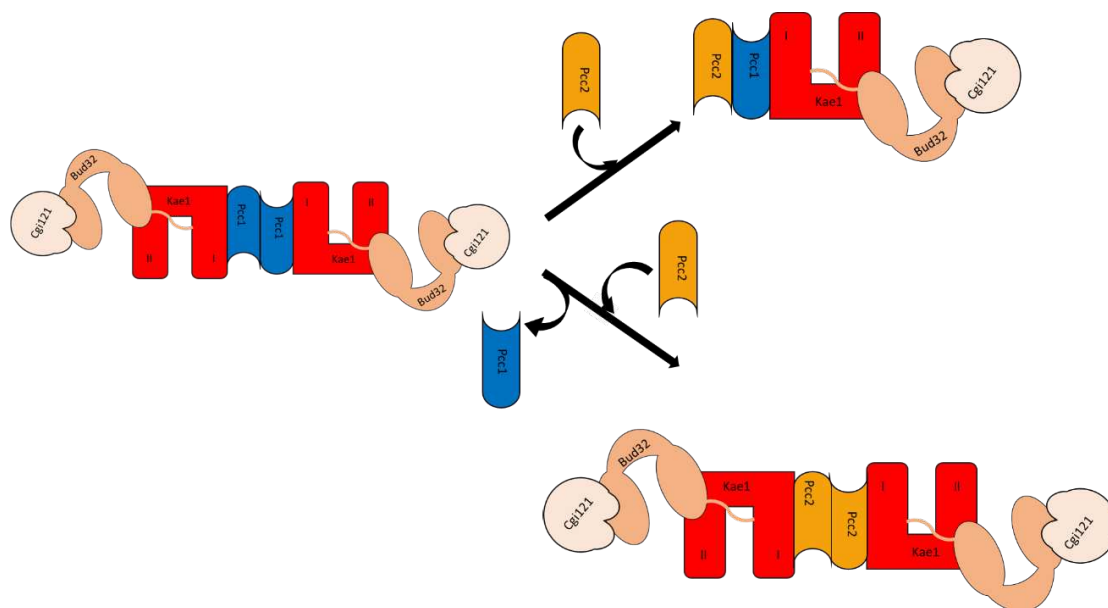


Figure 107 Représentation schématique de l'effet de Pcc2 sur l'agencement des complexes Kae1-Bud32-Cgi121-Pcc1 et Kae1-Bud32-Cgi121 dans *Pyrococcus abyssi*.

Il a été montré que PaKae1 interagit faiblement avec l'ARNt substrat ( $K_D$  apparent  $\sim 1 \mu\text{M}$ ), mais que l'affinité augmente pour le sous-complexe PaKae1-Pcc1, suggérant que Pcc1 contribue à la fixation de l'ARNt ( $K_D$  apparent  $\sim 100 - 500 \text{ nM}$ ) (Perrochia *et al.*, 2013). La structure de Pcc1 est similaire à celles des protéines possédant des domaines KH. Ce motif structural est connu pour interagir avec les acides nucléiques (Valverde *et al.*, 2008). En revanche, Pcc1 ne possède pas le motif GxxG caractéristique à cette famille de protéine. La boucle AC de l'ARNt interagit probablement avec la surface électropositive de Kae1 tout en présentant l'adénosine 37 dans son site actif et Pcc1 pourrait interagir avec le corps de l'ARNt. Pcc2 pourrait également contribuer à une meilleure fixation de l'ARNt. La recherche d'homologues structuraux a mis en évidence une analogie avec la protéine hnRNP K (RMSD 2.764 Å sur 64 C $\alpha$  superposés, PDB code 1J5K). La structure du complexe hnRNP K – ADN simple brin (ADNs)

montre que cet ADN interagit avec la protéine dans une poche constituée de 4 acides aminés basiques (K-K/R-K-R), encadrée par des acides aminés hydrophobes (Braddock *et al.*, 2002). PaPcc2 possède trois acides aminés basiques localisés sur l'hélice  $\alpha_1'$  (K<sub>36</sub>, K<sub>38</sub>, K<sub>39</sub>) encadrés par des acides aminés hydrophobes (S<sub>32</sub> and S<sub>40</sub>). Ce site de PaPcc2 pourrait participer à la fixation de l'ARNt.

Il faudrait cependant confirmer que le complexe KBCP<sub>1</sub>P<sub>2</sub> existe *in vivo*. Dans un premier temps, il faudra vérifier si on peut isoler le complexe à 5 sous-unités à partir de cultures de cellules. Dans un deuxième temps on pourrait tester si le gène Pcc2 est essentiel ou si l'inactivation du gène PCC2 a un impact sur la physiologie des cellules d'archées et mettre ces résultats en perspective avec des observations sur les cellules délétées pour Pcc1. Il sera aussi important de mesurer l'effet de Pcc2 sur les activités t<sup>6</sup>A *in vitro*.

L'hétérodimère PaPcc1-Pcc2 montre des ressemblances avec les hétérodimères Pcc1/Gon7 et LAGE3/GON7. Il serait donc intéressant de tester la complémentation de la levure, délétée de Pcc1 et Gon7, par Pcc1 et Pcc2.

En conclusion nous avons observé que tout comme les sous-unités GON7/Gon7, PaPcc2 présente un rôle de régulateur de l'état oligomérique de KEOPS *in vitro*. De façon surprenante, son analogie structurale avec PaPcc1 lui permet d'interagir avec Kae1 en absence de Pcc1, assurant la dimérisation du complexe KEOPS et reconstituant l'octamère, ce qui suggère un rôle compensatoire.

Il avait été montré chez *Haloferax volcanii* que l'absence de Pcc1 ne provoquait pas de diminution de la modification t<sup>6</sup>A *in vivo*, suggérant un rôle accessoire de cette protéine (Naor *et al.*, 2012). Il est très probable que l'absence de phénotype dans la souche délétée du gène PCC1 est imputable à la présence de Pcc2. En absence de Pcc1, Pcc2 pourrait prendre le relais et avoir une fonction redondante, mais la fonction de KBCP2 n'est pas encore bien comprise.

Il est possible que l'interaction de Pcc1 avec Kae1 soit plus forte que celle de Pcc2 avec Kae1. Ainsi Pcc1 et Pcc2 sont en compétition et Kae1 semble privilégier Pcc1 pour former un motif "4 helix bundle" plus stable. Ces observations sont basées sur les modèles *ab initio* et devront être testées par des expériences de calorimétrie par exemple, si l'on arrive à purifier les sous-unités séparément. Également des expériences de thermo-stabilité des sous-

complexes KP1 et KP2 peuvent être envisagées. Et enfin la structure du sous-complexe KP2 serait intéressante pour vérifier si l'interface est similaire à celle de KP1.

### **Adaptation des enzymes t<sup>6</sup>A synthases selon les règnes du vivant.**

Bien que la famille d'enzyme Kae1/OSGEP/TsaD/Qri7 soit universelle, différents partenaires ont été recrutés au cours de l'évolution pour assurer la seconde étape de la modification t<sup>6</sup>A (figure 108) (El Yacoubi *et al.*, 2012). Ainsi, nos études structurales ont permis d'affiner les structures quaternaires des différents complexes des trois domaines du vivant.

Dans les mitochondries, seule la sous-unité catalytique Qri7 assure la thréonylcarbamoylation des ARNt. Il a été mis en évidence qu'elle se dimérise via ses domaines N-terminaux.

Dans les procaryotes, TsaBD interagit avec la protéine TsaE, et cette interaction est stabilisée par l'ATP entre les hélices du domaine C-terminale de TsaD et la boucle P de TsaE. Cependant, ce mode d'interaction est totalement différent de celui réalisé par Kae1 et OSGEP avec Bud32 et TP53RK, respectivement, dont ces dernières sont uniques chez les archées et les eucaryotes. Leurs plis tridimensionnels sont reliés à celui des kinases RIO et une activité ATPase *in vitro* a été observée chez Bud32 de *Pyrococcus abyssi* (Perrochia *et al.*, 2013). Des mutations dans le site actif de Bud32 de *Saccharomyces cerevisiae* présentent des effets délétères pour la survie de la cellule, similaires à ceux observés par la délétion du gène *BUD32*, ce qui montre que la fonction de Bud32 pourrait être reliée à son activité ATPasique. Cependant, la fonction de Bud32 reste pour le moment inconnue. De plus, les sous-unités Bud32 et Cgi121 ne semblent pas améliorer la fixation de l'ARNt (Perrochia *et al.*, 2013). La structure du complexe KEOPS entier avec l'ARNt est indispensable pour répondre à ces questions.

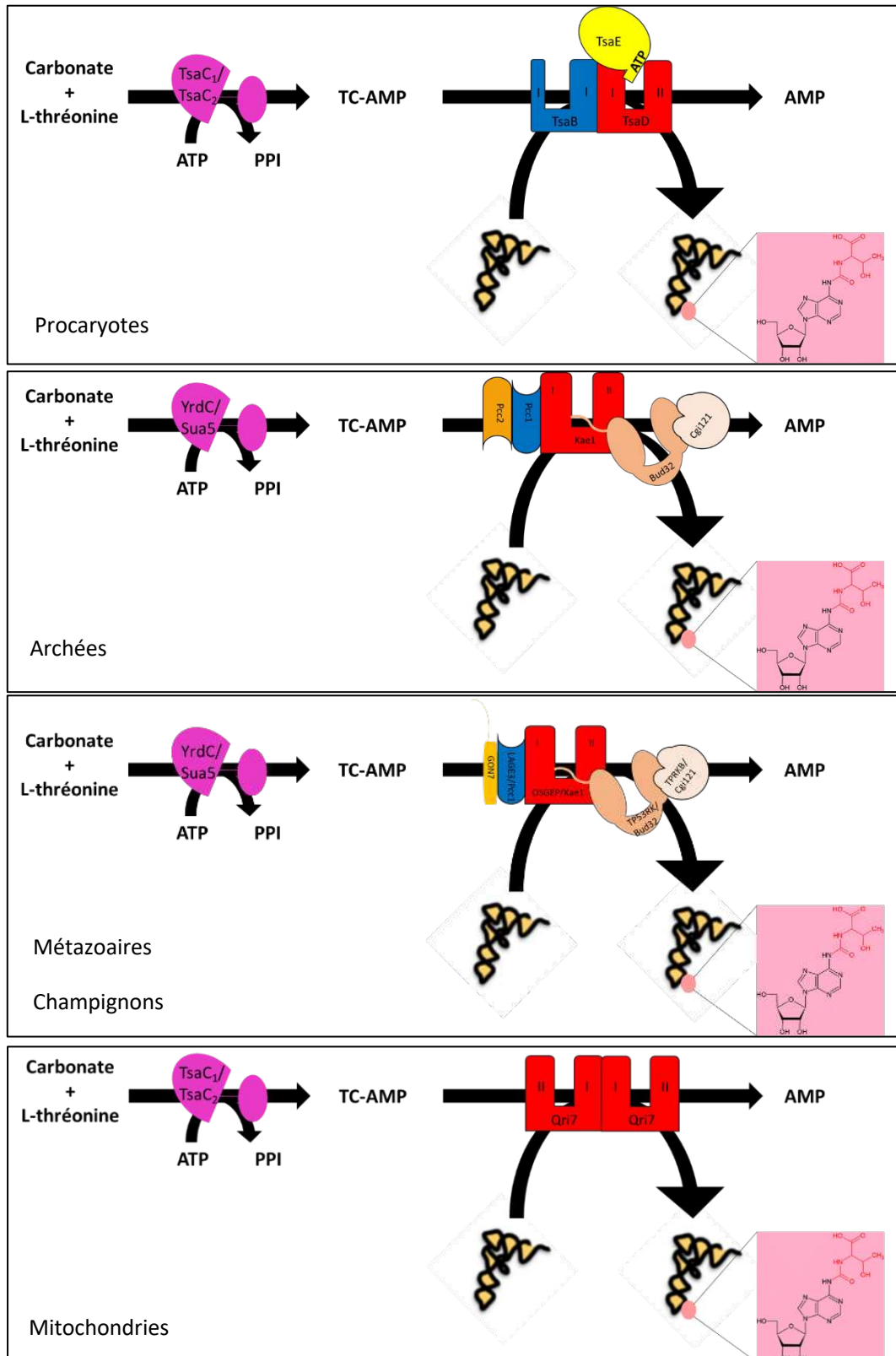


Figure 108 Représentation schématique des différentes machineries de synthèse de la modification t<sup>6</sup>A dans les 3 domaines du vivant et dans les mitochondries de *Saccharomyces cerevisiae*. La catalyse est assurée par les familles d'enzymes ubiquitaires YrdC/Sua5 (TsaC<sub>1</sub>/TsaC<sub>2</sub>) et Kae1/OSGEP/TsaD/Qri7. Ces dernières nécessitent des partenaires différents selon les espèces.

Les enzymes Kae1/OSGEP/TsaD/Qri7 interagissent tous via leurs domaines N-terminaux avec des partenaires différents. Les structures Kae1/Pcc1 ou OSGEP/LAGE3 ont déjà montré que leurs interfaces étaient semblables à celle de TsaD/TsaB et Qri7/Qri7. Il semblerait que les sous-unités catalytiques ont sélectionnées au cours de l'évolution un mode d'interaction "4 helix bundle" malgré des partenaires divers.

Enfin l'étude structurale de Pcc2 chez les archées et de GON7 chez l'Homme montre que tout, comme dans la levure, ces cinquièmes sous-unités changent drastiquement la structure quaternaire du complexe KEOPS, le stabilisant en pentamère. Malgré un impact identique sur l'état d'oligomérisation des complexes, Pcc2 ne présente pas de structure identique à la famille GON7/Gon7. Il n'est pas impossible que ces sous-unités interagissent avec d'autres partenaires au sein des cellules.

La diversité des partenaires conduisant probablement à des mécanismes de la modification t<sup>6</sup>A différents pourrait décrire des fonctions alternatives du complexe KEOPS chez les eucaryotes.

Il a été mis en évidence que le complexe KEOPS présente une fonction dans la réparation de l'ADN chez *Saccharomyces cerevisiae*. En effet dans un premier temps Cgi121 et Bud32 ont été identifiés comme des régulateurs de *cdc13-1*, un allèle des télomères (Downey *et al.*, 2006). Leurs délétions ainsi que les autres sous-unités du complexe KEOPS de la levure conduisent au raccourcissement des télomères et à l'inhibition de la recombinaison télomérique. De façon similaire, la délétion de la protéine Sua5 chez *Saccharomyces cerevisiae* conduit également à un défaut des télomères (Meng *et al.*, 2010). Ce phénotype commun du complexe KEOPS et de la protéine Sua5 est un argument supplémentaire pour suggérer leur implication dans l'homéostasie des télomères. Récemment, il a été montré qu'en absence du complexe KEOPS, la surexpression de la sous-unité Qri7, délétée de sa séquence d'adressage aux mitochondries, restitue la modification t<sup>6</sup>A et un phénotype sauvage de la levure. En revanche, Qri7 ne restitue pas la longueur optimale des télomères, confirmant l'implication du complexe KEOPS dans la régulation de l'ADN (He *et al.*, 2019). Cette fonction alternative pourrait expliquer la diversité observée de la cinquième sous-unité GON7/Gon7 chez les métazoaires et les champignons et de Pcc2 des archées, dont ces organismes ne présentent pas de télomères.

Il a été également mis en évidence que le complexe KEOPS joue un rôle dans les mécanismes de transcription des gènes essentiels dans la levure (Kisseleva-Romanova *et al.*, 2006).

Chez l'Homme, l'étude de l'interactome de KEOPS ou de ses sous-complexes a révélé près de 152 partenaires spécifiques (Costessi *et al.*, 2012; Wan *et al.*, 2017). Notamment, GON7 et LAGE3 chez l'Homme pourraient interagir avec des protéines chaperonnes et pourraient ainsi réguler le repliement des protéines. Il a été suggéré que LAGE3 et GON7 pourraient participer à la stabilisation de OSGEP dans le contexte du complexe KEOPS (Wan *et al.*, 2017).

Ces résultats semblent indiquer que les sous-unités du KEOPS aient des rôles supplémentaires pour la cellule pouvant expliquer les nombreux phénotypes observés lorsque les organismes sont faces à un défaut de ces sous-unités.



# ANNEXES

## Annexe 1 : Expression, production et purification du complexe TsABD d'*Escherichia coli*.

### Matériels et Méthodes.

100 ng du plasmide pET28adtsaD et 100 ng du plasmide pET9dtsaB ont été inoculés dans 100 µL de souches chimio compétentes BL21 gold DE3 (Novagen) respectivement. L'expression des protéines a été réalisée dans un milieu 2YT riche, sélectionnée en présence de marqueurs : antibiotiques appropriés (50 µg.mL<sup>-1</sup>) et la production simultanée des protéines a été induite à l'IPTG (0,5 µM) sous le contrôle du promoteur T7. Après l'induction à l'IPTG, une centrifugation (4000 g, 20 minutes, 8°C) a été réalisée. Les culots bactériens sont suspendus dans 35 mL de tampon 20 mM Hepes pH 7,5, 200 mM NaCl, 5 mM β-mercaptoéthanol puis conservés à -20°. Après décongélation, la lyse des cellules est réalisée par sonication (Branson sonifier 250, 3 cycles de 1 min, 30% de puissance). Les extraits cellulaires totaux sont centrifugés 30 minutes à 20000 g et le surnageant, composé de la fraction protéique soluble dont les protéines recombinantes d'intérêt, est récupéré pour la suite de la purification (Quevillon-Cheruel *et al.*, 2004). Les surnageants sont incubés avec la résine de silice Ni<sup>2+</sup>NTA (NiNTA-Agarose Qiagen). Ainsi les protéines TsAB et TsAD sont purifiées selon la méthode IMAC séparément (Immobilized Metal Affinity Chromatography) et éluées par des sauts de concentrations croissantes d'imidazole (de 20 à 400 mM) qui rentrent en compétition avec l'affinité de l'étiquette 6 histidines avec les ions Ni<sup>2+</sup>. Les fractions d'élution contenant les protéines d'intérêts sont réunies et concentrées par centrifugation sur membrane de porosité limitée à des objets de 10 kDa (cut-off) (Centricon). Les protéines sont ensuite purifiées par chromatographie d'exclusion stérique utilisant des colonnes Superdex 200 16-600 de façon séparées (GE Healthcare) préalablement équilibrées dans un tampon 20 mM Hépes pH 7,5, 200 mM NaCl, 5 mM β-mercaptoéthanol. L'élution le long de la colonne se fait avec un débit de 1 ml.min<sup>-1</sup>. Des fractions de 1,5mL sont collectées en sortie de colonne. Enfin les fractions d'intérêt sont analysées en sortie de colonne par gel SDS-PAGE. Et les fractions contenant les protéines d'intérêts les plus pures sont réunies et concentrées par centrifugation sur membrane de 10 kDa (Centricon). Puis le complexe TsABD est reconstitué extemporanément à un ratio 1 :1 en molarité et est purifié par chromatographie d'exclusion stérique utilisant une



colonne superdex S200 16-600 (Ge Healthcare) préalablement équilibrées dans un tampon 20 mM Hépes pH 7,5, 200 mM NaCl, 5 mM  $\beta$ -mercaptoéthanol. Les fractions d'élution sont analysées par gel SDS-PAGE et celles contenant le complexe TsaBD sont réunies et concentrées par centrifugation à l'aide d'une membrane de porosité 30 kDa (Centricon).

