Biological and biochemical characterization of dihydrouridilation in bacterial ribosomal RNA

Sabrine Toubdji

Sorbonne Université

École doctorale 515 - Complexité du vivant

Laboratoire Adaptation biologique et vieillissement UMR8256 –Institut de Biologie Paris-Seine

Équipe de recherche Traduction eucaryote

Biological and biochemical characterization of dihydrouridylation in bacterial ribosomal RNA

Thèse de doctorat de Biologie

Par **Sabrine Toubdji**

Présentée et soutenue publiquement le 18 septembre 2024

Devant un jury composé de :

Directeurs de Thèse :
Damien Bregeon — Maître de conférences, Sorbonne Université
Djemel Hamdane — Directeur de Recherche, CNRS – Sorbonne Université

Présidente du jury :
Laure Teysset — Professeur, Sorbonne Université

Rapporteurs :
Marc Graille — Directeur de Recherche, Polytechnique
Jean-Jacques Diaz — Directeur de Recherche, Inserm - Université Claude Bernard Lyon 1

Examinateur.ice.s :
Isabelle Iost — Chargé de recherche hors classe - Université Bordeaux
Olivier Namy — Directeur de Recherche, CNRS – Université Paris-Saclay

Acknowledgements

I am grateful for the members of my PhD jury for taking the time to evaluate my work.

Dr. Marc Graille and **Dr. Jean-Jacques Diaz** for graciously accepting to review my manuscript, **Dr. Olivier Namy**, **Dr. Isabelle Ios**t and **Pr. Laure Teysset** for taking part in my jury as examiners.

Thanks to the members of my thesis committee for their helpful advice and discussions: **Dr. Zeynep Baharuglu**, **Pr. Nicolas Leulliot** for your input to the D-2449 project and **Dr. Marina Pinskaya** for the crucial role you played in my career, when along with Pr. Huber Becker to whom I am also grateful, you allowed my admittance to the wonderful RNA Master of Sorbonne Université, and the interest I developed in RNA modifications, and for meeting **Dr. Damien Brégeon**, who welcomed me in his lab.

"Thank you" doesn't even begin to express the gratitude I have for you, **Damien**. A series of wonderful and challenging adventures led me to your lab, and I am convinced that I could not have hoped for a better mentor. I am immensely grateful for the opportunity to learn from you and work alongside you for four years, which feels like just the blink of an eye! Beyond opening the door to exciting research, your patience and kindness were pivotal in helping me reach the finish line, especially when I was too overwhelmed to function. Thank you for always teaching me, even when I sent you work over the weekends, for trusting and encouraging me, and for putting up with my terrible morning person habits!

What to say to my thesis supervisor **Dr. Djemel Hamdane**? First and foremost, a huge thank you for taking me under your wing despite my background in genetics being far from chemistry. Thank you for your trust and patience. I am extremely grateful for all our formative discussions that changed my perspective on science and dynamics! Thank you for the constant flow of ideas you suggested. Thank you for broadening my thinking to encompass biochemistry and structural biology. You have been a model of an honest, creative, and dedicated scientist.

For both of you **Damien** and **Djemel** I will be forever grateful for the freedom you granted me in my research and pushing my limits to become the scientist I am today.

A heartfelt thank you to **Dr. Olivier Jean-Jean** for welcoming me into his laboratory and for his constant interest in my work. Our discussions have been a continuous source of inspiration and encouragement. I am also grateful for his support during challenging times and for being a model of the scholar I aspire to be.

A big thank you to all the ET lab members, the amazing team that I had the privilege to work with, Thanks **Beatrice** and **Verene** for your wise advice, and stimulating discussions, for being

so helpful and kind everytime I come to you with a question and for overall being cool and nice people to be around. Thank you, **Catherine** our lab engineer for your support and for initiating me in every way to the D-2449 project, and to former members **Roba**, **Sarah, Chakib** and **Eliane** the first interns I supervised for the pleasure of working with you, for being dedicated, perseverant, and nice students and also **Nathalie**, the former and maybe future member of the lab! I wish you all the best in your careers and personal life. **Claudia t**hanks for being a great labmate and friend and for being immensely positive hahaha! and caring from day one for the great trips and outings sharing pizzas and drinks. I am glad our friendship upgraded from labmate to real life! I have great hope you'll kill it this last year, but still wish you good fortune for the future, professionally and personally. Thank you to the B2A PhD friends that made my world a better place. **Hiba** for being a great coffee/Lunch break companion and partner, for our Italian and Algerian restaurant outings. I'm still waiting for the Lebanese! I wish you all the best in your PhD adventure, and for the future.

I am grateful to **Pr. Yuri Motorin** and **Dr. Virginie Marchand** for opening your lab to me and contributing to the advancement of our project and initiating me to the AlkAniline-Seq, and for the team members **Valerie Igel-Bourguignon, Quentin Thuillier, Kasturi Raorane** for making me feel very welcomed**.** A huge thank you **Pr. Mark Helm**, **Dr. Lea-Marie Vogt** for your immense help with the LC-MS, availability for discussion, and active collaboration. To **Dr. Mariette Matondo** and **Rayen Elj** for the proteomic experiments. A big thank you **Pr. Nicolas Leulliot** and **Bruno Bonettaz** for your precious help in the polysome profiling experiments and initiating me to Cryo-Em despite your busy schedule, best of luck for the remainder of your PhD. An extended thank you to all the people of the A2 (The Muchard team and the Weil team), I couldn't cite you all by name, but know that I had a great time working and growing alongside you all.

To finish, I am grateful to all my family for their never-ending support. To my big sister my cousin my friend **Radus** thanks for always being there since day one, for supporting me through this adventure, thanks for always looking out for me and putting up with my messy self, I wish you a bright future! To **Nabil** where do I start... Well, thanks for being the person that I don't even need to speak to in order to be understood since forever best friend! We overcame little and big challenges together through the years, and I'm happy to report, we made it! thank you for supporting me and putting up with the best, but also the worst version of myself, thanks for sharing this wonderful moment with thesis writing! for our crazy outings, burgers, sushi, crepes

au Nutella, movie sessions, and trips, were the brightest these last two years, no need to wait for brighter days, these are the best days of our lives, and I am so glad we found each other.

Enfin, je dédie ce travail à ma famille, à celle qui est là, loin, ou d'un autre monde, à mes grands-parents, mon **grand-père** la plus belle et inspirante personne que j'ai connu. Enfin à mes **parents**, merci de m'avoir toujours poussé à aller plus loin dans mes études, à votre manière et depuis le début, vous aurez toujours cru en moi, même quand moi-même j'ai douté, merci pour tout. Je te souhaite une belle et longue vie remplie de bonheur et santé à mes côtés **Maman**, et un éternel repos à toi **Papa**.

Abstract

This manuscript explores the complex world of ribosomal dihydrouridylation in *Escherichia coli*. Dihydrouridine (D), an RNA modification found primarily in tRNA and less frequently in rRNA and mRNA, plays a crucial role in RNA stability and function. During my PhD, I contributed to the development of a novel innovative RhoRT-PCR method for detecting dihydrouridine in rRNA, that allows me with other methods such as MALDI-TOF mass spectrometry and AlkAniline-Seq to identify the gene responsible for ribosomal dihydrouridine synthase (RdsA). During my PhD, I contributed to the biochemical characterization of this enzyme. The findings highlight the evolutionary conservation and prevalence of dihydrouridine, particularly its presence near the critical peptidyl transferase site in the 23S rRNA of *E. coli*. I also contributed to the biological significance of ribosomal dihydrouridylation and the role of dihydrouridine synthase RdsA in the translation process of the ribosome. Polysome profiling analysis and ribosome Cryo-EM studies were employed to investigate how dihydrouridine modification influences ribosomal function and cellular processes. The results suggest that D-2449 play significant roles in the structural dynamics and architectural stability of rRNA, thereby impacting translation efficiency and accuracy. Overall, these results emphasize the importance of dihydrouridine in RNA folding dynamics and its broader implications for cellular processes and evolutionary adaptations. This research not only enhances our understanding of RNA modifications but also opens new avenues for exploring the chemical mechanism that RdsA employs to reduce the U2449 residue, to understand how RdsA binds its substrate, and investigating the specificity of its enzymatic activity.

French Abstract

Ce manuscrit explore le monde complexe de la dihydrouridylation ribosomique chez *Escherichia coli*. La dihydrourdine (D), une modification de l'ARN principalement trouvée dans les ARNt et plus rarement dans les ARNr et ARNm, joue un rôle crucial dans la stabilité et la fonction de l'ARN. Au cours de mon doctorat, j'ai contribué au développement d'une méthode innovante de RhoRT-PCR pour détecter la dihydrourdine dans les ARNr. Cette méthode, combinée à d'autres techniques telles que la spectrométrie de masse MALDI-TOF et l'AlkAniline-Seq, m'ont permis d'identifier le gène responsable de la synthase de dihydrourdine ribosomique (RdsA). J'ai également contribué à la caractérisation biochimique de cette enzyme. Les résultats mettent en évidence la conservation évolutive et la prévalence de la dihydrourdine, en particulier sa présence près du centre peptidyl transférase dans l'ARNr 23S de *E. coli*. J'ai aussi étudié l'importance biologique de la dihydrouridylation ribosomique et dans le processus de traduction du ribosome. Des analyses de profilage des polysomes et des études de Cryo-EM sur les ribosomes ont été utilisées pour examiner comment la modification de la dihydrourdine influence la fonction ribosomique et les processus cellulaires. Les résultats suggèrent que D-2449 joue un rôle significatif dans la dynamique structurelle et la stabilité architecturale de l'ARNr, impactant ainsi l'efficacité et la précision de la traduction. En somme, ces résultats soulignent l'importance de la dihydrourdine dans la dynamique de repliement de l'ARN et ses implications plus larges pour les processus cellulaires et les adaptations évolutives. Cette recherche non seulement améliore notre compréhension des modifications de l'ARN, mais ouvre également de nouvelles voies pour explorer le mécanisme chimique employé par RdsA pour réduire le résidu U2449, comprendre comment RdsA se lie à son substrat, et investiguer la spécificité de son activité enzymatique.

Introduction

1. Preface

In the intricate tapestry of life on Earth, spanning Eukaryotes, Archaea, and Prokaryotes, fundamental organizational features unite all living organisms:
i. A protective membrane delineates internal from external environments.
ii. Deoxyribonucleic acid (DNA) encodes the genetic blueprint that defines an organism.
iii. Ribonucleic acid (RNA) serves versatile roles in transcription and translation processes.
iv. Proteins, synthesized through the translation of mRNA, execute diverse biological functions.

The origin of life remains a profound scientific inquiry, particularly the transition from simple chemistry to complex biology. Synthetic efforts to produce amino acids and nucleotides from prebiotic molecules like methane and ammonia have yielded partial success, underscoring the challenges in replicating early Earth conditions in laboratory settings.
The RNA world hypothesis proposes RNA as the primordial molecule preceding the DNA-protein paradigm. RNA's capacity to store genetic information, self-replicate, and catalyze reactions akin to modern enzymes supports this hypothesis.

Francis Crick's central dogma of molecular biology posits a unidirectional flow from DNA to mRNA and to protein, yet discoveries like reverse transcriptase and pervasive transcription challenge this linear model. The genome, especially in complex eukaryotes, features not only protein-coding genes but also non-coding RNA transcripts, unveiling new dimensions in gene regulation and cellular function.
Beyond the genome, the epitranscriptome, comprising post-transcriptional RNA modifications, influences RNA stability, translation efficiency, and cellular identity. Recent advancements in sequencing technologies have unveiled the role of RNA modifications in health and disease, offering insights into therapeutic strategies, including RNA-based vaccines.

In this thesis, I delve into the intricate world of ribosomal dihydrouridylation in *Escherichia coli*. Through innovative methodologies, genetic analyses, biochemical characterizations, and exploration of biological implications, this work aims to unravel the role of dihydrouridine in RNA folding dynamics and architectural stability, contributing to our broader understanding of cellular processes and evolutionary adaptations.

The following Introduction is divided into four main sections, starting with ribosome, detailing their structure and function, then defining modified nucleotide, exemplifying what are posttranscriptional modifications and their occurrence in biological systems. The third section will focus on posttranscriptional modifications in rRNA. Roles of modified nucleotides will be extensively discussed in this section. Finally, a focus on RNA modification dihydrouridine is presented with an emphasis on dihydrouridine synthases (flavoenzymes).

2. The ribosome

Translation is the fundamental mechanism that converts genetic information, encoded by messenger RNAs (mRNAs), into proteins. This process occurs within the ribosome, a complex composed of non-coding RNAs and ribosomal proteins, which has two primary functions: (i) decoding the mRNA and facilitating its interaction with transfer RNAs (tRNAs); (ii) forming peptide bonds between the amino acids carried by the tRNAs, allowing the assembly of a polypeptide chain. This polypeptide is then matured into a functional protein. The translation mechanism is highly conserved throughout evolution. However, the eukaryotic ribosome is significantly more complex than its prokaryotic counterpart. As we will discuss, the recent acquisition of crystal structures of both prokaryotic and eukaryotic ribosomes has enhanced our understanding of the function and synthesis of these complexes.

2.1 The discovery of the ribosome

Ribosomes and translation were independently discovered by different groups in the 1950s. The definitive biochemical characterization of these ribonucleoprotein complexes as key players in translation occurred in the 1960s.

The first description of ribosomes is attributed to Georges E. Palade (**Figure 1A**). In 1955, using transmission electron microscopy, he described high-density, approximately spherical particles, 10 to 15 nm in diameter, in the cytoplasm of eukaryotic cells (Palade, 1955). These corpuscles, found in the 40 types of vertebrate cells examined in the study, are more abundant in the cytoplasm of actively dividing cells. In contrast, in tissues less prone to proliferation, these particles are preferentially associated with the outer membrane of the endoplasmic reticulum. The identification of similar particles in the cytoplasm of prokaryotes (**Figure 1B**) occurred later (Tissières & Watson, 1958)

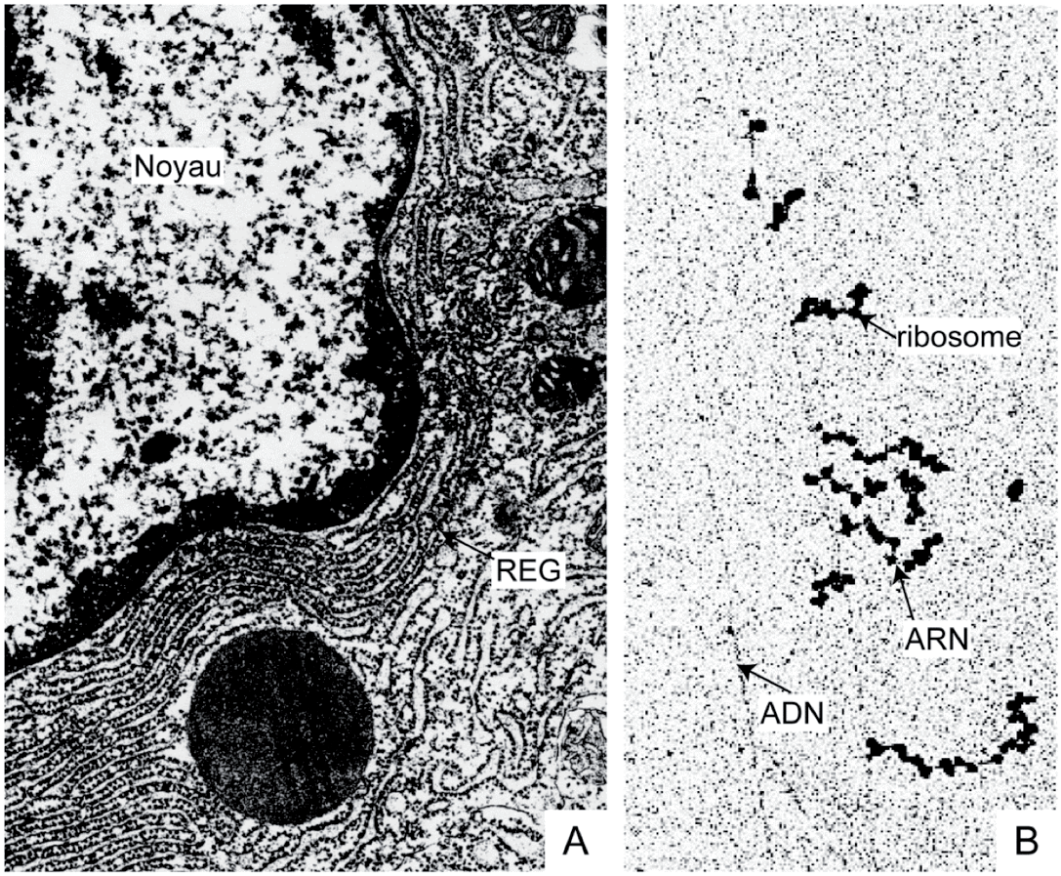


Figure 1: (A) Transmission electron microscopy image of a eukaryotic cell section. The nucleus is visible in the upper left, and adjacent to the nuclear envelope are the compartments of the rough endoplasmic reticulum (RER). Dense particles, identified as ribosomes, are associated with the outer surface of these saccules. (B) Transmission electron microscopy image of a bacterial DNA preparation. The fine vertical filaments represent DNA, while the denser, shorter filaments are mRNAs in the process of being synthesized, with ribosomes attached. This illustrates that in prokaryotes, transcription and translation occur in the same cellular compartment and are thus coupled.

At that time, the role of RNA in information transfer was still unknown. It wasn't until 1961 that messenger RNA (mRNA) was identified as the carrier of genetic information from DNA to proteins (Jacob & Monod, 1961)**.** Subsequent biochemical studies demonstrated that ribosomes were responsible for protein synthesis from mRNAs, in both bacteria (and eukaryotes. The decoding of the genetic code began the same year and was completed by the mid-1960s.These findings provided physiological significance to polysomes, previously referred to as ergosomes (Tissières & Hopkins, 1961).

2.2 The ribosome structure and function

Ribosomes are composed of two subunits across all domains of life: the large ribosomal subunit (60S in eukaryotes and 50S in prokaryotes) and the small subunit (40S in eukaryotes and 30S in prokaryotes). These two subunits assemble on mRNAs to form a functional ribosome (80S in eukaryotes and 70S in prokaryotes).

Each ribosomal subunit is composed of ribosomal RNAs (rRNAs) and ribosomal proteins. The ribosomal RNAs are named according to their sedimentation coefficient, 16S for the small subunit and 23S and 5S for the large subunit.

2.2.1 Structure of the ribosome

Prokaryotic ribosomes sediment as 70S particles and are formed by two subunits, 30S and 50S. In *E. coli*, the 70S ribosome is a 210-Å particle that consists of roughly two-thirds RNA and one-third protein (Schuwirth et al., 2005). The small subunit, 30S, is made of 16S rRNA (1,542 nt) and 21 ribosomal proteins (r- proteins), while the large subunit, 50S, is composed of two rRNAs, 23S (2,904 nt) and 5S (120 nt) rRNA, and 33 proteins (M. Nomura, 1996).

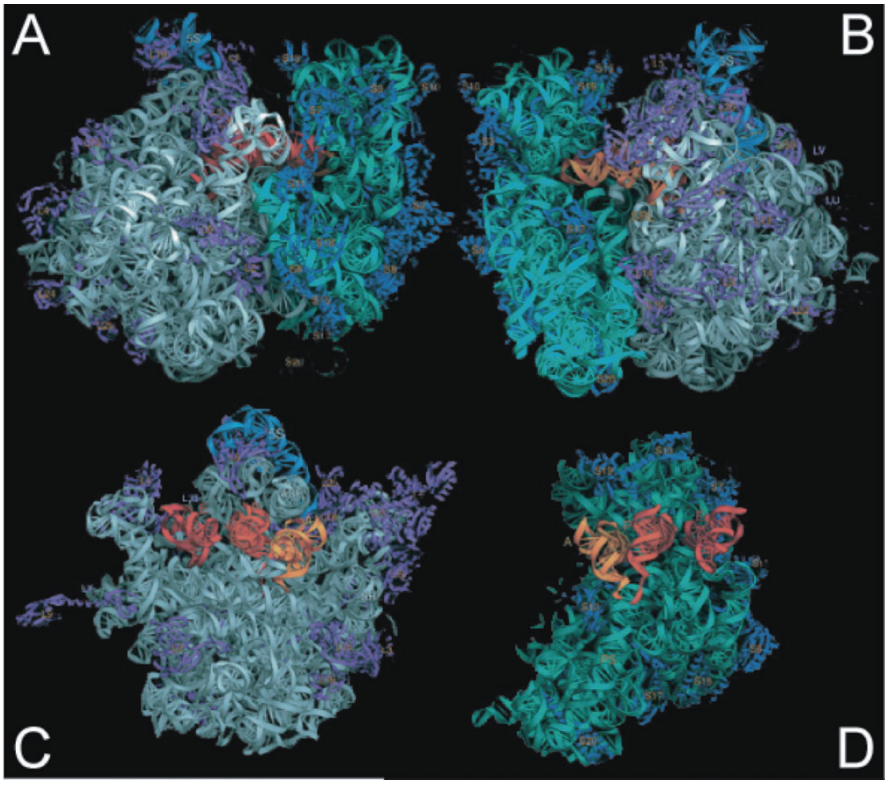


Figure 2: **Structures of Prokaryotic Ribosomes** (A-D) The structure of the T. thermophilus ribosomes was resolved by crystallography at a resolution of 5.5 Å. In the 50S subunit (C), the rRNAs are shown in gray, and the ribosomal proteins are in purple. In the 30S subunit (D), the rRNA is depicted in green, and the proteins are in blue. (A) and (B) show the entire 70S ribosome. The tRNA positioned at the A site is colored orange, and those at the P and E sites are colored red (Noller et al., 2001).

The rRNAs occupy a central position in the ribosome, forming a structure into which ribosomal proteins are embedded. Notably, the interface between the two subunits, where the ribosome's catalytic site is located, is primarily composed of RNA. Due to this composition, ribosomes are ribonucleoprotein complexes (RNPs). They are the largest RNPs in the cell (approximately 2.5 MDa) and by far the most abundant.

During translation, ribosomes move along the mRNA, interact with tRNAs and various protein factors, and undergo significant conformational changes (Druzina & Cooperman, 2004; Dudzińska-Bajorek et al., 2006; Rodnina et al., 2002). Despite the complexity of these macromolecular assemblies, their structures have been resolved (Figure 2). The structure of the ribosome from the thermophilic bacterium *Thermus thermophilus* was determined by X-ray crystallography at a resolution of 5.5 Å (Yusupov et al., 2001). Beyond providing crucial structural data, this work visualized the positioning of tRNAs at the A, P, and E sites (Aminoacyl-tRNA, Peptidyl-tRNA, and Exit sites, respectively), during protein synthesis.

Throughout evolution, the overall structure of ribosomes has been remarkably conserved. This conservation has enabled the resolution of several other ribosome structures across all domains of life, high-resolution X-ray crystallography, the overall shape and positioning of most ribosomal proteins have been determined (with a resolution of about 2.0 Å) for the ribosomes of *Saccharomyces cerevisiae*, using cryo-electron microscopy and comparisons with prokaryotic structures.

2.2.2 Function of the ribosome

Ribosomes are the site of protein synthesis. They catalyze the addition of an extra amino acid to a growing polypeptide chain with an efficiency far exceeding what would be possible through the random encounter of the various molecules involved, thus achieving a speed compatible with life. On average, the elongation rate of a polypeptide chain by a eukaryotic ribosome is 3 to 5 amino acids per second, allowing for the synthesis of short cellular proteins in less than a minute and longer ones in several hours. This addition of amino acids is achieved by decoding the genetic code carried by mRNAs, using aminoacyl-tRNAs specific to each nucleotide triplet, or codon.

Below, I will briefly present the different steps of the translation process in *E. coli,* and, as it is an important part of my work, I will introduce the peptidyl transferase activity of the ribosomal large subunit in more details.

2.2.2.1 The translation process in *E. coli*

The initiation step

Initiation of protein synthesis in *E. coli* is promoted by three initiation factors: IF1, IF2, and IF3. It involves interaction between the Shine-Dalgarno (SD) sequence in the mRNA, about 6 to 9 nucleotides (nt) upstream of the translation initiation codon, and the anti-SD sequence at the 3' end of the 16S rRNA (Shine & Dalgarno, 1974). After primary association of the mRNA to the 30S subunit via the SD–anti-SD interaction, the initiation factors are involved in start codon accommodation at the P site, contributing to the fidelity of translation initiation (Gualerzi et al., 2001). Initiation is thought to be the rate-limiting step in protein synthesis, and several ways of regulating translation through initiation are used.

The elongation step

After association of the 30S and 50S subunits at the end of the initiation step, the P site holds the aminoacylated initiator fMet-tRNA, while the A site is empty and ready to receive an aminoacylated tRNA. Aminoacyl tRNAs are delivered to the A site of the ribosome by the GTPase elongation factor Tu (EF-Tu). At this step, known as decoding, the appropriate or cognate tRNA must be selected from the pool of cellular tRNAs. The binding of a ternary complex with a cognate tRNA to the ribosome triggers the hydrolysis of GTP by EF-Tu and the dissociation of the factor. Following EF-Tu dissociation, the aminoacyl end of the A site tRNA swings into the P site in a process termed accommodation (**Figure 3**). Hence, the ends of the A and P site tRNAs are positioned at the peptidyl transferase center (PTC) on the 50S subunit, and peptide bond formation occurs resulting in addition of the new amino acid to the growing peptide chain (Pape, 1998). The deacylated tRNA is moved from the P site to the E site, to eventually be ejected from the ribosome, while the peptidyl tRNA repositions from the A site to the P site and the mRNA advances by one codon. This process, termed translocation, occurs in two main steps, the first involving the movement of the tRNAs with respect to the 50S subunit and the second, which is driven by GTP hydrolysis and EF-G, involving the movement of the mRNA and the anticodon ends of the tRNA relative to the 30S subunit. Translocation brings a new mRNA codon into the A site and prepares the ribosome for another round of the elongation cycle (Rodnina et al., 1997; Voorhees & Ramakrishnan, 2013).

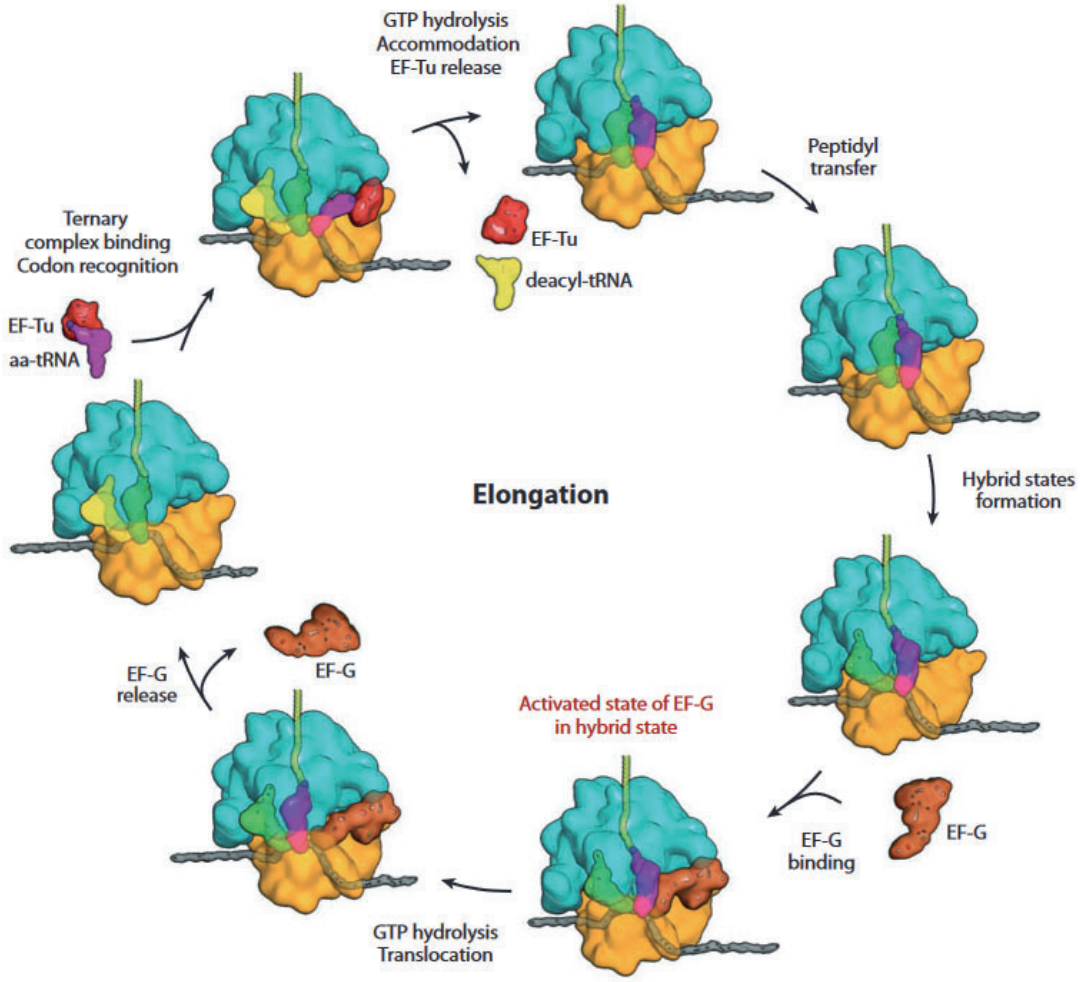


Figure 3: Schematic of the bacterial elongation cycle. At the heart of protein synthesis is the elongation cycle, which involves the sequential addition of amino acids to the growing peptide chain, facilitated by the GTPase factors elongation factor G (EF-G) and elongation factor Tu (EF-Tu). From *Voorhees and Ramakrishnan 2013.*

Direct physical evidence for the spontaneous, peptidyl-transferase–coupled translocation step has been obtained (Deng et al., 1986). On the basis of their observations, they proposed that the peptidyl chain remains in a constant position relative to the ribosome structure, while tRNA moves, during the peptidyl transferase reaction. The movements of tRNA-ribosome complexes corresponding to different intermediate states of the translational cycle led to the proposal of the hybrid states model for translation (Moazed & Noller, 1989).