## Résultats.

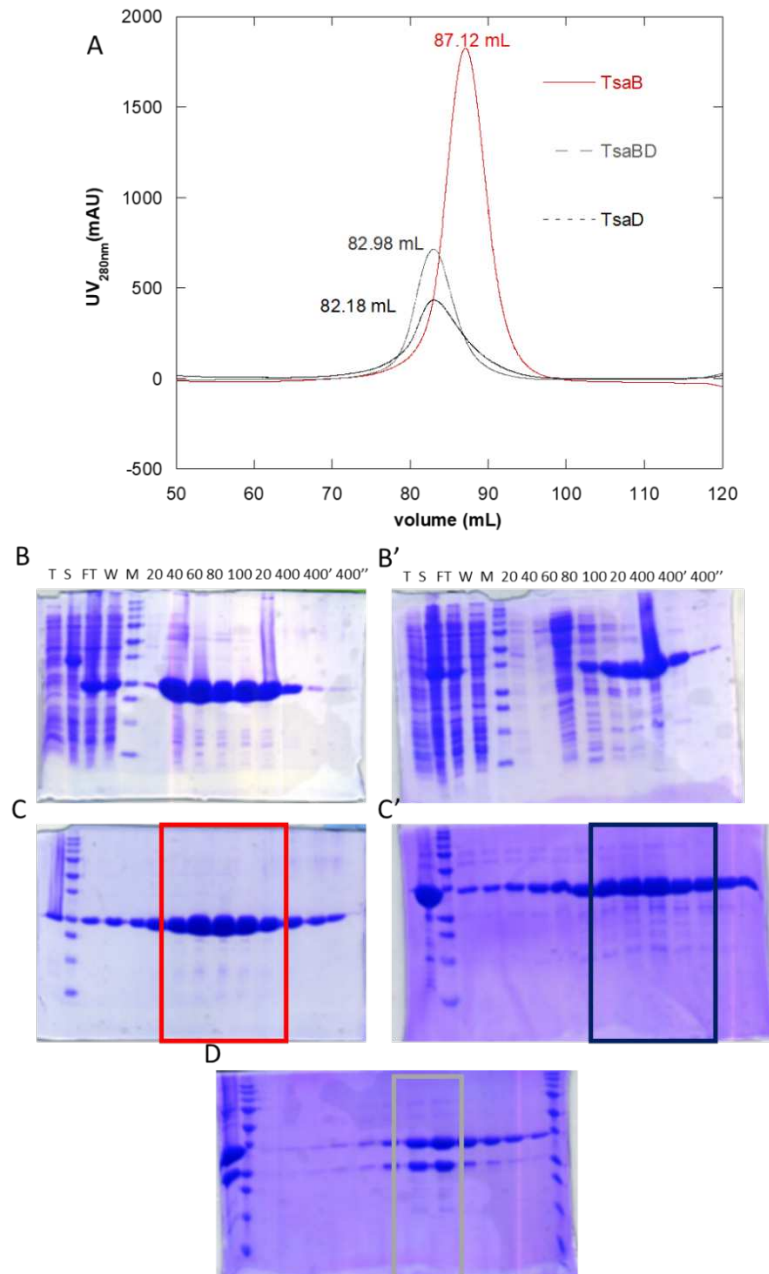


Figure 109 Purification du complexe TsaBD d'*Escherichia coli*. A. Le suivi de l'élution sur colonne S200 16-600 est représenté sur le chromatogramme. B.B' Analyse SDS-Page des protéines TsaB et TsaD en sortie de purification par chromatographie d'affinité sur ions  $Ni^{2+}$ . C. C'. Analyse SDS-Page des protéines TsaB et TsaD en sortie de colonne Superdex 200 16-600 (GE Healthcare). D. Analyse SDS-Page du complexe TsaBD en sortie de colonne Superdex 200 16-600 (GE Healthcare).

Les protéines TsaB et TsaD ont été purifiées par chromatographie d'affinité sur ions métalliques ( $\text{Ni}^{2+}$ ) immobilisés suivie d'une chromatographie d'exclusion stérique utilisant une colonne Superdex 200 16-600 (GE Healthcare) séparément.

L'analyse des fractions d'élution de la chromatographie d'affinité sur ions  $\text{Ni}^{2+}$  montre que TsaB est éluée à une concentration de 20 mM d'imidazole. Les fractions les plus pures contenant TsaB sont celles correspondant à 80 mM d'imidazole jusqu'à 400 mM (figure 109 B). Le chromatogramme de TsaB montre que des protéines sont éluées à un volume d'élution de 87,2 mL (figure 109 A). Le volume d'élution n'est pas compatible avec le poids moléculaire de TsaB monomérique ( $M = 25,180$  kDa) mais avec la taille d'un homodimère. Les fractions les plus pures, encadrées en rouge (figure 109 C) sont réunies et concentrées par centrifugation sur membrane.

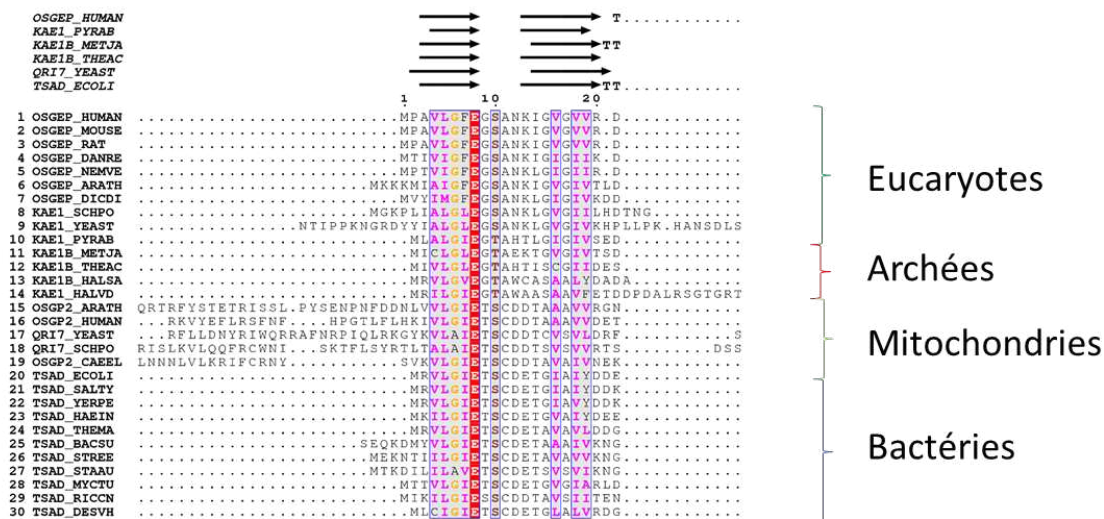
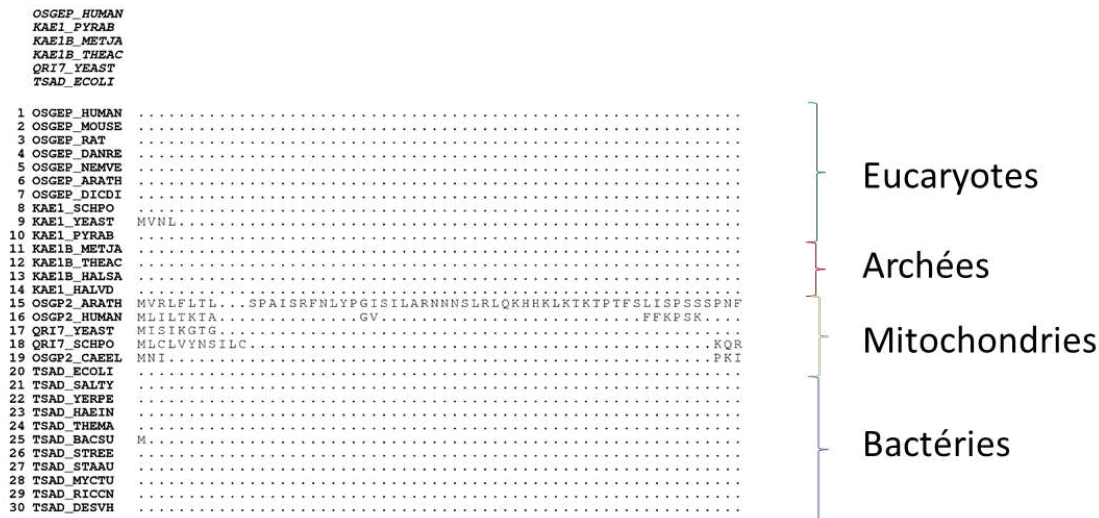
L'analyse des fractions d'élution de la chromatographie d'affinité sur ions  $\text{Ni}^{2+}$  montre que TsaD est élué à une concentration de 20 mM d'imidazole. Les fractions les plus pures contenant TsaD sont celles correspondant à 200 mM d'imidazole jusqu'à 400 mM (figure 109 B'). Le chromatogramme de TsaD montre que des protéines sont éluées à un volume d'élution de 82,18 mL (figure 109 A). Le volume d'élution n'est pas compatible avec le poids moléculaire de TsaD monomérique ( $M = 36$  kDa) mais avec la taille d'un homodimère. Les fractions les plus pures, encadrées en noir (figure 109 C') sont réunies et concentrées par centrifugation sur membrane.

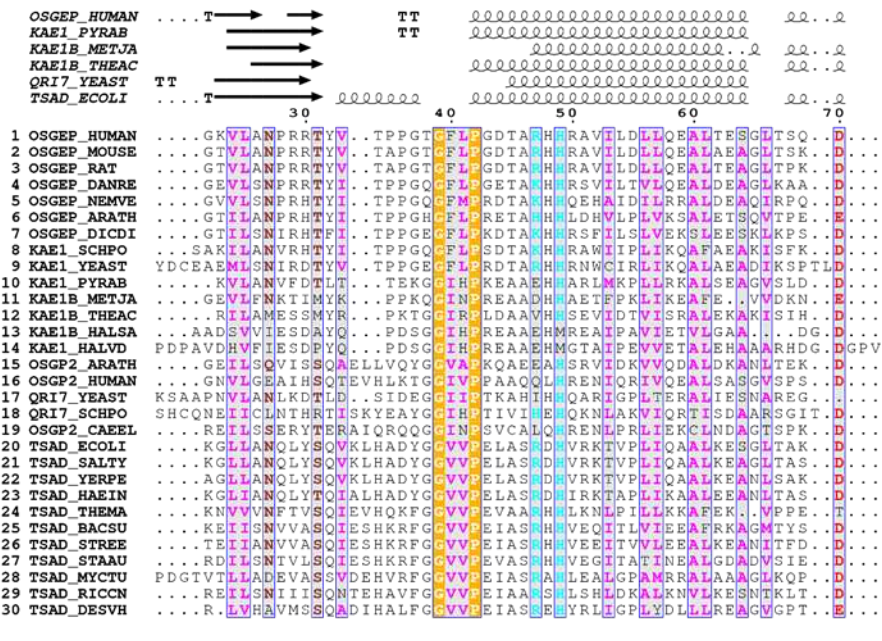
Le complexe TsaBD est reconstitué extemporanément avec un ratio 1 :1 et est purifié sur chromatographie d'exclusion stérique. Le chromatogramme (figure 109 A) montre que les protéines sont éluées à un volume de 82,98 mL, le volume d'élution est compatible avec le poids moléculaire de l'hétérodimère ( $M = 61,171$  kDa). L'analyse des fractions correspondant au pic sur gel SDS PAGE confirme la présence des deux protéines TsaBD (figure 109 C) montrant que le complexe est suffisamment stable pour être co-élué sur chromatographie d'exclusion stérique.

Les protéines les plus pures (cadre gris sur la figure 109 C) sont réunies et concentrées par centrifugation sur membrane de porosité de 30 kDa. Les protéines pures ont été ensuite utilisées pour la cristallogénèse du complexe TsaBD avec la molécule BK951 (cf. chapitre II pages 121).

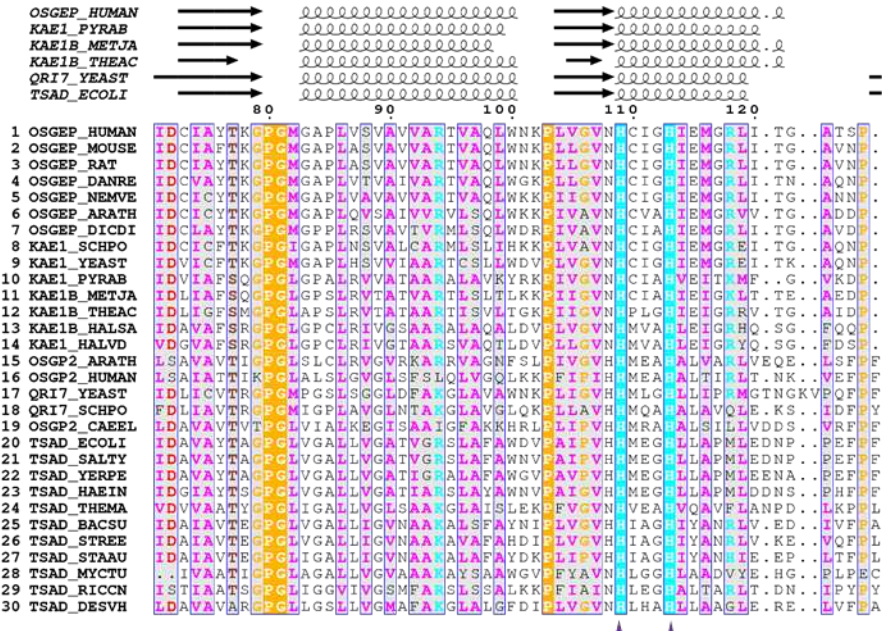
## Annexe 2 : Comparaison des séquences de la sous-unité catalytique.

Alignement multiple de séquences de la sous-unité catalytique Kae1/OSGEP/TsaD/Qri7 chez les eucaryotes, archées, bactéries et mitochondries. L'alignement a été réalisé par MAFFT en utilisant la matrice BLOSUM 60 (Kato *et al.*, 2019). La région 10 à 15 est indiquée par une barre grise. La sérine 10 est conservée ou substituée par une thréonine chez les archées. Les acides aminés suivants forment une triade d'acides aminés "ANK" conservée chez les eucaryotes, "AHT" ou "AWA" chez les archées et "CDD" ou "CDE" chez les procaryotes et dans les mitochondries. Les acides aminés qui stabilisent la fixation du métal dans le site actif sont indiquées par une étoile violette. Seuls les acides aminés la tyrosine 130 et l'histidine 139 sont spécifiques aux eucaryotes/archées et aux bactéries/mitochondries respectivement.



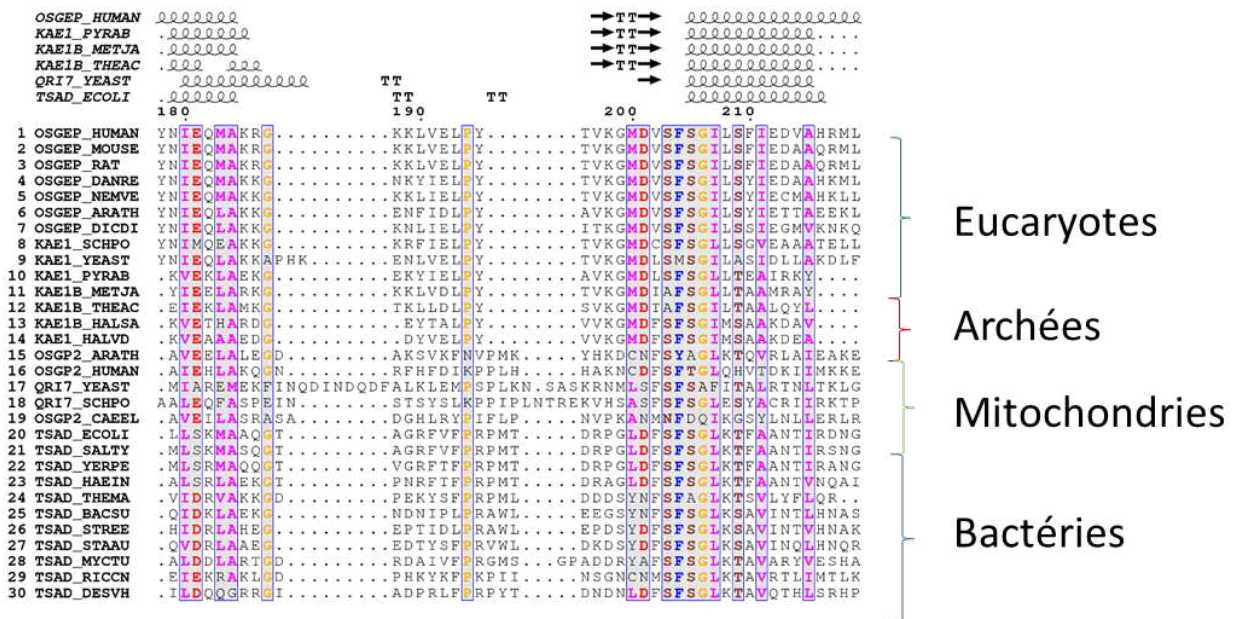
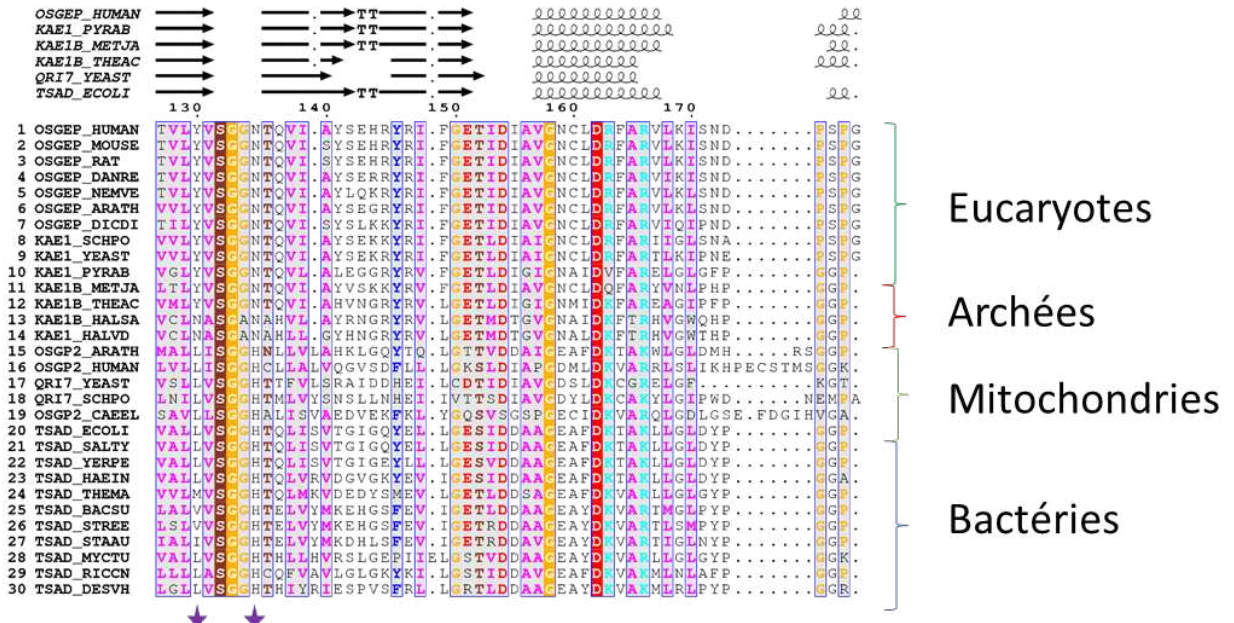


Eucaryotes  
 Archées  
 Mitochondries  
 Bactéries



Eucaryotes  
 Archées  
 Mitochondries  
 Bactéries





```

OSGEP_HUMAN .....00 0000000000000000000000000000
KAE1_PYRAB .....00 0000000000000000000000000000
KAE1B_METJA .....0 0000000000000000000000000000
KAE1B_THEAC .....0 0000000000000000000000000000
QRI7_YEAST .....0000000000000000000000000000 0000
TSAD_ECOLI .....0000000000000000000000000000

                220      230      240      250
1 OSGEP_HUMAN .....ATGECTPEDLCFSLQETVFFAMLVEITERAMAHC.....G
2 OSGEP_MOUSE .....ATGECTPEDLCFSLQETVFFAMLVEITERAMAHC.....G
3 OSGEP_RAT .....ATGECTPEDLCFSLQETVFFAMLVEITERAMAHC.....G
4 OSGEP_DANRE .....STDQCTPEDLCFSLQETVFFAMLVEITERAMAHC.....G
5 OSGEP_NEMVE .....SSEECTPEDLCFSLQETVFFAMLVEITERAMAHC.....G
6 OSGEP_ARATH .....KNNECTPADLCYSLQETVFFAMLVEITERAMAHC.....G
7 OSGEP_DICDI .....NKTQHSVEDLCYSLQETVFFAMLVEITERAMAHC.....G
8 KAE1_SCHPO DPK.....NPSSTVKQDLCYSLQETVFFAMLVEITERAMAHC.....G
9 KAE1_YEAST KGNKKNKILFDKTT...GEQKVTVEDLCYSLQENLFAMLVEITERAMAHV.....N
10 KAE1_PYRAB .....RSGKYRVEDLAYSFQETAFAALLVEITERAVAHT.....E
11 KAE1B_METJA .....DAGE.RLEDLICYSLQEYAFSMLTEITERALAHT.....N
12 KAE1B_THEAC .....KTGQ.AIEDLISYSIQEYAFAMLVEVLERALVYS.....G
13 KAE1B_HALSA .....DDGVPVAVDCRGLETVFAMLTEVAERALALT.....G
14 KAE1_HALVD .....DAGTVPVDLCAAGLETVFAMLTEVAERALSLT.....G
15 OSGP2_ARATH IDAKCPVS...SAT...NEDRRNRADLAASFRORVVLVHLLEEKCERADNWALELEP...S
16 OSGP2_HUMAN KEEGLEK...GQILSSAADLAAAVOHTMACHLVKRTHRAILFPCKQRDLLPQN
17 QRI7_YEAST KTEIQ...ELPEREIRSIAYVOESVFFDHIINKLKVLKSQPERFK...N
18 QRI7_SCHPO .....LNLSEKKFFAYLOYAAFQHICQKTLLALKRLDLS...K
19 OSGP2_CAEEEL KN.....SETSIDIPDFCASLONTWARHISSKLHIFFESLSEQEK...L
20 TSAD_ECOLI TD.....DQTRADLIARAFEDAVVDTLMIKCKRALDQT.....G
21 TSAD_SALTY GD.....EQTRADLIARAFEDAVVDTLMIKCKRALEST.....G
22 TSAD_YERPE DD.....DQTRADLIARAFEDAVVDTLLAIKSKRALDQT.....G
23 TSAD_HAEIN KNEG.....ELIEQTKADLIAYAFODAVVDTLLAIKSKRALKET.....G
24 TSAD_THEME .....EKGYKVEDVAASAFQKAVVDTLVEKTFLLARNL.....D
25 TSAD_BACSU Q.....KGQEIAPEDLSASFQNSVIDVLVTKAAAAKEY.....D
26 TSAD_STREE Q.....RGIEIAPEDLSASFQNSVIDVLVTKAAAAKEY.....N
27 TSAD_STAUA Q.....KNPILEANVATSFFQNSVVEVLTFKAIQAKEY.....G
28 TSAD_MYCTU A.....DPGFRTADLAAAFQEAVADVLTMKAVRALTAL.....G
29 TSAD_RICCN E.....INDTVINDLAAAFQETIGELLSSKVQAIRAYEQITNN.FD
30 TSAD_DESVH ELVPSFDA...AQAVAGGHDAPEGLRNCASFNEVAETLCIKLGRALDGDDGV.....G

```

Eucaryotes  
Archées  
Mitochondries  
Bactéries

```

OSGEP_HUMAN  >>>> 0000 0000000000.0000 >>>> 0000 0000000000
KAE1_PYRAB   >>>> 0000 0000000000.0000 >>>> 0000 0000000000
KAE1B_METJA >>>> 0000 0000000000.00 >>>> 0000 0000000000
KAE1B_THEAC >>>> 000 0000000000.0000 >>>> 0000 0000000000
QRI7_YEAST  >>>> 0000 0000000000.0000 >>>> 0000 0000000000
TSAD_ECOLI  >>>> 0000 0000000000.0000 >>>> 0000 0000000000

                260      270      280      290      300
1 OSGEP_HUMAN SQEALIVGGVGNVRLQEMMATM.CQER.....GARLFATDERFIDNGAMIAQAQWE
2 OSGEP_MOUSE SKEALIVGGVGNLRLQEMMGTM.CQER.....GAQLFATDERFCVNGAMIAQAQWE
3 OSGEP_RAT   SKEALIVGGVGNLRLQEMMATM.CQER.....GAQLFATDERFIDNGAMIAQAQWE
4 OSGEP_DANRE SQEVLIVGGVGNLRLQEMMGVM.CKER.....GARLFATDESFCIDNGAMIAQAQWE
5 OSGEP_NEMVE SNEVLIVGGVGNLRLQEMMDVM.AKER.....GAKLFATDERFIDNGAMIAQAQWE
6 OSGEP_ARATH KKDVLIVGGVGNLRLQEMMRTM.CSER.....DGKLFATDDRYCIDNGAMIAQTGLL
7 OSGEP_DICDI QNEVLAVGGVGNLRLQEMTQQM.ISQR.....NGKLFATDERYCIDNGAMIAWAQYL
8 KAE1_SCHPO ADSVLIVGGVGNLRLQEMMAEM.SSDR.....GADVFSTDERFCIDNGIMIAQAQGLL
9 KAE1_YEAST SNQVLIVGGVGNLRLQEMMAQM.CKDR.....NGQVHATDNRFIDNGVMIAQAQGLL
10 KAE1_PYRAB KDEVVLVGGVAANNRLREMIRIM.TEDR.....GIKEFFVPPYDLRDNAMIAQTGLR
11 KAE1B_METJA KGEVVLVGGVAANNRLREMIRIM.CEGQ.....NVDEFVPPYDLRDNAMIAQTGLL
12 KAE1B_THEAC KDELVLVGGVAANNRLRDMVTNM.AREA.....GIRSYLTDREYMDNIMIAQAQALL
13 KAE1B_HALSA RDELVLVGGVGNLRLRGMLEAM.CAAR.....GASFHAFEPFRIDNAGMIAVLAQAK
14 KAE1_HALVD TDELVLVGGVGNLRLRGMLEAM.CDQR.....GAKFHAFEPFRIDNAGMIAVLAQAK
15 OSGP2_ARATH IKHMVISGGVANSKYVRLRLNNTI.VENK.....NLKLVCPFPSTIDNGVMVAWTGLE
16 OSGP2_HUMAN NAVLVASGGVANSFYIRRALEIL.TNAT.....QCTLLCPPPRLIDNIMIAWNGIE
17 QRI7_YEAST VRELVCSGGVANSRQLRTRKLELE.LGTLN.....STSFNRYVPPMDLSDNIMIGWAQIE
18 QRI7_SCHPO VKYLVCSGGVANSRQLRTRKLELE.LMVLQFEHQPTDIKLVYSPDIDSNAAMIGYTAIQ
19 OSGP2_CAEEEL FKQLVIGGGVANSRQLTRAKLAEM.MKRR.....NVTIKVLLSLIDNAGMIAYSGLL
20 TSAD_ECOLI FKRVLVMAGGVANSRTRAKLAEM.MKRR.....RGEVFYARPEFTDNGAMIAVAQMV
21 TSAD_SALTY FKRVLVIAGGVANSRTRAKLAEM.MKRR.....RGEVFYARPEFTDNGAMIAVAQMV
22 TSAD_YERPE YKRVLVIAGGVANSRTRAKLAEM.MKRR.....RGEVFYARPEFTDNGAMIAVAQMV
23 TSAD_HAEIN YKRVLVIAGGVANSRTRAKLAEM.MKRR.....RGEVFYARPEFTDNGAMIAVAQMV
24 TSAD_THEME IRKLVAFVGGVAANRGRAREEVRKR.AERN.....NYEYVFPPLSLIDNALMVVAQYE
25 TSAD_BACSU VKQVLLVGGVAANRGRAREEVRKR.AERN.....NYEYVFPPLSLIDNALMVVAQYE
26 TSAD_STREE VKQVLLVGGVAANRGRAREEVRKR.AERN.....NYEYVFPPLSLIDNALMVVAQYE
27 TSAD_STAUA VQRLLVAGGGVANSKGRQSLADQ.CKVN.....DIQLTIPGPKLIDNAGMIGVAQHY
28 TSAD_MYCTU VSTLLVAGGGVANSRRELATOR.CGEA.....GRTLRIPSPRLIDNAGMIAAFAAQ
29 TSAD_RICCN KKNLVIAAGGGVANSKYQKILSSQ.AKTY.....GYRLIYPPHLLIDNAGMIAVAQLE
30 TSAD_DESVH VRAVLVAGGGVANSRREHARL.AVSR.....GLELIVPSPPLIDNAGSMIAVAQWL

```

Eucaryotes  
Archées  
Mitochondries  
Bactéries



*OSGEP\_HUMAN*    000                    000                    000                    T.....T  
*KAE1\_PYRAB*    000                    000                    TT    000                    T.....T  
*KAE1\_METJA*    000                    000                    TT    TT  
*KAE1\_THEAC*    000                    000                    TT    000  
*QRI7\_YEAST*    000                    000                    000  
*TSAD\_ECOLI*    000                    000                    000

310                    320                    330

1 *OSGEP\_HUMAN*    MFRAGH.RTPVSDSGV.....TQRVRTDEVVEVT.....WRD  
 2 *OSGEP\_MOUSE*    MFAAH.H.RTPVSDSGV.....TQRVRTDEVVEVT.....WRD  
 3 *OSGEP\_RAT*       MFAAH.H.RTPVSDSGV.....TQRVRTDEVVEVT.....WRD  
 4 *OSGEP\_DANRE*    MFRSH.H.VTELSDSGV.....TQRVRTDEVVEVT.....WRD  
 5 *OSGEP\_NEMVE*    MFTQS.S.VTFLSDSGV.....TQRVRTDEVVEVT.....WRD  
 6 *OSGEP\_ARATH*    AFVNI.I.ETPLSDSGV.....TQRVRTDEVVEVT.....WRD  
 7 *OSGEP\_DICDI*    IFKNI.T.TTFLSDSGV.....TQRVRTDEVVEVT.....WRD  
 8 *KAE1\_SCHPO*    AYKTD.D.RCAVVESTI.....TQRVRTDEVVEVT.....WRD  
 9 *KAE1\_YEAST*      EYRMI.GIVKDFSEITV.....TQRVRTDEVVEVT.....WRD  
 10 *KAE1\_PYRAB*     MYKAI.I.SFRLSDSGV.....TQRVRTDEVVEVT.....WRD  
 11 *KAE1\_METJA*     MHKNI.R.WMSLDEETKI.....IPNVRIDMVEVN.....WIK  
 12 *KAE1\_THEAC*     MYKSI.V.RMSVVEETAV.....DPNVRIDMVEVN.....WIT  
 13 *KAE1\_HALSA*     MAAAI.A.TTFLVADSAI.....NSQVPRDEVSVT.....WRD  
 14 *KAE1\_HALVD*     MAAAI.D.TTFLVADSAI.....NSQVPRDEVSVT.....WRG  
 15 *OSGP2\_ARATH*    HFRVY.R.YDPPPPATEPEDYVYDLRPRVPLGVEYAKG.....RSEARS  
 16 *OSGP2\_HUMAN*    RLRAI.L.GLLHDIETI.....RYEPKPLGVDISKEVGEA.....SIKVPQLK  
 17 *QRI7\_YEAST*     IWESI.R.LVSLDIDCP.....IRKWLPLNDLLEVD.....GWRT  
 18 *QRI7\_SCHPO*     MFKAI.Y.TSSFDVEP.....IRKWLPLNDLLEVD.....GWL  
 19 *OSGP2\_CAELI*    MLVNI.S.....E.....AIWVPRNDIPDIYAHARSDIGTDASSEIIDTPRK  
 20 *TSAD\_ECOLI*     RPKAI.A.TADLVGVSV.....RPRVPLAELPAAA.....  
 21 *TSAD\_SALTY*     RPKAI.V.TADLVGVTV.....RPRVPLAELPAAA.....  
 22 *TSAD\_YERPE*     .....  
 23 *TSAD\_HAEIN*     RLKAI.Q.HSDLAIDV.....KPRVAMAEVPAI.....  
 24 *TSAD\_THEMA*     KAKRI.M.FSPLSLNA.....DPNINV.....  
 25 *TSAD\_BACSU*     AFEKI.I.RGAYDMNG.....QPGLELTSYQSLTR.....  
 26 *TSAD\_STREE*     AVEQI.K.RATLALNA.....NSGLDIEA.....  
 27 *TSAD\_STAAU*     LVQAI.R.FADLALNG.....HSNLDIEA.....  
 28 *TSAD\_MYCTU*     LVAAI.A.PPSELDVPS.....DPGLPVMGQVR.....  
 29 *TSAD\_RICCN*     RYNNI.L.FTFLNFCP.....KARWLEDLSH.....  
 30 *TSAD\_DESVH*     LRQGI.L.SHSVLDLEA.....VPRCRAIPDDYR.....QGP

Eucaryotes  
 Archées  
 Mitochondries  
 Bactéries

*OSGEP\_HUMAN*  
*KAE1\_PYRAB*  
*KAE1\_METJA*  
*KAE1\_THEAC*  
*QRI7\_YEAST*  
*TSAD\_ECOLI*

1 *OSGEP\_HUMAN*    .....  
 2 *OSGEP\_MOUSE*    .....  
 3 *OSGEP\_RAT*       .....  
 4 *OSGEP\_DANRE*    .....  
 5 *OSGEP\_NEMVE*    .....  
 6 *OSGEP\_ARATH*    KEAVLLGD.....KKVAAN  
 7 *OSGEP\_DICDI*    .....  
 8 *KAE1\_SCHPO*     .....  
 9 *KAE1\_YEAST*      .....  
 10 *KAE1\_PYRAB*     .....  
 11 *KAE1\_METJA*     .....  
 12 *KAE1\_THEAC*     .....  
 13 *KAE1\_HALSA*     .....  
 14 *KAE1\_HALVD*     .....  
 15 *OSGP2\_ARATH*    MRTARIHPSLTSLIRADSL.....QQQTOT  
 16 *OSGP2\_HUMAN*    MEI.....  
 17 *QRI7\_YEAST*     DQL.....  
 18 *QRI7\_SCHPO*     KKNKKV.....  
 19 *OSGP2\_CAELI*    LVVSTIHG..TERIRFNLDLDFKKPKSPKTE  
 20 *TSAD\_ECOLI*     .....  
 21 *TSAD\_SALTY*     .....  
 22 *TSAD\_YERPE*     .....  
 23 *TSAD\_HAEIN*     .....  
 24 *TSAD\_THEMA*     .....  
 25 *TSAD\_BACSU*     .....  
 26 *TSAD\_STREE*     .....  
 27 *TSAD\_STAAU*     .....  
 28 *TSAD\_MYCTU*     .....  
 29 *TSAD\_RICCN*     .....  
 30 *TSAD\_DESVH*     AGCS.....

Eucaryotes  
 Archées  
 Mitochondries  
 Bactéries

**Annexe 3 : Séquences des gènes codant les protéines du complexe KEOPS humain.**

Gènes/ Plasmides/ Sites de clonage	Séquences nucléotidiques
<i>TPRKB</i> pET24b NdeI-BamHI	CATATGCATCACCCACCACCGAGAACCTGTACTTCCAGGGTCACTGACCACAGCTGGAACCTGTTCCGGAAATGCCGTGTGACCTGCTGCTTTAAGACGTTAAAAACCGGGGCA TCTGCGTCTGAAGCGGATGGAGGGTACCATCGACGGCAGCTGATTAAACCCAGCCGATGTCGGAAGCGCGTGTTCCTGGTCTGTTTCAAGCCGTGCGCGGTGATCAAGCACCGTTTTC ATGAAAAACCGTACCTGAGCACCAGATCATTTTCACTGAGCCCAACAACATTAGCGAAGCGCTGAAGAAATTTGGCATCAGCGCAACGATACAGCATCTGATTGTGTACATCGA GGAAAGTGAAAAAGCAGATTAACCAAGAATATCTGATCAGCAGGTGGAAAGGCCAACAGTTAGCTGAAGAACCTGCCGGAGATCATGAACATCACCAAGTGAAGAAAAATTAACAAGTGTAG CAGCCAGGAAGAGATCGTACCCTGCTGGACGCGATTCATTCGGTATGAGCACCAGAAATGTTCTGTAAGGATCC
<i>TP53RK</i> pET21a NdeI-BamHI	CATATGCATCACCCACCACCGAGAACCTGTACTTCCAGGGCAGCGCGCTGCGACCAACCCGCGGACCGGAGGAACCGCGCCGGAGGCGGAAGCGTGGCGCGG CGCGTGAACGTAGCAGCCGTTTCTGAGCGGTCTGGAGCTGTTGAAGCAGGGTGCAGAAAGCGCGTGTTCCTGGTCTGTTTCAAGCCGTGCGCGGTGATCAAGCACCGTTTTC CGAAGGGTTCACCTGATCCGGCGCTGGAGCGCGCTGCTGGTCTGCTGCTGACCGTGAAGAAAGCGCGTGTGCTGCTGCGTTCGCGTGTGCGGGTATTAGCGCCGCGTGGTTTTCT TTGTTGACTACCGAGCACTGCCGTATATGGAAGAAATCGAGGGCAGCGTGACCGTTCGTGATTATTCAGAGCACCATGGAGACCGAAAAAGACCCCGCAAGGTCTGAGCAAC TGGCGAAAAACATCGCCAGGTGCTGGCGGTATGACACGAGGATCTGATTACCGGTGACCTGACCACGACGAAACATGCTGCTGAACCCGCGCTGGAACCACTGAACATCGTTC TGATTGATTCGCTGAGCTTTATTAGCGCGCTGCCGGAGGACAAGGGCGGTGATCTGACGTTCTGGAAGAAAGCGTTTCTGAGCACCACCCGCAACCGAGACCGTGTTCGAAGC GTTTCTGAAGAGCTATAGCACGACGCAAGAAAGCGCGTCCGGTGTGAAGAACTGGATGAGTTCGCTGCTGCTGCTGTAAGCTAGCATGTTGGTGGCTAAGGATCC
<i>hKEOPS</i> pET24d NdeI-XhoI	CATATGCCGCGGTGCTGGTTTTGAAGCAGCGCCAAACAAATGGCGTGGCGTGGTGCAGATGGCAAAGTGTGGCGAACCCGCGTGCACCTATGTGACCGCCGGGACCGCGCTT CTGCCGGGTGATACCGCCGCCATCACCTGCGGTTATCTGGAATCTGTCGAGGAAGCACTGACGGAAGCGGCTTAACAGCCAGGATATGCACTGCACTTCAACCAAAGGCCGGCA TGGGTGCCCGCTGGTTAGCGTGGCGTTGCGCCGCTACCGTGGCCAGCTGTGGAATAAACCGTGGTGGTGTGAACCACTGTATTGGCCACATGAAATGGGCGCGCTGATCACCGCGCC CACCAAGCCCGACCGTGTATATGTAGCGCGGTAACACGAGGTGATTGCTACAGCGAACATCGTTATCGTATCTTTGGTGAACCATGATATGGCGTGGGTAATGGCTGGATCGTTTTGC CGGTGTGCTGAATAATTCGAACGATCCGAGCCCGGCTACAACATTAACAGATGGCCAAACCGGCAAAAAAATAGTTGAAGTCCGCTACACCGTGAAGGATGATGATGAGCTTTTGGCG ATCCTGAGCTTATTGAAGATGTTGCCATCGATGCTGGCCACCGGGCAGTGTACCCGGAAGATCTGTTTCAGCCTGAGGAACCGTGTTCATCGTGGGATGATGAGTGGAGTCAACGAAACCGC GATGGACATGTTGGTCCCAGGAAGCCGTTGTTGGTGGCGTGGTGTAAATGCGCTGTCGAGGAAATGATGGCCACCATGTGTCAGGAACCGTGGTCCGCGCTGTTGCGACCGATGAA CGCTTCTGATTGACAATGGTGGATGATTGCCAGCGCGCTGGGAAATGTTCTGTCGCGTACCGTACCCCGTTAAAGTATTCTGGCTTACCCAGCGTATCGACCGATGAAGTGGAGT GACCTGGCGGATTAAGTCTAGGATCTCTAGAAATTTTGTAACTTTAAGAAAGAGATATAACATGGCGCGCCGCGCCACACCGCGCGGATGGCGAAGAGCCGCGCCGAAAGCCG AGGCGTGGCGCAGCCGTAACGACGAGCGCGCTCTGAGCGCGCTGGAACGAGGTTGCGAGCGCGCGTTCCTGTTGGCGCTTTCAGGGCCGCGCGCGGTGATCAAA CACCGCTTCCGAAAGGCTACCGTACCCGCGCTGGAAGCGCGTATAGCGCTGCGCGTACCGTGCAGGAAGCCGCGCGCTTACGCTGTGCGCGCGGGCATAGCGCCCGGTTGTGT TTTTGTGGATTATGCGCAGCACTGCTGATATGGAAGAAATTAAGGCTCGGTGACCGTTCGCGATTATTCAGAGCACGATGGAACCGGAAAAAACCGCGAGGCTGTCTCAACTAGCC AAACCACTGGTCAAGTTTACGCGGATGACGATGAAGACTGATTATGTTGATCTGACCACAGCAACATGCTGTGAAACCGCGCTGGAACGATGAAACATTTGTTAATGACTTTGG TCTGAGCTTCATTTGCGCCGCGGAAAGATAAAGCGGTGACTTATATGTTCTGGAAGAAAGCTTCTGAGCACCATCCGAAACCGGAAACCGTGTGTAAGGCTTTCGAAAAAGCTACAGCA CCAGCTCGAAAAAGCCGCTCGGTGCTGAAAAATAGATGAAGTGCCTGCGCGGTGTAACCGCAGCATGTTGGTCACTAGAGAGCATTTTGTAACTTTAAGAAAGAGATATA ATATGCAAGTACCCATCAGCTGACCTGTTTCCGAAATCCGCGTGAACCTTACTTATTAAGATGTTAAAAATGCGGGGATCTGCTGCGCAAGCGATGGAAGGACCATCGATGGTAGC CTGTAATACCGACCTGATTGTTGATCGTTTCAAGATCTGGTGGCGGCAACAAAGCGGTTCACTTATACAACTGGTAAAAATGAAACCCGCGCTGAGCAGCAAAATTTTCAACTTA AGCCGAAATAACAATATTAGCGAAGCCCTGAAAAATTTGGTATCAGCGCAATGACACGACATCTGATTGTTTACATTAAGAGGGTGAACAAATAATCAGGAATACCTGATTAGCCA AGTGGAAAGTCACTAGGTTTCTGAAAAATTCGCGGAAATCATGAACATACGGAAGTGAACAAAAATTAATAATTAAGCTCGCAGGAAGAAAGTATTGGCACCTGTAGATGCGATATT GTCCGATGAGCACCAGATGTGCTGAACCTAAGCTTAATAATTTGTTAACTTTAAGAAAGAGATATAACATGCGTGTGCGGATGCGGATGAGCGGCGGCGCGGATGGCGGTGACCG CCGCGGTGGCCACAGCTGCCGCGCGGCGTGGATACCGCGCGCGCCGCGCGTGGCGCGCCGCGCGGCGCACGCGCGGGTCCGGCGCTGATGCCGAGCGCGCCCGCGGACCGCT ATGCGCCGACATCTTCCCTGAGCGTGGCTTCCGACCCCGTGGAAAGCGAAATCGCCACCGTTCGCTGGCGCGGATGCCAAACCGCACCGCGTGGTGGTGAAGATCTGACCG TGTGCGGCGCTATCTGGTGTGCGCTGGAAGCCGAAGATTGCGCTGTTACGATTTTCGCTCAACCTTTGAGCAGTAAAGCTGTTGGTGGCGCACCATGCAAGCGTGGTGGCGCGG GTTAGCCGTAATAACTGCGCCCGCGCTAGCTTTTAACTTTAAGAAAGAGATATAGTTATGAACTGCTGGTGAATACGTTGTTGAGGAAAGCAACCGCAAGAACTGCTGAGCTG TGAAGCGCGGGTATGGCGACCGTTCAGGGCTGCTGAGCGCGTGGCCAGATGAAAGATATGTTGACGGAATTAATCGATCCGCTGTCGAGGGTGAAGTGCAGCACCGCGTGGCGG CGGCCCCGACGAGATCTGGATGGTATGATGAGATGATGCGGAAGATGAAATAACATTTGATAACCGTACCACTTCATGTTGCTGAGCGCAACCGCCGAAAAACCCCTGCGACCATCA CCATCACCTAATCTGAG