In the hybrid states model (**Figure 4**), binding to and movement between the A and P sites of the small ribosomal subunit by tRNA is uncoupled from its movement between the A, P, and E sites of the large subunit. EF-Tu.GTP. Aminoacyl-tRNA ternary complex is delivered to the peptidyl-tRNA-ribosome complex and binds initially in the A/T hybrid state. In this state, the anticodon end is bound to the A site of the 30S subunit, and the CCA end is sequestered by EF-

Tu, which is bound to the large subunit (T site). EF-Tu released permits the movement of the CCA end of the aminoacyl-tRNA into the large subunit A site; this establishes the A/A state. The nascent polypeptide chain carried by the peptidyl-tRNA is in the P/P state. Peptidyl transferase then catalyzes peptide bond formation and the aminoacyl-tRNA, now carrying the nascent chain, moves into the large subunit P site, and peptidyl-tRNA, now deacylated, moves into the large subunit E site. As a result, the newly formed peptidyl-tRNA is in the A/P hybrid state (i.e. its anticodon end remains in the A site of the small subunit, but its CCA end occupies the large subunit P site), and the deacylated tRNA is in the P/E hybrid state (i.e. its anticodon end remains in the P site of the small subunit, but its CCA end occupies the large subunit E site). Binding of the acceptor end of the deacylated tRNA to the 50S subunit E site may provide the thermodynamic driving force for the spontaneous transition. EF-G then catalyzes movement of the anticodon ends of the hybrid-bound tRNAs, together with mRNA, relative to the small subunit. The peptidyl-tRNA is then in the P/P state, and the deacylated tRNA is in the E state interacting only with the large subunit E site at this stage.

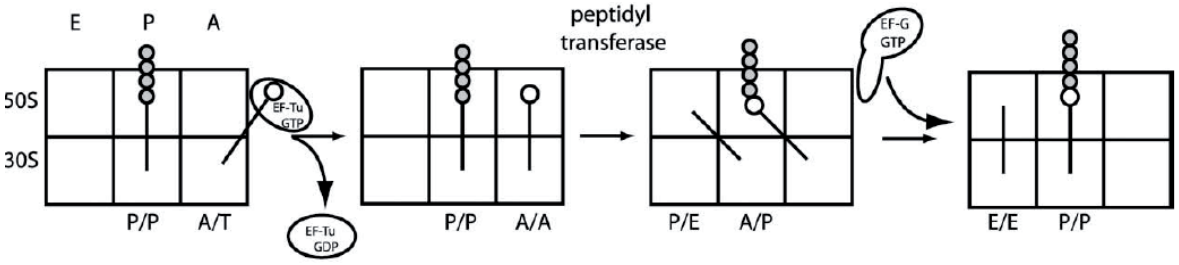


Figure 4: The Hybrid States Model for Translocation as proposed by Moazed and Noller (1989).

The hybrid-states model has several interesting consequences. First, the peptidyl moiety remains stationary; it is the tRNAs that move. Second, during translocation, both tRNAs move along with their bound mRNA with respect to the small and large subunits. This movement implicitly suggests that translocation involves relative movement of the two different ribosomal subunits.

The termination and recycling step

Translation termination begins when a stop codon in the mRNA enters the ribosomal A site. The termination codon is recognized by either release factor 1 (RF1) or RF2; RF1 terminates at stop codons UAA and UAG, while RF2 terminates at UAA and UGA. Binding of RF1/RF2 to the ribosome triggers hydrolysis and release of the peptide chain from the P site tRNA (Chavatte et al., 2003). Upon peptide bond hydrolysis, the third release factor, RF3, which is a GTPase,

binds to the ribosome and promotes dissociation of RF1/RF2 from the A site (Freistroffer et al., 1997; Klaholz et al., 2004).

After peptide release, the ribosome is left with bound mRNA and a deacylated tRNA in the P site. This complex needs to be disassembled to prepare the ribosomal subunits for a new round of protein synthesis. Ribosome recycling factor, along with EFG, is required for the process of subunit dissociation (Kiel et al., 2003). Subsequently, IF3 replaces the deacylated tRNA on the 30S subunit and allows the mRNA to either detach from the complex or form a new stable SD–anti-SD interaction with the downstream ribosome-binding site (Karimi et al., 1999).

2.2.2.2 The peptidyl transferase activity

The key catalytic step in protein synthesis is peptide bond formation, which occurs through the nucleophilic attack of the α-amino group of the aminoacyl tRNA on the aminoacyl ester of the peptidyl tRNA. This reaction leads to formation of a peptide bond and release of an alcohol product. The ribosome enhances the rate of aminolysis $\sim 10^7$-fold compared with the spontaneous rate in solution. Initial biochemical experiments suggested that the catalytic power of the ribosome was derived from its RNA components (Noller et al., 1992). Indeed, although proteins have been observed in the peptidyl transferase center of the bacterial ribosome, structural evidence suggests that no ribosomal proteins are directly involved in catalysis (Ban et al., 2000; Nissen et al., 2000).

Upon binding of an A-site substrate, a series of conformational changes in the 23S rRNA exposes the peptidyl tRNA ester for nucleophilic attack by the α-amine. These conformational changes are induced by the binding of any A-site substrate containing at least residue C75 (Brunelle et al., 2006), consistent with observations that binding of even a deacylated tRNA in the A site increases the rate of hydrolysis of the peptidyl tRNA.

Several catalytic mechanisms for peptidyl transfer have been proposed including general acid base catalysis (Muth et al., 2000; Nissen et al., 2000), and substrate-assisted catalysis (Dorner, 2003) (Weinger et al., 2004). Both mechanisms are described below.

Acid base catalysis: Role of A2451 of 23S rRNA

Early structural studies ascribed a catalytic role to a highly conserved 23S rRNA residue, A2451, which was within hydrogenbonding distance of the nucleophilic amino group (Nissen et al., 2000) (Muth et al., 2000). In the reaction scheme described by Nissen et al. (2000), the substrates of the reaction catalyzed by the large subunit are an aminoacyl-tRNA (aatRNA) and

a peptidyl-tRNA. The former binds in the ribosome's A-site and the latter in its P-site. The α-amino group of the aatRNA attacks the carbon of the carbonyl acylating the 3' hydroxyl group of the peptidyl-tRNA, and a tetrahedral intermediate is formed at the carbonyl carbon (**Figure 5**). In their crystallographic studies revealing the atomic structure of the large ribosomal subunit from *Haloarcula marismortui*, Nissen et al. (2000) observed that the N3 of A2486 (A2451 in *E. coli*) is the titratable group nearest to the peptide bond being synthesized and is likely functioning as a general base to facilitate the nucleophilic attack by the α-amino group of the A-site substrate.

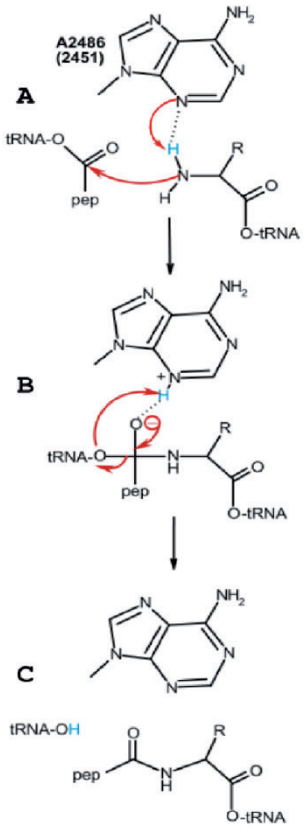


Figure 5: The acid base catalysis mechanism proposed for the peptide transfer reaction catalyzed by the ribosome. The sequence number of the catalytic A is 2486 in *H. marismortui* and 2451 in *E. coli*. In step A, the unprotonated N3 of A2486 removes a proton from the amino group of an incoming aminoacyl tRNA as it attacks the carbonyl carbon of the ester bond linking the nascent polypeptide to a second tRNA. In step B, the protonated A helps stabilize the tetrahedral intermediate formed by hydrogen bonding to its oxygen. In step C, the A donates its proton to the hydroxide ion formed when the tetrahedral intermediate resolves itself. In this mechanism A2451 plays a role similar to that of the active site histidines in serine proteases (from Nissen et al., 2000).

In order to function in this capacity, the pKa of the N3 of A2451 has to be roughly 5 units higher than normal. Ordinarily, the pKa the N1 of adenosine monophosphate is about 3.5 and that of its N3 is 2 pH units lower. In order to function as a general base, the pKa of the N3 would have

to be raised to 7 or higher. Because the crystal is at pH 5.8, this implies that the pKa of the N3 is >6. Nissen et al. (2000) observed that the crystal structure itself suggests that its pKa is, in fact, quite unusual. They suggest that, at physiological pH, several features of A2451 environment might affect its pKa. The unusual pKa may be produced by an interaction between A2451 and G2447 which in turn hydrogen bonds with the buried phosphate of A2450 (note that A2450 follows the modified D2449 described in our study, see Results section).

The N3 of A2451 can only hydrogen bond to the carbonyl oxygen of the peptidyl tRNA, as observed, if it is protonated. The distance between these two atoms is about 3 Å, indicating that a hydrogen bond does, indeed, exist between them. The tetrahedral intermediate resolves to yield a peptide extended by one amino acid esterified to the A site-bound tRNA and a deacylated tRNA in the P-site. (**Figure 5**).

However, mutation of A2451 and several other active site rRNA bases had only modest effects on the rate of peptidyl transfer (Polacek et al., 2001). These observations suggested that either the ribosome does not employ general acid-base catalysis; or catalysis occurs through a functional group that is unchanged by mutation, such as a backbone phosphate or ribosyl group; or that catalysis is performed by a functional group located not on the rRNA but on the substrates themselves.

Substrate-assisted catalysis: Role of the 2' OH of A76 of the Peptidyl-tRNA

A primary candidate to perform substrate-assisted catalysis was the 2' OH of A76 of the peptidyl-tRNA. Crystal structures indicated that the 2' OH of A76 is one of the few functional groups properly positioned to function directly in peptide bond formation (Hansen et al., 2002). Biochemical studies initially reported that substitution of this hydroxyl moiety with either a fluorine or hydrogen results in a $\sim 10^6$-fold reduction in the rate of peptidyl transfer (Weinger et al., 2004). The 2' OH was hypothesized to be unlikely to act as a general base. Instead, this moiety was widely accepted to function as part of a proton shuttle mechanism (Dorner, 2003) that catalyzed peptidyl transfer by facilitating proton exchange and orienting the substrates for reactivity. The primary role of the 2' OH of A76 of the peptidyl tRNA is thought to induce an active conformation of the peptidyl transferase center, which slowly rearranges when bound to a 2' deoxy substrate (Zaher et al., 2011). Based on this, peptide bond formation was hypothesized to be catalyzed by a fully concerted eight-membered proton-shuttle mechanism involving protons originating from the α-amine, a crystallographic water molecule, and the 3'OH of A76 of the peptidyl tRNA (**Figure 6**).

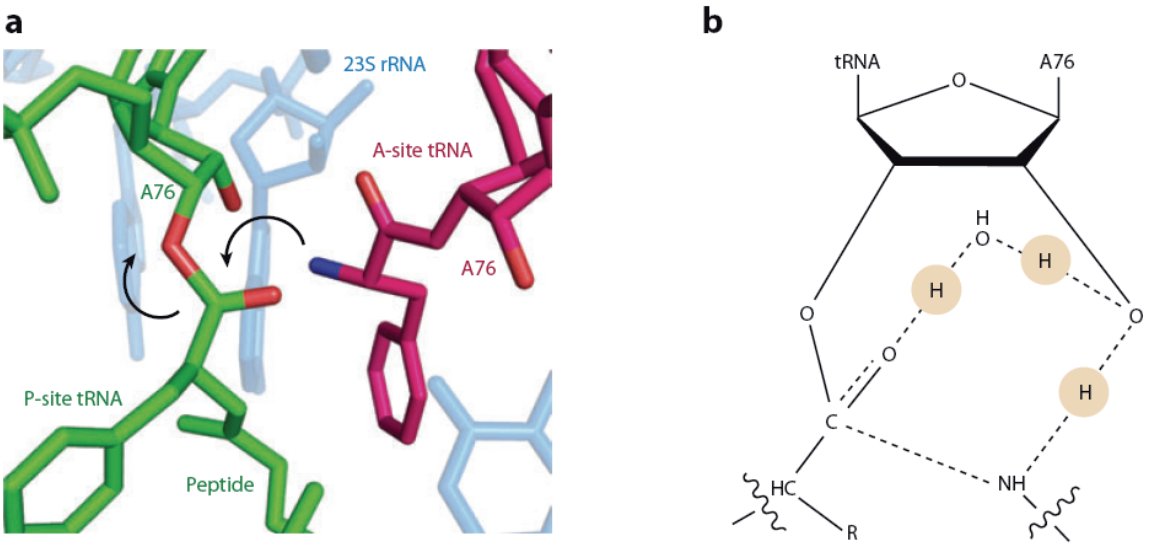


Figure 6: The chemical mechanism of peptide bond formation. (a) Peptidyl transfer involves the nucleophilic attack of the α-amino group on the aminoacyl transfer RNA (tRNA) (magenta) on the peptidyl-ester on the peptidyl tRNA (green). (b) Fully concerted mechanism of peptidyl transfer involving an eight-membered proton shuttle in the transition state. From Voorhees and V. Ramakrishnan 2013.

In conclusion, as pointed out by Voorhees and V. Ramakrishnan in their review (2013), what is clear is that the mechanism of peptide bond formation catalyzed by ribosomes differs significantly from that of the uncatalyzed reaction, in in which nucleophilic attack and deprotonation are not concerted(Voorhees & Ramakrishnan, 2013). Therefore, the ribosome not only increases the rate of peptide bond formation, but also significantly disrupts the energy landscape of the reaction. This observation cannot be explained by catalysis via substrate positioning alone, and there is growing evidence that the ribosome must also play a role in orchestrating the chemical mechanism of peptide bond formation.

2.3 Ribosome biogenesis

The assembly of bacterial ribosomes requires the transcription of rRNAs from operons dedicated to ribosome synthesis, followed by their maturation through cleavage and base modifications. The 30S and 50S subunits are then assembled by associating rRNAs with ribosomal proteins. This entire process takes place within a single compartment, the cytoplasm.

2.3.1 Transcription of rRNA operons

The biogenesis of ribosomes begins with transcription of the ribosomal RNA 16S, 23S, and 5S rRNA, which is carried out within specialized operons and synthesized as one primary transcript. In *E. coli*, there are 7 operons dedicated to rRNA synthesis. Their structures vary slightly around a common pattern. For example, the structure of the first sequenced operon, rrnB (Woese et al., 1980), is shown in **Figure 7**. The regions coding for the mature RNAs 16S, $tRNA^{Glu}$, 23S, and 5S are depicted in light blue, green, purple, and orange, respectively. The promoter (P1, P2) and terminator (T1, T2) regions are indicated. The P1 region is enlarged to show the consensus promoter region with the -10 and -35 boxes, forming the extended promoter with the UP element (in yellow). The UAS (Upstream Activating Sequence) region contains three binding sites for the Fis activator protein.

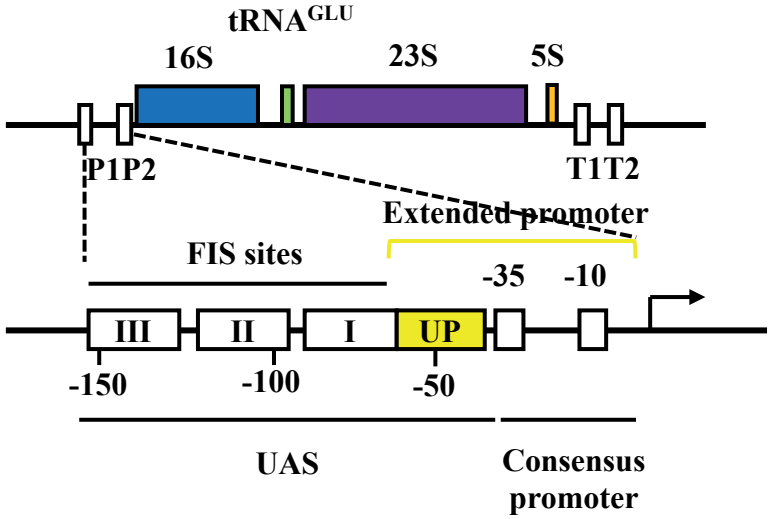


Figure 7: **Schematical representation of the operon rrnB adapted from (Woese et al., 1980).** The regions coding for the mature RNAs 16S, $tRNA^{Glu}$, 23S, and 5S are represented in light blue, green, purple, and orange respectively. The promoter (P1, P2) and terminator (T1, T2) regions are indicated. The P1 region is enlarged, showing the consensus promoter region with the -10 and -35 boxes, forming the extended promoter with the UP element (in yellow) The UAS (Upstream Activating Sequence) region contains three binding sites for the Fis activator protein.

The transcribed sequence contains the 16S, 23S, and 5S rRNA sequences, separated by spacer regions. Between the 16S and 23S rRNAs, the spacer region contains a copy of the $tRNA^{GLU}$ gene. Variations in other operons affect the number and type of tRNAs, which can be positioned either centrally, as in rrnB, or terminally. Additionally, the rrnD operon has two cistrons for synthesizing 5S rRNA. The exceptional efficiency of transcription of these operons is ensured

by two particular mechanisms, increased initiation frequency and more processive transcription elongation.

Maturation of the transcript begins before transcription is completed, with instant formation of local secondary structures and, as soon as their binding sites emerge from the polymerase, binding of ribosomal proteins. Simultaneously, rRNA becomes chemically modified at several positions and is processed by several RNases to generate mature rRNA species (Williamson, 2003).

2.3.2 Nucleolytic processing of rRNA

From the primary transcripts produced by operons, a succession of endonucleolytic and exonucleolytic cleavages leads to the release of mature rRNAs 16S, 23S, and 5S, as well as tRNAs, and the elimination of spacer sequences (**Figure 8**).

Initially, the double-stranded endonuclease RNase III recognizes a pairing between the 5' region of the primary transcript and the first spacer sequence, leading to the release of the 16S rRNA precursors. The 5' end of the 16S rRNA is released by the endonucleolytic action of RNase M16 (maturase). The enzyme responsible for forming the 3' end is still unknown.

The 5' end of the tRNA contained within the sequence separating the 16S and 23S rRNAs is released by the action of a ribonucleoprotein complex, RNase P. RNase III then acts to separate the tRNA precursors from the 23S and 5S rRNA precursors. The release of the 5S rRNA from the terminal part of the transcript is ensured by double endonucleolytic cleavage by RNase E, upstream and downstream of the mature RNA.

The definitive 3' ends of the RNAs are formed by the exonucleolytic activities of RNases X (for tRNA) and T (for rRNAs). The enzymes involved in forming the 5' ends of the 23S and 5S rRNAs are still unknown (D. L. J. Lafontaine & Tollervey, 2001).

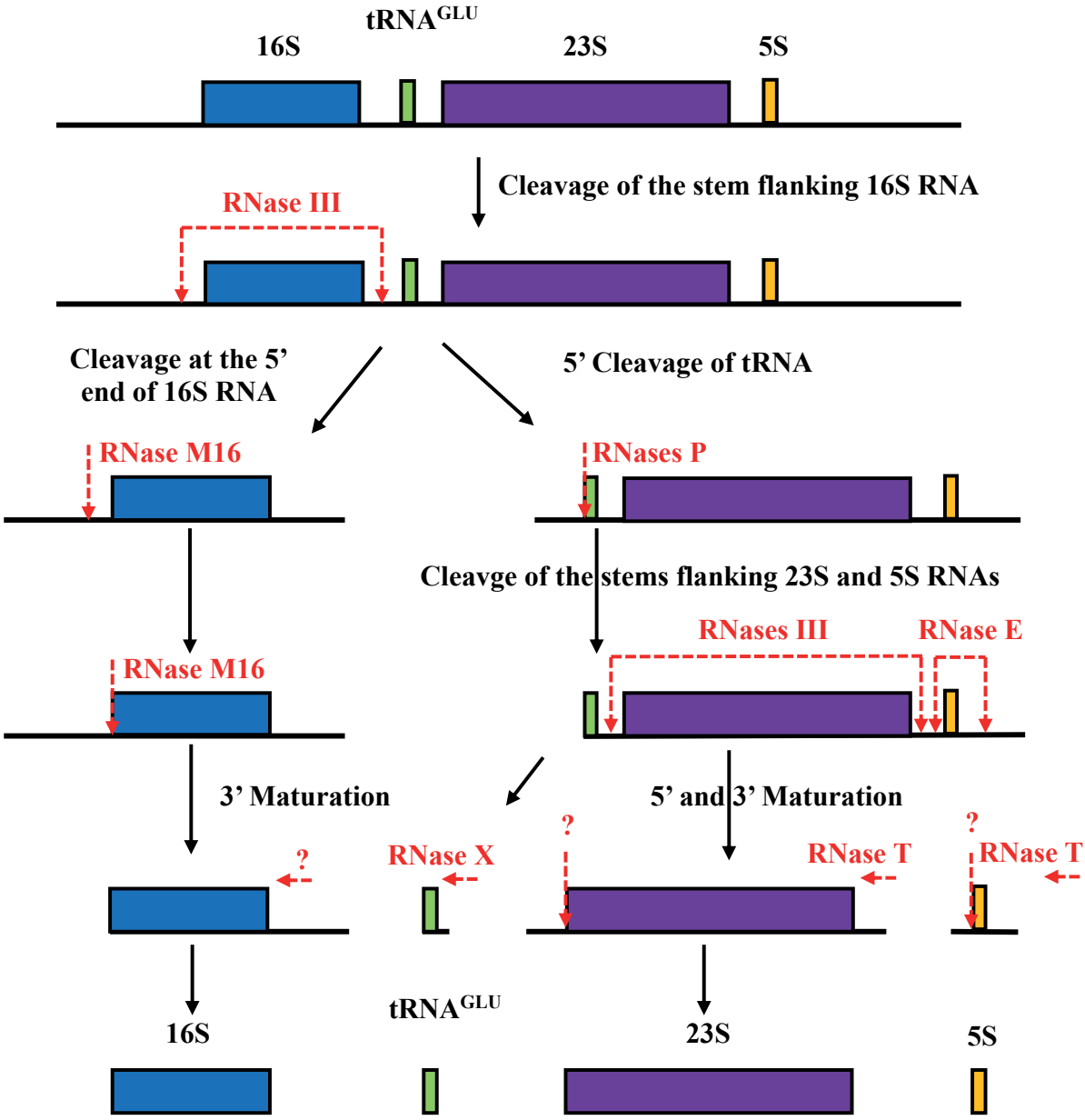


Figure 8: Cleavage Pathways of rRNA Precursors in *E. coli* for the rrnB Operon According to Lafontaine and Tollervey (D. L. J. Lafontaine & Tollervey, 2001). The enzymes involved in the different cleavages are indicated in red. Endonucleolytic cleavage sites are marked by vertical arrows, while exonucleolytic maturation steps are represented by horizontal arrows.

2.3.3 Chemical modification of rRNA

In *E. coli*, tRNA and rRNA (except for the 5S rRNA) are covalently modified during maturation. Both the base and ribose entities can be subject to different modifications.

Most of the modified bases are located, in the final structure of the ribosome, at the active site specifically, at the decoding site of genetic information in the small ribosomal subunit and at the peptidyl transferase activity site in the large ribosomal subunit (Boccaletto et al., 2022).This section will be elaborated in detail below.

3. RNA modifications

Adenosine (A), guanosine (G), cytidine (C), and uridine (U) are the four ribonucleosides typically found in ribonucleic acid (RNA) molecules. These ribonucleosides are classified into two categories based on the shape and chemical properties of their nucleobases: pyrimidines and purines. Pyrimidines, which include cytidine (C) and uridine (U), feature a six-membered heterocyclic ring similar to pyridine. Purines, which include adenosine (A) and guanosine (G), consist of a fused heterocyclic structure combining a pyrimidine ring with an imidazole ring (**Figure 9**). When these pyrimidines and purines are attached to ribose molecules, they form ribonucleosides (McCown et al., 2020)

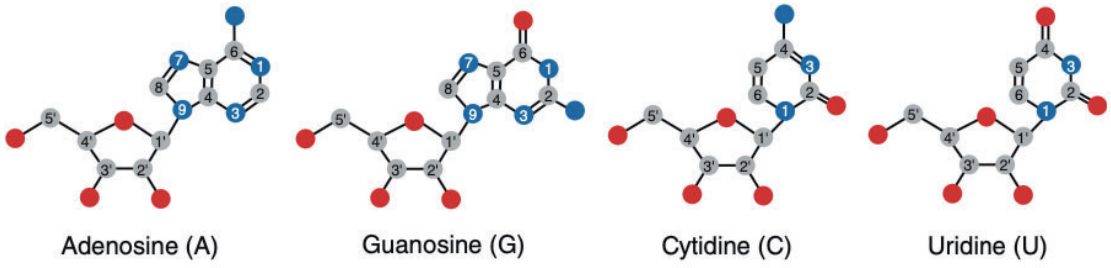


Figure 9: Numbering of ribonucleosides for adenine, guanine, cytosine, and uracil bases, along with the numbering of the ribose sugar for adenosine (A), guanosine (G), cytidine (C), and uridine (U), is detailed as follows. Carbon atoms are represented in gray, oxygen atoms in red, and nitrogen atoms in blue. Single lines between atoms indicate single bonds, double lines indicate double bonds, and hydrogen atoms are omitted for simplicity.

From these fundamental pyrimidines and purines, various derivatives such as thiamin pyrophosphate (TPP), flavin mononucleotide (FMN), and cyclic di-guanosine monophosphate (c-di-GMP) are believed to have originated from the RNA world a period in which genes, genomes, and enzymes were composed solely of RNA (Crick, 1968; Joyce & Szostak, 2018) (Nelson & Breaker, 2017; Robertson & Joyce, 2012). In addition to their primary forms, pyrimidines and purines serve as precursors for numerous modifications (Cantara et al., 2011; Cohn & Volkin, 1957; Holley et al., 1965; Yanas & Liu, 2019).

The first modification of RNA was discovered in 1957, namely the pseudouridine (Ψ) (**Figure 10**), also called the "fifth nucleotide" for its abundance in the transcriptome of all organisms (F. F. Davis & Allen, 1957).

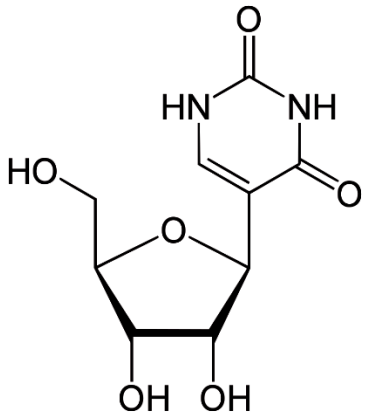


Figure 10: Chemical structure of pseudouridine.

Forty years later, more than ninety naturally occurring modified nucleosides were identified primarily in two large classes of RNA, rRNA and tRNA but also in other types of RNA such as mRNA and lncRNA, miRNA…. The frequency and function of the majority of these modifications was limited. However, the advent of technologies for next-generation sequencing (NGS) and improvements to liquid chromatography–mass spectrometry (LC-MS/MS), have help tremendously to study various modifications across all RNA species and domains of life (Frye et al., 2016). Today more than 170 chemical modifications are reported in databases such as MODOMICS (Boccaletto et al., 2022).

As evidenced by a steady increase in the number of relevant articles in recent years on RNA modifications, this research is an actively developing area with N6-methyladenosine (m6A) being the most studied, which is not surprising considering its abundance in mRNA, its now-established strong connection to cancer processes, and a plethora of high-throughput methods for its detection (see **Figure 11**). These discoveries and characterization of RNA modifications highlight their biological significance and represent significant advancements in RNA biology (Jung & Goldman, 2018; H. Shi et al., 2019; Yanas & Liu, 2019).