**Annexe 4 : séquence du pUC18DNA<sup>ARNt</sup>**

Séquence du gène codant le promoteur T7 (en noir), l'ARNt<sup>le</sup><sub>AUU</sub> et du ribozyme sous cloné entre les sites : BglII et XhoI dans le plasmide puc18. Entre la séquence du gène codant l'ARNt<sup>le</sup><sub>AUU</sub> (en rouge) et la séquence du gène codant le ribozyme (en vert) un site de restriction a été inséré : NdeI (séquence soulignée) :

AGATCTTAATACGACTCACTATAGGGCCGGTTAGCTCAGTTGGTTAGAGCGTGGTGTCTAATAACGCC  
AAGGTCGCGGGTTCGATCCCCGTACGGGCCATGGGTCGGCATGGCATCTCCACCTCCTCGCGGTCCG  
ACCTGGGCTACTTCGGTAGGCTAAGGGAGAAGCTTGGCACTGGCCGTCGTTTCTCGAG



## **Annexe 5 : Expression, production et purification des complexes archées.**

### **Matériels et Méthodes.**

Des séquences polycistroniques contenant différentes combinaisons de sous-unités PaKEOPS codant pour les gènes codant Kae1, Bud32, Cgi121, Pcc1 et Pcc2 ont été clonées dans les plasmides pETt26b entre les sites de restriction NdeI-XhoI annotés respectivement pKBC, pBC, pKP1His6, pKP2His6 et pKP2strep-P1his6. Les gènes codant pour Pcc1 et Pcc2 sont fusionnés à leur extrémité 3' à une séquence codant pour un marqueur hexa-histidine indiqué comme His6.

Les protéines recombinantes ont été exprimées dans les souches de Rosetta2 (DE3) pLYS E coli (Novagen). Les souches bactériennes transformées par les plasmides pKBC, pBC, pKP1-His6, pKP2-His6 ou pKP2strep-P1 his6 ont été cultivées à 37°C dans un milieu LB complet avec de la kanamycine et du chloramphénicol jusqu'à une densité optique (600 nm) de 0,6-0,7. Après induction avec l'IPTG (1mM), l'incubation a été poursuivie à 37°C pendant 2h. Les cellules ont été collectées par centrifugation et remises en suspension dans un tampon de lyse froid + des inhibiteurs de protéase complets sans EDTA (Roche). Le tampon de lyse est composé de 20 mM de Tris-Hcl (pH 8), 300 mM de NaCl, 5 mM de  $\beta$ -mercaptoéthanol, 10 % de Glycérol. Selon les expériences, la concentration de NaCl et de glycérol utilisé peut varier, des variations seront indiquées. Les cellules ont été lysées à 4°C avec un broyeur cellulaire à un coup et le lysat a été clarifié par centrifugation à 30 000 g pendant 30 min à 18°C. Comme les protéines recombinantes proviennent d'archées thermophiles, le lysat clarifié a été incubé pendant 10 à 15 minutes à 65°C pour précipiter les protéines d'*Escherichia coli*. Les précipités de protéines ont été éliminés par centrifugation à 30 000 g pendant 20 minutes à 18°C. La fraction soluble a été ajustée à 10 mM d'imidazole et les protéines ont été purifiées par chromatographie d'affinité à écoulement par gravité sur une colonne de résine NiNTA (Quiagen) à température ambiante. Les étapes d'équilibrage, de lavage et d'élution ont été réalisées avec un tampon de lyse contenant respectivement 10, 40 et 400 mM d'imidazole. Les fractions d'intérêt ont été mises en commun et chargées sur une colonne d'affinité d'héparine HiTRAP prééquilibrée dans un tampon de lyse. L'élution a été réalisée avec un gradient linéaire de NaCl (de 150 mM à 1M). Les fractions d'intérêt ont été regroupées, concentrées à 18°C et chargées sur une chromatographie d'exclusion de taille sur un HiLoadR

16/600 SuperdexR 200pg ou un HiLoadR 16/600 SuperdexR 75pg selon la taille des complexes protéiques.

Chaque étape de la purification a été contrôlée par analyse SDS-PAGE des aliquotes sur des gels à gradient de 4 à 20 %, les protéines ont été révélées par le colorant Commassie InstantBlueTm (Sigma-Aldrich).

Les complexes CBKP1P2, CBKP1 et CBKP2 utilisés pour l'analyse de SAXS ont été purifiés chacun à partir du mélange suivant de cellules Rosetta : 2 volumes pKBC + 1 volume pBC et respectivement 1 volume pKP1His6 ou 1 volume pKP2His6 ou 1 volume pKP2strep-P1His6. Dans ces conditions, la stœchiométrie des différentes sous-unités est la meilleure obtenue et permet la purification d'une plus grande quantité de complexes KEOPS complets. En revanche, dans ces conditions, on obtient un excès significatif des complexes P1P2 et P2 par rapport aux complexes KEOPS complets dont les volumes d'élutions par chromatographie d'exclusion stérique sont différents et ont permis d'isoler ces sous-unités. Les complexes P1P2 et la protéine P2 pures ont été mis en cristallisation.

## Résultats.

### Purification du complexe KBCP1P2 :

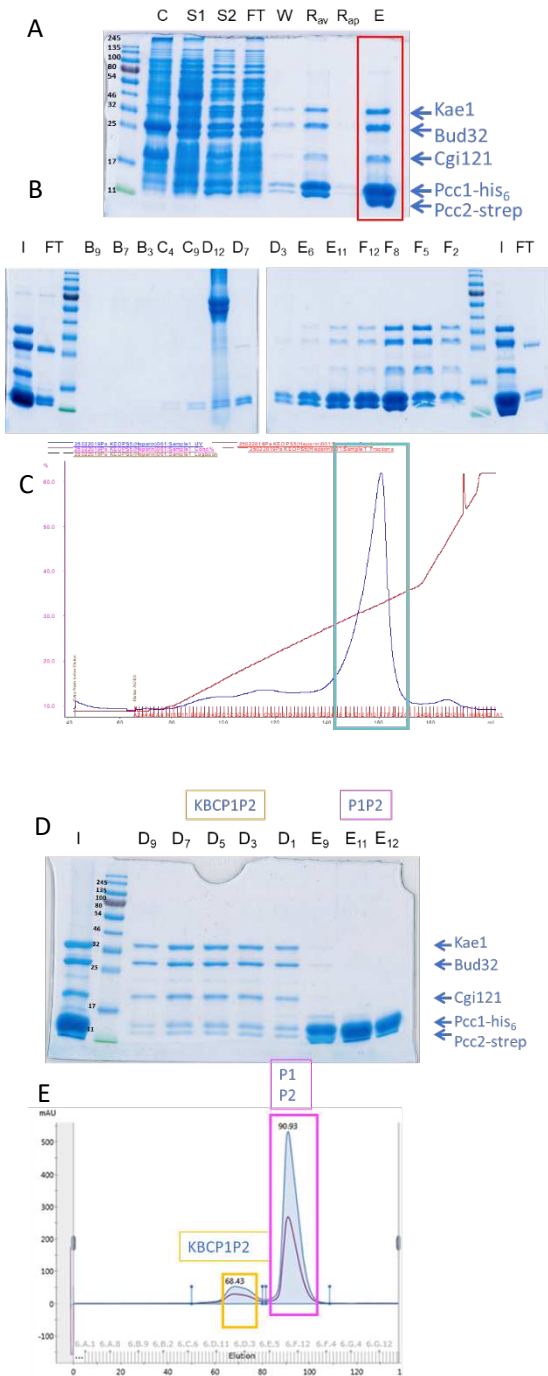


Figure 110 Purification du complexe KBCP1P2.

Les protéines du complexe KBCP1P2 ont été purifiées par chromatographie d'affinité sur ions métalliques ( $\text{Ni}^{2+}$ ) immobilisés suivie d'une chromatographie d'affinité et d'exclusion stérique utilisant une colonne Superdex 200 16-600 (GE-Healthcare).

L'analyse des fractions d'élution de la chromatographie d'affinité sur ions  $\text{Ni}^{2+}$  montre que les sous-unités du complexe sont éluées à 400 mM d'imidazole (figure 110 A). Les protéines du complexe sont ensuite chargées sur une colonne de chromatographie d'affinité et sont éluées selon un gradient de concentration croissante en sels. Le chromatogramme montre que des protéines sont éluées à une concentration de 300 mM NaCl (figure 110 C). L'analyse SDS PAGE confirme que les protéines correspondent à celles du complexe KBCP1P2 (figure 110 B). Les fractions les plus pures sont réunies et concentrées puis chargées sur une colonne de chromatographie d'exclusion stérique Superdex 200 16-600 (GE-Healthcare). Le chromatogramme montre que 2 pics différents sont élués à des volumes de 68,43 mL et 90,93 mL (figure 110 E). L'analyse SDS PAGE de ces fractions confirme que le pic 1 (encadré en jaune) correspond aux protéines du complexe KBCP1P2 et que le pic 2 (encadré en rose) correspond aux protéines P1P2 (figure 110 D). Les fractions les plus pures sont réunies puis concentrées par centrifugation sur membrane et enfin conservées à  $-80^{\circ}\text{C}$ .

## Purification du complexe KBCP2 :

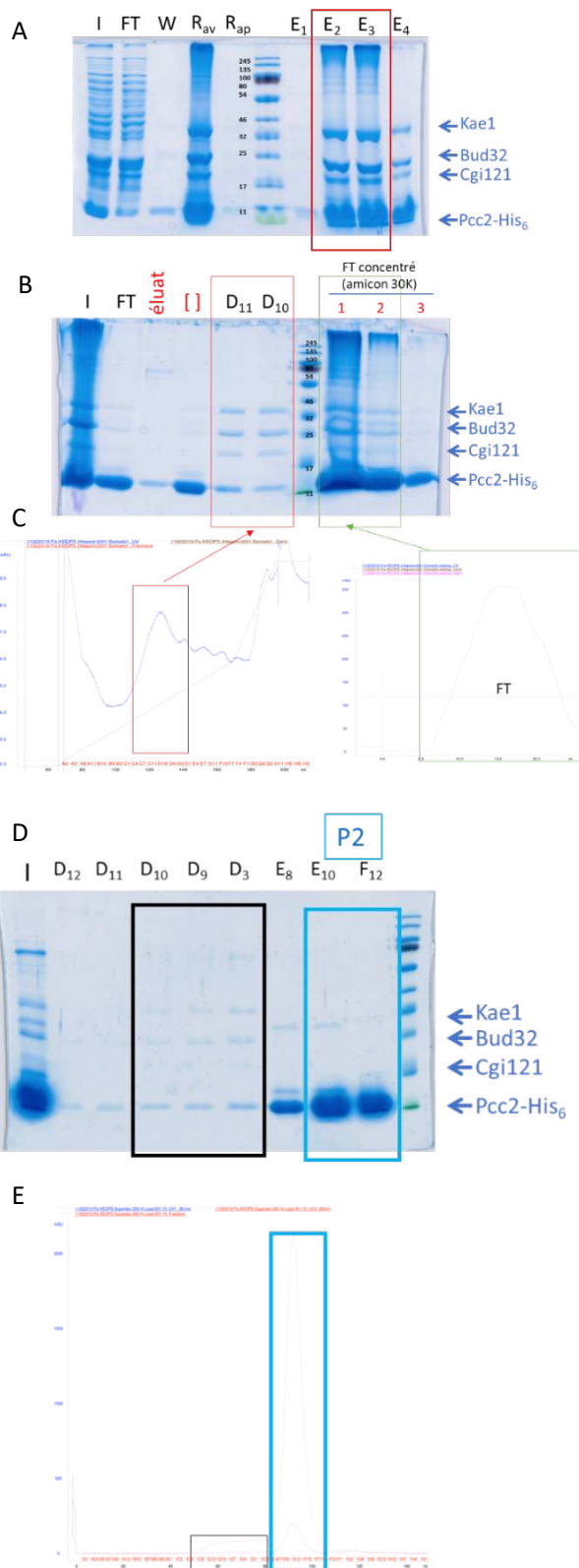
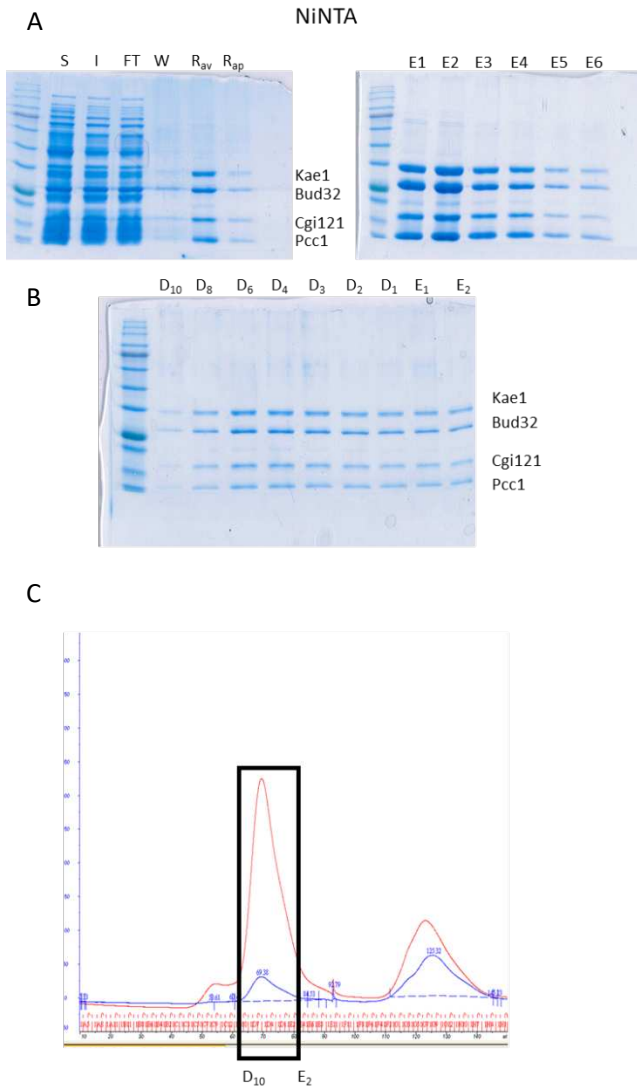


Figure 111 Purification du complexe KBCP2.

Les protéines du complexe KBCP2 ont été purifiées de façon identique que les protéines KBCP1P2 (figure 111 A-C).

Le chromatogramme montre que 2 pics différents sont élués à des volumes de 60 mL et 90 mL (figure 111 E). L'analyse SDS PAGE de ces fractions confirme que le pic 1 (encadré en noir) correspond aux protéines du complexe KBCP2 et que le pic 2 (encadré en bleu) correspond aux protéines P2 (figure 111 D). Les fractions les plus pures sont réunies puis concentrées par centrifugation sur membrane et enfin conservées à -80°C.

Purification du complexe KBCP1 :



Les protéines du complexe KBCP1 ont été purifiées par chromatographie d'affinité sur ions métalliques (Ni<sup>2+</sup>) immobilisés suivie d'une chromatographie d'exclusion stérique utilisant une colonne Superdex 200 16-600 (GE-Healthcare).

L'analyse des fractions d'élution de la chromatographie d'affinité sur ions Ni<sup>2+</sup> montre que les sous-unités du complexe sont éluées de 20 à 400 mM d'imidazole (figure 112 A). Les protéines du complexe sont ensuite chargées sur une colonne de chromatographie d'exclusion stérique Superdex 200 16-600 (GE-Healthcare). Le chromatogramme montre que 2 pics différents sont élués à des volumes de 69,34 mL et 125,32 mL (figure 112 C). L'analyse SDS PAGE de ces fractions confirment que le pic 1 (encadré en noir)

Figure 112 Purification du complexe KBCP1.

correspond aux protéines du complexe KBCP1 (figure 112 B). Les fractions les plus pures sont réunies puis concentrées par centrifugation sur membrane et enfin conservées à -80°C.

## Liste des abréviations et acronymes :

aaRS = Aminoacyl-tRNA-synthetase  
AC = Anticodon  
ACP = Acylphosphatase  
ADP = Adénosine diphosphate  
ADN = acide désoxyribonucléique  
AMPCPP =  $\alpha,\beta$ -Methyleneadenosine 5'-triphosphate  
ATPyS = adenosine-5'-o-(3-thio-triphosphate)  
AMP = Adénosine monophosphate  
ARN = Acide ribonucléique  
ARNm = ARN messenger  
ARNr = ARN ribosomique  
ARNt = ARN de transfert  
ATP = Adénosine triphosphate  
ATPase = Adénosine triphosphate hydrolase  
Ca-AMP = carbamoyl-AMP  
CCAse = ARNt nucléotidyl-transférase  
COG = Cluster of Orthologous Genes  
CTF = Contrast Transfer Function  
EF = Elongation factor  
HIV = Human Immunodeficiency Viruses  
IPTG = Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
KO = knock-out  
MET = Microscopie Electronique en Transmission  
RMSD = Root-Mean-Square Deviation  
t6A = N<sup>6</sup>-thréonyl-carbamoyl-adenosine  
 $\chi^2$  = khi-deux



## BIBLIOGRAPHIE

- Abelson, J., Trotta, C. R., & Li, H. (1998). tRNA splicing. *The Journal of Biological Chemistry*, 273(21), 12685-12688. <https://doi.org/10.1074/jbc.273.21.12685>
- Absalon, C., Obuchowski, M., Madec, E., Delattre, D., Holland, I. B., & Séror, S. J. (2009). CpgA, EF-Tu and the stressosome protein YezB are substrates of the Ser/Thr kinase/phosphatase couple, PrkC/PrpC, in *Bacillus subtilis*. *Microbiology*, 155(3), 932-943. <https://doi.org/10.1099/mic.0.022475-0>
- Adrian, M., Dubochet, J., Lepault, J., & McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, 308(5954), 32-36. <https://doi.org/10.1038/308032a0>
- Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., ... Shinkai, A. (2008). X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins*, 70(3), 1108-1111. <https://doi.org/10.1002/prot.21794>
- Agris, P. F. (2004). Decoding the genome : A modified view. *Nucleic Acids Research*, 32(1), 223-238. <https://doi.org/10.1093/nar/gkh185>
- Agris, Paul F., Narendran, A., Sarachan, K., Väre, V. Y. P., & Eruysal, E. (2017). The Importance of Being Modified. Dans *The Enzymes* (Vol. 41, pp. 1-50). (S.l.): Elsevier. <https://doi.org/10.1016/bs.enz.2017.03.005>
- Agris, Paul F., Vendeix, F. A. P., & Graham, W. D. (2007). tRNA's Wobble Decoding of the Genome : 40 Years of Modification. *Journal of Molecular Biology*, 366(1), 1-13. <https://doi.org/10.1016/j.jmb.2006.11.046>
- Alexandrov, A., Martzen, M. R., & Phizicky, E. M. (2002). Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA*, 8(10), 1253-1266. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1370335/>
- Alford, B., & Hecht, S. M. (1978). 2'-Versus 3'-OH specificity in tRNA aminoacylation. Further support for the « secondary cognition » proposal. *Journal of Biological Chemistry*, 253(14), 4844-4850. Repéré à <http://www.jbc.org/content/253/14/4844>
- Allali-Hassani, A., Campbell, T. L., Ho, A., Schertz, J. W., & Brown, E. D. (2004). Probing the active site of YjeE : A vital *Escherichia coli* protein of unknown function. *Biochemical Journal*, 384(Pt 3), 577-584. <https://doi.org/10.1042/BJ20041082>
- Antczak, M., Popena, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., ... Szachniuk, M. (2016). New functionality of RNAComposer : Application to shape the axis of miR160 precursor structure. *Acta Biochimica Polonica*, 63(4), 737-744. [https://doi.org/10.18388/abp.2016\\_1329](https://doi.org/10.18388/abp.2016_1329)
- Arigoni, F., Talabot, F., Peitsch, M., Edgerton, M. D., Meldrum, E., Allet', E., ... Loferer, H. (1998). A genome-based approach for the identification of essential bacterial genes, 16, 6.
- Arimbasseri, A. G., Iben, J., Wei, F.-Y., Rijal, K., Tomizawa, K., Hafner, M., & Maraia, R. J. (2016). Evolving specificity of tRNA 3-methyl-cytidine-32 (m3C32) modification : A subset of tRNAs<sup>Ser</sup> requires N6-isopentenylation of A37. *RNA (New York, N.Y.)*, 22(9), 1400-1410. <https://doi.org/10.1261/rna.056259.116>
- Arragain, S., Handelman, S. K., Forouhar, F., Wei, F.-Y., Tomizawa, K., Hunt, J. F., ... Atta, M. (2010). Identification of Eukaryotic and Prokaryotic Methylthiotransferase for Biosynthesis of 2-Methylthio-N6-threonylcarbamoyladenine in tRNA. *The Journal of Biological Chemistry*, 285(37), 28425-28433. <https://doi.org/10.1074/jbc.M110.106831>



- Arrondel, C., Missouri, S., Snoek, R., Patat, J., Menara, G., Collinet, B., ... Mollet, G. (2019). Defects in t6A tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome. *Nature Communications*, 10. <https://doi.org/10.1038/s41467-019-11951-x>
- Ashraf, S. S., Ansari, G., Guenther, R., Sochacka, E., Malkiewicz, A., & Agris, P. F. (1999). The uridine in « U-turn » : Contributions to tRNA-ribosomal binding. *RNA*, 5(4), 503-511. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1369777/>
- Auffinger, P., & Westhof, E. (2001). An extended structural signature for the tRNA anticodon loop. *RNA (New York, N.Y.)*, 7(3), 334-341. <https://doi.org/10.1017/s1355838201002382>
- Augustin, M. A., Reichert, A. S., Betat, H., Huber, R., Mörl, M., & Steegborn, C. (2003). Crystal Structure of the Human CCA-adding Enzyme : Insights into Template-independent Polymerization. *Journal of Molecular Biology*, 328(5), 985-994. [https://doi.org/10.1016/S0022-2836\(03\)00381-4](https://doi.org/10.1016/S0022-2836(03)00381-4)
- Avery, O. T., Macleod, C. M., & McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *The Journal of Experimental Medicine*, 79(2), 137-158. <https://doi.org/10.1084/jem.79.2.137>
- Aydin, I., Dimitropoulos, A., Chen, S. H., Thomas, C., & Roujeinikova, A. (2011). Purification, crystallization and preliminary X-ray crystallographic analysis of the putative *Vibrio parahaemolyticus* resuscitation-promoting factor YeaZ. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 67(5), 604-607. <https://doi.org/10.1107/S1744309111010219>
- Aydin, I., Saijo-Hamano, Y., Namba, K., Thomas, C., & Roujeinikova, A. (2011). Structural Analysis of the Essential Resuscitation Promoting Factor YeaZ Suggests a Mechanism of Nucleotide Regulation through Dimer Reorganization. *PLoS ONE*, 6(8). <https://doi.org/10.1371/journal.pone.0023245>
- Bai, X., McMullan, G., & Scheres, S. H. W. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1), 49-57. <https://doi.org/10.1016/j.tibs.2014.10.005>
- Barraud, P., & Tisné, C. (2019). To be or not to be modified : Miscellaneous aspects influencing nucleotide modifications in tRNAs. *Iubmb Life*, 71(8), 1126-1140. <https://doi.org/10.1002/iub.2041>
- Bartz, J., Söll, D., Burrows, W. J., & Skoog, F. (1970). Identification of the Cytokinin-Active Ribonucleosides in Pure *Escherichia coli* tRNA Species. *Proceedings of the National Academy of Sciences of the United States of America*, 67(3), 1448-1453. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC283372/>
- Bendory, T., Bartesaghi, A., & Singer, A. (2020). Single-particle cryo-electron microscopy : Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2), 58-76. <https://doi.org/10.1109/msp.2019.2957822>
- Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, 334(6062):1524-9. doi: 10.1126/science.1212642. Epub 2011 Nov 17. PMID: 22096102.
- Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron

- micrographs. *Nature Methods*, 16(11), 1153-1160. <https://doi.org/10.1038/s41592-019-0575-8>
- Björk, G. R. ., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jönsson, Y. H., & Wikström, P. M. (1987). Transfer Rna Modification. *Annual Review of Biochemistry*, 56(1), 263-285. <https://doi.org/10.1146/annurev.bi.56.070187.001403>
- Blokesch, M., Paschos, A., Bauer, A., Reissmann, S., Drapal, N., & Böck, A. (2004). Analysis of the transcarbamoylation-dehydration reaction catalyzed by the hydrogenase maturation proteins HypF and HypE. *European Journal of Biochemistry*, 271(16), 3428-3436. <https://doi.org/10.1111/j.1432-1033.2004.04280.x>
- Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., ... Bujnicki, J. M. (2018). MODOMICS : A database of RNA modification pathways. 2017 update. *Nucleic Acids Research*, 46(Database issue), D303-D307. <https://doi.org/10.1093/nar/gkx1030>
- Boguta, M., & Reese, J. C. (2018). Regulation of tRNA synthesis and modification in physiological conditions and disease. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, 1861(4), 283-284. <https://doi.org/10.1016/j.bbagr.2018.03.004>
- Boschi-Muller, S., & Motorin, Y. (2013). Chemistry enters nucleic acids biology : Enzymatic mechanisms of RNA modification. *Biochemistry (Moscow)*, 78(13), 1392-1404. <https://doi.org/10.1134/S0006297913130026>
- Boutin, J. A., Li, Z., Vuillard, L., & Vénien-Bryan, C. (2016). La cryo-microscopie, une alternative à la cristallographie aux rayons X ? *Med Sci (Paris)*. 2016 8-9;32(8-9):758-67. French. doi: 10.1051/medsci/20163208025. .
- Braddock, D. T., Baber, J. L., Levens, D., & Clore, G. M. (2002). Molecular basis of sequence-specific single-stranded DNA recognition by KH domains : Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *The EMBO Journal*, 21(13), 3476-3485. <https://doi.org/10.1093/emboj/cdf352>
- Braun, D. A., Rao, J., Mollet, G., Schapiro, D., Daugeron, M.-C., Tan, W., ... Hildebrandt, F. (2017). Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nature Genetics*, 49(10), 1529-1538. <https://doi.org/10.1038/ng.3933>
- Braun, D. A., Shril, S., Sinha, A., Schneider, R., Tan, W., Ashraf, S., ... Hildebrandt, F. (2018). Mutations in WDR4 as a new cause of Galloway-Mowat syndrome. *American journal of medical genetics. Part A*, 176(11), 2460-2465. <https://doi.org/10.1002/ajmg.a.40489>
- Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica Et Biophysica Acta*, 34, 103-110. [https://doi.org/10.1016/0006-3002\(59\)90237-9](https://doi.org/10.1016/0006-3002(59)90237-9)
- Brookes, E., Vachette, P., Rocco, M., & Pérez, J. (2016). US-SOMO HPLC-SAXS module : Dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of Applied Crystallography*, 49(Pt 5), 1827-1841. <https://doi.org/10.1107/S1600576716011201>
- Brown, C. J., Johnson, A. K., Dunker, A. K., & Daughdrill, G. W. (2011). Evolution and disorder. *Current Opinion in Structural Biology*, 21(3), 441-446. <https://doi.org/10.1016/j.sbi.2011.02.005>
- Brzezicha, B., Schmidt, M., Makalowska, I., Jarmolowski, A., Pienkowska, J., & Szwejkowska-Kulinska, Z. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Research*, 34(20), 6034-6043. <https://doi.org/10.1093/nar/gkl765>

- Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., ... Emili, A. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, *433*(7025), 531-537. <https://doi.org/10.1038/nature03239>
- Byrne, R. T., Konevega, A. L., Rodnina, M. V., & Antson, A. A. (2010). The crystal structure of unmodified tRNA Phe from *Escherichia coli*. *Nucleic Acids Research*, *38*(12), 4154-4162. <https://doi.org/10.1093/nar/gkq133>
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., ... Agris, P. F. (2011). The RNA modification database, RNAMDB : 2011 update. *Nucleic Acids Research*, *39*(Database), D195-D201. <https://doi.org/10.1093/nar/gkq1028>
- Chan, P. P., & Lowe, T. M. (2016). GtRNADB 2.0 : An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, *44*(D1), D184-189. <https://doi.org/10.1093/nar/gkv1309>
- Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, *6*(6), 201-209. <https://doi.org/10.1007/BF02173653>
- Chatterjee, K., Nostramo, R. T., Wan, Y., & Hopper, A. K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface : Location, location, location. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, *1861*(4), 373-386. <https://doi.org/10.1016/j.bbagr.2017.11.007>
- Chheda, G. B., Hall, R. H., Magrath, D. I., Mozejko, J., Schweizer, M. P., Stasiuk, L., & Taylor, P. R. (1969). Aminoacyl nucleosides. VI. Isolation and preliminary characterization of threonyladenine derivatives from transfer ribonucleic acid. *Biochemistry*, *8*(8), 3278-3282. <https://doi.org/10.1021/bi00836a022>
- Claude, A. (1938). A Fraction from Normal Chick Embryo Similar to the Tumor Producing Fraction of Chicken Tumor I. *Proceedings of the Society for Experimental Biology and Medicine*, *39*, n<sup>o</sup> 2, 398-403. <https://doi.org/10.3181>
- Colin, E., Huynh Cong, E., Mollet, G., Guichet, A., Gribouval, O., Arrondel, C., ... Antignac, C. (2014). Loss-of-Function Mutations in WDR73 Are Responsible for Microcephaly and Steroid-Resistant Nephrotic Syndrome : Galloway-Mowat Syndrome. *American Journal of Human Genetics*, *95*(6), 637-648. <https://doi.org/10.1016/j.ajhg.2014.10.011>
- Costessi, A., Mahrouf, N., Sharma, V., Stunnenberg, R., Stoel, M. A., Tijchon, E., ... Stunnenberg, H. G. (2012). The Human EKC/KEOPS Complex Is Recruited to Cullin2 Ubiquitin Ligases by the Human Tumour Antigen PRAME. *PLoS ONE*, *7*(8), e42822. <https://doi.org/10.1371/journal.pone.0042822>
- Cramer, P. (2020). Rosalind Franklin and the Advent of Molecular Biology. *Cell*, *182*(4), 787-789. <https://doi.org/10.1016/j.cell.2020.07.028>
- Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, *12*, 138-163.
- Crick, F. H. C. (1966). Codon—Anticodon pairing : The wobble hypothesis, *J Mol Biol.* *19*(2):548-55. doi: 10.1016/s0022-2836(66)80022-0. PMID: 5969078.
- Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., & Phizicky, E. M. (1997). A 2'-phosphotransferase implicated in tRNA splicing is essential in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, *272*(20), 13203-13210. <https://doi.org/10.1074/jbc.272.20.13203>
- Cusack, S., Härtlein, M., & Leberman, R. (1991). Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucleic Acids Research*, *19*(13), 3489-3498. <https://doi.org/10.1093/nar/19.13.3489>

- Dare, K., & Ibba, M. (2012). Roles of tRNA in cell wall biosynthesis. *Wiley interdisciplinary reviews. RNA*, 3(2). <https://doi.org/10.1002/wrna.1108>
- Daugeron, M.-C., Lenstra, T. L., Frizzarin, M., El Yacoubi, B., Liu, X., Baudin-Baillieu, A., ... Libri, D. (2011). Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. *Nucleic Acids Research*, 39(14), 6148-6160. <https://doi.org/10.1093/nar/gkr178>
- de Crécy-Lagard, V., & Jaroch, M. (2020). Functions of Bacterial tRNA Modifications : From Ubiquity to Diversity. *Trends in Microbiology*. <https://doi.org/10.1016/j.tim.2020.06.010>
- de la Sierra-Gallay, I. L., Pellegrini, O., & Condon, C. (2005). Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature*, 433(7026), 657-661. <https://doi.org/10.1038/nature03284>
- Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V., & Iwata-Reuyl, D. (2012). Biosynthesis of threonylcarbamoyl adenosine (t6A), a universal tRNA nucleoside. *The Journal of Biological Chemistry*, 287(17), 13666-13673. <https://doi.org/10.1074/jbc.M112.344028>
- Downey, M., Houlsworth, R., Maringele, L., Rollie, A., Brehme, M., Galicia, S., ... Durocher, D. (2006). A Genome-Wide Screen Identifies the Evolutionarily Conserved KEOPS Complex as a Telomere Regulator. *Cell*, 124(6), 1155-1168. <https://doi.org/10.1016/j.cell.2005.12.044>
- Drew, K., Müller, C. L., Bonneau, R., & Marcotte, E. M. (2017). Identifying direct contacts between protein complex subunits from their conditional dependence in proteomics datasets. *PLoS Computational Biology*, 13(10). <https://doi.org/10.1371/journal.pcbi.1005625>
- Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, A. W., & Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2), 129-228. <https://doi.org/10.1017/s0033583500004297>
- Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3), 197-208. <https://doi.org/10.1038/nrm1589>
- Edelmann, F. T., Niedner, A., & Niessing, D. (2014). Production of pure and functional RNA for in vitro reconstitution experiments. *Methods (San Diego, Calif.)*, 65(3), 333-341. <https://doi.org/10.1016/j.ymeth.2013.08.034>
- Edmonds, C. G., Crain, P. F., Gupta, R., Hashizume, T., Hocart, C. H., Kowalak, J. A., ... McCloskey, J. A. (1991). Posttranscriptional modification of tRNA in thermophilic archaea (Archaeobacteria). *Journal of Bacteriology*, 173(10), 3138-3148. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC207908/>
- Edvardson, S., Prunetti, L., Arraf, A., Haas, D., Bacusmo, J. M., Hu, J. F., ... Elpeleg, O. (2017). tRNA N6-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3 (OSGEP) mutation manifest by neurodegeneration and renal tubulopathy. *European journal of human genetics : EJHG*, 25(5), 545-551. <https://doi.org/10.1038/ejhg.2017.30>
- El Yacoubi, B., Bailly, M., & de Crécy-Lagard, V. (2012). Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics*, 46(1), 69-95. <https://doi.org/10.1146/annurev-genet-110711-155641>
- El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., ... de Crécy-Lagard, V. (2011). A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. *The EMBO Journal*, 30(5), 882-893. <https://doi.org/10.1038/emboj.2010.363>

- El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., ... de Crécy-Lagard, V. (2009). The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Research*, 37(9), 2894-2909. <https://doi.org/10.1093/nar/gkp152>
- Elkins, B. N., & Keller, E. B. (1974). Enzymic synthesis of N-(purin-6-ylcarbamoyl)threonine, an anticodon-adjacent base in transfer ribonucleic acid. *Biochemistry*, 13(22), 4622-4628. <https://doi.org/10.1021/bi00719a024>
- Endicott, J. A., Noble, M. E. M., & Johnson, L. N. (2012). The Structural Basis for Control of Eukaryotic Protein Kinases. *Annual Review of Biochemistry*, 81(1), 587-613. <https://doi.org/10.1146/annurev-biochem-052410-090317>
- Eriani, G., Cavarelli, J., Martin, F., Ador, L., Rees, B., Thierry, J. C., ... Moras, D. (1995). The class II aminoacyl-tRNA synthetases and their active site : Evolutionary conservation of an ATP binding site. *Journal of Molecular Evolution*, 40(5), 499-508. <https://doi.org/10.1007/BF00166618>
- Evrard, G., Mareuil, F., Bontems, F., Sizun, C., & Perez, J. (2011). DADIMODO : A program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *Journal of Applied Crystallography*, 44(6), 1264-1271. <https://doi.org/10.1107/S0021889811035758>
- Ferreira-Cerca, S., Sagar, V., Schäfer, T., Diop, M., Wesseling, A.-M., Lu, H., ... LaRonde-LeBlanc, N. (2012). ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nature Structural & Molecular Biology*, 19(12), 1316-1323. <https://doi.org/10.1038/nsmb.2403>
- Frank, J. (1995). Approaches to large-scale structures. *Current Opinion in Structural Biology*, 5(2), 194-201. [https://doi.org/10.1016/0959-440x\(95\)80075-1](https://doi.org/10.1016/0959-440x(95)80075-1)
- Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., ... Svergun, D. I. (2017). ATSAS 2.8 : A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, 50(Pt 4), 1212-1225. <https://doi.org/10.1107/S1600576717007786>
- Freiberg, C., Wieland, B., Spaltmann, F., Ehlert, K., Brötz, H., & Labischinski, H. (2001). Identification of novel essential Escherichia coli genes conserved among pathogenic bacteria. *Journal of Molecular Microbiology and Biotechnology*, 3(3), 483-489.
- Freier, S. M., & Tinoco, I. (1975). Binding of complementary oligoribonucleotides to yeast initiator transfer RNA. *Biochemistry*, 14(15), 3310-3314. <https://doi.org/10.1021/bi00686a004>
- Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G., & Suzuki, T. (2016). RNA modifications : What have we learned and where are we headed? *Nature Reviews Genetics*, 17(6), 365-372. <https://doi.org/10.1038/nrg.2016.47>
- Galloway, W. H., & Mowat, A. P. (1968). Congenital microcephaly with hiatus hernia and nephrotic syndrome in two sibs. *Journal of Medical Genetics*, 5(4), 319-321. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1468664/>
- Galperin, M. Y. (2004). « Conserved hypothetical » proteins : Prioritization of targets for experimental study. *Nucleic Acids Research*, 32(18), 5452-5463. <https://doi.org/10.1093/nar/gkh885>
- Ganoza, M. C., Kiel, M. C., & Aoki, H. (2002). Evolutionary Conservation of Reactions in Translation. *Microbiology and Molecular Biology Reviews*, 66(3), 460-485. <https://doi.org/10.1128/MMBR.66.3.460-485.2002>
- Garcia GA, Goodenough-Lashua DEM. (1998). Mechanisms of RNA-modifying and -editing enzymes. *Grosjean H, Benne R (eds) Modification and Editing of RNA, ASM Press, Washington DC, Chap 8*, 135-168.