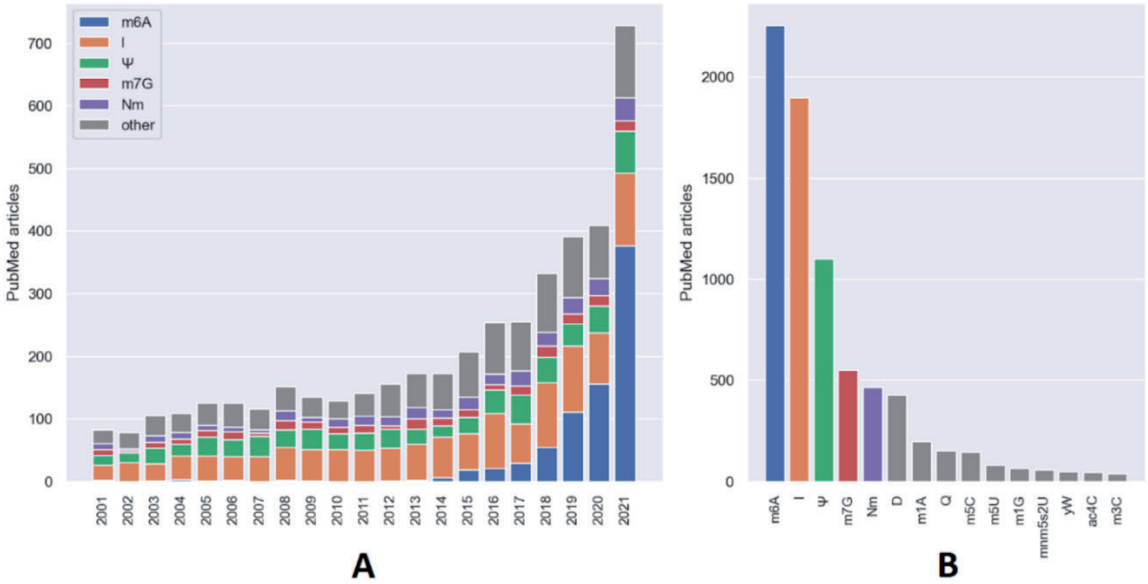


Figure 11: (A) Number of published articles that are strongly related to specific RNA modifications in the PubMed database per year. (B) Total number of articles on specific RNA modifications in the PubMed database. This figure is taken from (Arzumanian et al., 2022).

3.1 RNA modifications and chemical diversity

Ribonucleoside modifications include a wide variety of chemical moieties that are added to adenosine, guanosine, cytidine, or uridine (**Figure 12**). These modifications range from simple methylations or hydrations of double bonds to more complex changes like ring closures of the nitrogenous base or the addition of large groups such as amino acids or monosaccharides. Most of these modifications are catalyzed by enzymes. Some enzymes form multimers with auxiliary proteins (e.g., helper methyltransferases), which expand the range of substrates or allow for the integration of cellular activities and enzymatic processes (Guy & Phizicky, 2014). However, some modifications result from nonenzymatic processes or oxidative damage. RNA editing, initially described as the addition of polyuridine residues within the coding regions of certain RNAs, has been expanded to include RNA base excisions or additions (e.g., deletion of uridine residues in pre-mRNAs or addition of guanosine to certain tRNAs) as well as RNA base conversions (e.g., A-to-I or C-to-U editing). However, further distinctions in RNA editing remain somewhat unclear (Grosjean, H & Benne, R, 1998).

Currently, 111 modifications have been identified in transfer RNAs (tRNAs), 36 in ribosomal RNAs (rRNAs), 17 in messenger RNAs (mRNAs), and 11 in long non-coding RNAs (lncRNAs) and other non-coding RNAs (ncRNAs) (Boccaletto et al., 2022; Lorenz et al., 2017; Yanas & Liu, 2019). The nature of the chemical modification is abbreviated with a letter code

followed by the number in superscript of the position harboring the modification. If more than one identical modification at the same position is present, then their number is reported in subscript after the position. When the modification occurs on the sugar then its letter code is reported after the nucleoside symbol. However, some modifications are named with an individual letter instead (see below).

Figure 12 shows examples of simple post-transcriptional modifications mostly methylations, isomerization or sulfuration for example. A single enzymatic step is usually involved in their synthesis, carried by a single enzyme. These simple modified nucleotides are the most abundant ones in vivo.

The most abundant modification, Ψ derives from the isomerization of uridine via breakage of N1-C1 bond linking the base to the ribose followed by a 180° rotation of the base thus forming a C5-Cl bond instead (Ge & Yu, 2013; Hamma & Ferré-D'Amaré, 2006). Ψ increases base stacking, rigidify the sugar-phosphate backbone (C-C bond is more stable than N-C) and improves base-pairing via the additional imino group providing an extra hydrogen bond donor (Charette & Gray, 2000; D. R. Davis, 1995).

Thiolation (s) has been detected in pyrimidines (s4U, s2U, s2C), replacing the carbonyls at position C2 or C4, as well as in cytosine and adenosine both at position C2. The reaction consists in incorporation of a thiol that strengthens U:A or C:G pairing while destabilizing U:G or C:A pairing. (Sheng et al., 2014; Y. Zhou et al., 2013). Furthermore, s4U nucleotide has a unique UV absorption spectrum with a maximum at 330 nm thus conferring cross-linking properties upon near-UV light irradiation (Shigi, 2014).

Methylation (m) consists in incorporation of a methyl group (-CH_3) either on the base or the ribose at its 2' hydroxyl (Srinivasan & Borek, 1964). Methylations deriving from the four canonical bases have been identified with the majority of nitrogen positions targeted as well as C5 of pyrimidine, C2 and C8 of adenosine (Motorin et Helm 2011). Methylated nucleotides usually lead to higher stability of the RNA (C.-H. Lee & Tinoco, 1977; Watanabe et al., 1979). Furthermore, depending on the methylation position disruption of hydrogen bonding can occur, restricting the pairing possibilities for the methylated nucleotide (Helm et al., 1999).

Modified nucleotides are classified in two categories: simple modifications or hypermodified bases (McCown et al., 2020). This classification depends on the chemical nature of the modification.

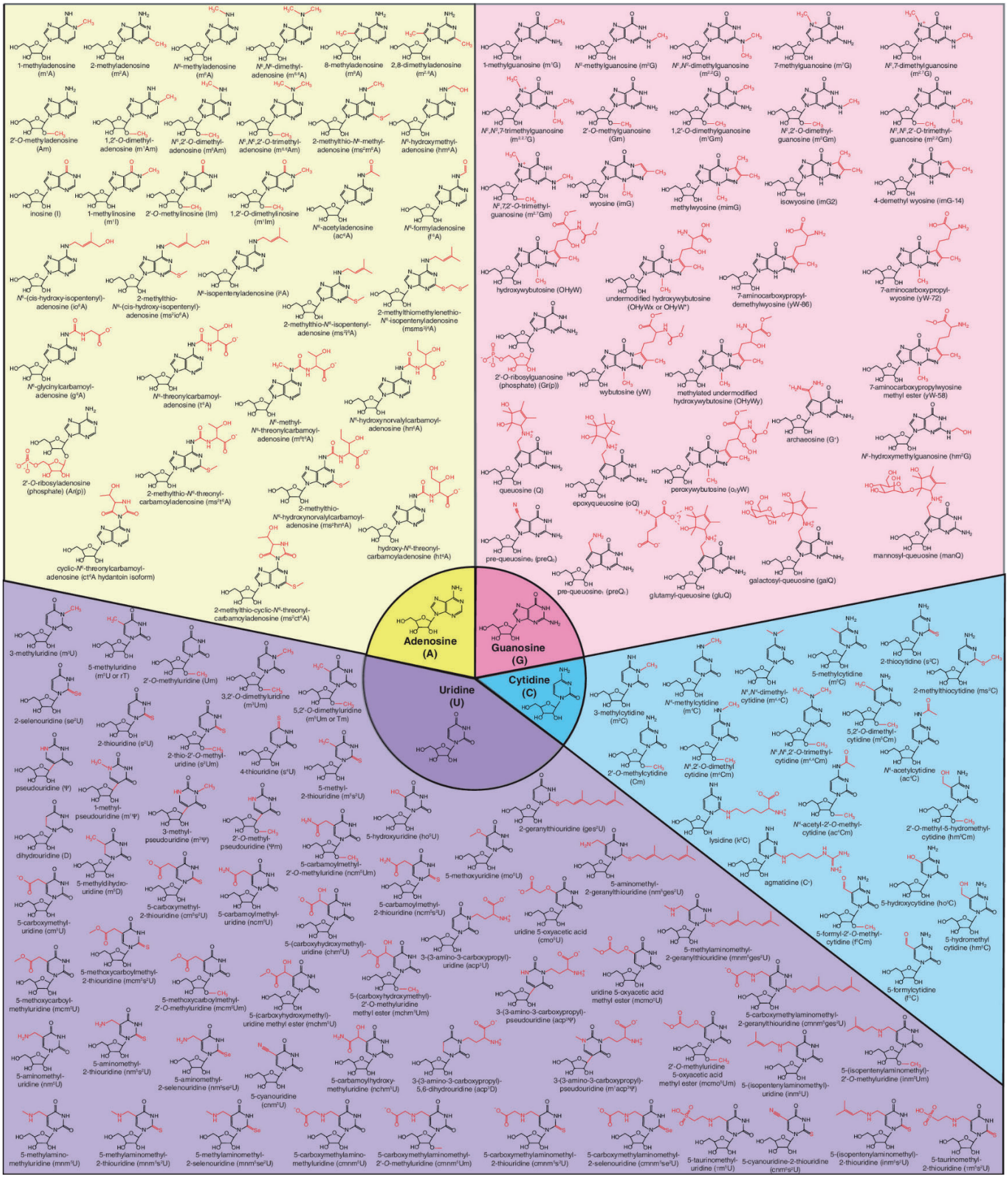


Figure 12: Chemical structures of all currently known RNA modifications. Adenosine-derived (yellow), guanosine-derived (pink), uridine-derived (purple), and cytidine-derived

(cyan) modifications are classified based on the parent ribonucleoside. Red moieties indicate which portion of the modified ribonucleoside is different from the parent ribonucleoside, whose structures are shown in the central circle. This figure is taken from (McCown et al., 2020)

Transfer RNAs of *E. coli* have been the model molecules in the studies of these modifications and all the tRNA modifying enzymes have been identified (**Figure 13**). Regarding rRNA modifications, I will fully describe them in the next chapter of my thesis. While mRNA, rRNA contains mostly simple modifications, tRNA molecules contain both types of modifications. However, the complex modifications are observed exclusively at position 34 (wobble position) and 37 of tRNAs (**Figure 13**).

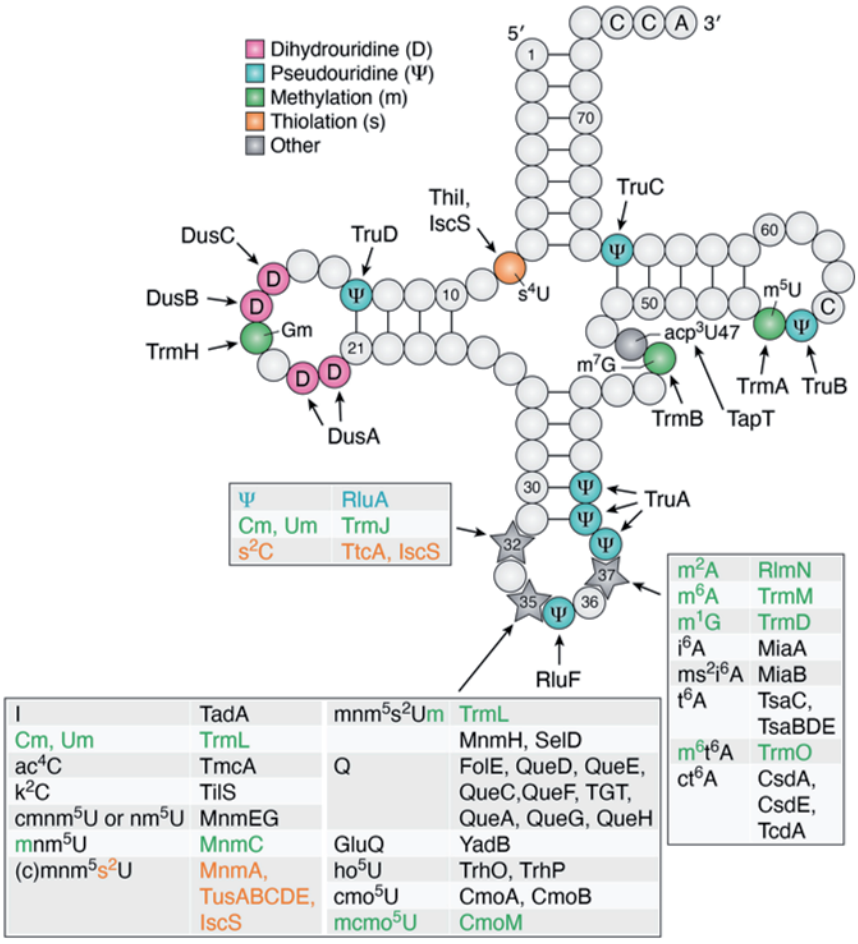


Figure 13: adapted from (Schultz & Kothe, 2024) tRNA structure and modification. Locations and identities of all modifications in *E. coli* tRNAs on the cloverleaf representation and the corresponding catalyzing RNA modifying enzymes. Common modifications are denoted by color, as indicated in the legend, which also indicates the abbreviations for these common modifications. Modifications are systematically abbreviated with letters preceding the nucleotide indicating a base modification. Multiple different modifications can occur at positions 32, 34, and 37 in different tRNAs.

Dihydrouridine (D) is the only non-aromatic nucleotide detected in vivo. It derives from the reduction of C5-C6 double bond in uridine. D favors the C2' endo conformation while disrupting base stacking (J. Dalluge, 1996b; Dyubankova et al., 2015). In addition, D is less thermally stable and is prone to ring opening especially at alkaline pH (House & Miller, 1996). Inosine (I) is produced by hydrolytic deamination of adenosine (Grosjean et al., 1996). An identical mechanism can occur on cytidine thus forming uridine. The conversion of A to I is referred to as editing since base pairing properties are changed (Gerber & Keller, 2001; Torres et al., 2014). Indeed, inosine shares similar hydrogen bond edges with guanosine despite deriving from adenosine.

Figure 12 also illustrates examples of nucleosides containing complex modifications (cmnm5U, t6A…. or hypermodified bases (double, triple modified nucleosides with two simple modifications like cmnm5s2U or m2s2i6A, or m2,2,7G...). The formation of these bases involves intricate biochemical mechanisms, requiring multiple steps carried out by different enzymes (see **figure 13**) in a coordinated manner or in a specific sequence. Additionally, nucleotides can undergo multiple modifications at different positions through orthogonal transformations. In these cases, the sequence in which the modifications occur can be interchangeable.

Some hypermodified bases arise directly from nucleotides in the RNA chain with the incorporation of amino acids. Examples include glycine in 5-carboxymethylaminomethyl uridine (cmnmU)(Moukadiri et al., 2009; R. Shi et al., 2009), taurine in 5-taurinomethyl uridine (tmU)(Suzuki, 2002), threonine in N^6-threonylcarbamoyl adenosine (t^6A)(X. Chen et al., 2007; U. F. Müller, 2006), and lysine in 2-lysidine cytosine (k^2C) (Ge & Yu, 2013; House & Miller, 1996). Others result from the addition of sugars, such as ribose in 2'-O-ribosyladenosine (Ar(p)) (Åström & Byström, 1994; Czerwoniec & Bujnicki, 2011), galactose in galactosyl-queuosine (galQ), and mannose in mannosyl-queuosine (manQ) (Y.-C. Chen et al., 2011; Kasai et al., 1976). Additionally, N^6-isopentenyladenosine (i^6A) is derived from mevalonic acid through the incorporation of dimethylallyl pyrophosphate (DMAPP) onto adenosine (Chimnaronk et al., 2009; Persson et al., 1994; Soderberg & Poulter, 2000; C. Zhou & Huang, 2008). Furthermore, hypermodified nucleotides containing 5'-nicotinamide adenine dinucleotide (NAD) have also been identified.

Heterocycle exchange within the RNA chain can also result in hypermodifications. In this process, the modified nucleotide or a precursor is synthesized before being incorporated into

the RNA polymer. Archaeosine (G+) and Queuosine (Q) are usually synthesized from guanosine triphosphate (GTP) or obtained from nutrients, and then introduced into RNA by breaking a glycosidic bond, replacing the original nucleotide(Carell et al., 2012; El Yacoubi et al., 2012; Helm & Alfonzo, 2014). Most hypermodified nucleotides exhibit increased hydrophobic properties and often favor the C3' endo conformation. Additionally, the bulky modifications create steric hindrance, affecting the base pairing of neighboring nucleotides in the RNA chain.

3.2 Prevalence of modified Nucleotides

Posttranscriptional modifications of nucleotides have been observed across all three domains of life (**Figure 14**) (Machnicka et al., 2012; Motorin & Helm, 2011). Some modifications, such as dihydrouridine (D), pseudouridine (Ψ), and various methylations, are ubiquitous across all organisms, suggesting convergent evolution towards these modifications or their presence in the last universal common ancestor (LUCA)(Martínez Giménez et al., 1998), highlighting their fundamental importance. Conversely, certain modifications are species-specific. For instance, archaeosine is exclusive to archaea (after which it is named), yet shares a precursor, $preQ_0$, with queuosine found in eukarya and bacteria (Czerwoniec & Bujnicki, 2011). Similarly, 5-taurinomethyluridine (tm^5U) is restricted to mitochondria, while wybutosine (yW), which requires up to five enzymes for its synthesis (De Crécy-Lagard et al., 2010; Noma et al., 2006; Perche-Letuvée et al., 2014), is specific to eukarya and shares common precursors with archaea. Inosine is present in both bacteria and eukarya. Modified nucleotides have been detected in a wide range of RNA species as shown in **Figure 14**. Transfer RNA (tRNA) exhibits the greatest diversity of posttranscriptional modifications. Some modifications are common across different RNA species. For instance, 2'-O-methylation on the ribose hydroxyl group is found in tRNA, messenger RNA (mRNA), ribosomal RNA (rRNA), and other small non-coding RNAs (snRNAs). In contrast, modifications targeting the nucleotide base are less widespread, except for N^6-methyladenosine (m^6A) and, more recently, pseudouridine and dihydrouridine, which has also been detected in mRNA (Carlile et al., 2014; Dixit & Jaffrey, 2022; Draycott et al., 2022; Finet, Yague-Sanz, Krüger, et al., 2022; Li et al., 2015). These classifications of modified nucleotides are expected to evolve with the development of new quantitative tools with higher sensitivity, such as liquid chromatography coupled with mass spectrometry (LC-MS), allowing for the detection of novel modifications.

Furthermore, current representations lack information about the temporal and spatial dynamics of modified bases. For example, m⁶A has been shown to occur reversibly in vivo, highlighting the discovery of demethylation (Jia et al., 2013; Wang & He, 2014; Yi & Pan, 2011). Similarly, environmental conditions can alter the profile of RNA modifications in cells (Chan et al., 2010; Gu et al., 2014; Nawrot et al., 2011; Wang & He, 2014). Additionally, there is a lack of information regarding the sequence in which these modifications are introduced and their interdependencies. Expanding on these observations is crucial to investigate the mechanisms underlying the regulation of these modifications, their specific roles in various biological processes, and their impact on cellular function under different environmental conditions. Understanding the interplay between different modifications and their regulatory networks could provide deeper insights into RNA biology and its evolution.

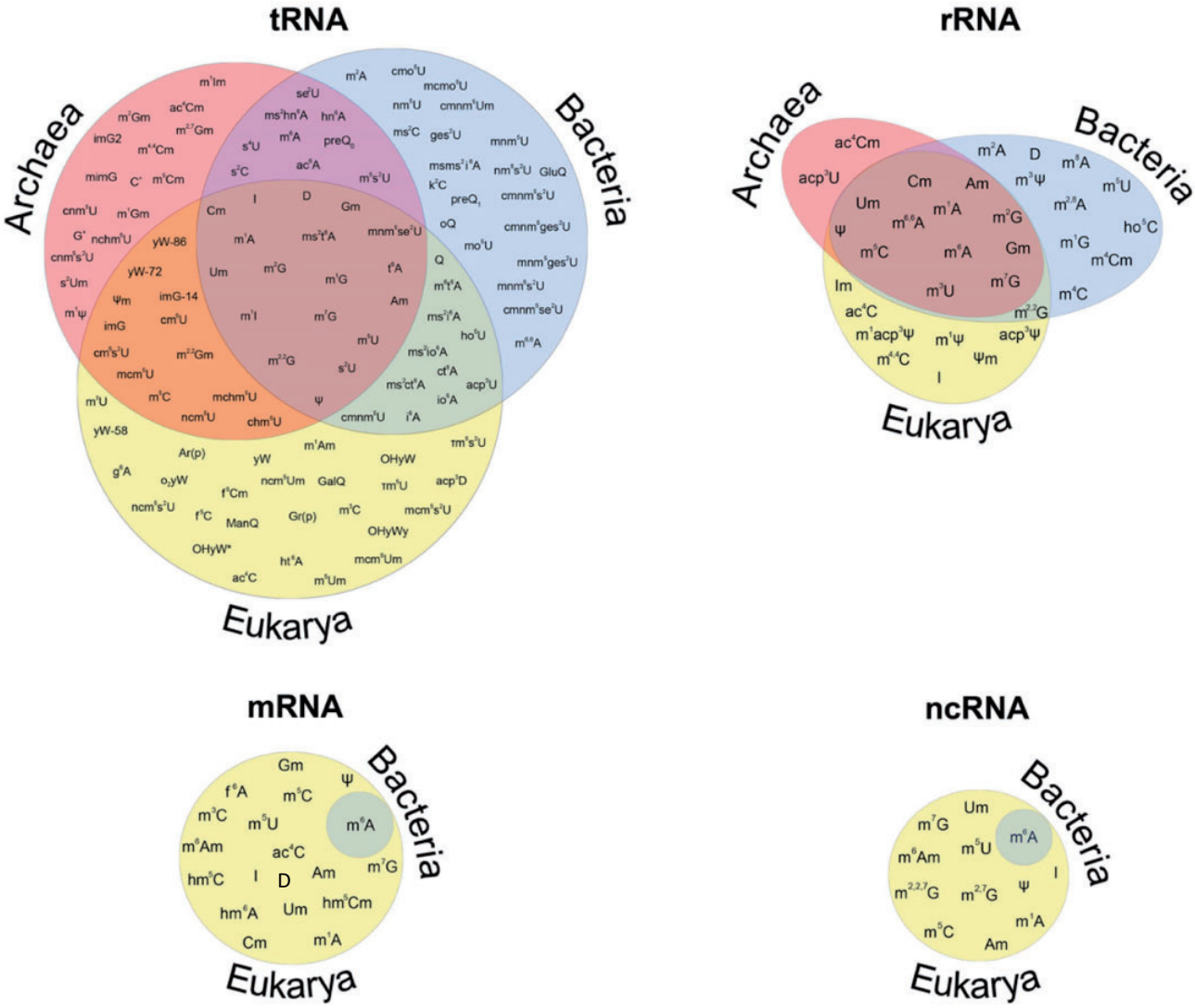


Figure 14: Euler diagrams showing the currently known phylogenetic distribution of ribonucleoside modifications in tRNA, rRNA, mRNA, and ncRNA classes. Archaeal modifications are in pink, bacterial modifications are in blue, and eukaryotic modifications are in yellow. Adapted from (McCown et al., 2020).

3.3 RNA modifications and functional roles

RNA modifications play huge number of biological roles, and I am not going to mention all of them. In particular the RNA modifications that were of particular interest for my thesis are those observed in rRNA (see next chapter).

RNA modifications are well known to provide stabilizing effects in the conformations of bases and ribose by altering local environments (Yokoyama, 1987). These effects then contribute to the formation and stabilization of higher-order structures of RNA molecules (Lorenz et al., 2017). RNA modifications regulate RNA-binding interactions with proteins, RNA processing, decoding of the genetic code, alteration of genetic information, and translational regulation (Suzuki, 2021). The dynamic nature of some of these RNA modifications (m6A, m5C, m1A…) have a direct impact on gene expression regulation and have hence generated a sense of the birth of a new field referred to as 'epitranscriptomic'. Modified 5′-cap structure protects the mRNA from degradation and assists in the recruitment of ribosomes. In addition, RNA modifications can function as determinant for RNA transport and localization in the cell (Lamond, 1990) or antagonize RNA sensors to mediate escape from the innate immune system (Karikó et al., 2005). Obviously, because of these important cellular functions, it is well documented that aberrant RNA modification can lead to various diseases such as cancer, diabetes, neuromuscular diseases, developmental disorders, autoimmune diseases, and mitochondrial diseases (Suzuki, 2021).

In bacteria, deletion of RNA modification enzymes can decrease expression of bacterial virulence factors and have been suggested as a strategy to make live-attenuated vaccines. rRNA modifications also adjust the strength of ribosome subunits contacts, maintain rRNA folds at the functional interfaces, and influence the action of antibiotics. Thermophilic organisms exhibit characteristic tRNA modifications that are dynamically regulated in response to varying growth temperatures, thereby bolstering fitness in extreme environments (Ohira & Suzuki, 2024) Unexpectedly, 5′-cap can include coenzymes such as nicotinamide adenine dinucleotide (NAD) or flavin adenine dinucleotide (FAD), and the coenzyme derivatives dephospho-

coenzyme A (dpCoA) and thiamine adenine dinucleotides. They have been shown to be accepted by RNA polymerases and can serve as noncanonical initiating nucleotides during transcription.

Beyond their biological role, RNA modifications can be repurposed for RNA-based therapies. For instance, post-transcriptional modifications can reduce the immunogenicity of RNA therapies by preventing activation of Toll-like receptors (Ψ, 2sU, m5C, m6A, m5U) (Karikó et al., 2005). More recently, N1-methyl-pseudouridine (m1Ψ) modification has been used in COVID-19 mRNA vaccines, which increased protein expression (Andries et al., 2015).

3.4 Post-transcriptional modification in rRNA

During ribosome biogenesis, precursor ribosomal RNAs (rRNAs) undergo post-transcriptional processing, which includes cleavage and trimming of the leader and trailer sequences, as well as chemical modifications, post-transcriptional modifications of ribosomal RNA have been shown to be another source of ribosome heterogeneity and can either target base or ribose moieties such as methylation, pseudouridylation, and base modification (Trahan & Oeffinger, 2023).

The rRNAs are the 2nd heavely modified RNA spicies. These rRNA modifications are concentrated in the functional sites of the ribosome, including the peptidyl transferase center (PTC), the nascent polypeptide chain exit tunnel, the subunit interface, and the binding sites for mRNA and tRNAs (Noeske et al., 2015; Polikanov et al., 2015).

In *E. coli,* 17 species of known modified nucleosides are present at 36 positions in the 16S and 23S rRNAs. The complete 23S rRNA modification pattern has only been determined for a few bacterial species, and it is challenging to ascertain when a modification dataset is truly comprehensive. It is generally accepted that the full rRNA modification pattern in *E. coli* has been elucidated, revealing 25 modifications in the 23S rRNA and 11 in the 16S rRNA. Specifically, *E. coli* 23S rRNA comprises 14 methylations, nine pseudouridines, one methylpseudouridine, one dihydrouridine, and one hydroxycytidine. The 16S rRNA count one pseudouridine Ψ and ten methylations. All these modifications are categorized and classified in **Table 1**.

Table 1: List of *E. coli* ribosomal 23S RNA post-transcriptional modifications and their modifying enzymes. In light blue 16S rRNA modifications, green 23S rRNA modifications.

Modification	Enzyme	Modification	Enzyme	Modification	Enzyme
Ψ -516	RsuA	m^1G -745	RlmA	Gm -2251	RlmB
m^7 -G527	RsmG	Ψ -746	RluA	m2G -2445	RlmKL
m^2 -G966	RsmD	m^5 -U747	RlmC	D -2449	RdsA (see chapter II)
m^5 -C967	RsmB	Ψ -955	RluC	Ψ -2457	RluE
m^2 -G1207	RsmC	m^6A -1618	RlmF	Cm -2498	RlmM
m^4 -C1402	RsmH	m^2G -1835	RlmG	ho^5C -2501	RlhA
Cm -1402	RsmI	Ψ -1911	RluD	m^2A -2503	RlmN
m^5 -C 1407	RsmF	$m^3Ψ$ -1915	RlmH	Ψ -2504	RluC
m^3 -U1498	RsmE	Ψ -1917	RluD	Um -2552	RlmE
m^2 -G1516	RsmJ	m^5U -1939	RlmD	Ψ -2580	RluC
m_6^2 -A1518	RsmA	m^5C -1962	RlmI	Ψ -2604	RluF
m_6^2 -A1519	RsmA	m^6A -2030	RlmJ	Ψ -2605	RluB
		m^7G -2069	RlmKL		

Over the last decades, the biogenesis of these rRNA modifications has been studied extensively, and most of the responsible methyltransferases and pseudouridylases have been identified (Golovina et al., 2012). In bacteria, RNA modifying enzymes are typically specific to a single site and vary in their substrate preferences, such as unmodified rRNA, assembly intermediates, or mature ribosomal subunits. Different stages of ribosome biogenesis may dictate the actions of specific modification enzymes, with most 23S rRNA modifications occurring during the early steps of assembly. This contrasts with 16S rRNA modifications, many of which are applied to the assembled 30S subunit (Kirpekar et al., 2018).Collectively, these modifications are essential for fine-tuning ribosomal function to ensure optimal cellular fitness (Ge & Yu, 2013; Sergeeva et al., 2015). Genetic and biochemical studies have characterized many rRNA modifications, uncovering their roles in translation fidelity(Baudin-Baillieu et al., 2009; Baxter-Roshek et al., 2007; Kimura & Suzuki, 2010), efficient rRNA processing (Connolly et al., 2008; Ito, Akamatsu, et al., 2014; Ito, Horikawa, et al., 2014; D. Lafontaine et al., 1995; Meyer et al., 2016), subunit assembly (Arai et al., 2015; Kimura et al., 2012), antibiotic resistance (Wilson, 2014), the virulence of pathogenic bacteria (Garbom et al., 2004; Kyuma et al., 2015; Su et al., 2007), and the evasion of innate immunity in eukaryotic hosts (Oldenburg et al., 2012).

4. rRNA modifying enzymes: Mechanisms, Substrate(s) Recognitions, Phenotypes and beyond

In this chapter, we will introduce the enzymology of rRNA modifications, primarily in *E. coli*. We will demonstrate that it is crucial to study the enzymatic mechanism of these rRNA modification enzymes to understand the molecular role of rRNA modification in the ribosomal function.

As previously mentioned, the 70S ribosome in *E. coli* comprises 36 modified nucleotides: 23 methylated, 10 pseudouridines, 1 methylated pseudouridine, 1 dihydrouridine, and 1 hydroxycytidine. These modifications are primarily located in the functional centers of the ribosome, such as the decoding and peptidyl transferase centers, as well as at the ribosomal subunit interface. The conservation of many of these modified nucleotides across bacterial species indicates their functional importance. Modification enzymes, which exhibit specific activity toward assembly intermediates, can block further assembly until the preceding steps are accurately completed. Before my PhD, all rRNA modification enzymes in *E. coli* were identified except for the rRNA dihydrouridine synthase (See later). These enzymes differ in their substrate preferences, targeting rRNA, assembly intermediates, or mature ribosomal subunits. Each stage of ribosome biogenesis is timed to coincide with the action of specific modification enzymes, emphasizing their crucial role in ribosome assembly and related processes.

Interestingly, while most 16S rRNA modifications occur during the final stages of 30S subunit assembly, the majority of 23S rRNA modifications happen at the early stages. None of the 16S rRNA modifications are required for the functional activity of in vitro reconstituted ribosomes. However, within the 23S rRNA, only six modifications near the peptidyl transferase center in the V domain are necessary for the reconstruction of functional ribosomal particles. An 80 nucleotide region in the 23S rRNA (positions 2445 to 2523) has been identified as essential for the 50S subunit assembly in *E. coli.* This region requires only six specific post-transcriptional modifications (m2G2445, D2449, ψ2457, Cm2498, m2A2503, and ψ2504) for the assembly of the 50S subunit or for peptidyl transferase activity. Notably, no single rRNA modification is indispensable for cell survival when considered individually.

4.1 The pseudouridines and their synthases

4.1.1 Pseudouridine synthase families

Pseudouridine synthases are the enzymes responsible for the most abundant posttranscriptional modification of cellular RNAs. These enzymes catalyze the site-specific isomerization of uridine residues that are already part of an RNA chain and appear to employ both sequence and structural information to achieve site specificity.

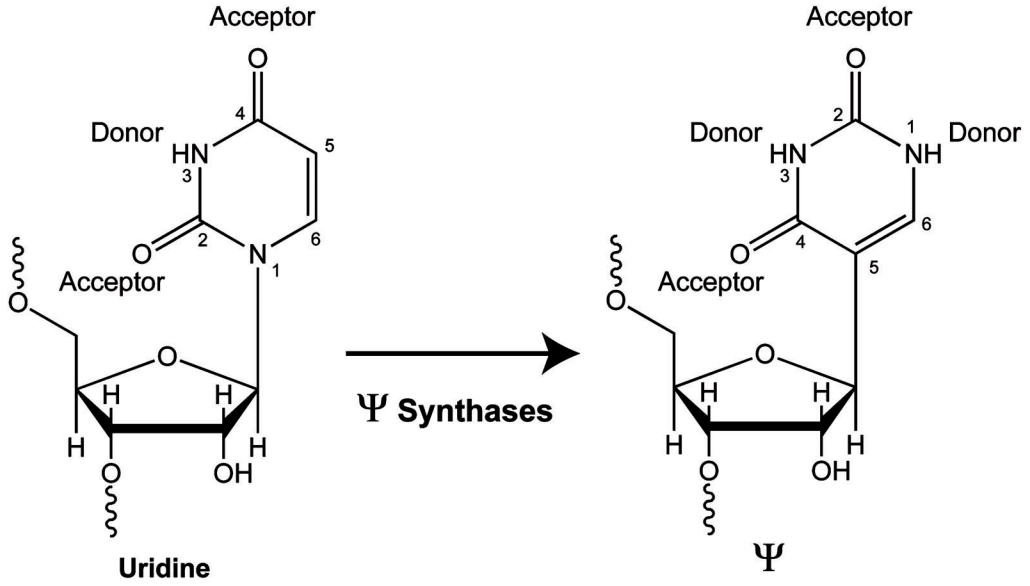


Figure 15: Catalyzed Reaction of Pseudouridine synthases.

The 70S ribosome in *E. coli* contains 10 pseudouridines and 1 methylated pseudouridine, 1 in 16S rRNA and 10 in 23S rRNA **(Figure 16)**. Pseudouridine synthases from archaea, bacteria, and eukarya are categorized into five families: RluA, RsuA, TruA, TruB, and TruD (**Table 2 & Figure 17**) (Hamma & Ferré-D'Amaré, 2006) Despite limited sequence similarity, all these synthases share a conserved core structure with an eight-stranded mixed β sheet and a catalytic cleft containing a crucial aspartate residue. Each family features additional secondary structures and distinct N- or C-terminal extensions. Notably, some RluA and RsuA enzymes have S4-like N-terminal domains. RluA and RsuA are closely related, sharing three conserved motifs. TruB enzymes align with RluA and RsuA sequences but lack one specific motif, indicating evolutionary relationships among the families.

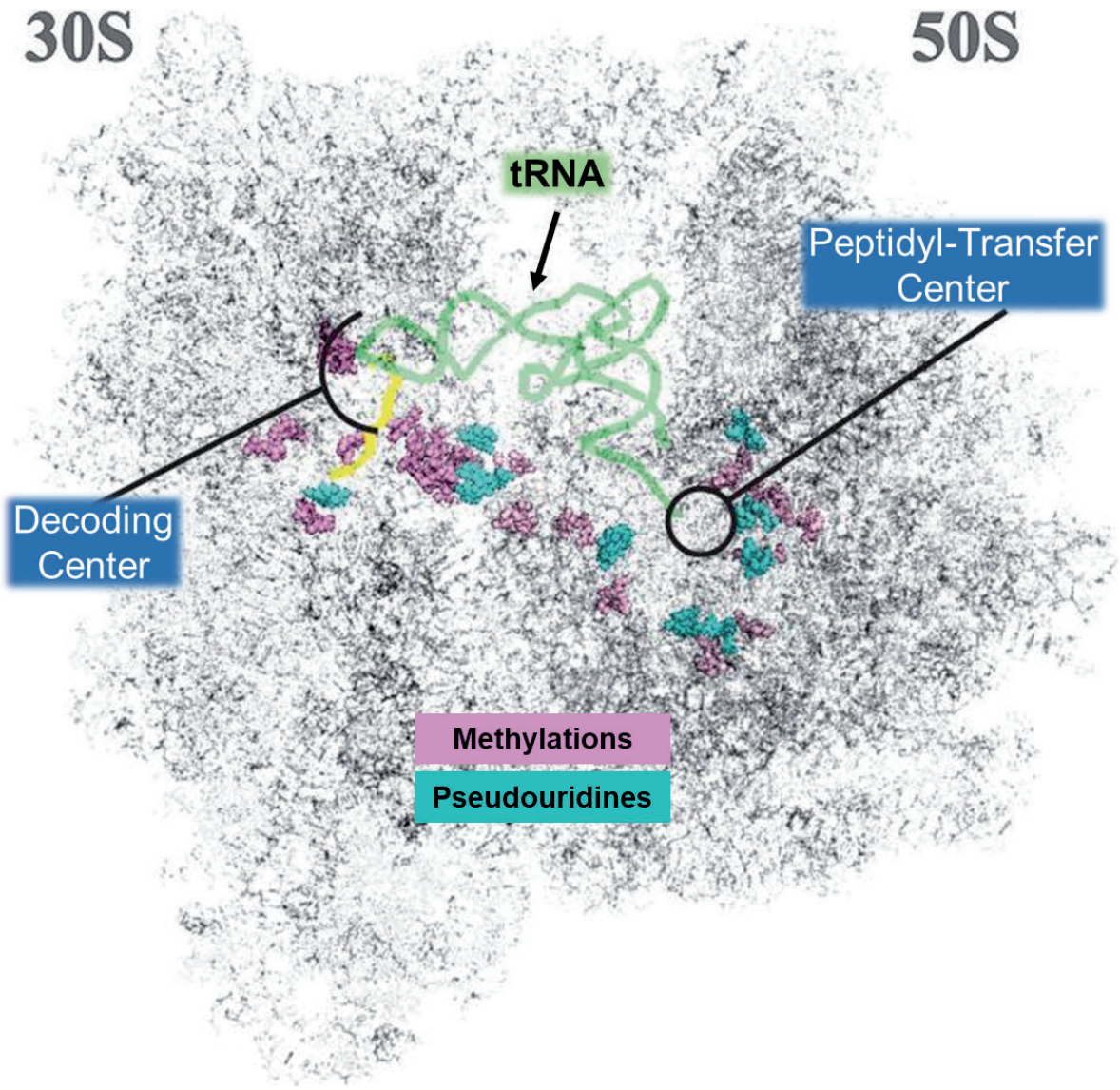


Figure 16: adapted from (O'Connor et al., 2018)**Modified nucleosides in the *E. coli* ribosome** (PDB 5AFI). Pseudouridines (cyan) and methylations (magenta) are clustered around the decoding site (DC), peptidyltransferase center (PTC), and tRNA binding sites (P-site tRNA in green). The mRNA is depicted in yellow. The Ψ modifications, with the relevant PUS responsible for the isomerization(s) in parentheses, are as follows: 16S rRNA, Ψ516 (RsuA); 23S rRNA, Ψ746 (RluA), Ψ2605 (RluB), Ψ955, Ψ2504, Ψ2580 (RluC), Ψ1911, Ψ1915, Ψ1917 (RluD), Ψ2557 (RluE), and Ψ2604 (RluF).

Table 2 Pseudouridine synthase families, adapted from (Mueller EG & Ferré-D'Amaré AR, 2000)**.**

Name	Substrate RNA	Modification Site	Catalytic Aspartate	PDB ID
TruD Family				
TruD	tRNA	13	Asp80	1SB7, 1SI7, 1SZW (1Z2Z)
TruA Family				
TruA	tRNA	38, 39, 40	Asp60	1DJ0, 2NRE, 2NQP, 2NR0 (1VS3)
TruB Family				
TruB	tRNA	55	Asp48	1K8W, 1R3F, 1ZL3, (1R3E, 1SGV, 2AB4, 1ZE1, 1ZE2)
Cbf5p	rRNA, etc.	Many	Asp81	(2APO, 2AUS, 2EY4, 2HVY, 2RFK)
RsuA Family				
RsuA	16S rRNA	516	Asp102	1KSK, 1KSL, 1KSV, (1VIO)
RluB	23S rRNA	2605	Asp110	
RluE	23S rRNA	2457	Asp69	2OML
RluF	23S rRNA	2604	Asp107	2GML
RluA Family				
RluA	23S rRNA	746	Asp64	2I82
	tRNA	32		
RluC	23S rRNA	955, 2504, 2580	Asp144	1V9K, 1XPI
RluD	23S rRNA	1911, 1915, 1917	Asp139	1QYU, 1PRZ, IV9F, 2IST
TruC	tRNA	65	Asp54	
Pus10 Family				
Pus10p	tRNA	55	Asp344	(2V9K)

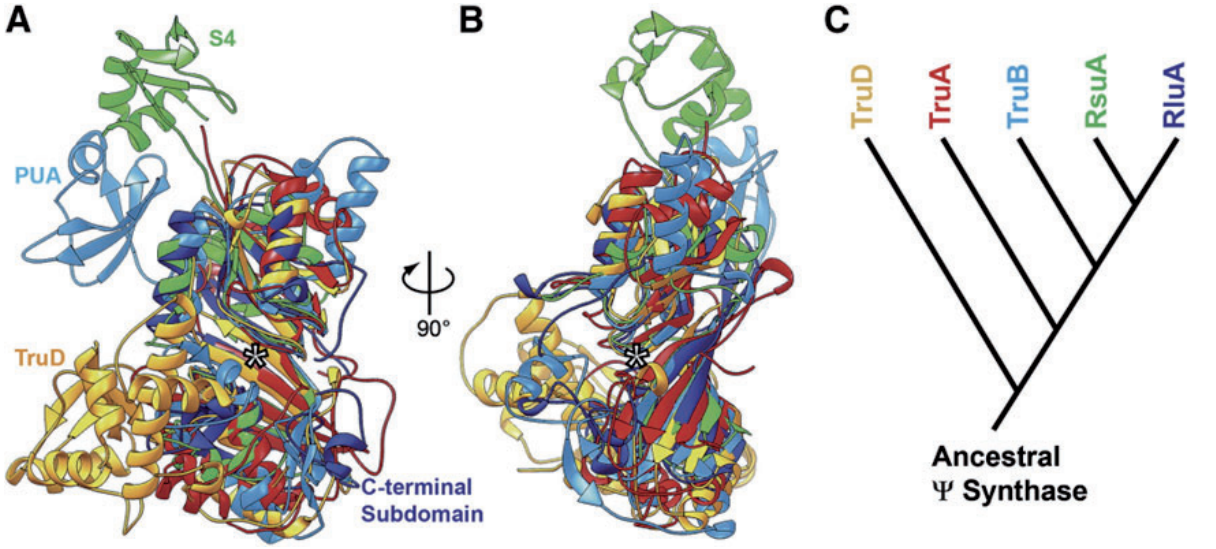


Figure 17: adapted from (Hamma & Ferré-D'Amaré, 2006) **Structural Comparison of Members of the Five Families of Ψ Synthases.** (A) Superposition of ribbon representations of the structures of TruD (yellow), TruA (red), TruB (cyan), RsuA (green), and RluA (navy) all from *E. coli*. The core domain of all enzymes is very similar. In this front view, the active site cleft bisects the core domain horizontally. Peripheral domains (TruD, PUA, S4, and C-terminal subdomain) unique to specific enzymes are also indicated (see Table 2). The active

site cleft is indicated by an outlined black asterisk. (B) View rotated 90° along the vertical axis. (C) Hypothetical evolutionary relationship between Ψ synthase families based on their sequences and structures.

4.1.2 The chemical mechanism of uridine isomerization: a still open issue

The conserved structure and active site residues of pseudouridine synthases (Ψ synthases) suggest a shared chemical mechanism. Mutation of the essential aspartic acid residue results in complete loss of activity, indicating its critical role beyond typical acid/base functions. Two proposed mechanisms involve this aspartate acting as a nucleophile (Mueller EG & Ferré-D'Amaré AR, 2000) one suggests a Michael addition at C6 of uracil, forming a covalent intermediate that repositions for a new C-C bond (**Figure 15**), while the other proposes nucleophilic attack at C1′, forming an acylal intermediate (**Figure 15**). Experimental evidence from fluoro-5-uridine (f5U) suggests the first mechanism, but cocrystal structures indicate conversion to a C-glycoside rather than a covalent adduct. Both mechanisms accommodate observed hydrated products, leaving the exact pathway uncertain. The identity of general acid/base residues remains unclear, with aspartate and tyrosine being potential candidates.

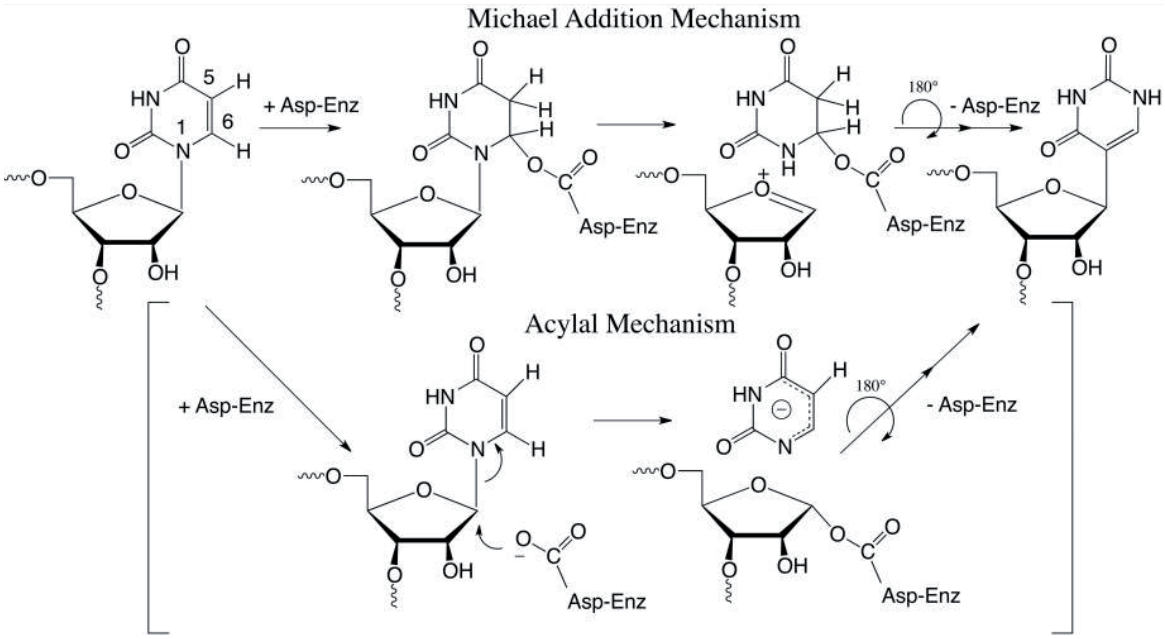


Figure 18: Proposed mechanisms for Ψ synthases

More recently the group of Stroud studied the mechanism of pseudouridine synthase RluB using crystallography as a tool (Czudnochowski et al., 2014). The 1.3 Å crystal structure of RluB, which catalyzes the conversion of U2605 to Ψ2605 in the 23S rRNA of *E. coli*, with a modified substrate containing the 5-fluorouridine at the target site revealed a covalent bond between the

isomerized base and tyrosine 140. This structure, compared to the RluB catalytic domain alone (2.5 Å), shows the stem-loop with 5-FU2605 in the active site. According to the authors, these results are rather in favor of the proposed Michael addition mechanism. Again, further research is needed to confirm the precise mechanism and the roles of active site residues in Ψ synthases.

4.1.3 The ribosomal pseudouridine synthases and substrate(s) specificities

Pseudouridine (Ψ) synthases in *E. coli* exhibit a remarkable diversity in their molecular recognition mechanisms, reflecting the complex structural contexts in which pseudouridine sites are found within ribosomal RNA (rRNA). We will discuss some of these enzymes, such as RusA, RluA, RluD, RluE, and RluF, which demonstrate both similarities and distinct differences in how they recognize and modify specific uridine residues.

RsuA (Ψ516 Synthase)

The RsuA Ψ516 synthase in *E. coli* exemplifies a dependence on RNA-protein interactions for substrate recognition (Wrzesinski et al., 1995). Studies have shown that while protein-free or truncated 16S rRNA transcripts are not substrates, the assembly of ribonucleoprotein (RNP) particles, containing both RNA and ribosomal proteins, facilitates efficient Ψ516 modification (Denman et al., 1989). This suggests that RsuA requires a structured RNA environment, likely facilitated by ribosomal proteins, for optimal recognition and activity. Recently Mg^{2+} was shown to be crucial for stabilizing RNA structures, including the 30S 5′-domain rRNA, which is necessary for RsuA binding (Jayalath et al., 2020). This study investigates the binding affinity of RsuA to the 5′-domain rRNA at varying Mg^{2+} concentrations using a FRET-based assay. The highest affinity was observed at 8 mM Mg^{2+} (Kd = 0.56 ± 0.04 nM), while higher Mg^{2+} concentrations decreased binding stability. RsuA also binds more tightly to helix 18 pseudoknot mutants than the wild type, indicating a preference for intermediates over native structures. Furthermore, the binding of RsuA and ribosomal protein S17 shows thermodynamic cooperativity, enhancing RsuA's binding and pseudouridylation activity, with the highest activity seen in the presence of S17 alone (Jayalath et al., 2020). The peripheral S4-like domain of RsuA increases binding stability, suggesting it aids in recognizing specific RNA conformations. Again, all these findings suggest RsuA’s role in ribosome assembly involves targeting specific RNA intermediates.

RluA (Dual-Specificity Synthase)

RluA is a dual specificity pseudouridine synthase capable of modifying Ψ746 in 23S rRNA and Ψ32 in tRNAPhe. Unlike RsuA, RluA can use an in vitro synthesized 23S RNA transcript as a substrate, with higher reactivity observed in truncated RNA fragments compared to full-length transcripts. Notably, the presence of EDTA enhances the reaction rate, indicating that RluA's substrate recognition might be influenced by metal ion concentration rather than RNA structure. The dual specificity of RluA is attributed to structural and sequence similarities between the stem-loop regions of 23S rRNA and tRNAs. Specifically, the consensus sequence UUN(A/C)AAA, located 3′ to the Ψ site, serves as a primary recognition determinant for RluA, underscoring its ability to discern specific RNA sequences across different RNA molecules.

In that respect, an overall structure of the RluA-RNA complex, focusing on how the enzyme RluA interacts with its RNA substrate, specifically the anticodon stem-loop (ASL) of *E. coli* $tRNA^{Phe}$ was obtained, helping the understanding of the molecular mechanism of RluA specificities (Hoang et al., 2006) (**Figure 19**). The interaction between RluA and its RNA substrate involves intricate molecular recognition mechanisms that ensure both specificity and efficiency in the pseudouridylation process. At the molecular level, RluA makes extensive contacts with the major groove of the ASL RNA. These interactions are crucial for stabilizing the RluA-RNA complex and positioning the substrate for catalysis. Upon binding, uridine 32 (U32) is flipped out from the RNA helical stack and inserted into the active site pocket of RluA. Here, interactions occur between the guanidino group of RluA residue R62 and the bases of adjacent nucleotides (A31 and U33), involving cation-π interactions, salt bridges, and hydrogen bonds.

These interactions not only stabilize the enzyme-substrate complex but also play a critical role in positioning the RNA for the catalytic pseudouridylation reaction. Furthermore, structural rearrangements induced upon binding are essential for the enzymatic activity of RluA. These include flipping out several nucleotides from their canonical positions and establishing new base-pairing interactions that are pivotal for the modification process. Mutational and structural analyses have underscored the importance of specific residues, particularly U33 and A36 of its consensus sequence ΨUXXAAA, in RNA recognition and enzymatic activity (**Figure 19C**). Mutations at these sites significantly impair RluA's ability to catalyze pseudouridylation, highlighting their role in substrate recognition and catalysis.

Although this discovery remains to be taken cautiously (that is why we still preferred to leave in the text the fact that it is a dual tRNA and rRNA enzyme), recent study from Pr Wendy

Gilbert lab reports the presence of pseudouridines in mRNA in *E. coli,* wherein RluA was identified as the predominant mRNA-modifying enzyme (Schaening-Burgos et al., 2023).
This pseudouridine synthase was shown to modify at least 31 of the 44 high-confidence sites identified in *E. coli* mRNAs. Using RNA structure probing data to inform secondary structures, they showed that the target sites of RluA occur in a common sequence and structural motif comprised of a ΨURAA sequence located in the loop of a short hairpin, recognition element is shared with identified target sites of RluA in tRNAs and rRNA.

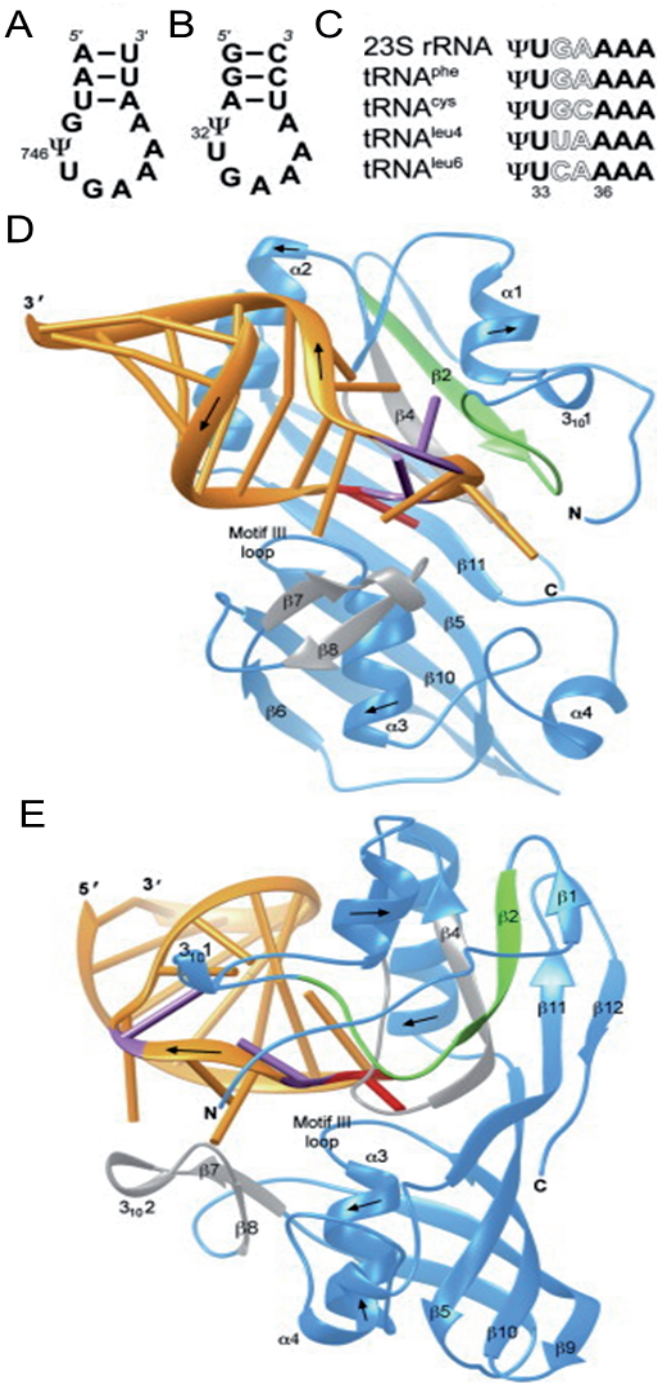


Figure 19: adapted from (Hoang et al., 2006) Overall Structure of the RluA-RNA Complex. (A and B) Sequences of the RluA targets in 23S rRNA (A) and in the ASL of *E. coli* tRNAPhe (B). (C) Comparison of sequences following sites of pseudouridylation by *E. coli* RluA. (D) Ribbon representation of the RluA-ASL complex structure. The protein is colored cyan, with motif I green and motif II and the thumb gray. RNA is yellow, the site of modification red, and nucleotides participating in a reverse-Hoogsteen base pair magenta. Except where noted, this color scheme is used throughout. (E) View rotated 90° along the vertical axis.

RluD (a multisite specificities Ψ1911, Ψ1915, Ψ1917 Synthase)

RluD is responsible for modifying Ψ1911, Ψ1915, and Ψ1917 in 23S rRNA. Early in vitro studies have demonstrated that RluD can specifically modify full-length 23S RNA transcripts under optimal conditions (Wrzesinski et al., 2000). However, at higher Mg2+ concentrations, RluD's specificity diminishes, leading to nonspecific modification of additional sites. This sensitivity to Mg2+ concentration suggests that RluD's substrate recognition is influenced by divalent metal ions, which can alter RNA conformation and accessibility of uridine residues. However recent study demonstrated that RluD showed higher efficiency on 50S ribosomal subunits compared to free 23S rRNA (Vaidyanathan et al., 2007) (**Figure 20**). This suggested that free RNA might not be the natural substrate of RluD. To explore this phenomenon, this research group examined RluD's activity on 50S subunits from a mutant strain lacking three pseudouridines in helix 69 (H69).

Even at low enzyme concentrations, RluD efficiently modified H69 within these subunits, indicating that RluD's physiological substrate is likely a structured ribonucleoprotein particle rather than free RNA (**Figure 20A**). Further assays revealed that RluD acts specifically on its target sites within the 50S subunits and does not induce nonspecific modifications. In contrast, activity on free 23S RNA was significantly lower and nonspecific, supporting the notion that RluD recognizes specific structural contexts present in the assembled subunits. Structural docking models show that RluD's catalytic and RNA binding domains interact with the H69 region in the 50S subunit, which could explain its specificity and efficiency (**Figure 20B**). This interaction is hindered in 70S ribosomes, where H69 forms a bridge with 16S RNA, blocking RluD access. Hence, RluD requires a properly folded RNA structure within ribosomal subunits for precise and efficient activity. This aligns with evidence that such enzymes also function as ribosomal assembly factors, ensuring proper ribosome formation and function (see next section).