- Gerdes, S., El Yacoubi, B., Bailly, M., Blaby, I. K., Blaby-Haas, C. E., Jeanguenin, L., ... de Crécy-Lagard, V. (2011). Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC genomics*, *12 Suppl 1*, S2. <https://doi.org/10.1186/1471-2164-12-S1-S2>
- Gil, R., Silva, F. J., Peretó, J., & Moya, A. (2004). Determination of the Core of a Minimal Bacterial Gene Set. *Microbiology and Molecular Biology Reviews*, *68*(3), 518-537. <https://doi.org/10.1128/MMBR.68.3.518-537.2004>
- Goto-Ito, S., Ito, T., & Yokoyama, S. (2017). Trm5 and TrmD : Two Enzymes from Distinct Origins Catalyze the Identical tRNA Modification, m<sup>1</sup>G37. *Biomolecules*, *7*(1). <https://doi.org/10.3390/biom7010032>
- Grishin, N. V. (2001). KH domain : One motif, two folds. *Nucleic Acids Research*, *29*(3):638-43. doi: 10.1093/nar/29.3.638. PMID: 11160884; PMCID: PMC30387.
- Grosjean, H., Söll, D. G., & Crothers, D. M. (1976). Studies of the complex between transfer RNAs with complementary anticodons. *Journal of Molecular Biology*, *103*(3), 499-519. [https://doi.org/10.1016/0022-2836\(76\)90214-X](https://doi.org/10.1016/0022-2836(76)90214-X)
- Grosjean, H., Sprinzl, M., & Steinberg, S. (1995). Posttranscriptionally modified nucleosides in transfer RNA : Their locations and frequencies. *Biochimie*, *77*(1), 139-141. [https://doi.org/10.1016/0300-9084\(96\)88117-X](https://doi.org/10.1016/0300-9084(96)88117-X)
- Grosjean, Henri, & Björk, G. R. (2004). Enzymatic conversion of cytidine to lysidine in anticodon of bacterial tRNA<sup>Ile</sup> – an alternative way of RNA editing. *Trends in Biochemical Sciences*, *29*(4), 165-168. <https://doi.org/10.1016/j.tibs.2004.02.009>
- Grosjean, Henri, Henau, S. D., Doi, T., Yamane, A., Ohtsuka, E., Ikehara, M., ... Cedergren, R. (1987). The in vivo stability, maturation and aminoacylation of anticodon-substituted Escherichia coli initiator methionine tRNAs. *European Journal of Biochemistry*, *166*(2), 325-332. <https://doi.org/10.1111/j.1432-1033.1987.tb13518.x>
- Grosjean, Henri, & Oshima, T. (2007). How Nucleic Acids Cope with High Temperature.
- Grosjean, Henri, & Westhof, E. (2016). An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Research*, *44*(17), 8020-8040. <https://doi.org/10.1093/nar/gkw608>
- Gustilo, E. M., Vendeix, F. A. P., & Agris, P. F. (2008). tRNA's Modifications Bring Order to Gene Expression. *Current opinion in microbiology*, *11*(2), 134-140. <https://doi.org/10.1016/j.mib.2008.02.003>
- Guy, M. P., & Phizicky, E. M. (2015). Conservation of an intricate circuit for crucial modifications of the tRNA<sup>Phe</sup> anticodon loop in eukaryotes. *RNA (New York, N.Y.)*, *21*(1), 61-74. <https://doi.org/10.1261/rna.047639.114>
- Guy, M. P., Podyma, B. M., Preston, M. A., Shaheen, H. H., Krivos, K. L., Limbach, P. A., ... Phizicky, E. M. (2012). Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNA<sup>Phe</sup> anticodon loop. *RNA (New York, N.Y.)*, *18*(10), 1921-1933. <https://doi.org/10.1261/rna.035287.112>
- Guy, M. P., Shaw, M., Weiner, C. L., Hobson, L., Stark, Z., Rose, K., ... Phizicky, E. M. (2015). Defects in tRNA Anticodon Loop 2'-O-Methylation Are Implicated in Nonsyndromic X-Linked Intellectual Disability due to Mutations in FTSJ1. *Human Mutation*, *36*(12), 1176-1187. <https://doi.org/10.1002/humu.22897>
- Haag, S., Sloan, K. E., Ranjan, N., Warda, A. S., Kretschmer, J., Blessing, C., ... Bohnsack, M. T. (2016). NSUN3 and ABH1 modify the wobble position of mt-tRNA Met to expand codon recognition in mitochondrial translation. *The EMBO Journal*, *35*(19), 2104-2119. <https://doi.org/10.15252/embj.201694885>
- Hampsey, M., Na, J. G., Pinto, I., Ware, D. E., & Berroteran, R. W. (1991). Extragenic suppressors of a translation initiation defect in the *cyc1* gene of *Saccharomyces*

- cerevisiae. *Biochimie*, 73(12), 1445-1455. [https://doi.org/10.1016/0300-9084\(91\)90177-3](https://doi.org/10.1016/0300-9084(91)90177-3)
- Han, L., & Phizicky, E. M. (2018). A rationale for tRNA modification circuits in the anticodon loop. *RNA*, 24(10), 1277-1284. <https://doi.org/10.1261/rna.067736.118>
- Handford, J. I., Ize, B., Buchanan, G., Butland, G. P., Greenblatt, J., Emili, A., & Palmer, T. (2009). Conserved Network of Proteins Essential for Bacterial Viability. *Journal of Bacteriology*, 191(15), 4732-4749. <https://doi.org/10.1128/JB.00136-09>
- Hanks, S. K., Quinn, A. M., & Hunter, T. (1988). The protein kinase family : Conserved features and deduced phylogeny of the catalytic domains. *Science (New York, N.Y.)*, 241(4861), 42-52. <https://doi.org/10.1126/science.3291115>
- Harada, F., & Nishimura, S. (1972). Possible anticodon sequences of tRNA His , tRNA Asm , and tRNA Asp from Escherichia coli B. Universal presence of nucleoside Q in the first position of the anticodons of these transfer ribonucleic acids. *Biochemistry*, 11(2), 301-308. <https://doi.org/10.1021/bi00752a024>
- Harris, J. R. (1991). Negative staining-carbon film technique : New cellular and molecular applications. *Journal of Electron Microscopy Technique*, 18(3), 269-276. <https://doi.org/10.1002/jemt.1060180309>
- Hartmann, E., & Hartmann, R. K. (2003). The enigma of ribonuclease P evolution. *Trends in Genetics*, 19(10), 561-569. <https://doi.org/10.1016/j.tig.2003.08.007>
- Hartmann, R. K., Gössringer, M., Späth, B., Fischer, S., & Marchfelder, A. (2009). The making of tRNAs and more—RNase P and tRNase Z. *Progress in Molecular Biology and Translational Science*, 85, 319-368. [https://doi.org/10.1016/S0079-6603\(08\)00808-8](https://doi.org/10.1016/S0079-6603(08)00808-8)
- He, M.-H., Liu, J.-C., Lu, Y.-S., Wu, Z.-J., Liu, Y.-Y., Wu, Z., ... Zhou, J.-Q. (2019). KEOPS complex promotes homologous recombination via DNA resection. *Nucleic Acids Research*, 47(11), 5684-5697. <https://doi.org/10.1093/nar/gkz228>
- Hecker, A., Leulliot, N., Gadelle, D., Graille, M., Justome, A., Dorlet, P., ... Forterre, P. (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 Research*, 35(18), 6042-6051. <https://doi.org/10.1093/nar/gkm554>
- Hecker, A., Lopreiato, R., Graille, M., Collinet, B., Forterre, P., Libri, D., & van Tilbeurgh, H. (2008). Structure of the archaeal Kae1/Bud32 fusion protein MJ1130 : A model for the eukaryotic EKC/KEOPS subcomplex. *The EMBO Journal*, 27(17), 2340-2351. <https://doi.org/10.1038/emboj.2008.157>
- Heimowitz, A., Andén, J., & Singer, A. (2020). Reducing bias and variance for CTF estimation in single particle cryo-EM. *Ultramicroscopy*, 212, 112950. <https://doi.org/10.1016/j.ultramic.2020.112950>
- Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Research*, 34(2), 721-733. <https://doi.org/10.1093/nar/gkj471>
- Hershey, A. D. (1956). The organization of genetic material in bacteriophage T2. *Brookhaven Symposia in Biology*, (8), 6-16.
- Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., & Zamecnik, P. C. (1958). A Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *Journal of Biological Chemistry*, 231(1), 241-257. Repéré à <http://www.jbc.org/content/231/1/241>
- Högenauer, G., Turnowsky, F., & Unger, F. M. (1972). Codon-anticodon interaction of methionine specific tRNAs. *Biochemical and Biophysical Research Communications*, 46(6), 2100-2106. [https://doi.org/10.1016/0006-291x\(72\)90765-6](https://doi.org/10.1016/0006-291x(72)90765-6)

- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., ... Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*, *147*(3664), 1462-1465. <https://doi.org/10.1126/science.147.3664.1462>
- Honda, S., Kirino, Y., Maragkakis, M., Alexiou, P., Ohtaki, A., Murali, R., ... Kirino, Y. (2013). Mitochondrial protein BmPAPI modulates the length of mature piRNAs. *RNA*, *19*(10), 1405-1418. <https://doi.org/10.1261/rna.040428.113>
- Hopper, A. K., & Phizicky, E. M. (2003). tRNA transfers to the limelight. *Genes & Development*, *17*(2), 162-180. <https://doi.org/10.1101/gad.1049103>
- Illergård, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins : Structure, Function, and Bioinformatics*, *77*(3), 499-508. <https://doi.org/10.1002/prot.22458>
- Ishida, K., Kunibayashi, T., Tomikawa, C., Ochi, A., Kanai, T., Hirata, A., ... Hori, H. (2011). Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium *Thermus thermophilus*. *Nucleic Acids Research*, *39*(6), 2304-2318. <https://doi.org/10.1093/nar/gkq1180>
- Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., & Yokoyama, S. (2003). Alternative Tertiary Structure of tRNA for Recognition by a Posttranscriptional Modification Enzyme. *Cell*, *113*(3), 383-394. [https://doi.org/10.1016/S0092-8674\(03\)00280-0](https://doi.org/10.1016/S0092-8674(03)00280-0)
- Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions, 32.
- Jackman, J. E., & Alfonzo, J. D. (2013). Transfer RNA modifications : Nature's combinatorial chemistry playground. *Wiley interdisciplinary reviews. RNA*, *4*(1), 35-48. <https://doi.org/10.1002/wrna.1144>
- Jahn, D., Verkamp, E., & Söll, D. (1992). Glutamyl-transfer RNA : A precursor of heme and chlorophyll biosynthesis. *Trends in Biochemical Sciences*, *17*(6), 215-218. [https://doi.org/10.1016/0968-0004\(92\)90380-r](https://doi.org/10.1016/0968-0004(92)90380-r)
- Janczarek, M., Vinardell, J.-M., Lipa, P., & Karaś, M. (2018). Hanks-Type Serine/Threonine Protein Kinases and Phosphatases in Bacteria : Roles in Signaling and Adaptation to Various Environments. *International Journal of Molecular Sciences*, *19*(10). <https://doi.org/10.3390/ijms19102872>
- Jensen, M. H., Toft, K. N., David, G., Havelund, S., Pérez, J., & Vestergaard, B. (2010). Time-resolved SAXS measurements facilitated by online HPLC buffer exchange. *Journal of Synchrotron Radiation*, *17*(6), 769-773. <https://doi.org/10.1107/S0909049510030372>
- Jia, J., Lunin, V. V., Sauvé, V., Huang, L.-W., Matte, A., & Cygler, M. (2002). Crystal structure of the YciO protein from *Escherichia coli*. *Proteins*, *49*(1), 139-141. <https://doi.org/10.1002/prot.10178>
- Jia, W., & Higgs, P. G. (2008). Codon Usage in Mitochondrial Genomes : Distinguishing Context-Dependent Mutation from Translational Selection. *Molecular Biology and Evolution*, *25*(2), 339-351. <https://doi.org/10.1093/molbev/msm259>
- Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., & Pütz, J. (2009). tRNADB 2009 : Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research*, *37*(Database issue), D159-D162. <https://doi.org/10.1093/nar/gkn772>
- Kahn, A. (1989). Le code génétique et la spécificité des ARN de transfert. *Med Sci (Paris)*, *Vol. 5, N° 2; p.119-120*. Repéré à



- [http://www.ipubli.inserm.fr/bitstream/handle/10608/3928/MS\\_1989\\_2\\_119.pdf?sequence=2&isAllowed=y](http://www.ipubli.inserm.fr/bitstream/handle/10608/3928/MS_1989_2_119.pdf?sequence=2&isAllowed=y)
- Kahn, A. (2003). L'hélice de la vie. *Médecine/sciences*, 19(4), 491-495.  
<https://doi.org/10.1051/medsci/2003194491>
- Kapur, M., & Ackerman, S. L. (2018). mRNA Translation Gone Awry : Translation Fidelity and Neurological Disease. *Trends in Genetics*, 34(3), 218-231.  
<https://doi.org/10.1016/j.tig.2017.12.007>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service : Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160-1166. <https://doi.org/10.1093/bib/bbx108>
- Kilpatrick, M. W., & Walker, R. T. (1981). The nucleotide sequence of the tRNAMet from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Research*, 9(17), 4387-4390. <https://doi.org/10.1093/nar/9.17.4387>
- Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J.-C., Ilan, L., Hofmann, K., ... Libri, D. (2006). Yeast homolog of a cancer-testis antigen defines a new transcription complex. *The EMBO Journal*, 25(15), 3576-3585.  
<https://doi.org/10.1038/sj.emboj.7601235>
- Kline, L. K., Fittler, F., & Hall, R. H. (1969). N6-(delta-2-isopentenyl) adenosine. Biosynthesis in transfer ribonucleic acid in vitro. *Biochemistry*, 8(11), 4361-4371.  
<https://doi.org/10.1021/bi00839a021>
- Knüppel, R., Christensen, R. H., Gray, F. C., Esser, D., Strauß, D., Medenbach, J., ... Ferreira-Cerca, S. (2018). Insights into the evolutionary conserved regulation of Rio ATPase activity. *Nucleic Acids Research*, 46(3), 1441-1456.  
<https://doi.org/10.1093/nar/gkx1236>
- Koonin, E. V. (2005). Systemic determinants of gene evolution and function. *Molecular Systems Biology*, 1, 2005.0021. <https://doi.org/10.1038/msb4100029>
- Körner, A., & Söll, D. (1974). N-(purin-6-ylcarbamoyl)threonine : Biosynthesis in vitro in transfer RNA by an enzyme purified from *Escherichia coli*. *FEBS letters*, 39(3), 301-306. [https://doi.org/10.1016/0014-5793\(74\)80135-3](https://doi.org/10.1016/0014-5793(74)80135-3)
- Krutyholowa, R., Zakrzewski, K., & Glatt, S. (2019). Charging the code—TRNA modification complexes. *Current Opinion in Structural Biology*, 55, 138-146.  
<https://doi.org/10.1016/j.sbi.2019.03.014>
- Kurland, C. G. (1992). Translational accuracy and the fitness of bacteria. *Annual Review of Genetics*, 26, 29-50. <https://doi.org/10.1146/annurev.ge.26.120192.000333>
- LaRonde, N. A. (2014). The Ancient Microbial RIO Kinases. *Journal of Biological Chemistry*, 289(14), 9488-9492. <https://doi.org/10.1074/jbc.R113.538090>
- Lauhon, C. T. (2012). Mechanism of N6-Threonylcarbamoyladenonsine (t<sup>6</sup>A) Biosynthesis : Isolation and Characterization of the Intermediate Threonylcarbamoyl-AMP. *Biochemistry*, 51(44), 8950-8963. <https://doi.org/10.1021/bi301233d>
- Lei, T., Yang, J., Zheng, L., Markowski, T., Witthuhn, B. A., & Ji, Y. (2012). The Essentiality of Staphylococcal Gcp Is Independent of Its Repression of Branched-Chain Amino Acids Biosynthesis. *PLoS ONE*, 7(10), e46836.  
<https://doi.org/10.1371/journal.pone.0046836>
- Lepault, J., & Dubochet, J. (1986). Electron microscopy of frozen hydrated specimens : Preparation and characteristics. *Methods in Enzymology*, 127, 719-730.  
[https://doi.org/10.1016/0076-6879\(86\)27056-1](https://doi.org/10.1016/0076-6879(86)27056-1)
- Lescrinier, E., Nauwelaerts, K., Zanier, K., Poesen, K., Sattler, M., & Herdewijn, P. (2006). The naturally occurring N6-threonyl adenine in anticodon loop of

- Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Research*, 34(10), 2878-2886. <https://doi.org/10.1093/nar/gkl081>
- Leulliot, N., Chaillet, M., Durand, D., Ulryck, N., Blondeau, K., & van Tilbeurgh, H. (2008). Structure of the Yeast tRNA m7G Methylation Complex. *Structure*, 16(1), 52-61. <https://doi.org/10.1016/j.str.2007.10.025>
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., ... Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6), 584-590. <https://doi.org/10.1038/nmeth.2472>
- Lin, C. A., Ellis, S. R., & True, H. L. (2010). The Sua5 Protein Is Essential for Normal Translational Regulation in Yeast. *Molecular and Cellular Biology*, 30(1), 354-363. <https://doi.org/10.1128/MCB.00754-09>
- Lin, H., Miyauchi, K., Harada, T., Okita, R., Takeshita, E., Komaki, H., ... Suzuki, T. (2018). CO2-sensitive tRNA modification associated with human mitochondrial disease. *Nature Communications*, 9. <https://doi.org/10.1038/s41467-018-04250-4>
- Littlefield, J. W., Keller, E. B., Gross, J., & Zamecnik, P. C. (1955). Studies on Cytoplasmic Ribonucleoprotein Particles from the Liver of the Rat. *Journal of Biological Chemistry*, 217(1), 111-124. Repéré à <http://www.jbc.org/content/217/1/111>
- Liu, Y.-Y., He, M.-H., Liu, J.-C., Lu, Y.-S., Peng, J., & Zhou, J.-Q. (2018). Yeast KEOPS complex regulates telomere length independently of its t6A modification function. *Journal of Genetics and Genomics*, 45(5), 247-257. <https://doi.org/10.1016/j.jgg.2018.03.004>
- Lopreiato, R., Facchin, S., Sartori, G., Arrigoni, G., Casonato, S., Ruzzene, M., ... Carignani, G. (2004). Analysis of the interaction between piD261/Bud32, an evolutionarily conserved protein kinase of Saccharomyces cerevisiae, and the Grx4 glutaredoxin. *Biochemical Journal*, 377(2), 395-405. <https://doi.org/10.1042/bj20030638>
- Lorenz, C., Lünse, C. E., & Mörl, M. (2017). tRNA Modifications : Impact on Structure and Thermal Adaptation. *Biomolecules*, 7(2). <https://doi.org/10.3390/biom7020035>
- Luthra, A., Paranagama, N., Swinehart, W., Bayooz, S., Phan, P., Quach, V., ... Swairjo, M. A. (2019). Conformational communication mediates the reset step in t6A biosynthesis. *Nucleic Acids Research*, 47(12), 6551-6567. <https://doi.org/10.1093/nar/gkz439>
- Luthra, A., Swinehart, W., Bayooz, S., Phan, P., Stec, B., Iwata-Reuyl, D., & Swairjo, M. A. (2018). Structure and mechanism of a bacterial t6A biosynthesis system. *Nucleic Acids Research*, 46(3), 1395-1411. <https://doi.org/10.1093/nar/gkx1300>
- Machnicka, Milanowska, Oglou, Purta, Kurkowska1, Olchowik, ... Grosjean, H. (2013). MODOMICS: a database of RNA modification pathways—2013 update, pp. D262–D267. Repéré à doi:10.1093/nar/gks1007
- Magasanik, B., & Chargaff, E. (1951). Studies on the structure of ribonucleic acids. *Biochimica et Biophysica Acta*, 7, 396-412. [https://doi.org/10.1016/0006-3002\(51\)90043-1](https://doi.org/10.1016/0006-3002(51)90043-1)
- Mannella, C. A., & Frank, J. (1984). Negative staining characteristics of arrays of mitochondrial pore protein : Use of correspondence analysis to classify different staining patterns. *Ultramicroscopy*, 13(1-2), 93-102. [https://doi.org/10.1016/0304-3991\(84\)90060-3](https://doi.org/10.1016/0304-3991(84)90060-3)
- Mao, D. Y. L., Neculai, D., Downey, M., Orlicky, S., Haffani, Y. Z., Ceccarelli, D. F., ... Sicheri, F. (2008). Atomic Structure of the KEOPS Complex : An Ancient Protein Kinase-Containing Molecular Machine. *Molecular Cell*, 32(2), 259-275. <https://doi.org/10.1016/j.molcel.2008.10.002>

- Marquet, R., Isel, Ehresmann, & Ehresmann. (1995). TRNAs as primer of reverse transcriptases, 12.
- Meier, F., Suter, B., Grosjean, H., Keith, G., & Kubli, E. (1985). Queuosine modification of the wobble base in tRNA<sup>His</sup> influences « in vivo » decoding properties. *The EMBO Journal*, 4(3), 823-827. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC554263/>
- Melton, D. A., Cortese, R., de Robertis, E. M., Trendelenburg, M. F., & Gurdon, J. B. (1980). Gene injection into amphibian oocytes. *Results and Problems in Cell Differentiation*, 11, 8-14. [https://doi.org/10.1007/978-3-540-38267-6\\_2](https://doi.org/10.1007/978-3-540-38267-6_2)
- Meng, F.-L., Chen, X.-F., Hu, Y., Tang, H.-B., Dang, W., & Zhou, J.-Q. (2010). Sua5p is required for telomere recombination in *Saccharomyces cerevisiae*. *Cell Research*, 20(4), 495-498. <https://doi.org/10.1038/cr.2010.34>
- Miyauchi, K., Kimura, S., & Suzuki, T. (2013). A cyclic form of N<sup>6</sup>-threonylcarbamoyladenine as a widely distributed tRNA hypermodification. *Nature Chemical Biology*, 9(2), 105-111. <https://doi.org/10.1038/nchembio.1137>
- Mohler, K., Mann, R., Kyle, A., Reynolds, N., & Ibba, M. (2017). Aminoacyl-tRNA quality control is required for efficient activation of the TOR pathway regulator Gln3p. *RNA Biology*, 15(4-5), 594-603. <https://doi.org/10.1080/15476286.2017.1379635>
- Moore, P. B., & Asano, K. (1966). Divalent cation-dependent binding of messenger to ribosomal RNA. *Journal of Molecular Biology*, 18(1), 21-37. [https://doi.org/10.1016/s0022-2836\(66\)80073-6](https://doi.org/10.1016/s0022-2836(66)80073-6)
- Morin, A., Auxilien, S., Senger, B., Tewari, R., & Grosjean, H. (1998). Structural requirements for enzymatic formation of threonylcarbamoyladenine (t<sup>6</sup>A) in tRNA : An in vivo study with *Xenopus laevis* oocytes. *RNA*, 4(1), 24-37. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1369593/>
- Motorin, Y., Bec, G., Tewari, R., & Grosjean, H. (1997). Transfer RNA recognition by the *Escherichia coli* delta2-isopentenyl-pyrophosphate:tRNA delta2-isopentenyl transferase : Dependence on the anticodon arm structure. *RNA*, 3(7), 721-733. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1369520/>
- Motorin, Yuri, & Helm, M. (2010). TRNA Stabilization by Modified Nucleotides. *Biochemistry*, 49(24), 4934-4944. <https://doi.org/10.1021/bi100408z>
- Murphy, F. V., Ramakrishnan, V., Malkiewicz, A., & Agris, P. F. (2004). The role of modifications in codon discrimination by tRNA<sup>Lys</sup>UUU. *Nature Structural & Molecular Biology*, 11(12), 1186-1191. <https://doi.org/10.1038/nsmb861>
- Na, J. G., Pinto, I., & Hampsey, M. (1992). Isolation and characterization of SUA5, a novel gene required for normal growth in *Saccharomyces cerevisiae*. *Genetics*, 131(4), 791-801.
- Nakamura, A., Wang, D., & Komatsu, Y. (2018). Molecular mechanism of substrate recognition and specificity of tRNA<sup>His</sup> guanylyltransferase during nucleotide addition in the 3'-5' direction. *RNA*, 24(11), 1583-1593. <https://doi.org/10.1261/rna.067330.118>
- Nakamura, Y., Ito, K., & Isaksson, L. A. (1996). Emerging Understanding of Translation Termination. *Cell*, 87(2), 147-150. [https://doi.org/10.1016/S0092-8674\(00\)81331-8](https://doi.org/10.1016/S0092-8674(00)81331-8)
- Naor, A., Thiaville, P. C., Altman-Price, N., Cohen-Or, I., Allers, T., de Crécy-Lagard, V., & Gophna, U. (2012). A Genetic Investigation of the KEOPS Complex in Halophilic Archaea. *PLoS ONE*, 7(8), e43013. <https://doi.org/10.1371/journal.pone.0043013>
- Nguyen, H.-A., El Khoury, T., Guiral, S., Laaberki, M.-H., Candusso, M.-P., Galisson, F., ... Jault, J.-M. (2017). Expanding the Kinome World : A New Protein Kinase Family

- Widely Conserved in Bacteria. *Journal of Molecular Biology*, 429(20), 3056-3074.  
<https://doi.org/10.1016/j.jmb.2017.08.016>
- Nicastro, G., Taylor, I. A., & Ramos, A. (2015). KH–RNA interactions : Back in the groove. *Current Opinion in Structural Biology*, 30, 63-70.  
<https://doi.org/10.1016/j.sbi.2015.01.002>
- Nichols, C. E., Lamb, H. K., Thompson, P., Omari, K. E., Lockyer, M., Charles, I., ... 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 Science*, 22(5), 628-640.  
<https://doi.org/10.1002/pro.2247>
- Nirenberg, M., & Leder, P. (1964). RNA CODEWORDS AND PROTEIN SYNTHESIS. THE EFFECT OF TRINUCLEOTIDES UPON THE BINDING OF SRNA TO RIBOSOMES. *Science (New York, N.Y.)*, 145(3639), 1399-1407.  
<https://doi.org/10.1126/science.145.3639.1399>
- Nishikura, K., & De Robertis, E. M. (1981). RNA processing in microinjected *Xenopus* oocytes. Sequential addition of base modifications in the spliced transfer RNA. *Journal of Molecular Biology*, 145(2), 405-420. [https://doi.org/10.1016/0022-2836\(81\)90212-6](https://doi.org/10.1016/0022-2836(81)90212-6)
- Oberto, J., Breuil, N., Hecker, A., Farina, F., Brochier-Armanet, C., Culetto, E., & Forterre, P. (2009). Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance. *Nucleic Acids Research*, 37(16), 5343-5352. <https://doi.org/10.1093/nar/gkp557>
- Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., & Dunker, A. K. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins*, 53 Suppl 6, 566-572.  
<https://doi.org/10.1002/prot.10532>
- Onodera, T., Satoh, K., Ohta, T., & Narumi, I. (2013). *Deinococcus radiodurans* YgjD and YeaZ are involved in the repair of DNA cross-links. *Extremophiles*, 17(1), 171-179.  
<https://doi.org/10.1007/s00792-012-0506-4>
- Pal, S., Ganesan, K., & Eswaran, S. (2018). Chemical Crosslinking-Mass Spectrometry (CXL-MS) for Proteomics, Antibody-Drug Conjugates (ADCs) and Cryo-Electron Microscopy (cryo-EM). *IUBMB Life*, 70(10), 947-960.  
<https://doi.org/10.1002/iub.1916>
- Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J., & Stroud, R. M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex : RNA recognition through a combination of rigid docking and induced fit. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), 12648-12653.  
<https://doi.org/10.1073/pnas.2135585100>
- Pang, Y. L. J., Poruri, K., & Martinis, S. A. (2014). tRNA synthetase : TRNA aminoacylation and beyond: Aminoacyl-tRNA synthetases. *Wiley Interdisciplinary Reviews : RNA*, 5(4), 461-480. <https://doi.org/10.1002/wrna.1224>
- Parthasarathy, R., Ginell, S. L., De, N. C., & Chheda, G. B. (1978). Conformation of N4-acetylcytidine, a modified nucleoside of tRNA, and stereochemistry of codon-anticodon interaction. *Biochemical and Biophysical Research Communications*, 83(2), 657-663. [https://doi.org/10.1016/0006-291X\(78\)91040-9](https://doi.org/10.1016/0006-291X(78)91040-9)
- Parthasarathy, R., Ohrt, J. M., & Chheda, G. B. (1974). Conformation and possible role of hypermodified nucleosides adjacent to 3'-end of anticodon in tRNA : N-(purin-6-ylcarbonyl)-L-threonine riboside. *Biochemical and Biophysical Research Communications*, 60(1), 211-218. [https://doi.org/10.1016/0006-291X\(74\)90193-4](https://doi.org/10.1016/0006-291X(74)90193-4)

- Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Bräuer, U., Fandrich, U., ... Stubbs, M. T. (2012). The *O*-Carbamoyltransferase TobZ Catalyzes an Ancient Enzymatic Reaction. *Angewandte Chemie International Edition*, 51(17), 4046-4052. <https://doi.org/10.1002/anie.201108896>
- Paschos, A., Bauer, A., Zimmermann, A., Zehelein, E., & Böck, A. (2002). HypF, a carbamoyl phosphate-converting enzyme involved in [NiFe] hydrogenase maturation. *The Journal of Biological Chemistry*, 277(51), 49945-49951. <https://doi.org/10.1074/jbc.M204601200>
- Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., ... Basta, T. (2013). In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Research*, 41(3), 1953-1964. <https://doi.org/10.1093/nar/gks1287>
- Perrochia, L., Guetta, D., Hecker, A., Forterre, P., & Basta, T. (2013). Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification. *Nucleic Acids Research*, 41(20), 9484-9499. <https://doi.org/10.1093/nar/gkt720>
- Petkun, S., Shi, R., Li, Y., Asinas, A., Munger, C., Zhang, L., ... Cygler, M. (2011). Structure of Hydrogenase Maturation Protein HypF with Reaction Intermediates Shows Two Active Sites. *Structure*, 19(12), 1773-1783. <https://doi.org/10.1016/j.str.2011.09.023>
- Phelps, S. S., Jerinic, O., & Joseph, S. (2002). Universally Conserved Interactions between the Ribosome and the Anticodon Stem-Loop of A Site tRNA Important for Translocation. *Molecular Cell*, 10(4), 799-807. [https://doi.org/10.1016/S1097-2765\(02\)00686-X](https://doi.org/10.1016/S1097-2765(02)00686-X)
- Pichard-Kostuch, A., Zhang, W., Liger, D., Daugeron, M.-C., Létoquart, J., Li de la Sierra-Gallay, I., ... Basta, T. (2018). Structure–function analysis of Sua5 protein reveals novel functional motifs required for the biosynthesis of the universal t<sup>6</sup>A tRNA modification. *RNA*, 24(7), 926-938. <https://doi.org/10.1261/rna.066092.118>
- Pierrel, F., Björk, G. R., Fontecave, M., & Atta, M. (2002). Enzymatic Modification of tRNAs MiaB IS AN IRON-SULFUR PROTEIN. *Journal of Biological Chemistry*, 277(16), 13367-13370. <https://doi.org/10.1074/jbc.C100609200>
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., ... Adamiak, R. W. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, 40(14), e112-e112. <https://doi.org/10.1093/nar/gks339>
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server : 2018 update. *Nucleic Acids Research*, 46(W1), W200-W204. <https://doi.org/10.1093/nar/gky448>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017a). cryoSPARC : Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), 290-296. <https://doi.org/10.1038/nmeth.4169>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017b). cryoSPARC : Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), 290-296. <https://doi.org/10.1038/nmeth.4169>
- Pütz, J., Dupuis, B., Sissler, M., & Florentz, C. (2007). Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. *RNA*, 13(8), 1184-1190. <https://doi.org/10.1261/rna.588407>
- Quevillon-Cheruel, S., Liger, D., Leulliot, N., Graille, M., Poupon, A., De la Sierra-Gallay, I. L., ... Van Tilbeurgh, H. (2004). The Paris-Sud yeast structural genomics pilot-project : From structure to function. *Biochimie*, 86(9-10), 617-623. <https://doi.org/10.1016/j.biochi.2004.09.013>

- Quigley, G. J., & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science (New York, N.Y.)*, *194*(4267), 796-806. <https://doi.org/10.1126/science.790568>
- Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell*, *108*(4), 557-572. [https://doi.org/10.1016/s0092-8674\(02\)00619-0](https://doi.org/10.1016/s0092-8674(02)00619-0)
- Ramos, J., & Fu, D. (2019). The emerging impact of tRNA modifications in the brain and nervous system. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, *1862*(3), 412-428. <https://doi.org/10.1016/j.bbagr.2018.11.007>
- Ramsay, E. P., & Vannini, A. (2018). Structural rearrangements of the RNA polymerase III machinery during tRNA transcription initiation. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, *1861*(4), 285-294. <https://doi.org/10.1016/j.bbagr.2017.11.005>
- Reddy, D. M., Crain, P. F., Edmonds, C. G., Gupta, R., Hashizume, T., Stetter, K. O., ... McCloskey, J. A. (1992). Structure determination of two new amino acid-containing derivatives of adenosine from tRNA of thermophilic bacteria and archaea. *Nucleic Acids Research*, *20*(21), 5607-5615. <https://doi.org/10.1093/nar/20.21.5607>
- Redko, Y., Li de la Sierra-Gallay, I., & Condon, C. (2007). When all's zed and done : The structure and function of RNase Z in prokaryotes. *Nature Reviews Microbiology*, *5*(4), 278-286. <https://doi.org/10.1038/nrmicro1622>
- Reece, J. B., & Campbell, N. A. (Éds). (2011). *Campbell biology* (9th ed). Boston: Benjamin Cummings / Pearson.
- Rich, A., & RajBhandary, U. L. (1976). Transfer RNA : Molecular Structure, Sequence, and Properties. *Annual Review of Biochemistry*, *45*(1), 805-860. <https://doi.org/10.1146/annurev.bi.45.070176.004105>
- Rich, Alexander, & Kim, S. H. (1977). The Three-dimensional Structure of Transfer RNA. *SCIENTIFIC AMERICAN*, *12*.
- Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., & Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, *250*(467), 546-551. <https://doi.org/10.1038/250546a0>
- Rojas-Benítez, D., Eggers, C., & Glavic, A. (2017). Modulation of the Proteostasis Machinery to Overcome Stress Caused by Diminished Levels of t6A-Modified tRNAs in *Drosophila*. *Biomolecules*, *7*(1). <https://doi.org/10.3390/biom7010025>
- Rozov, A., Westhof, E., Yusupov, M., & Yusupova, G. (2016). The ribosome prohibits the G•U wobble geometry at the first position of the codon–anticodon helix. *Nucleic Acids Research*, *44*(13), 6434-6441. <https://doi.org/10.1093/nar/gkw431>
- Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., & Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins*, *23*(3), 318-326. <https://doi.org/10.1002/prot.340230306>
- Salinas-Giegé, T., Giegé, R., & Giegé, P. (2015). tRNA Biology in Mitochondria. *International Journal of Molecular Sciences*, *16*(3), 4518-4559. <https://doi.org/10.3390/ijms16034518>
- Scheres, S. H. W. (2012). RELION : Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519-530. <https://doi.org/10.1016/j.jsb.2012.09.006>
- Scheres, S. H. W. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology*, *189*(2), 114-122. <https://doi.org/10.1016/j.jsb.2014.11.010>
- Scheres, S. H. W. (n.d.). Single-particle processing in relion-3., 72.