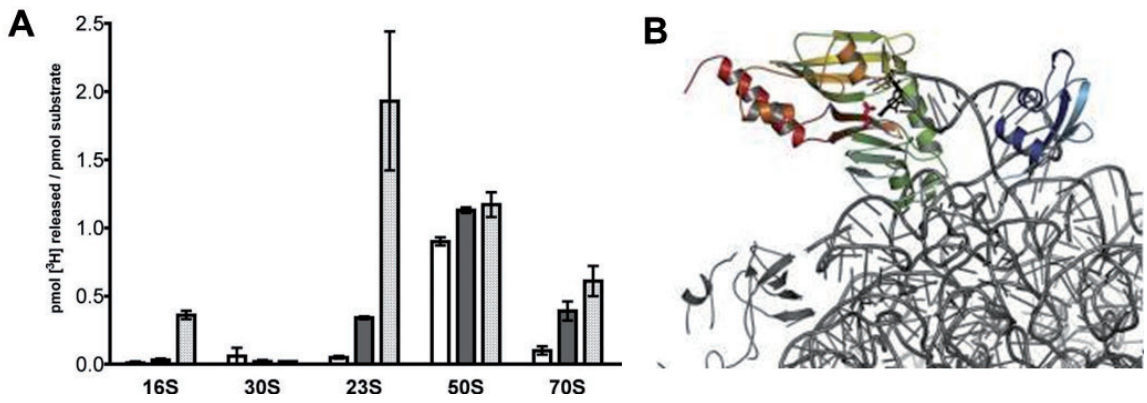


Figure 20: adapted from (Vaidyanathan et al., 2007)Substrate specificity of RluD and molecular recognition. (A) The pseudouridylation activity of increasing amounts (20 nM [open], 200 nM [gray], and 2 μM [stippled]) of purified, His-tagged, full-length RluD was compared on different RNA species extracted from ΔBED mutant cells. The substrates were maintained at a constant concentration of 200 nM of RNA. The substrates included 16S RNA (16S), 30S subunits (30S), 23S RNA (23S), 50S subunits (50S), and 70S ribosomes (70S). (B) Model of RluD docked on the H69 stem–loop in the 50S subunit, poised to modify U1915. A stereo view shown focuses on H69 in the 50S (gray cartoon) with RluD represented by a colored ribbon. Full-length RluD (PDB ID: 2IST) manually docked on to the H69 stem–loop of the *E. coli* 50S subunit (PDB ID: 2AWB). The catalytic Asp (D139) of RluD and U1915 of 23S RNA are depicted as red and black stick models, respectively.

RluE (Ψ2457 Synthase)

In contrast to RluD, RluE exhibits a simpler recognition mechanism. It can modify Ψ2457 in vitro on free 23S RNA without requiring additional RNA-binding proteins or complex RNA structures. This suggests that RluE may recognize and modify specific uridine residues based primarily on local RNA sequence context rather than extensive tertiary RNA structures. A recent study identified key regions in 23S rRNA and the pseudouridine synthase RluE crucial for RNA recognition and enzymatic activity (Tillault et al., 2018) (**Figure 21**). The single-stranded regions flanking helix 89 (H89) in 23S rRNA, along with the upper portion of H89, are essential for optimal binding and pseudouridine formation by RluE (**Figure 21**). The large RNA element recognized by RluE, which spans more than 60 nucleotides, explains the enzyme's high specificity for a single uridine in rRNA. This complex requirement distinguishes U2457 from other uridines. Despite RluE's moderate size (25 kDa), it efficiently contacts the entire RNA, particularly the upper portion of H89 and its flanking single-stranded regions. RluE likely acts early in ribosome biogenesis when the target site is accessible. The interaction may involve the single-stranded regions wrapping around RluE and the N-terminal region reaching the tip of H89. Several structural elements within RluE, such as loops L7–8 and L1–2 and the N-terminal region, play significant roles in RNA recognition and catalysis. RluE's rapid

catalytic rate (1.7 s−1) makes it the fastest pseudouridine synthase characterized, likely acting quickly during early ribosome biogenesis (**Figure 21i**).

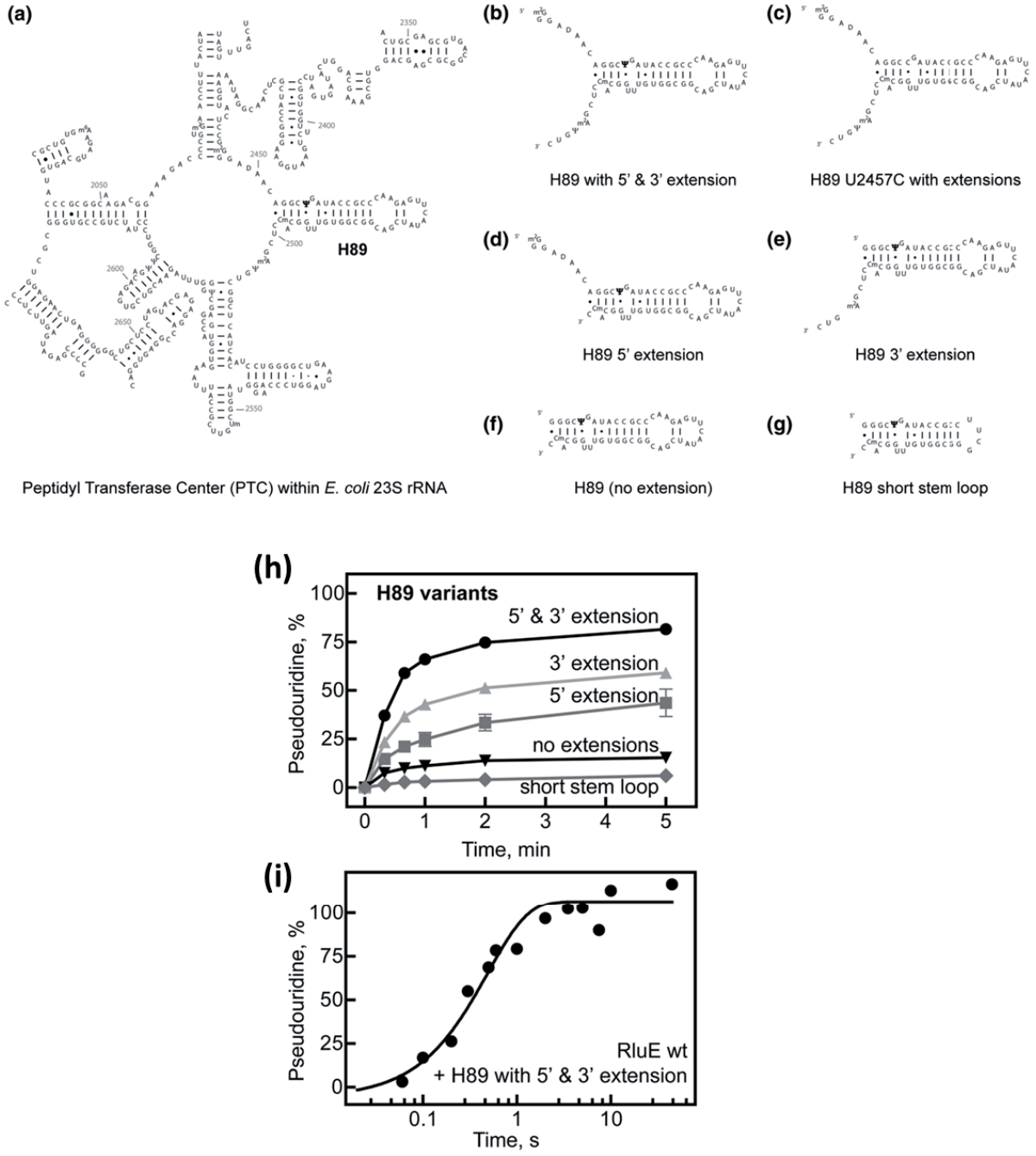


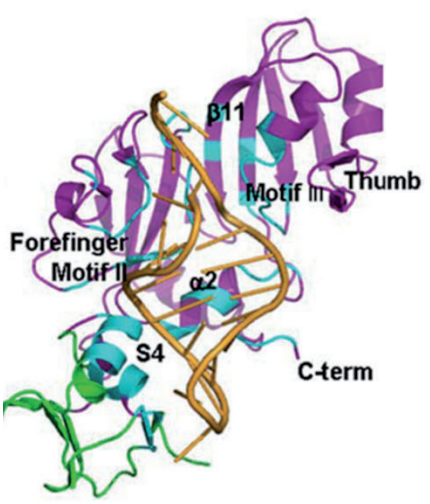
Figure 21: adapted from (Tillault et al., 2018) RNA constructs comprising H89 of the PTC of *E. coli* 23S rRNA to determine the minimal RNA recognition site for RluE and pseudouridylation activities (a) Representation of the secondary structure of the PTC of *E. coli* 23S rRNA. The positions of pseudouridines are indicated by Ψ. The target of RluE (Ψ2457) in H89 is represented in bold. (b) H89 (H89) with 5′ and 3′ single-stranded extensions (nt 2445–2507); (c) H89 U2457C with 5′ and 3′ extensions (nt 2445–2507) that cannot be modified by RluE; (d) H89 with 5′ extension only (nt 2445–2499); (e) H89 with 3′ extension only (nt 2453–2507); (f) H89 without single-stranded extensions (nt 2453–2499) and (g) H89 short stem–loop where the upper stem–loop of H89 was deleted and replaced with a tetraloop (nt 2453–2499: Δ2467–2483). (h) Comparison of pseudouridylation of different RNAs comprising H89 catalyzed by RluE wt. Time course of pseudouridine formation by RluE wt in different 23S

rRNA fragments under multiple-turnover conditions at 37 °C. H89 with 5′ and 3′ extensions (black circles); H89 5′ extension (dark gray squares), H89 3′ extension (gray triangles), H89 (no extensions) (black inverted triangles), and H89 short stem–loop (dark gray diamonds). (i) Pseudouridine formation in H89 with 5′ and 3′ extensions by RluE under single-turnover conditions. Single-exponential fitting of the time course determined a rate constant of pseudouridine formation of 1.7 ± 0.5 s− 1.

RluF (Ψ2604, Ψ2605 Synthase)

RluF is another pseudouridine synthase in *E. coli* that demonstrates high specificity for modifying Ψ2604 in 23S rRNA. Studies have shown that RluF can efficiently recognize and modify U2604 within both 23S RNA and 50S ribosomal subunits isolated from *rluB- and rluF*-deleted cells. Notably, RluF exhibits remarkable discrimination between U2604 and its adjacent nucleotide U2605, similar to the specificity observed in RluB. This specificity highlights RluF's ability to distinguish subtle differences in RNA sequence and structure, emphasizing its role in precisely modifying specific uridine residues within complex ribosomal RNA assemblies. Also, RluF exhibits varying levels of activity on different lengths of RNA stem–loop substrates. Longer RNA fragments, particularly those with extra nucleotides at the 5′ and 3′ ends of the stem–loop, show higher pseudouridylation activity compared to shorter ones. This highlights the importance of substrate length and sequence context in RluF activity.

An, overall Structure of RluF–RNA (22-mer) complex provided further information on the mechanism of substrate recognition and rearrangement (Alian et al., 2009) . Upon binding, RluF induces a significant rearrangement of the RNA stem–loop, including the refolding of A2602 into the stem and translation of nucleotides 3′ to A2602. This rearrangement aligns U2604 with the active site, facilitating its modification. The interaction between RluF and RNA involves multiple hydrogen bonds and ionic interactions, stabilizing the rearranged RNA structure and directing the site specificity. The flexibility of the central β-sheet and helix α2 in RluF allows for conformational changes necessary for substrate binding and catalysis. Structural similarities with other Ψ synthases (TruB, RluA) suggest a common mode of substrate recognition and catalysis, despite variations in substrate specificity and RNA-binding modes.



Take home message on pseudouridine synthases molecular recognition capacities

Overall, pseudouridine synthases in *E. coli* showcase a spectrum of molecular recognition capacities tailored to the diverse structural and sequence landscapes of ribosomal RNA. While some enzymes (like RsuA and RluA) rely on RNA-protein interactions and specific RNA sequences for substrate recognition, others (such as RluD RluE and RluF) exhibit simpler mechanisms that may be influenced by local RNA sequence context or metal ion concentrations. These findings underscore the adaptability and specificity of pseudouridine synthases in modifying uridine residues within the intricate framework of ribosomal RNA, essential for maintaining ribosomal function and protein synthesis fidelity in *E. coli* probably via molecular fine tuning.

4.1.4 The ribosomal pseudouridine synthases, phenotypes and other roles

Pseudouridines are known to enhance RNA stability by forming more stable base pairs with the four major bases compared to uridine (Chow et al., 2007). In the crucial H69 region of the *E. coli* ribosome, three pseudouridine residues strengthen both base-pairing and stacking interactions (Jiang et al., 2014).

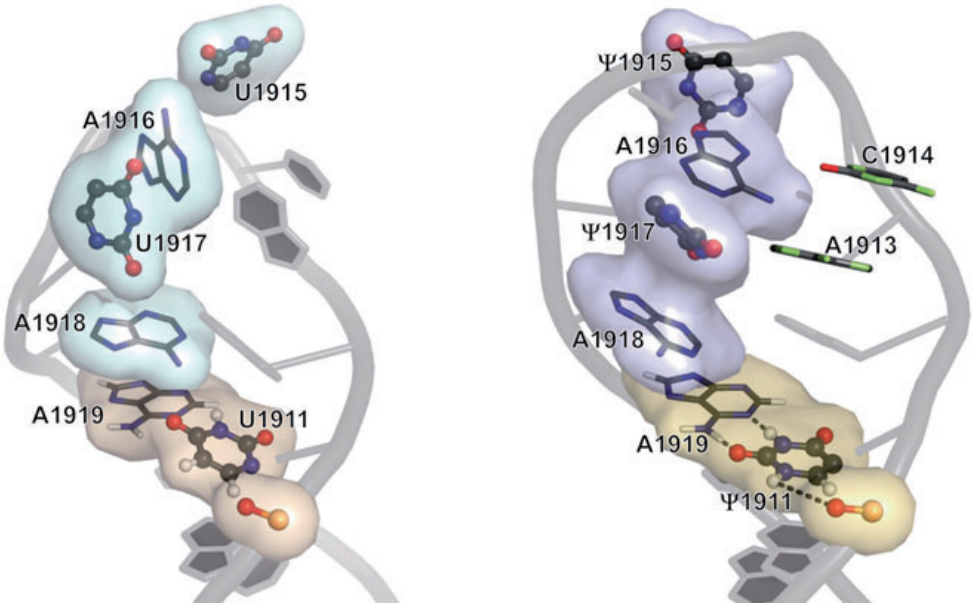


Figure 22: Structural comparisons of the loop regions of H69 UUU (left) and ΨΨΨ (right) taken from (Jiang et al., 2014). Ψ1915 and Ψ1917 alter the loop folding by promoting local base-stacking interactions, and the conformational effects are locked within the loop region by a stabilized Ψ1911–A1919 base pair, together with possible water-mediated hydrogen-bonding interactions involving Ψ1911N1H. The conformational behavior of residues in the 5' half of H69 loop region is affected by pseudouridylation.

High-resolution crystal structures of the *E. coli* ribosome reveal that six of the seven well-ordered pseudouridines make water-mediated contacts between their N1 imino groups and the rRNA phosphate backbone, which likely contributes to their stabilizing effects (Noeske et al., 2015).

Curiously, the removal of pseudouridines, achieved through deletion of corresponding synthase genes (**Table 3**), typically results in mild phenotypic effects under standard growth conditions (Ofengand & Del Campo, 2004). For instance, deletion of multiple Ψ synthase genes had minimal impact on growth in rich and minimal media, suggesting that pseudouridines may not be essential for basic cellular functions under these conditions.

Recently, Jaanus Remme's team at Ohio State University found that eliminating all pseudouridine (Ψ) modifications from *E. coli* ribosomes surprisingly has little effect on cell growth, ribosome biogenesis, or function (O'Connor et al., 2018). This discovery is unexpected given the high conservation of Ψ modifications in the ribosome's critical functional regions. In contrast, the loss of these modifications in eukaryotic systems greatly impairs ribosome assembly and function. The findings in *E. coli* are remarkably similar to those observed in the archaeon Haloferax volcanii (Blaby et al., 2011), prompting the question of why Ψ

modifications are nonessential in bacterial and archaeal rRNAs but crucial in eukaryotes (Jack et al., 2011). One significant difference is the much higher number of Ψ modifications in eukaryotic rRNAs (Del Campo et al., 2005). It is suggested that during evolution, eukaryotic ribosomes have become heavily reliant on the stabilizing properties of Ψ, rendering the loss of 40 to 90 such modifications intolerable.

Table 3: Effect of deletion of *E. coli* rRNA pseudouridine synthases on cell growth obtained from (Ofengand & Del Campo, 2004)

Deletion strain	RNA substrate	Ψ site(s)	Exponential growth rate (% of WT)					
			Rich (LB) medium			Poor (M9) medium		
			25°C	37°C	42°C	25°C	37°C	42°C
ΔrsuA	16S RNA	516	108	100	100	101	102	102
ΔrluA	23S RNA	746	102	109	110	94	105	106
	tRNA	32						
ΔrluB	23S RNA	2605	97	100	100	—	89	—
ΔrluC	23S RNA	955, 2504, 2580	99	101	99	92	94	98
ΔrluD	23S RNA	1911, 1915, 1917	—	16	—	—	—	—
ΔrluE	23S RNA	2457	103	104	106	—	100	—
ΔrluF	23S RNA	2604	97	104	111	—	98	—
ΔrluB, ΔrluC, ΔrluE, ΔrluF	23S RNA	955, 2457, 2504, 2580, 2604, 2605	—	98	—	—	85	—

The absence of RluD synthase, responsible for Ψ1911, Ψ1915, and Ψ1917 in helix 69 of the ribosome, significantly impairs large ribosomal subunit assembly and stability (Gutgsell et al., 2005) (**Table 3**). This defect can be mitigated by reintroducing the rluD gene, highlighting the direct involvement of RluD synthase and its products in ribosome assembly processes. Ψ1917,

in particular, is universally conserved across organisms, underscoring its critical role. However, and quite interestingly, O'Connor & Gregory isolated and characterized further ΔrluD mutants in *E. coli* and *S. enterica* (O'Connor & Gregory, 2011). While all reported genetic and biochemical analyses of RluD function in the literature have been carried out with K-12 strains of *E. coli* (used as a wild-type model organism), it carries a variant and defective RF2 protein, with a Thr residue at position 246, in contrast to all other bacteria (including other *E. coli* strains) which have Ala or Ser at this position. These authors examined the effect of Thr246 RF2 on the phenotypes conferred by the *ΔrluD* mutation, by replacing the K-12 (Thr246) prfB allele with the *E. coli* B (Ala246) allele. Surprisingly, deletion of the *rluD* gene in a strain carrying the *E. coli* B RF2 (Ala246) allele had no effect on growth, ribosomal subunit association, or stop codon readthrough (O'Connor & Gregory, 2011). Moreover, deletion of the *rluD* gene in *Salmonella enterica,* which also carries the typical bacterial Ala246 RF2 had negligible effects on these same parameters. These data indicated that the severe phenotypes of *ΔrluD* mutants reported in the literature are largely the result of a defective RF2 protein in the *E. coli* K-12 strains used for these analyses and are not typical of bacteria carrying fully active RF2 proteins.

In contrast, non-physiological pseudouridylation of ribosomal RNA can be detrimental. In 2017, Leppik et al. judiciously engineered a chimeric pseudouridine synthase (RluCD) containing N-terminal S4 domain of the RluC and C-terminal catalytic domain of the RluD that is able to introduce excessive pseudouridines into rRNA at nonnative positions (Leppik et al., 2017). The chimeric enzyme (RluCD) was used as a tool to study an effect of over-modification of rRNA on the ribosome biogenesis. Excessive pseudouridines in rRNA was shown to inhibit ribosome assembly (**Figure 23A**). Ten-fold increase of pseudouridines in the 16S and 23S rRNA made by a chimeric pseudouridine synthase leads to accumulation of the incompletely assembled large ribosome subunits. Eighteen positions of 23S rRNA were identified where isomerization of uridines interferes with ribosome assembly. Most of the interference sites are located in the conserved core of the large subunit around the peptide exit tunnel (**Figure 23B**).

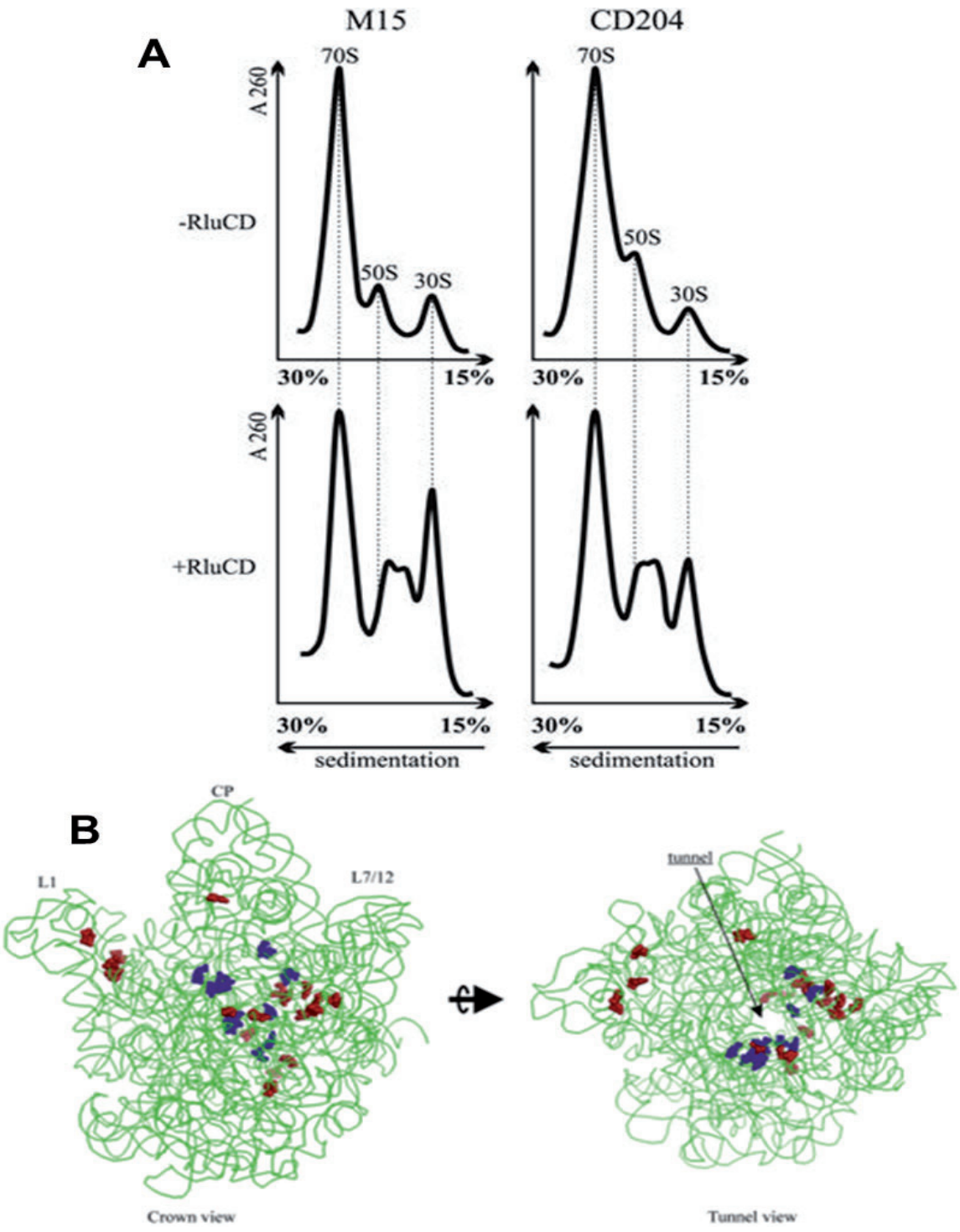


Figure 23: Effect of non-canonical pseudouridylations modified from (Leppik et al., 2017). (A) The chimeric protein was expressed under the control of inducible phage T5 promoter in the *E. coli* wild-type strain M15 and in the CD204 strain lacking genes for RluC and RluD. Sucrose gradient profiles of ribosomes of M15 and CD204 strains. M15 or CD204 cells grown without expression of chimeric RluCD pseudouridine synthase represent normal ribosomal profile in sucrose density gradient (upper panels). Dotted lines indicate the location of normal ribosomal particles in sucrose density gradients. (B) Distribution of RluCD dependent pseudouridines in 23S rRNA secondary and tertiary structures. Location of modification interference sites in the 23S rRNA tertiary structure. 50S subunit (coordinates from PDB ID:

3R8S) ‘crown’ view and ‘tunnel entry’ view. Blue—native pseudouridine positions; red—interfering pseudouridine positions.

Although RsuA is not vital under normal conditions, this enzyme had been nonetheless identified as a survival protein that plays a crucial role in the survival of bacteria under various environmental stress conditions (Amitai et al., 2009). A comparison of the growth curves of wildtype and RsuA knock-out *E. coli* strains illustrates that RsuA renders a survival advantage to bacteria under streptomycin stress (Abedeera et al., 2023). Moreover, the RsuA-dependent growth advantage for bacteria was found to be dependent on its pseudouridylation activity. Circular dichroism spectroscopy measurements and RNase footprinting studies have demonstrated that pseudouridine at position 516 influences helix 18 structure, folding, and streptomycin binding exemplifying the importance of bacterial rRNA modification enzymes during environmental stress.

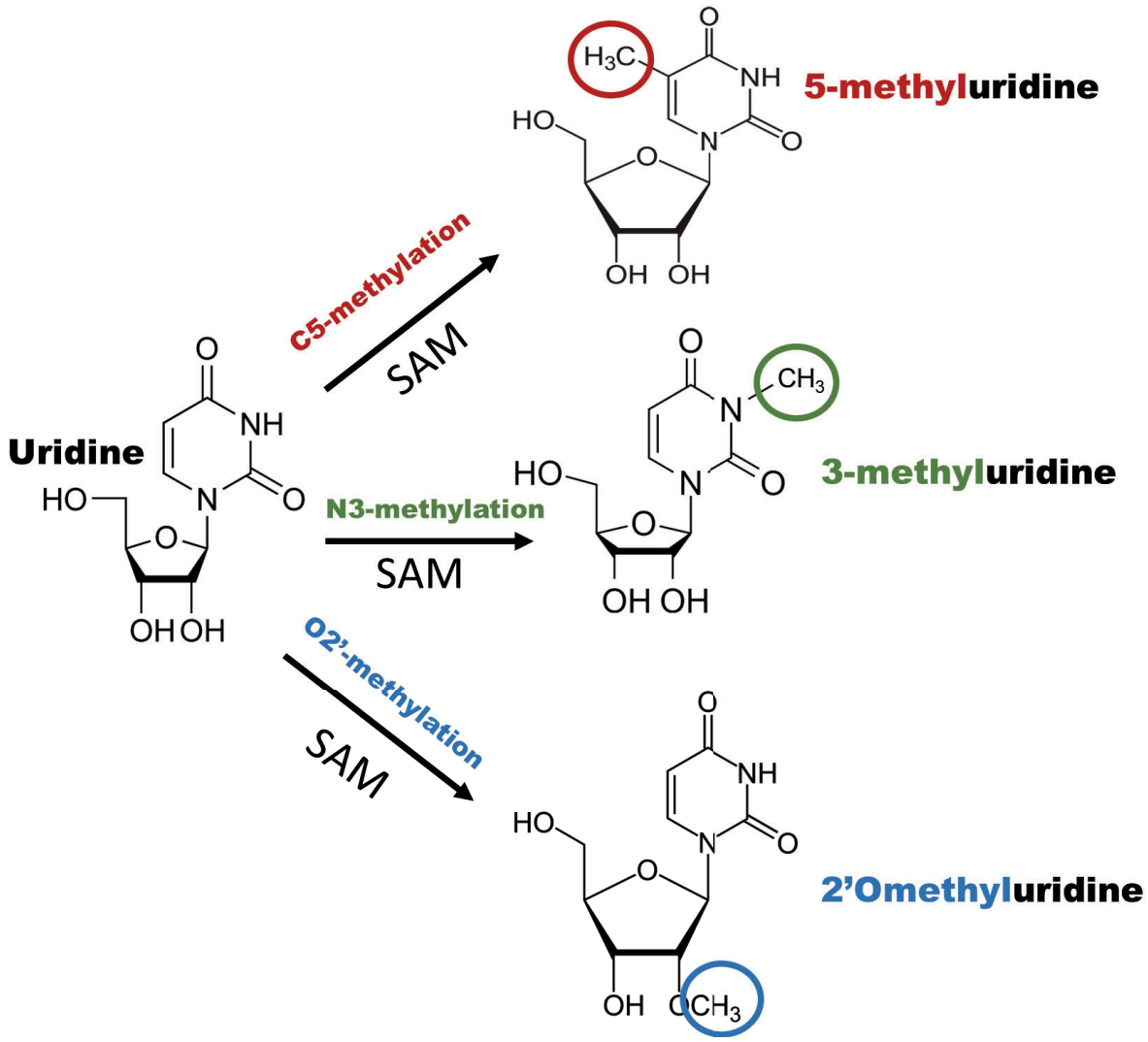
Overall, while the exact functional significance of pseudouridines in ribosomes remains complex and context-dependent, their presence appears crucial for optimal ribosomal function under various physiological conditions. Further exploration under different stress conditions is essential to fully elucidate their roles in cellular physiology and adaptability.

4.2 Methylations and rRNA methyltransferases

rRNA methyltransferases transfer a methyl group from a methyl donor to either a carbon (C-methylation) or nitrogen (cyclic or exocyclic; N-methylation) of the base, or to the oxygen of the ribose 2’OH group (O-methylation). Unlike pseudouridine synthases, which do not require other substrates or cofactors for the modification, methyltransferases, in addition to their RNA substrate, require a methyl group donor, typically SAM. However, there is an exception in mollicutes where methylation is folate-dependent (Sirand-Pugnet et al., 2020). In *E. coli*, methylation is solely SAM-dependent.

The 70S ribosome in *E. coli* comprises 23 methylated nucleosides. In bacteria, most methylated rRNA nucleotides are located on the surface of the ligand-free small ribosomal subunit, which render them accessible to the rRNA modifying enzymes (**Figure 16 & 24**).

♦ **Examples of rRNA methylations**



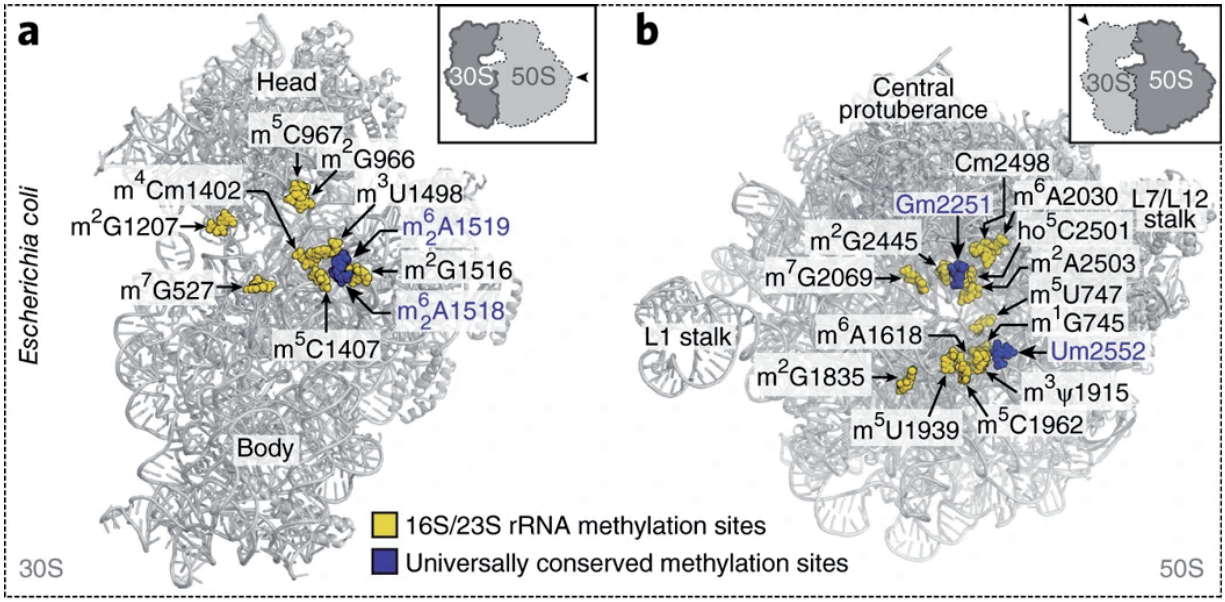


Figure 24: Spatial distribution of modified nucleotides in the structures of small (left) and large (right) ribosomal subunits from *E. coli* (PDB entry 4YBB19) adapted from (Sergiev et al., 2018). Both small and large subunits are viewed from their interface sides, as indicated by the insets. Methylation sites are highlighted in yellow. Universally conserved methylation sites that are the same among all taxonomic groups are shown in dark blue.

4.2.1 The rRNA methyltransferases fold

Five structurally distinct folds or classes have evolved to bind the SAM and perform methylation (Schubert et al., 2003). Seven of the eight known *E. coli* rRNA Methyltransferases (MTs) belong to class I and are grouped into six subfamilies. Class I MTs consists of alternating α-helices and β-strands leading to seven-stranded, parallel except for one strand, β-sheet flanked on both sides by α-helices (**Figure 25**). RsmC belongs to the RNA m2G MT family (Bujnicki, 2000); RsmB belongs to the RNA m5C MT family (Reid et al., 1999); KsgA belongs to the Erm family of m6A MT (Van Buul & Van Knippenberg, 1985); RlmAI (m1G745), previously named RrmA, belongs to the RlmAI/II family of RNA m1G MT (Liu & Douthwaite, 2002), both RlmC and RlmD (**Figure 25**) belong to the RNA m5U MT family (Agarwalla et al., 2002; Gustafsson, 1996); and RlmE (ex RrmJ) belongs to the RrmJ/fibrillarin family of 2′-O-MT (Feder et al., 2003). The SAM-binding site is in general in a cleft but the orientation/conformation of SAM in each structure can be different according to the different kinds of methylation. The MTs can have additional domains that are presumably used for substrate recognition/specificity like the N-terminal TRAM domain of RlmD (**Figure 25**) composed entirely of β-sheet and involved in RNA binding and its additional central domain with a [Fe4S4] cluster of unknown function (T. T. Lee et al., 2004).

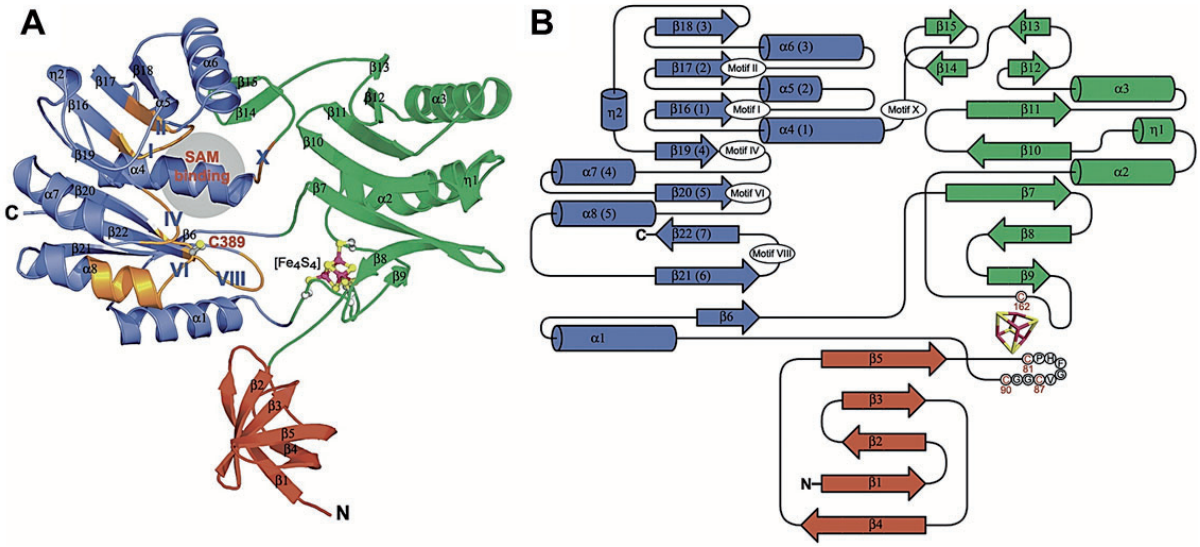


Figure 25: Overall Structure of RlmD adapted from (T. T. Lee et al., 2004) The model includes residues 15–431. The N-terminal, central, and C-terminal domains are colored in red, green, and blue, respectively. The α helices and β strands are each numbered sequentially. (A) Ribbon diagram of RumA. The catalytic cysteine (389), the [Fe4S4] cluster, and the side chains of its coordinating cysteines are shown in ball-and-stick model (magenta for Fe, yellow for S, and silver for C). The locations of the conserved MTase motifs are colored in gold and indicated in roman numerals. The potential SAM binding region is also marked. (B) Planar representation showing the topology of the same model. Classic numbering of the secondary structure elements in the MTase domain is also shown in parentheses.

All known class IV methyltransferases are members of the trefoil knot superfamily of methyltransferases (referred as SPOUT) identified by sequence similarity (Anantharaman et al., 2002). The SPOUT superfamily consists of the trmH family of RNA Gm MT, the trmD family of tRNA m1G methylase, and several other families of putative methylase. *E. coli* RlmB belongs to the trmH family, which consisting of a six-stranded, parallel β-sheet flanked by α-helices and an unusual C-terminal trefoil knot structure. RlmB also has an N-terminal domain that is most like ribosomal proteins rRNA-binding domain (**Figure 26**) (Michel et al., 2002).

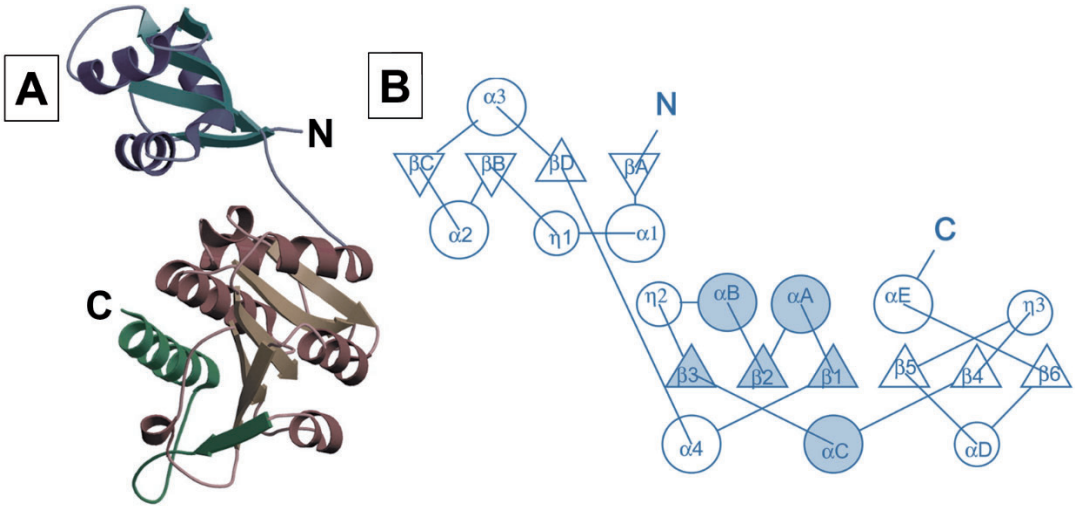


Figure 26: fold of *E. coli* RlmB adapted from (Michel et al., 2002) . (A) ribbon representation of the RlmB monomer. The N-terminal domain is in blue shades and the C-terminal domain is in red shades, with the knot region shown in green. (B) Topology of RlmB. The β strands are shown as triangles, α helices as large circles, and 3/10 helices as smaller circles.

4.2.2 The mechanism of rRNA methylations: A chemical diversity

Contrary to the shared chemical mechanism of Ψ synthases, the mechanism of rRNA methyltransferases differs according to the atom target on the nucleoside. While the N-methylation or O-methylation mechanism is rather a simple one involving a direct transfer of the methyl group from the electrophilic -CH3 group of SAM cofactor to the N or O atom of the base or ribose via an SN2 type mechanism, the mechanisms involving C-methylations are much more complex and require a preliminary activation of the sp2 carbon (C5 for U and C; or C2 for A) of the base. Here we are going to rather focus on C-methylations because of their more complex chemistries.

Activation of the C5 pyrimidine atom by Mikael addition

C5-pyrimidine atom is a poorly reactive sp2 carbon, thus requiring an activation step prior to its modification. Usually, this activation is carried by Michael-like addition of a cysteine residue on the adjacent C6 atom (C389 in *E. coli* RlmC) (**Figure 27**). All C5-pyrimidine methyltransferases acting on rRNA leading to m5U or m5C use this type of mechanism (Hamdane et al., 2016). This information is generally obtained using a substrate analog containing a stable carbon–fluorine bond at C5 atom, which enables the trapping of an irreversible 5,6-dihydropyrimidine covalent complex between the cysteine nucleophile and the C6 atom of the analog that is subsequently detected by either MS or SDS-PAGE (**Figure 27A**). This trapping is made possible because the fluorine group prevents β-elimination by the active

site base of the enzyme. The existence of such covalent adduct has been directly observed in the crystal structure of RlmC in complex with 5F-mini-RNA substrate (**Figure 27B & 27C**) (T. T. Lee et al., 2005)

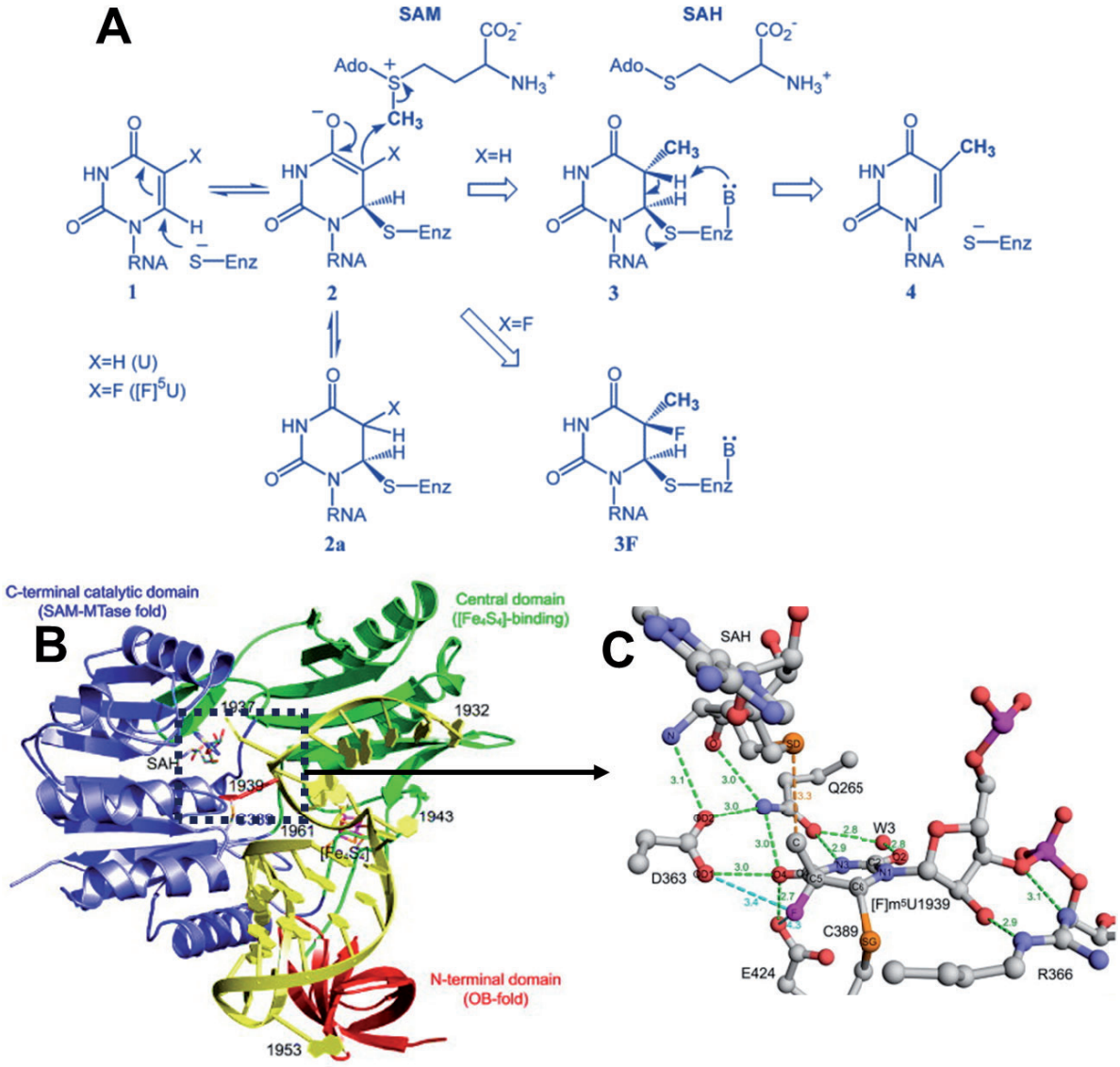


Figure 27: Mechanism of RlmC involved in m5U1939 rRNA methylation in *E. coli* adapted from (T. T. Lee et al., 2005) A) Proposed Catalytic Mechanism of RNA m5U MTases. (B) crystal structure of RlmC trapped with a mini-5FURNA substrate analog. (A) Ribbon representation of the complex. The N-terminal, central, and C-terminal domains of RlmC are colored in red, green, and blue, respectively. The observed RNA, corresponding to 1932–1961 of *E. coli* 23S rRNA, is colored yellow, with the exception of U1939, which is colored red. SAH, the catalytic cysteine (C389) and the [Fe4S4] cluster, and the side chains of its coordinating cysteines are shown in sticks (silver for C, red for O, blue for N, orange for S, and magenta for Fe). (C) Active Site of RumA showing interactions between amino acid side chains, the target uridine, and SAH. Possible hydrogen bonds are represented by green dashed lines. Distances between the C5 fluorine and the two general base candidates D363 and E424

are shown as cyan dashed lines. The distance between SAH-SD and the C5 methyl group, shown as a yellow dashed line.

Activation of the C2 of Adenine by a radical mechanism involving FeS clusters

RlmN catalyzes the C2 methylation of the A2503 in the 23S subunit of ribosomal RNA (Toh et al., 2008) but also that of the adenosine 37 position of some transfer RNAs (Benítez-Páez et al., 2012b). Its mechanism has been shown to be one of the most challenging one in RNA methylation reaction in which the methylated C2 is sp2-hybridized and consists of an homolytic C–H bond dissociation, which in term of energy is higher than that of a methyl group (**Figure 28C**) (Grove et al., 2013). RlmN transfers the methyl group to its conserved C355 before appending it to the rRNA (Grove et al., 2011). A first SAM molecule binds to the active site and gives its methyl group. Then a second SAM arrives and undergoes reductive cleavage by the 4Fe4S cluster of RlmN (**Figure 28A & C**) to generate the highly oxidant species, namely the 5′-dA• radical. This radical then abstracts hydrogen from methylcysteine to produce a protein-bound •CH2-SCys intermediate (**Figure 28C**) (Grove et al., 2011). The methylene radical performs an addition to the poorly nucleophilic aromatic ring. The X-ray structure of RlmN has shown that upon the binding of SAM to the active site, the cysteinyl sulfur atom is favorably located to be deprotonated and transform into the thiolate form, critical to initiating the reaction (Boal et al., 2011). In this conformation, the methylated cysteine is also ideally located for methyl transfer from SAM (**Figure 28 A & B**). An acid base reaction permits to protonate the C2-CH2 into a C2-CH3 and restore the aromaticity of the adenine.

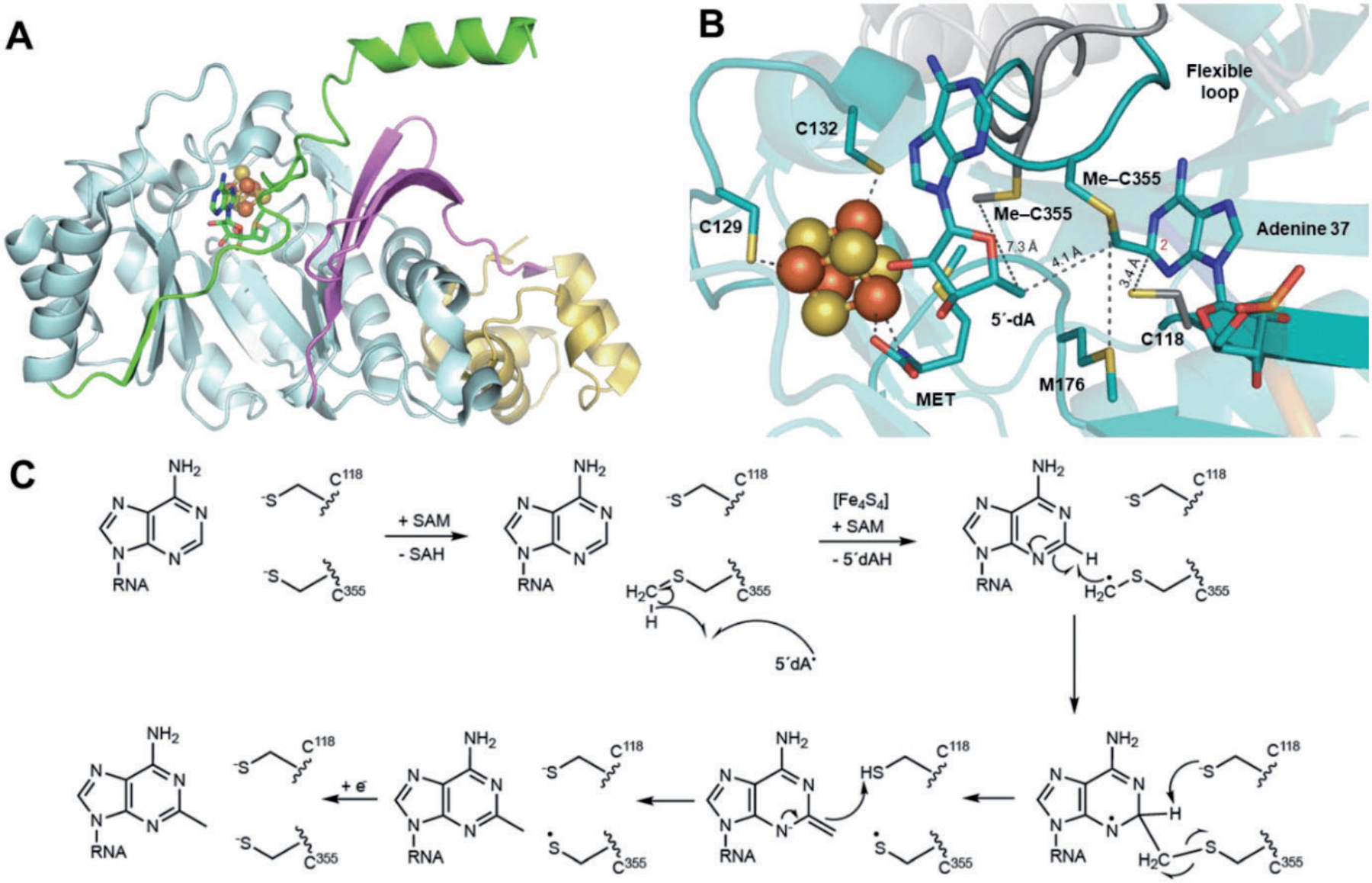

A
B
Flexible loop
C132
Me–C355
Me–C355
C129
Adenine 37
7.3 Å
4.1 Å
3.4 Å
5′-dA
C118
MET
M176
C
NH2
RNA
C118
C355
+ SAM
- SAH
[Fe4S4]
+ SAM
- 5′dAH
5′dA•
+ e-
HS

Figure 28: Structure of RlmN and its chemical mechanism adapted from (Nguyen & Nicolet, 2022). (A) Overall RlmN structure composed of a radical SAM domain (light cyan), additional β strands (pink), an N-terminal helical domain (gold) and a C-terminal stretch (green). (B) Close-up view of the RlmN active site in the tRNA-crosslinked structure (PDB 5HR7) in the presence of 5′-dA and methionine (MET) (SAM cleavage products), depicted in teal. This structure is superimposed with the free RlmN structure (PDB 3RFA), depicted in gray. Please note that cysteine residue C118 shown in the figure was mutated into alanine in order to stabilize the crosslink between the protein C335 and the RNA substrate and was therefore not present in the actual structure (Boal et al., 2011). (C) Proposed mechanism of C2 methylation on adenosine 2503 by RlmN.

4.2.3 Substrate(s) specificities determined by activity tests in vitro

Methylation of the 16S rRNA

KsgA catalyzes the methylation of $m^6{}_2A1518$ and $m^6{}_2A1519$ in the 16S rRNA's 3′-terminal helix, requiring the presence of certain ribosomal proteins for optimal activity (Desai & Rife, 2006). It prevents assembly intermediates from participating in translation, with ribosomal protein S21 and IF3 inhibiting its activity.

RsmB and RsmD have mutually exclusive specificities, modifying cytosine 967 (m^5C966) and guanosine 966 (m^2G966), respectively (Weitzmann et al., 1991). RsmB prefers free 16S rRNA, while RsmD works better with ribosomal proteins bound, particularly influenced by proteins S7, S9, and S19.

RsmG modifies guanosine 527 (m^7G527), and **RsmC** modifies guanosine 1207 (m^2G1207), both targeting the assembled 30S subunit (Benítez-Páez et al., 2012a; Sunita et al., 2007).

RsmI and RsmH collaboratively modify cytosine 1402 ($m^4Cm1402$) on the 30S subunit, demonstrating higher efficiency when acting together (Wei et al., 2012; Zhao et al., 2014).

RsmE modifies uridine 1498 (m^3U1498) (Basturea et al., 2006), and **RsmF** modifies cytosine 1407 (m^5C1407) (Andersen & Douthwaite, 2006), both in the assembled 30S subunit, with RsmF differing from RsmB which prefers free 16S rRNA.

RsmJ methylates guanosine 1516 (m^2G1516) in the assembled 30S subunit (Basturea et al., 2012), with its absence leading to cold sensitivity.

In summary, KsgA, RsmD, RsmG, RsmC, RsmI, RsmH, RsmE, RsmF, and RsmJ primarily act on assembled or partially assembled ribosomal subunits. RsmB is unique in preferring free 16S rRNA. RsmD's activity is enhanced by ribosomal protein binding, whereas RsmB's activity is inhibited by it.

Methylation of the 23S rRNA

m3Ψ1915: **RlmH** methylates pseudouridine at position 1915 of the 23S rRNA (Purta, Kaminska, et al., 2008). It strictly requires the associated 70S ribosome as a substrate and favor the presence of Ψ1915 suggesting a sequential modification process (Ero et al., 2010).

m1G745: **RlmA^{I}** methylates guanosine at position 745 in the 23S rRNA. It binds to free 23S rRNA, possibly aiding in 50S ribosomal subunit assembly (Gustafsson & Persson, 1998; Liu et al., 2004).

m5U747 and m5U1939: **RlmC** and **RlmD** methylate uridine residues at positions 747 and 1939 of the 23S rRNA, respectively, acting on free 23S rRNA (Madsen, 2003). This is comforted by the crystal structure of a miniARN substrate in complex with RlmC (see **Figure 27**)

m6A1618: **RlmF** methylates adenosine at position 1618 in the 23S rRNA. It acts on free 23S rRNA. Ribosomal proteins potentially block this modification (Sergiev et al., 2008).

m2G1835: **RlmG** methylates guanosine at position 1835 in the 23S rRNA, targeting deproteinized 23S rRNA and promoting ribosomal subunit association (Osterman et al., 2011; Sergiev et al., 2006)

m5C1962: **RlmI** methylates cytosine at position 1962 in the 23S rRNA, preferring deproteinized 23S rRNA, though C1962 does not contact proteins in the assembled 50S subunit (Purta, O'Connor, et al., 2008).

m6A2030: **RlmJ** modifies adenosine at position 2030 in the 23S rRNA, with a preference for deproteinized 23S rRNA (Golovina et al., 2012).

Gm2251, Cm2498, Um2552: **RlmB** and **RlmE** methylate the ribose of guanosine at position 2251 and uridine at position 2552 in the 23S rRNA, respectively (Lövgren & Wikström, 2001). These enzymes act on the 50S subunit and 70S ribosome, not on deproteinized 23S rRNA.

m2G2445 and m7G2069: **RlmKL** modifies guanosine residues at positions 2445 and 2069 in the 23S rRNA (Kimura et al., 2012; Lesnyak et al., 2006). It introduces two different modifications using its two domains, acting on free 23S rRNA.

m2A2503: **RlmN** is dual substrate methylase that methylates adenosine at position 2503 in the 23S rRNA and position 37 in tRNA, using a unique structure and catalytic mechanism (Benítez-Páez et al., 2012b). A small RNA can be used as substrate reinforcing the fact that this enzyme could acts locally.

4.2.4 rRNA methyltransferases, phenotypes, and ribosome maturation

Several studies have highlighted the critical roles of various rRNA methyltransferases in ribosomal subunit assembly and their broader implications in *E. coli* (**Table 4**). A recent comprehensive analysis of ribosomal subunits sedimentation profiles of a set of rRNA methyltransferase gene knockout strains from the Keio collection revealed a surprisingly small number of rRNA MT that significantly affected ribosome assembly (Pletnev et al., 2020) (**Figure 29**).