- Schimmel, P. (2018). The emerging complexity of the tRNA world : Mammalian tRNAs beyond protein synthesis. *Nature Reviews Molecular Cell Biology*, 19(1), 45-58. <https://doi.org/10.1038/nrm.2017.77>
- Schneider, T. D., & Stephens, R. M. (1990). Sequence logos : A new way to display consensus sequences. *Nucleic Acids Research*, 18(20), 6097-6100. <https://doi.org/10.1093/nar/18.20.6097>
- Schweizer, M. P., Chheda, G. B., Baczynskyj, L., & Hall, R. H. (1969). Aminoacyl nucleosides. VII. N-(purin-6-ylcarbamoyl)threonine. A new component of transfer ribonucleic acid. *Biochemistry*, 8(8), 3283-3289. <https://doi.org/10.1021/bi00836a023>
- Schweizer, U., Bohleber, S., & Fradejas-Villar, N. (2017). The modified base isopentenyladenosine and its derivatives in tRNA. *RNA Biology*, 14(9), 1197-1208. <https://doi.org/10.1080/15476286.2017.1294309>
- Shepherd, J., & Ibba, M. (2013a). Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEBS letters*, 587(18), 2895-2904. <https://doi.org/10.1016/j.febslet.2013.07.036>
- Shepherd, J., & Ibba, M. (2013b). Lipid II-independent trans Editing of Mischarged tRNAs by the Penicillin Resistance Factor MurM. *The Journal of Biological Chemistry*, 288(36), 25915-25923. <https://doi.org/10.1074/jbc.M113.479824>
- Shigi, N., Suzuki, T., Terada, T., Shirouzu, M., Yokoyama, S., & Watanabe, K. (2006). Temperature-dependent biosynthesis of 2-thioribothymidine of *Thermus thermophilus* tRNA. *The Journal of Biological Chemistry*, 281(4), 2104-2113. <https://doi.org/10.1074/jbc.M510771200>
- Shomura, Y., & Higuchi, Y. (2012). Structural Basis for the Reaction Mechanism of S-Carbamoylation of HypE by HypF in the Maturation of [NiFe]-Hydrogenases. *Journal of Biological Chemistry*, 287(34), 28409-28419. <https://doi.org/10.1074/jbc.M112.387134>
- Sigworth, F. J. (2016). Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1), 57-67. <https://doi.org/10.1093/jmicro/dfv370>
- Sokołowski, M., Klassen, R., Bruch, A., Schaffrath, R., & Glatt, S. (2018). Cooperativity between different tRNA modifications and their modification pathways. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, 1861(4), 409-418. <https://doi.org/10.1016/j.bbagr.2017.12.003>
- Sonawane, K. D., & Sambhare, S. B. (2015). The influence of hypermodified nucleosides lysidine and t<sup>6</sup>A to recognize the AUA codon instead of AUG : A molecular dynamics simulation study. *Integrative Biology*, 7(11), 1387-1395. <https://doi.org/10.1039/C5IB00058K>
- Sprinzi, M, Horn, C., Brown, M., Ioudovitch, A., & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 26(1), 148-153. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC147216/>
- Sprinzi, Mathias, & Vassilenko, K. S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 33(Database issue), D139-140. <https://doi.org/10.1093/nar/gki012>
- Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E. V., Karzai, A. W., & Sternglanz, R. (2011). The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t<sup>6</sup>A. *The EMBO Journal*, 30(5), 873-881. <https://doi.org/10.1038/emboj.2010.343>
- Stuart, J. W., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., & Agris, P. F. (2000). Functional Anticodon Architecture of Human tRNA<sup>Lys3</sup> Includes Disruption of Intraloop Hydrogen Bonding by the Naturally Occurring Amino Acid

- Modification, t<sup>6</sup>A<sup>†</sup>. *Biochemistry*, 39(44), 13396-13404.  
<https://doi.org/10.1021/bi0013039>
- Sundaram, M., Durant, P. C., & Davis, D. R. (2000). Hypermodified nucleosides in the anticodon of tRNA<sup>Lys</sup> stabilize a canonical U-turn structure. *Biochemistry*, 39(41), 12575-12584. <https://doi.org/10.1021/bi0014655>
- Svergun, D., Barberato, C., & Koch, M. H. J. (1995). CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, 28(6), 768-773.  
<https://doi.org/10.1107/S0021889895007047>
- Swinehart, W. E., Deutsch, C. W., Sarachan, K. L., Luthra, A., Bacusmo, J. M., Crécy-Lagard, V. de, ... Iwata-Reuyl, D. (2020). Specificity in the Biosynthesis of the Universal tRNA Nucleoside N6-Threonylcarbamoyl Adenosine (t6A)—TsaD is the Gatekeeper. *RNA*, rna.075747.120. <https://doi.org/10.1261/rna.075747.120>
- Takemura, S., Murakami, M., & Miyazaki, M. (1969). Nucleotide Sequence of Isoleucine Transfer RNA from *Torulopsis utilis*. *The Journal of Biochemistry*, 65(3), 489-491. <https://doi.org/10.1093/oxfordjournals.jbchem.a129040>
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007). EMAN2 : An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1), 38-46. <https://doi.org/10.1016/j.jsb.2006.05.009>
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database : A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1), 33-36. <https://doi.org/10.1093/nar/28.1.33>
- Teplova, M., Tereshko, V., Egli, M., Sanishvili, R., Joachimiak, A., Bushueva, T., & Anderson, W. F. (2000). The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. *Protein Science*, 9(12), 2557-2566. <https://doi.org/10.1110/ps.9.12.2557>
- Tepljakov, A., Obmolova, G., Tordova, M., Thanki, N., Bonander, N., Eisenstein, E., ... Gilliland, G. L. (2002). Crystal structure of the YjeE protein from *Haemophilus influenzae* : A putative Atpase involved in cell wall synthesis. *Proteins*, 48(2), 220-226. <https://doi.org/10.1002/prot.10114>
- Thiaville, P. C., El Yacoubi, B., Köhrer, C., Thiaville, J. J., Deutsch, C., Iwata-Reuyl, D., ... de Crécy-Lagard, V. (2015). Essentiality of threonylcarbamoyladenine (t(6)A), a universal tRNA modification, in bacteria. *Molecular Microbiology*, 98(6), 1199-1221. <https://doi.org/10.1111/mmi.13209>
- Thiaville, P. C., El Yacoubi, B., Perrochia, L., Hecker, A., Prigent, M., Thiaville, J. J., ... de Crécy-Lagard, V. (2014). Cross Kingdom Functional Conservation of the Core Universally Conserved Threonylcarbamoyladenine tRNA Synthesis Enzymes. *Eukaryotic Cell*, 13(9), 1222-1231. <https://doi.org/10.1128/EC.00147-14>
- Thiaville, P. C., Iwata-Reuyl, D., & de Crécy-Lagard, V. (2014). Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t(6)A), a universal modification of tRNA. *RNA biology*, 11(12), 1529-1539. <https://doi.org/10.4161/15476286.2014.992277>
- Thiaville, P. C., Legendre, R., Rojas-Benítez, D., Baudin-Baillieu, A., Hatin, I., Chalancon, G., ... de Crécy-Lagard, V. (2016). Global translational impacts of the loss of the tRNA modification t6A in yeast. *Microbial Cell (Graz, Austria)*, 3(1), 29-45. <https://doi.org/10.15698/mic2016.01.473>
- Tomikawa, C., Yokogawa, T., Kanai, T., & Hori, H. (2010). N7-Methylguanine at position 46 (m7G46) in tRNA from *Thermus thermophilus* is required for cell viability at high temperatures through a tRNA modification network. *Nucleic Acids Research*, 38(3), 942-957. <https://doi.org/10.1093/nar/gkp1059>



- Tominaga, T., Kobayashi, K., Ishii, R., Ishitani, R., & Nureki, O. (2014). Structure of *Saccharomyces cerevisiae* mitochondrial Qri7 in complex with AMP. *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 8), 1009-1014. <https://doi.org/10.1107/S2053230X14014046>
- Torres, A. G., Battle, E., & Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. *Trends in Molecular Medicine*, 20(6), 306-314. <https://doi.org/10.1016/j.molmed.2014.01.008>
- Towns, W. L., & Begley, T. J. (2012). Transfer RNA methyltransferases and their corresponding modifications in budding yeast and humans : Activities, predications, and potential roles in human health. *DNA and cell biology*, 31(4), 434-454. <https://doi.org/10.1089/dna.2011.1437>
- Tsutomu Suzuki, Asutaka Nagao, and Takeo Suzuki. (2011). Human Mitochondrial tRNAs : Biogenesis, Function, Structural Aspects, and Diseases.
- Valverde, R., Edwards, L., & Regan, L. (2008). Structure and function of KH domains. *The FEBS Journal*, 275(11), 2712-2726. <https://doi.org/10.1111/j.1742-4658.2008.06411.x>
- Väre, V., Eruysal, E., Narendran, A., Sarachan, K., & Agris, P. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules*, 7(4), 29. <https://doi.org/10.3390/biom7010029>
- Velázquez-Muriel, J. A., Sorzano, C. O. S., Fernández, J. J., & Carazo, J. M. (2003). A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment. *Ultramicroscopy*, 96(1), 17-35. [https://doi.org/10.1016/S0304-3991\(02\)00377-7](https://doi.org/10.1016/S0304-3991(02)00377-7)
- Vendeix, F. A. P., Murphy, F. V., Cantara, W. A., Leszczyńska, G., Gustilo, E. M., Sproat, B., ... Agris, P. F. (2012). Human tRNALys3UUU Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto–Enol Tautomerism. *Journal of Molecular Biology*, 416(4), 467-485. <https://doi.org/10.1016/j.jmb.2011.12.048>
- Vénien-Bryan, C., Li, Z., Vuillard, L., & Boutin, J. A. (2017). Cryo-electron microscopy and X-ray crystallography : Complementary approaches to structural biology and drug discovery. *Acta Crystallographica. Section F, Structural Biology Communications*, 73(Pt 4), 174-183. <https://doi.org/10.1107/S2053230X17003740>
- Vogel, A., Schilling, O., Späth, B., & Marchfelder, A. (2005). The tRNase Z family of proteins : Physiological functions, substrate specificity and structural properties. *Biological Chemistry*, 386(12), 1253-1264. <https://doi.org/10.1515/BC.2005.142>
- Voigts-Hoffmann, F., Hengesbach, M., Kobitski, A. Y., van Aerschot, A., Herdewijn, P., Nienhaus, G. U., & Helm, M. (2007). A Methyl Group Controls Conformational Equilibrium in Human Mitochondrial tRNALys, 2.
- Wagner, T., & Raunser, S. (2020). The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows. *Communications Biology*, 3(1), 61. <https://doi.org/10.1038/s42003-020-0790-y>
- Waldron, C., & Lacroute, F. (1975). Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *Journal of Bacteriology*, 122(3), 855-865. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC246135/>
- Walker, S. C., & Engelke, D. R. (2006). Ribonuclease P : The Evolution of an Ancient RNA Enzyme. *Critical reviews in biochemistry and molecular biology*, 41(2), 77-102. <https://doi.org/10.1080/10409230600602634>
- Wan, L. C. K., Maisonneuve, P., Szilard, R. K., Lambert, J.-P., Ng, T. F., Manczyk, N., ... Sicheri, F. (2017). Proteomic analysis of the human KEOPS complex identifies

- C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acids Research*, 45(2), 805-817. <https://doi.org/10.1093/nar/gkw1181>
- Wan, L. C. K., Mao, D. Y. L., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., ... Sicheri, F. (2013). Reconstitution and characterization of eukaryotic N6-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Research*, 41(12), 6332-6346. <https://doi.org/10.1093/nar/gkt322>
- Wan, L. C. K., Pillon, M. C., Thevakumaran, N., Sun, Y., Chakrabartty, A., Guarné, A., ... Sicheri, F. (2016). Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t6A biosynthesis. *Nucleic Acids Research*, 44(14), 6971-6980. <https://doi.org/10.1093/nar/gkw542>
- Wang, P. Z. T., Prasad, C., Rodriguez Cuellar, C. I., & Filler, G. (2018). Nephrological and urological complications of homozygous c.974G>A (p.Arg325Gln) OSGEP mutations. *Pediatric Nephrology (Berlin, Germany)*, 33(11), 2201-2204. <https://doi.org/10.1007/s00467-018-4060-x>
- Watson, J. D. (1964). THE SYNTHESIS OF PROTEINS UPON RIBOSOMES. *Bulletin De La Societe De Chimie Biologique*, 46, 1399-1425.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738. <https://doi.org/10.1038/171737a0>
- Wei, F.-Y., & Tomizawa, K. (2012). Development of type 2 diabetes caused by a deficiency of a tRNA(lys) modification. *Islets*, 4(1), 71-73. <https://doi.org/10.4161/isl.18262>
- Weissenbach, J., & Grosjean, H. (1981). Effect of Threonylcarbamoyl Modification (t6A) in Yeast tRNAArgIII on Codon-Anticodon and Anticodon-Anticodon Interactions. *European Journal of Biochemistry*, 116(1), 207-213. <https://doi.org/10.1111/j.1432-1033.1981.tb05320.x>
- Witts, R. N., Hopson, E. C., Koballa, D. E., Van Boening, T. A., Hopkins, N. H., Patterson, E. V., & Nagan, M. C. (2013). Backbone-base interactions critical to quantum stabilization of transfer RNA anticodon structure. *The Journal of Physical Chemistry. B*, 117(25), 7489-7497. <https://doi.org/10.1021/jp400084p>
- Xiao, J., Li, H., Wen, S., & Hong, W. (2014). Concentrated biosynthesis of tobramycin by genetically engineered *Streptomyces tenebrarius*. *The Journal of General and Applied Microbiology*, 60(6), 256-261. <https://doi.org/10.2323/jgam.60.256>
- Yarian, C., Townsend, H., Czestkowski, W., Sochacka, E., Malkiewicz, A. J., Guenther, R., ... Agris, P. F. (2002). Accurate Translation of the Genetic Code Depends on tRNA Modified Nucleosides. *Journal of Biological Chemistry*, 277(19), 16391-16395. <https://doi.org/10.1074/jbc.M200253200>
- Yarus, M. (1979). The accuracy of translation. *Progress in Nucleic Acid Research and Molecular Biology*, 23, 195-225. [https://doi.org/10.1016/s0079-6603\(08\)60134-8](https://doi.org/10.1016/s0079-6603(08)60134-8)
- Yokoyama, S. (1987). Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. *Advances in Biophysics*, 23, 115-147. [https://doi.org/10.1016/0065-227X\(87\)90006-2](https://doi.org/10.1016/0065-227X(87)90006-2)
- Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., & Miyazawa, T. (1985). Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proceedings of the National Academy of Sciences of the United States of America*, 82(15), 4905-4909. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390466/>
- Yoshizawa, D. F., Henri Grosjean, Satoko. (2003, 1 novembre). Le ribosome : L'usine à protéines. *Pour la science N°313*. Pour la Science. Repéré à

<https://www.pourlascience.fr/sd/biologie-moleculaire/le-ribosome-lusine-a-proteines-5314.php>

- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., & Noller, H. F. (2001). Crystal Structure of the Ribosome at 5.5 Å Resolution, *292*, 14.
- Zagryadskaya, E. I., Kotlova, N., & Steinberg, S. V. (2004). Key Elements in Maintenance of the tRNA L-shape. *Journal of Molecular Biology*, *340*(3), 435-444.  
<https://doi.org/10.1016/j.jmb.2004.05.002>
- Zaitseva, L., Myers, R., & Fassati, A. (2006). TRNAs Promote Nuclear Import of HIV-1 Intracellular Reverse Transcription Complexes. *PLoS Biology*, *4*(10).  
<https://doi.org/10.1371/journal.pbio.0040332>
- Zamecnik, P. C., & Keller, E. B. (1954). Relation Between Phosphate Energy Donors and Incorporation of Labeled Amino Acids into Proteins. *Journal of Biological Chemistry*, *209*(1), 337-354. Repéré à <http://www.jbc.org/content/209/1/337>
- Zhang, W., Collinet, B., Graille, M., Daugeron, M.-C., Lazar, N., Libri, D., ... van Tilbeurgh, H. (2015). Crystal structures of the Gon7/Pcc1 and Bud32/Cgi121 complexes provide a model for the complete yeast KEOPS complex. *Nucleic Acids Research*, *43*(6), 3358-3372. <https://doi.org/10.1093/nar/gkv155>
- Zhang, W., Collinet, B., Perrochia, L., Durand, D., & van Tilbeurgh, H. (2015). The ATP-mediated formation of the YgjD–YeaZ–YjeE complex is required for the biosynthesis of tRNA t6A in Escherichia coli. *Nucleic Acids Research*, *43*(3), 1804-1817.  
<https://doi.org/10.1093/nar/gku1397>
- Zhou, C., & Huang, R. H. (2008). Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: Insight into tRNA recognition and reaction mechanism. *Proc Natl Acad Sci U S A*, *105*(42):16142-7.  
[doi:10.1073/pnas.0805680105](https://doi.org/10.1073/pnas.0805680105).



**Titre :** Etude structurale et fonctionnelle des systèmes de modification t<sup>6</sup>A des ARNt dans les trois domaines du vivant.

**Mots clés :** modification post-transcriptionnelle ARNt, biosynthèse t<sup>6</sup>A, thréonyl-carbamoyle, structure.

**Résumé :** La thréonylcarbamoylation de la base A37 des ARNt de type ANN (t<sup>6</sup>A) est universelle au sein des trois domaines du vivant car essentielle au bon fonctionnement de la cellule. La synthèse se catalyse en deux étapes, dont la première est assurée par la famille d'enzymes YrdC/Sua5, qui forme l'intermédiaire thréonyl-carbamoyle-AMP et la seconde par Kae1/OSGEP/Qri7/TsaD qui transfère le groupement thréonyl-carbamoyle du TC-AMP sur les ARNt substrats. La seconde famille d'enzymes requiert des partenaires protéiques différents selon le domaine du vivant. Le but de cette thèse était de comprendre les différents systèmes de biosynthèse du t<sup>6</sup>A au niveau moléculaire en utilisant des techniques biochimiques et structurales (la diffraction des rayons X (cristallographie), la diffusion des rayons X en solution (SAXS) et la cryo-microscopie électronique). La structure quaternaire du complexe TsaBDE de *Thermotoga maritima* a été résolue mettant en lumière les changements conformationnels de la sous-unité catalytique TsaD qui se traduit par une ouverture et une perturbation

partielle du site actif induits par la présence de TsaE. Combinés à des données biochimiques, il apparaît que ce remodelage est nécessaire au recyclage du site catalytique. Chez l'Homme, la structure du sous-complexe OSGEP-LAGE3-GON7 a été résolue mettant en évidence que GON7, intrinsèquement désordonnée en solution, se structure au contact de LAGE3. L'étude structurale du complexe humain par SAXS combinée aux études génétiques et biochimiques montrent que GON7 stabilise le complexe sous forme pentamérique. Ces données ont été indispensables pour mieux comprendre les bases moléculaires d'une maladie génétique rare (le syndrome de Galloway-Mowat) causée par des mutations dans toutes les protéines de la voie de synthèse du t<sup>6</sup>A. Enfin, une cinquième sous-unité a été identifiée comme potentielle homologue de GON7 chez les Archées. Son étude structurale montre que la structure est différente de celle de GON7. Cependant, cette protéine stabilise le complexe sous forme pentamérique, mimant l'effet de GON7 chez les eucaryotes.

**Title:** Functional and structural studies of t<sup>6</sup>A tRNA modification systems in the three domains of life.

**Key words:** post-transcriptional tRNA modification, biosynthesis t<sup>6</sup>A, threonyl-carbamoyl, protein structure.

**Abstract:** The t<sup>6</sup>A modification of the A37 position of ANN-type tRNAs is totally conserved in the 3 domains of life as it is essential to the general functioning of the cell. It is carried out in two catalytic steps. the YrdC/Sua5 enzyme family, first, synthesizes the threonyl-carbamoyl-AMP (TC-AMP) intermediate, then the Kae1/OSGEP/Qri7/TsaD enzyme family transfers the threonyl-carbamoyl moiety of TC-AMP onto the tRNA substrate. The enzymes that are responsible for the transfer step use different protein partners in bacteria, archaea and eukaryotes. The aim of this thesis was to understand at a molecular level these various threonylcarbamoylation systems using a combination of biochemical and biophysical techniques (crystallography, X-ray scattering in solution (SAXS) and cryo-electron microscopy). The structure of the ternary TsaBDE complex of *Thermotoga maritima* was solved, highlighting conformational changes of TsaD that result in its opening and the partial melting of its active site induced by its interaction with TsaE. Combined with

biochemical data, it was concluded that this conformational change seems necessary to promote multiple catalytic turnovers. In Humans, the structure of the OSGEP-LAGE3-GON7 KEOPS sub-complex has been solved, showing that GON7, which is intrinsically disordered in solution, becomes structured upon interaction with LAGE3. The study of the structural properties of the whole human complex showed that GON7 stabilizes a pentameric form of the complex. These data were indispensable for a better comprehension of the molecular basis of a rare genetic disease (Galloway-Mowat syndrome) caused by mutations in all subunits of the t<sup>6</sup>A biosynthesis proteins. Finally, a fifth subunit has been identified as the potential homologue of GON7 in Archaea. However, its structural study shows that the fold is different from GON7's. Nevertheless, this protein stabilizes the archaeal complex in a pentameric form, as GON7 does in Eukaryotes.