Table 4: List of *E. coli* rRNA MT coding genes and the phenotypes of their knockouts

Nucleotide	Enzyme	Reference	Growth of knockout strain/ growth at overexpression*	CER expression**	FastFT expression***	Accumulation of assembly intermediates and 17S rRNA precursor	Expression level**** and timing
16S rRNA							
527 m^7G	RsmG	Okamoto et al., 2007	+++\++	+++	+++	–/–	+++, 3h
966 m^2G	RsmD	Lesnyak et al., 2007	+++\+++	+	+++	–/–	+++, 2h
967 m^5C	RsmB	Tscherne et al., 1999a; Gu et al., 1999	+++\+++	++	++	–/++	++++, 4h
1207 m^2G	RsmC	Tscherne et al., 1999b	+++\+++	++	+++	–/–	++++, 2h
1402 m^4Cm	RsmI, RsmH	Kimura and Suzuki, 2010	+++\+	+	+/+++#	–/–	+, 2h
1407 m^5C	RsmF	Andersen and Douthwaite, 2006	+++\+++	++++	++++	–/–	+++, 2h
1498 m^3U	RsmE	Basturea et al., 2006	+++\+++	+++	+++	–/–	+++, 2h
1516 m^2G	RsmJ	Basturea et al., 2012	+++\+++	+	++	–/–	+++, 2h
1518/9 $m^6{}_2A$	RsmA	Helser et al., 1972; Poldermans et al., 1979	+++\+++	+	+	–/++	++++, 2h
23S rRNA							
745 m^1G	RlmA	Gustafsson and Persson, 1998	+++\++	+	+/+++#	–/–.	++++, 2h
747 m^5U	RlmC	Madsen et al., 2003	+++\++	+	+++	–/++	++, 2h
1618 m^6A	RlmF	Sergiev et al., 2008	+++\++	+	+++	–/–	+++, 1–7h
1835 m^2G	RlmG	Sergiev et al., 2006	+++\+++	++	++	–/–	+++, 2h
1915 $m^3\Psi$	RlmH	Purta et al., 2008a; Ero et al., 2008	+++\+++	+	++	–/–	+++, 2h
1939 m^5U	RlmD	Agarwalla et al., 2002; Madsen et al., 2003	+++\+++	+++	+++	–/–	+++, 2h
1962 m^5C	RlmI	Purta et al., 2008b	+++\+	+	++	–/–	+++, 2h
2030 m^6A	RlmJ	Golovina et al., 2012	+++\+++	+	+++	–/–	++++, 2h
2069 m^7G, 2445 m^2G	RlmKL	Kimura et al., 2012	+++\+++	+	+++	–/–	+++, 2h
2251 Gm	RlmB	Lovgren and Wikstrom, 2001	+++\+++	++	+++	–/–	+++, 2h
2498 Cm	RlmM	Purta et al., 2009	+++\+++	+	++	+?/–	+++, 2h
2503 m^2A	RlmN	Toh and Mankin, 2008	+++\++	+	+++	+?/-	+++, 2h
2552 Um	RlmE	Caldas et al., 2000; Bugl et al., 2000	+\+	+	+/++#	+++/–	++++, 3h
	WT		+++	+++	+++	–/–	

*+++ corresponds to doubling times 40–60 min, ++ 60–70 min, + > 70 min. ** CER fluorescence after overnight growth at inducing conditions ++++ > $6{\cdot}10^5$, +++ 4–$6{\cdot}10^5$, ++ 2–$4{\cdot}10^5$, + < $4{\cdot}10^5$. ***Fluorescence intensity of the blue, newly produced form of the FastFT at the exponential growth conditions ++++ > $2{\cdot}10^3$, +++1–$2{\cdot}10^3$, ++$5{\cdot}10^2$–10^3, + < $5{\cdot}10^2$. ****Expression levels at a timepoint with maximal expression. ++++ > 10^{-4}, +++ > 10^{-5}, ++ > 10^{-6}, + < 10^{-6} relative to the 16S rRNA. #During growth of these strains transformed by FastFT expression plasmid a significant amount of non-fluorescent cells were accumulated.

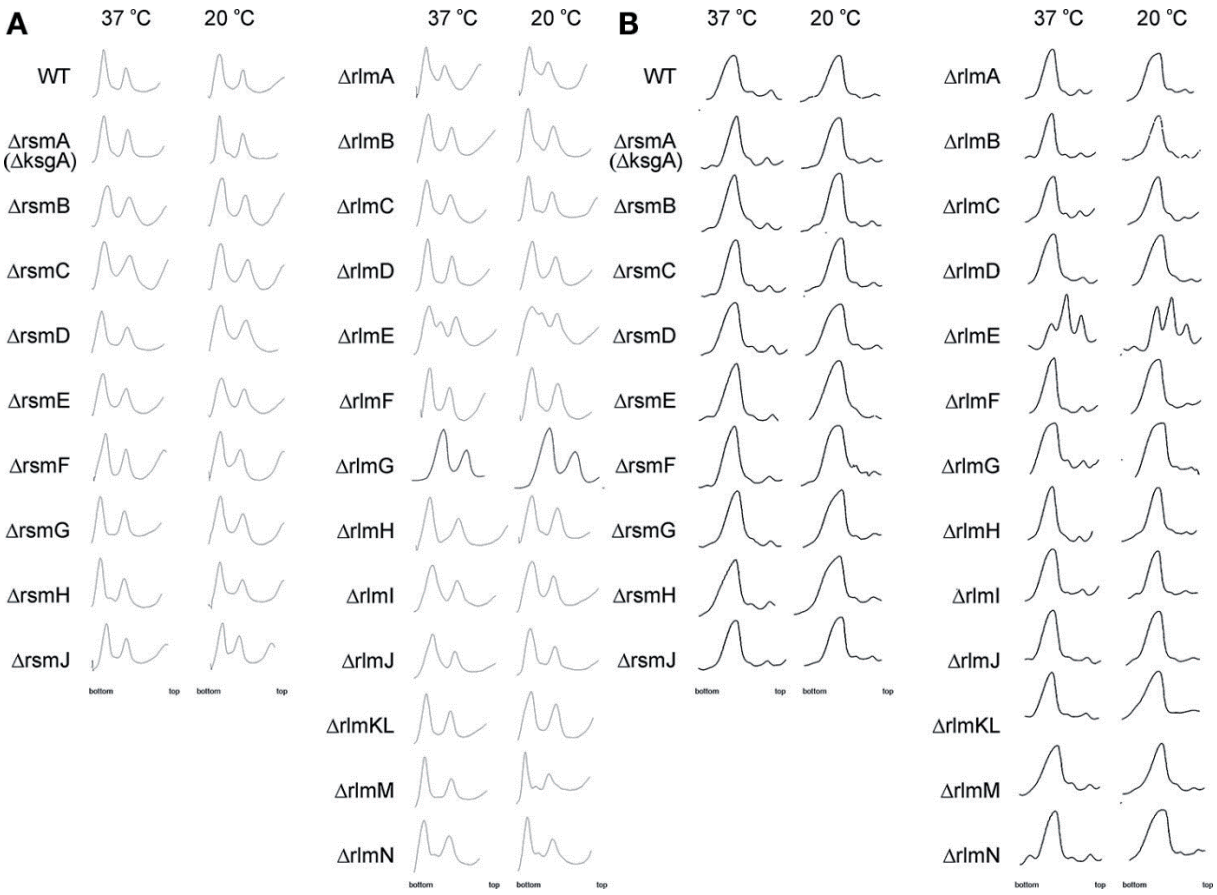


Figure 29: adapted from (Pletnev et al., 2020) Accumulation of ribosomal subunits assembly intermediates in the *E. coli* strains with inactivation of the 16S rRNA and 23S rRNA methyltransferase genes marked on the left side. Shown are sucrose gradient centrifugation profiles at subunit dissociation conditions (1 mM magnesium ions concentration) (A) and association conditions (10 mM magnesium ions concentration) (B). Left panels correspond to the cells grown at 37°C, while that to the right correspond to the cells grown at 20°C known to exacerbate ribosomal subunits assembly defects.

Several methylated rRNA nucleotides have been implicated in controlling ribosomal subunit assembly, with KsgA playing a significant role in the final stages of small subunit assembly. KsgA binds to small subunit intermediates, delaying its methyltransferase activity until the later stages of assembly, a process that may require higher levels of KsgA compared to other rRNA methylases (Connolly et al., 2008). No significant accumulation of assembly intermediates was observed in the *ΔKsgA* strain, there was an increase in the 17S rRNA precursor at low temperatures (**Figure 29**), indicating KsgA's role in small subunit assembly across bacterial, eukaryal (D. Lafontaine et al., 1995), and mitochondrial ribosomes (Metodiev et al., 2009). RsmB, acts on early intermediates of the 30S subunit before the S19 protein is incorporated (Weitzmann et al., 1991). Its role might involve accelerating assembly or preventing misassembly, with inactivation leading to the accumulation of the 17S rRNA precursor. RsmB orthologs are widespread in bacteria, and similar modifications occur in archaea and eukarya.

Inactivation of the 23S rRNA MT RlmC also resulted in the accumulation of the 17S rRNA precursor, a phenomenon paralleled by the inactivation of the large subunit pseudouridine synthase RluD (Gutgsell et al., 2005). This suggests that RlmC's influence on small subunit rRNA processing is likely indirect.

Finally, RlmE stands out due to the severe ribosome misassembly associated with its inactivation, leading to the most significant growth retardation among rRNA methylases knockouts (Arai et al., 2015) (**Figure 29**) (see below the corresponding chapter).

While the growth rates of most rRNA MT knockout strains are similar to wild-type strains (**Table 4**), the absence of rRNA methylation was shown to significantly reduce the capacity to synthesize exogenous proteins. This suggests that rRNA methylation is crucial under conditions of increased protein synthesis demand. In summary, all rRNA methylases knockouts reveals that while only a few affects bacterial growth, ribosome assembly, or the proteome, most strains exhibit a suboptimal capacity to synthesize exogenous proteins. This underscores the importance of rRNA methylation in maintaining efficient protein synthesis under varying cellular conditions and could explain why Nature has conserved these enzymes.

4.2.4.1 Regulatory role of the SAM methyl group donor in the ribosome assembly

A recent study from Pr Suzuki lab in Tokyo University explored the relationship between cellular SAM levels and ribosome biogenesis in *E. coli*, focusing on the role of the methyltransferase RlmE (Ishiguro et al., 2019). By utilizing an *E. coli Δmtn* strain with reduced intracellular SAM concentration (approximately one-third of wild-type levels) (Halliday et al., 2010), they observed several significant effects on rRNA and tRNA methylation, ribosome assembly, and cell growth.

RlmE's activity is sensitive to SAM concentration, particularly in modifying Um2552 in the 45S precursor, which is a critical step for 50S subunit assembly (Hager et al., 2002). Low SAM levels result in reduced Um2552 modification, leading to 45S precursor accumulation and less efficient translation due to fewer 50S subunits (Ishiguro et al., 2019). Quite interestingly, only the overexpression of RlmE substantially restored the growth phenotype and normalized 45S precursor accumulation in the *Δmtn* strain. This indicates that RlmE-mediated methylation of Um2552 is crucial for proper ribosome biogenesis and cell proliferation under low SAM conditions. The study identified hypomethylation at two sites in 16S rRNA and two sites in 23S rRNA in the *Δmtn* strain. Specifically, m5C1407 and m3U1498 in 16S rRNA, and Um2552 and

m5U747 in 23S rRNA, were affected. Hypomethylation of these sites is linked to ribosome quality and function, influencing translation under stress conditions.

The postulated mechanism of Suziki is that when SAM concentration is high, RlmE facilitates the methylation of Um2552 in the 45S precursor, promoting its maturation into the 50S subunit, which is necessary for efficient protein synthesis (**Figure 30**). In contrast, under low SAM conditions, reduced Um2552 methylation leads to the accumulation of the 45S precursor, thus slowing down ribosome assembly and translation (**Figure 30**). This regulatory mechanism ensures that protein synthesis is tightly coordinated with the cell's metabolic state, linking translational efficiency to the availability of a critical metabolite.

We can indeed propose that this mechanism of SAM-sensitive regulation of ribosome assembly by RlmE might be conserved across different organisms, including eukaryotes. Indeed, homologs of RlmE in yeast and humans play similar roles in rRNA methylation, essential for ribosome function and cellular activities (Kressler, 1999; Lapeyre & Purushothaman, 2004; Pintard, 2002).

These findings underscore the broader biological principle that cellular metabolism and gene expression are intricately connected, with implications for understanding growth and adaptation in various organisms.

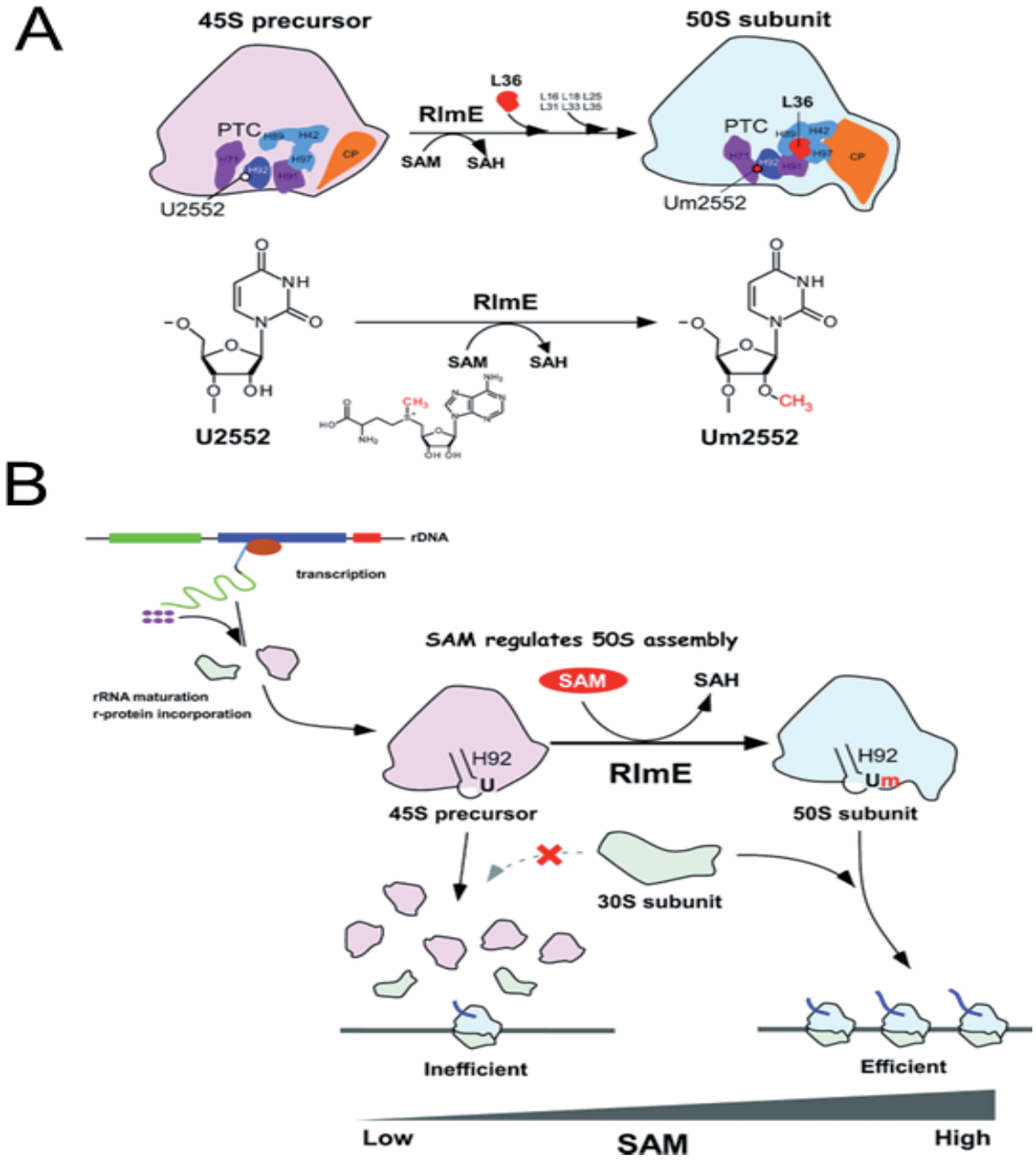


Figure 30: Taken from (Ishiguro et al., 2019) The regulatory role of SAM in 50S assembly. (A) Mechanistic model for late assembly of 50S subunits triggered by Um2552 formation. The 45S precursor is a flexible molecule that lacks Um2552 and several r-proteins including L36. RlmE-mediated Um2552 formation promotes the interaction of helices H92 and H71, and incorporation of L36 and other r-proteins. This process triggers 50S maturation in the peptidyl transferase center (PTC) and central protuberance (CP). The chemical scheme of Um2552 formation is shown in the lower figure. SAH, S-adenosyl-L-homocysteine. (B) Regulatory mechanism of 50S assembly mediated by Um2552 modification via sensing of the intracellular SAM concentration. When SAM is abundant, RlmE introduces Um2552 into the 45S precursor, which promotes 50S assembly and thereby maintains efficient translation. When SAM is less abundant, Um2552 remains hypomodified and the 45S precursor accumulates, leading to less efficient translation. As the 45S precursor is stably present in cells, 50S subunits can be generated when SAM levels in the cell recover.

4.2.4.2 rRNA methyltransferase role beyond methylation activity: A proofreading and chaperone activities

RNA modifying enzymes can have other functions than their biochemical catalysis function per se. Indeed, since they bind to RNA, they can in theory have additional functions like shared by other RNA binding proteins. In 2016, Keffer-Wilkes et al. introduced the concept of tRNA chaperone for tRNA modifying enzymes (Keffer-Wilkes et al., 2016). Using TruB as a model, they demonstrate that *E. coli* Ψ55 TruB synthase folds tRNA independent of its catalytic activity, thus increasing the fraction of tRNA that can be aminoacylated. Rapid kinetic showed that TruB binds and unfolds both misfolded and folded tRNAs thereby providing misfolded tRNAs a second chance at folding. More recently, the same group showed that m5U54 methyltransferase TrmA of *E. coli* enhances tRNA folding in vitro independent of its catalytic activity identifying TrmA as the second tRNA modifying enzyme acting as a tRNA chaperone supporting a functional link between RNA modification and folding (Keffer-Wilkes et al., 2020). Both TruB and TrmA act on a crucial region of tRNA, namely the elbow region that involved tertiary interaction networks crucial form maintaining this specific L-shaped tRNA structure.

Like for tRNA modifying enzymes, a recent Cryo-EM structure of KsgA in complex with an immature 30S subunits of *E. coli* ribosome obtained from a 30SΔksgA particle revealed additional functional features for KsgA rRNA m62A1518 and m62A1519 methyltransferase (Sun et al., 2023) (**Figure 31A & B**). Interestingly, KsgA binds to immature 30S subunits and causes structural destabilization by displacing critical helices, leading to partial subunit disassembly without dissociating ribosomal proteins. This destabilization allows the subunits to reassemble, giving them another chance to achieve an active conformation. The structures reveal crucial atomic interactions that facilitate substrate binding and the order of base methylation in the KsgA active site (**Figure 31B**). The KsgA-bound structure is incompatible with subunit joining, explaining the growth inhibition observed with catalytically inactive KsgA variants.

This study shows that KsgA plays a dual role in ribosome biogenesis, acting as both a methylase and an assembly factor (**Figure 31C**). KsgA recognizes and destabilizes nearly mature but inactive 30S subunits, enhancing the efficiency and correctness of ribosome assembly. The authors of this structures introduced the concept of assembly-factor-mediated proofreading model in which nearly mature but inactive particles are recognized and destabilized by KsgA binding, resulting in partial subunit disassembly. They hypothesize that upon methylation and subsequent KsgA dissociation, these particles reassemble and are thereby provided another

opportunity to adopt a translationally active conformation to maximize the efficiency of the assembly process for ribosomal small subunits, and to enforce the proper assembly order.

This proposed role of KsgA resembles that recently assigned to RbgA, a ribosome assembly factor in *Bacillus subtilis*, which ensures the 50S subunit follows a canonical maturation pathway where the functional sites are the last structural motifs to mature (Seffouh et al., 2022).

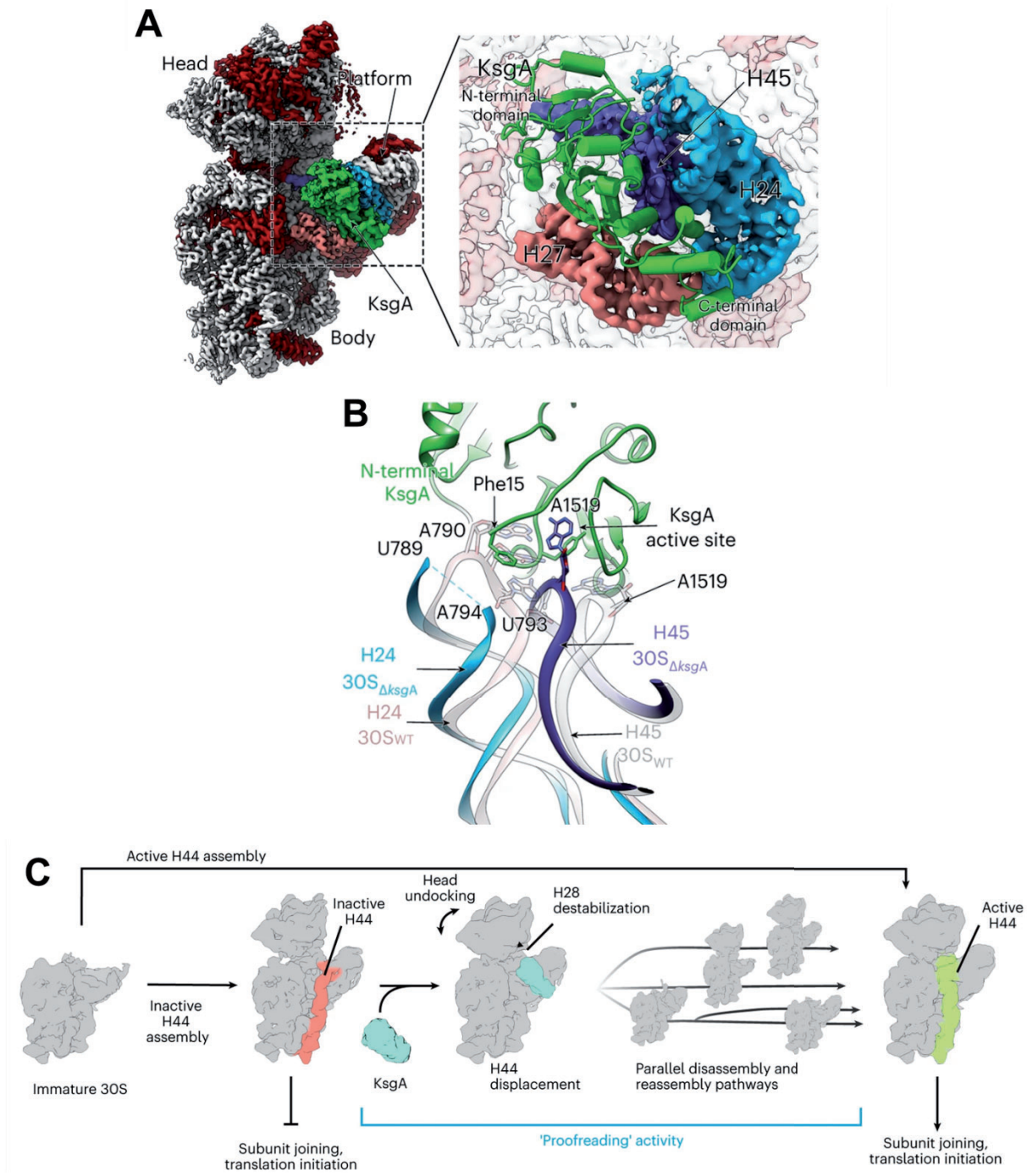


Figure 31: Substrate engagement by KsgA displaces a gatekeeping rRNA helix adapted from (Sun et al., 2023). (A), Interface view of the cryo-EM structure obtained for the immature 30S *ΔksgA* particle bound to KsgA (green). Ribosomal proteins are shown in red, the 16S rRNA is shown in light gray, and structural landmarks of the ribosomal subunit are indicated. Key

rRNA helices interacting with KsgA are colored pink (helix 27), cyan (helix 24), and blue (helix 45). The interaction area is enlarged at the right, which depicts a molecular model of KsgA derived from the cryo-EM map. (B), Overlay of rRNA helices 24 and 45 from the molecular model of the mature 30SWT subunit in the absence of KsgA (rose) and those from the 30S *ΔksgA* particle bound to KsgA (H24 in cyan; H45 in purple). Note positioning of A1519 in KsgA's active site necessitates displacement of helix to avoid steric clashes between helices 24 and 45. (C) KsgA recognizes and remodels inactive subunits. Integrated model depicting KsgA's proposed role in late-stage assembly of the small ribosomal subunit.

4.2.4.3 rRNA Methylations and Antibiotic Resistances

Beyond the physiological functions of rRNA modification enzymes in ribosome homeostasis, rRNA methyltransferases can have major roles in antibiotic resistance in bacteria (**Table 5**). A wealth of information on antibiotic resistance caused by methylation of rRNA has been revealed during the last ten years and I do not intend to describe all of them since it will require thousand more pages. I will just focus on some of them that are relevant in *E. coli* and for an exhaustive review on the subject please refer to the (Osterman et al., 2020).

Antibiotic resistance is a critical challenge in treating pathogenic bacterial infections. Bacteria have developed strategies to evade antibiotics, including modifying rRNA (**Table 5**). The ribosome is a primary antibiotic target, with key sites on the 30S and 50S subunits, and in which high-resolution ribosome structures have improved understanding of drug-binding sites. Methylation of rRNA is a significant mechanism for antibiotic resistance and prevents drug binding, hindering protein synthesis. Only methylation provides acquired antibiotic resistance among RNA modifications.

Ribosomal RNA (rRNA) methyltransferases were identified about 50 years ago. The ErmC methyltransferase confers resistance to macrolides, lincosamides, and streptogramin B. The Erm family has approximately 40 classes of methylases targeting A2058 in 23S rRNA. Eight classes of RNA methyltransferases act on 23S rRNA, causing antibiotic resistance. TlyA methyltransferase modifies nucleotides in inter-subunit bridge B2a, affecting capreomycin resistance. Lack of methylation can also confer antibiotic resistance, e.g., Ksg methyltransferase absence leads to kasugamycin resistance. RNA methyltransferases like Tsr and EmtA target GTPase center and orthosomycin resistance sites, respectively.

Table 5: rRNA methyltransferases affecting antibiotic resistance from (Osterman et al., 2020)

ID	Protein	Product	Occurrence	Effect of methylation on antibiotic resistance	Mechanism of the effect on antibiotic activity	Putative ancestral enzyme*
P36999	RlmA^{I}	m^1G745 23S rRNA	Gram-negative bacteria	susceptibility to viomycin (moderate)	presumably, conformational changes	RlmA^{II}**
Q9S1M6	RlmA^{II}	m^1G748 23S rRNA	tylosin producer *Streptomyces roseoflavus*, other Gram-positive bacteria	resistance to tylosin; susceptibility to telithromycin	direct steric clashes, hydrophobic interaction	RlmA^{I}**
P18644	TsnR	Am1067 23S rRNA	thiostrepton producer *Streptomyces azureus*	resistance to thiostrepton	direct steric clashes, hydrophobic interaction	TrmH, RlmB
P9WJ63	TlyA	Cm1920 23S rRNA (and 16S rRNA Cm1409)	pathogenic *Mycobacterium tuberculosis* and some other bacteria	susceptibility to capreomycin and viomycin	presumably, conformational changes or hydrophobic interaction	
P13956	Erm family	23S rRNA m_2^6A2058	numerous pathogens and producers	resistance to macrolides, lincosamides, and streptogramins B	direct steric clashes	RsmA(KsgA)
P97178	TlrD	23S rRNA m^6A2058	tylosin producer *Streptomyces roseoflavus*, other producers	resistance to macrolides and lincosamides (moderate)	direct steric clashes	RsmA(KsgA)
Q9FBG4	Cfr	m^8A2503 23S rRNA	pathogenic *Staphylococcus sciuri* and others	resistance to phenicol, lincosamides, oxazolidinones, pleuromutilins, streptogramins A	direct steric clashes	RlmN
P36979	RlmN	m^2A2503 23S rRNA	all bacteria	susceptibility to linezolid, sparsomycin, tiamulin (moderate)	presumably, conformational changes	
Q9F5K6	AviRb	Um2479 23S rRNA	avilamycin producer *Streptomyces viridochromogenes*	resistance to avilamycin	direct steric clashes	TrmH, RlmB
Q93CQ2	EmtA	mG2470 23S rRNA	pathogenic *Enterococcus faecium* and others	resistance to avilamycin, evernimicin	direct steric clashes	THUMPD3, RlmKL, Trm14
Q9F5K5	AviRa	mG2535 23S rRNA	avilamycin producer *Streptomyces viri-*	resistance to avilamycin (moderate)	direct steric clashes	
P0ADX9	RsmD	m^2G966 16S rRNA	all bacteria	resistance to tetracycline; (moderate) susceptibility to tigecycline (moderate)	direct steric clashes	
E9KIK3	EfmM	m^5C1404 16S rRNA	pathogenic *E. faecium*	resistance to kanamycin and tobramycin (moderate)	direct steric clashes	RsmF

ID	Protein	Product	Occurrence	Effect of methylation on antibiotic resistance	Mechanism of the effect on antibiotic activity	Putative ancestral enzyme*
Q6F5A0	ArmA	m^7G1405 16S rRNA	pathogenic *Klebsiella pneumoniae* and others	resistance to aminoglycosides with 4-6 substituted deoxystreptamine	direct steric clashes	
Q33DX5	RmtA-H	m^7G1405 16S rRNA	pathogenic *Pseudomonas aeruginosa* and others	resistance to aminoglycosides with 4-6 substituted deoxystreptamine	direct steric clashes	
Q53316	KgmB and others	m^7G1405 16S rRNA	nebramycin producer *Streptomyces tenebrarius* and others	resistance to aminoglycosides with 4-6 substituted deoxystreptamine	direct steric clashes	
A8C927	NpmA	m^1A1408 16S rRNA	pathogenic *Escherichia coli*	resistance to aminoglycosides with 4-6 and 4-5 substituted deoxystreptamine	direct steric clashes	TrmB
P25920	KamB	m^1A1408 16S rRNA	tobramycin producer *S. tenebrarius*	resistance to aminoglycosides with 4-6 and 4-5 substituted deoxystreptamine	direct steric clashes	TrmB
A6YEH1	CmnU	m^1A1408 16S rRNA	capreomycin producer *Streptomyces capreolus*	resistance to capreomycin	direct steric clashes	TrmB
P9WJ63	TlyA	Cm1409 (and 16S rRNA Cm1920)	pathogenic *Mycobacterium tuberculosis* and some other bacteria	susceptibility to capreomycin and viomycin	presumably, conformational changes or hydrophobic interaction	
P0AGL7	RsmE	m^3U1498 16S rRNA	all bacteria	resistance to aminoglycosides with 4-6 and 4-5 substituted deoxystreptamine (moderate)	presumably, conformational changes	
P06992	RsmA (KsgA)	m_2^6A1518/19 16S rRNA	all organisms	susceptibility to kasugamycin	conformational changes	

The Cfr methyltransferase confers resistance to five antibiotic classes by methylating A2503 in 23S rRNA located within the peptidyl transferase center (PTC), a region of the ribosome essential for catalyzing peptide bond formation and consequently, a common target for antibiotics to give m8A2503 in *E. coli* numbering (**Figure 32**). Cfr's resistance mechanism is linked to steric hindrance at the drug-binding site. The *cfr* gene has spread globally and is found in both Gram-positive and Gram-negative bacteria. The rRNA-methylating enzyme Cfr is the homologue of RlmN and acts on the same adenine substrate with or without the presence of m2A (**Figure 32**) using the same radical SAM methylation mechanism (**Figure 28**) (Grove et al., 2013)

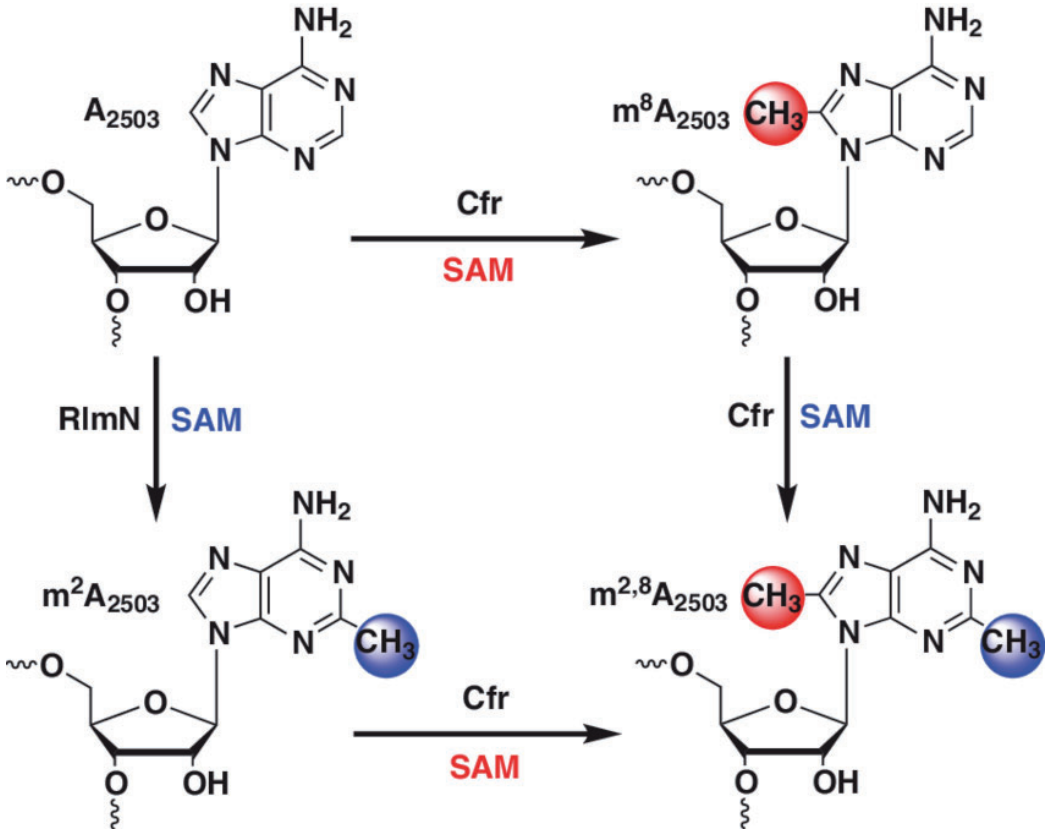


Figure 32: Reactions catalyzed by RlmN and Cfr. RlmN catalyzes uniquely methylation at C-2, whereas Cfr catalyzes methylation at C-8 and C-2, although C-8 is the preferred target.

Recent directed evolution study combined with structural determination have improved the mechanism of Cfr dependent antibiotic resistance (Tsai et al., 2022). In this study, they performed directed evolution of Cfr under antibiotic selection to generate Cfr variants that confer increased resistance by enhancing methylation of A2503 in cells. Using a variant that achieves near-stoichiometric methylation of rRNA, they determined a 2.2 Å cryo-electron microscopy structure of the Cfr-modified ribosome revealing the molecular basis for broad resistance to antibiotics (**Figure 33**).

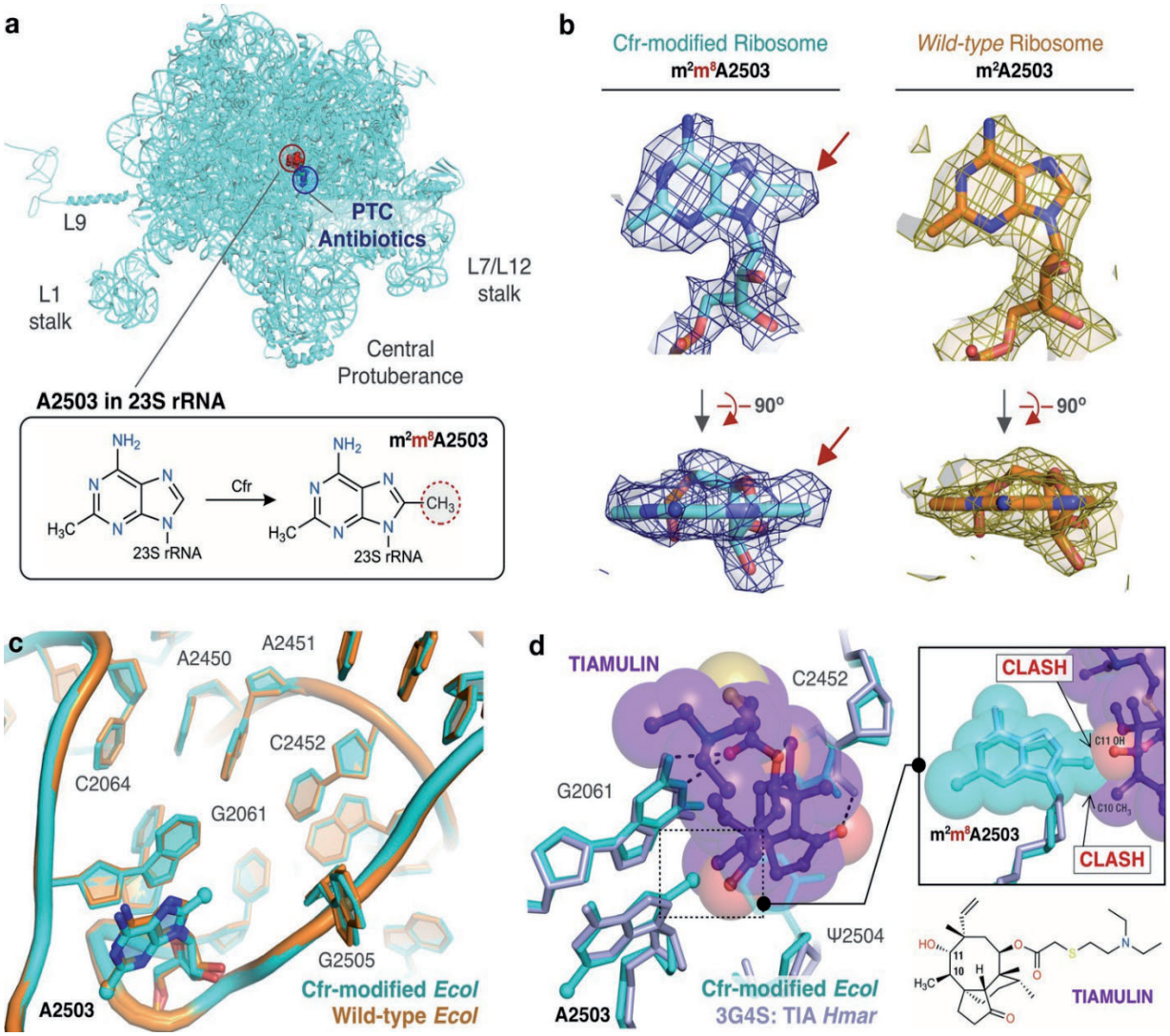


Figure 33 Near-stoichiometric ribosome methylation by CfrV7 enables structural understanding of Cfr-mediated resistance to antibiotics taken from (Tsai et al., 2022). (a) Cfr-modified 50S ribosomal subunit highlighting adenosine 2503 (A2503) within 23S rRNA and the binding site of PTC-targeting antibiotics. Cfr methylates A2503 at the C8 carbon to produce m2m8A2503. (b) Cryo-EM density maps of adenosine 2503 in 23S rRNA contoured to 3σ. Cfr-modified (m2m8A2503) in cyan. Wild type (m2A2503) in orange; PDB 6PJ6. (c) Close-up view of 23S rRNA nucleotides in the 50S ribosomal subunit. Cfr-modified Escherichia coli ribosome in cyan. Wild-type *E. coli* ribosome in orange; PDB 6PJ6. (d) Structural overlay of Cfr-modified *E. coli* ribosome (cyan) and *Haloarcula marismortui* 50S ribosome in complex with pleuromutilin antibiotic tiamulin (purple, PDB 3G4S) highlighting steric clashes between m8A2503 and the antibiotic. EM, electron microscopy.

Comparison of the Cfr-modified ribosome with the high-resolution cryo-EM structure of unmodified, wild-type *E. coli* ribosome published previously (Stojković et al., 2020) allowed them to identify with high confidence any structural changes due to the presence of m8A2503 (**Figure 33**). Modification of A2503 by Cfr does not affect the conformation or position of the

A2503 nucleotide. The adenine ring remains in the syn-conformation and places the newly installed C8-methyl group directly into the PTC to sterically obstruct antibiotic binding (**Figure 33c–d**). The presence of m8A2503 does not result in any additional structural changes to the PTC region of the ribosome (**Figure 33c**). Structural superposition of the Cfr-modified ribosome with ribosomes in complex with PhLOPSA antibiotics, hygromycin A, nucleoside analog A201A, and 16-membered macrolides enables direct identification of chemical moieties responsible for steric collision with m8A2503 for these eight antibiotic drug classes (**Figure 33**). Overlay of a bacterial ribosome in complex with the pleuromutilin derivative tiamulin, the selection antibiotic used during directed evolution, reveals steric clashes between the C10 and C11 substituents of the antibiotic with the Cfr-introduced methyl group (**Figure 33d**).

A more recently high-resolution crystal structures of the Cfr-modified 70S ribosome containing aminoacyl- and peptidyl-transfer RNAs from *T. thermophilus* were obtained (Aleksandrova et al., 2024). The structural study of Cfr-mediated resistance to PTC-targeting ribosomal antibiotics suggests that the underlying mechanism of resistance at the molecular level are likely due to two-component: (1) direct steric hindrance of the A2503-C8-methyl group with the ribosome-bound drugs; and (2) Cfr-methylation-induced rearrangement of the nucleotide A2062 to a conformation incompatible with drug binding. Although the relative contributions of these two mechanisms vary depending on the particular antibiotic, together they ensure that a single methyl group added to the A2503 residue in a 2.5-MDa ribosome renders many chemically unrelated classes of antibiotics unable to bind to such ribosomes.

4.3 The rRNA hydroxylase

Recently, the group of Suzuki conducted a genome-wide screen in *E. coli* to identify genes required for ho5C2501 formation, and found a previously-uncharacterized gene, *ydcP* renamed rlhA, iron-sulfur cluster (isc) genes, and a series of genes responsible for prephenate biosynthesis, indicating that iron-sulfur clusters and prephenate are required for ho5C2501 formation (Kimura et al., 2017)(**Figure 34**). No in vitro activity could be detected with the recombinant RlhA protein owing to the complex cofactors required and the unknown mechanism although it was shown through polysome profiling that RlhA interacted with precursors of the 50S ribosomal subunit, suggesting that this protein is directly involved in the formation of ho5C2501.

More recently, Fasnacht et al measured growth curves with the modification-deficient *ΔrlhA* strain and quantified the extent of the modification during different conditions by mass spectrometry and reverse transcription (Fasnacht et al., 2022). The levels of ho5C2501 in *E. coli* ribosomes turned out to be highly dynamic and growth phase-dependent, with the most effective hydroxylation yields observed in the stationary phase. High ho5C2501 levels reduced protein biosynthesis which however turned out to be beneficial for *E. coli* for adapting to oxidative stress. This functional advantage was small under optimal conditions or during heat or cold shock, but becomes pronounced in the presence of hydrogen peroxide, likely providing functional rRNA hydroxylation under oxidative stress. Still research is needed to uncover the molecular mechanism behind rRNA hydroxylation and its impact on the ribosome function.

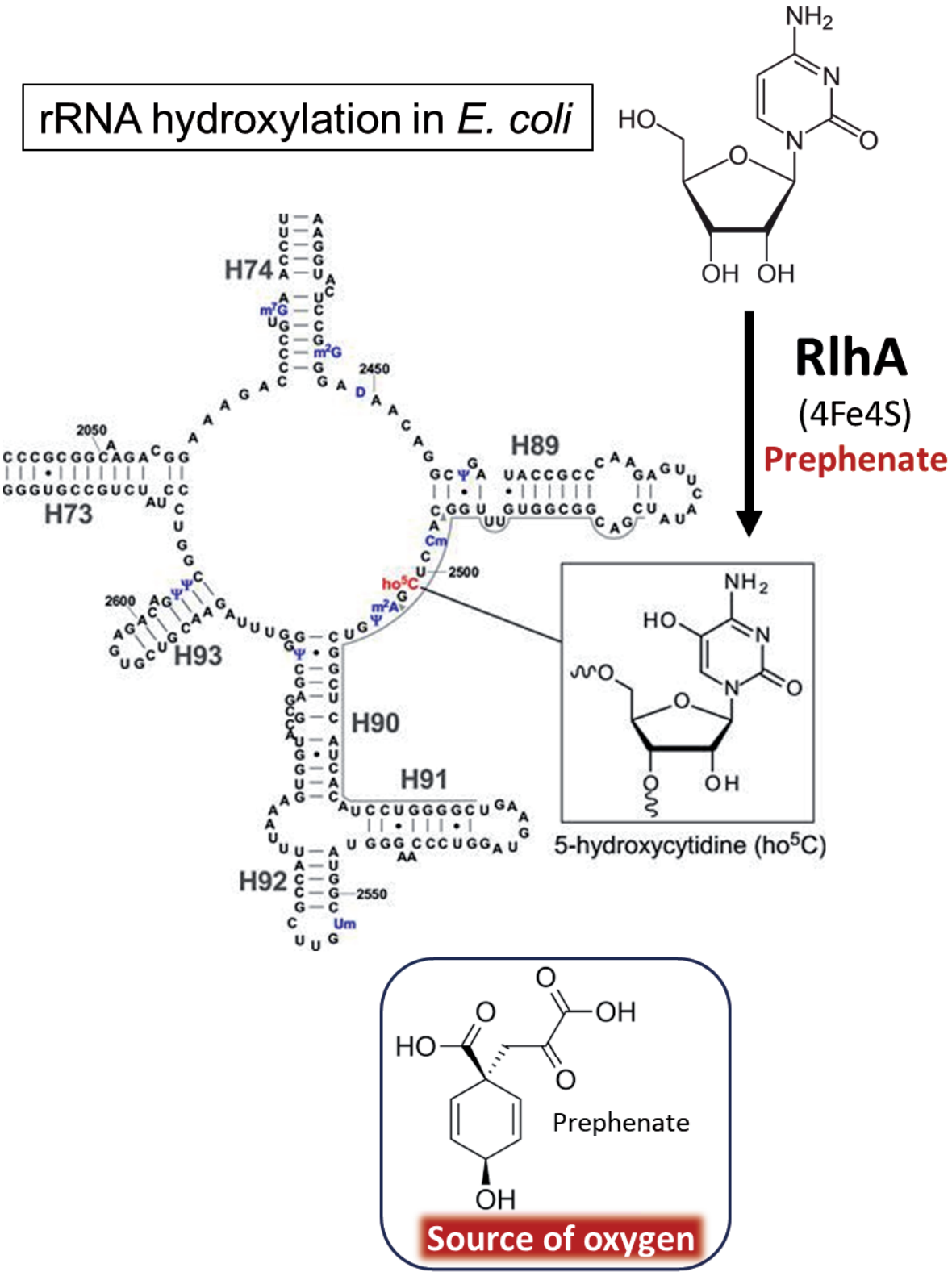


Figure 34: Biosynthetic pathway of 5-hydroxycytidine (ho5C) in rRNA *E. coli.* The enzyme responsible for this hydroxylation is RlhA and it is a FeS-dependent protein. In vivo deletion experiments of prephenate metabolic pathways indicate that this hydroxylation is dependent on the prephenate, which might be the oxygen donor.

4.3 Spatiotemporal action of rRNA modifying enzymes followed in vivo

Indeed, the majority of modified nucleotides in bacterial 16S rRNA are either modified during the late stages of subunit assembly or are introduced into already assembled subunits (Siibak & Remme, 2010). In addition, sequential 5′-to-3′ order methylation events in the 16S rRNA can occur during the late stages of the assembly of individual domains of the small ribosomal subunit, which assemble independently and concurrently rather than during the final assembly of the entire subunit (Popova & Williamson, 2014).

In contrast, modifications in the large subunit occur during the early stages of subunit assembly (Siibak & Remme, 2010), which is consistent with the fact that most of modified nucleotides in the 23S rRNA are inaccessible to any rRNA modifying enzymes (**Figure 16**) (Polikanov et al., 2015). Nonetheless, few surface-exposed modified nucleosides are observed and directly interact with key ligands, such as tRNAs and translation factors.

Results were obtained by following the incorporation of these modifications into rRNA during ribosome assembly by inducing the accumulation of incompletely assembled ribosomal particles (25S, 35S, and 45S) in *E. coli* cultures treated with chloramphenicol or erythromycin (Siibak & Remme, 2010). Isolated particles and their analyzed notably the content of the modified nucleosides of 16S and 23S rRNA using high-performance liquid chromatography (HPLC) and the chemical modification/primer extension method for pseudouridines categorized modifications into early, intermediate, and late assembly-specific groups and summarized in the **Figure 35**.

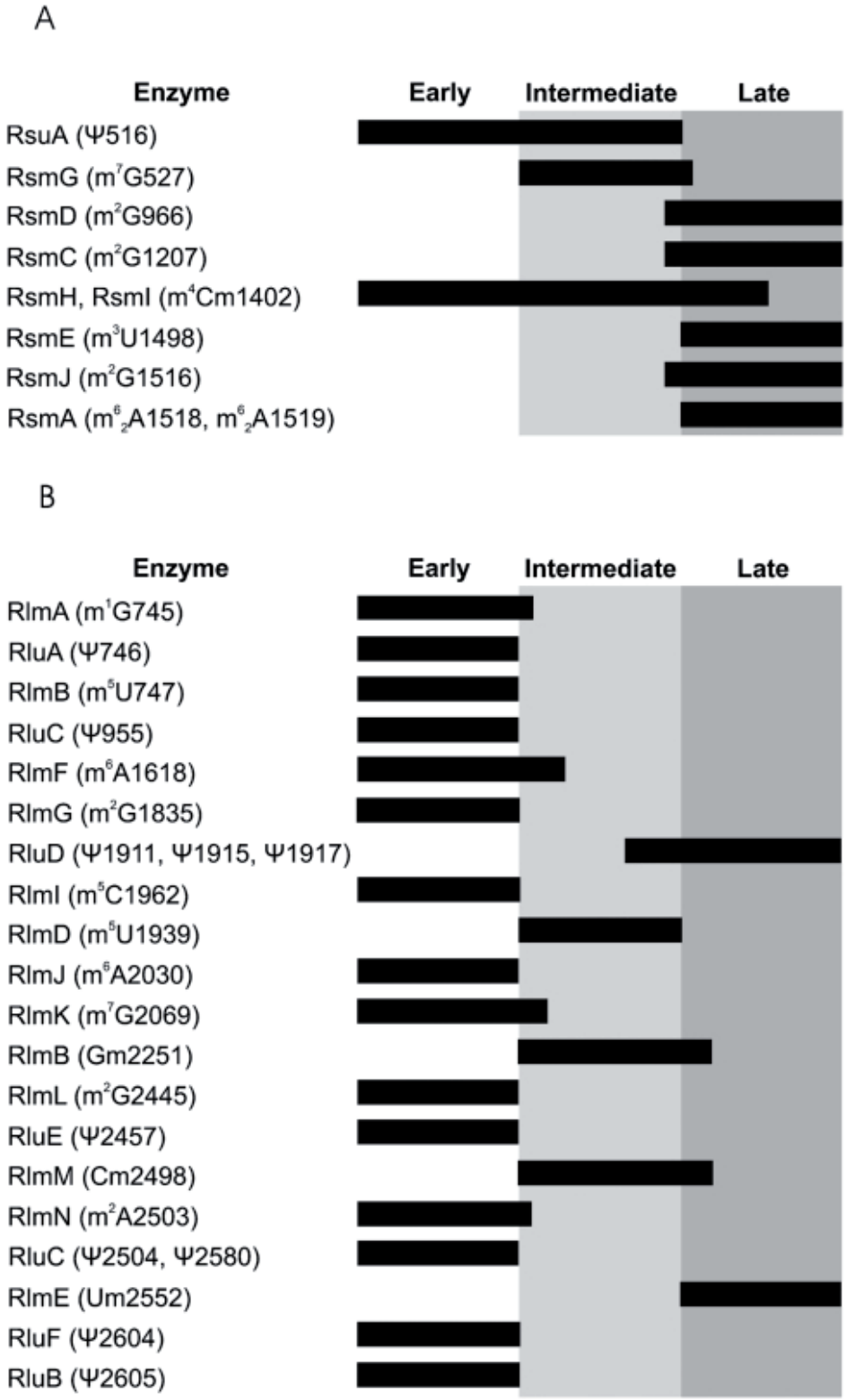


Figure 35: Summary of the specificity of the rRNA modification enzymes with respect to ribosome subunit assembly originated from(Siibak & Remme, 2010). Ribosome assembly is divided into three stages (early, intermediate, and late), which are shown by white and gray zones. Activity of rRNA modification enzymes is shown by black bars. (A) Modification of 16S rRNA. (B) Modification of 23S rRNA.

A powerful new 14N- and 15N-labeled stable isotope labeling - Mass Spectrometry approach has been recently developed for efficient quantitative monitoring of rRNA modifications (Popova & Williamson, 2014). This study presented a new approach that provided unprecedented quantitative profiling of rRNA modifications, highlighting their temporal and functional relationships during ribosome assembly and biogenesis. The study leveraged the sensitivity, accuracy, and automation of MS-based techniques, which are ideal for studying post-transcriptional modifications that are difficult to analyze through direct RNA sequencing. Extensive data sets were collected to quantify RNA modifications in preribosomal particles isolated from wild-type *E. coli* across 30S and 50S regions of the sucrose gradient (**Figure 36**).

For the 30S subunit, three groups of modifications were identified (**Figure 36**):

- Early modifications: m7G(527)
- Intermediate modifications: m2G(966), m5C(967), m2G(1207)
- Late modifications: m4Cm(1402), m5C(1407), m3U(1498), m2G(1516), m62A(1518), m62A(1519)

These groups are localized to the 5′ body domain, the 3′ head domain, and the 3′ minor domain of the 30S subunit, respectively, reflecting the in vivo and in vitro assembly process that proceeds by structural domains in a 5′ to 3′ direction.

For the 50S subunit, the modifications were categorized as follows (**Figure 36**):

- Early modifications: m1G(745), m5U(747), m6A(1618), m2G(1835), m5U(1939), m5C(1962), m6A(2030), m7G(2069), Gm(2251), Cm(2498), m2A(2503).
- Late modifications: Ψ(1911), m3Ψ(1915), Ψ(1917), Um(2552)

Some modifications, like Cm(2498) and m5U(1939), might occur at an early to intermediate stage of 50S assembly.

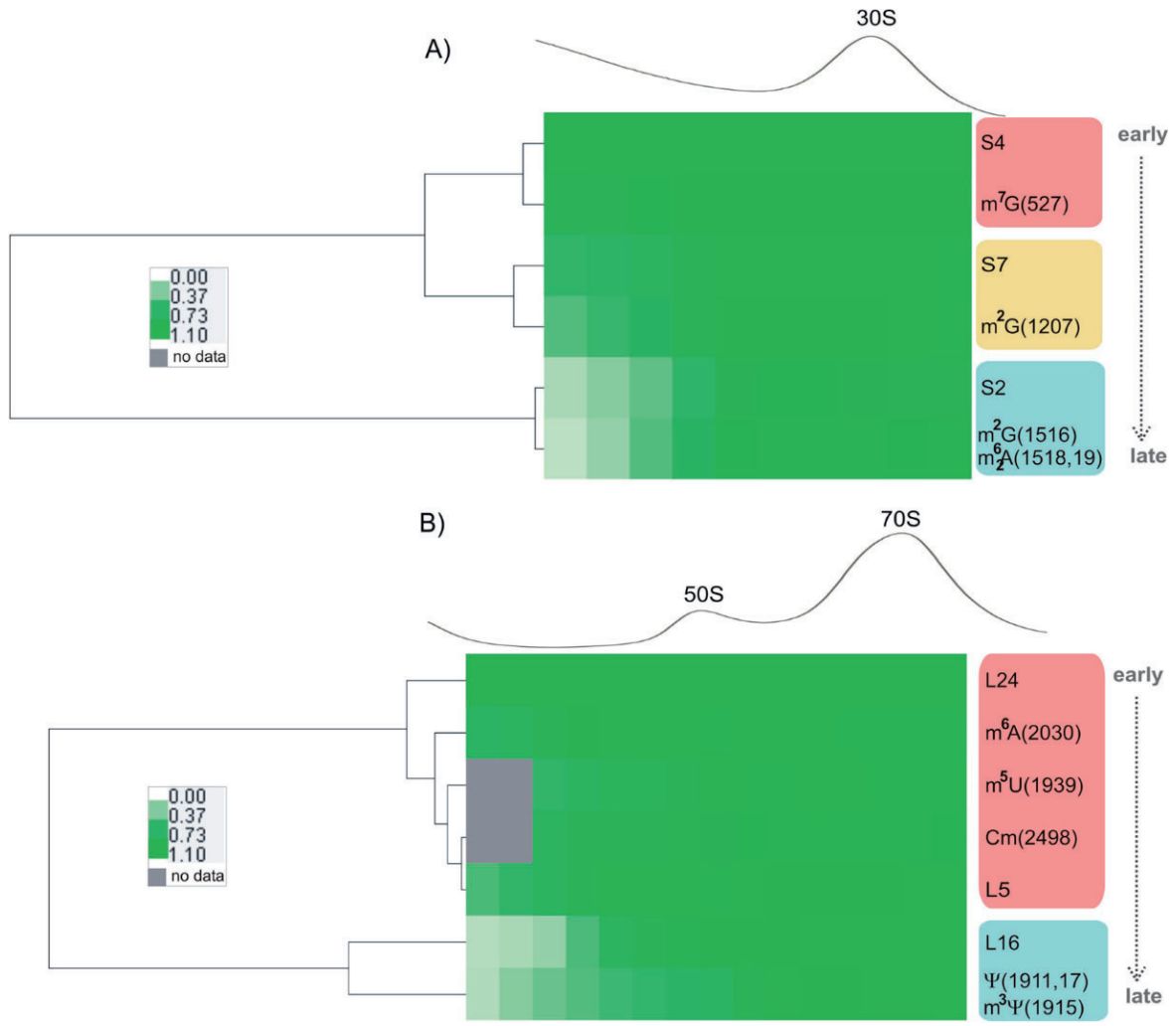
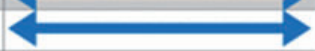


Figure 36: Inventory analysis of rRNA modifications and ribosomal proteins. Relative levels of RNA modifications and ribosomal proteins for the subset of the small (A) and large (B) ribosomal subunit components are shown as a heat map along the corresponding polysome traces. The figure is from (Popova & Williamson, 2014).

More recently Narayan et al. employed another strategy to study the spatiotemporal rRNA modifications (Narayan et al., 2023). The study determined the level of incorporations of 2-methyl adenosine, 3-methyl pseudouridine, 5-hydroxycytosine, and seven pseudouridines in an early-stage *E. coli* large-subunit assembly intermediate with a sedimentation coefficient of 27S. The 27S intermediate is one of three large subunit intermediates accumulated in *E. coli* cells lacking the DEAD-box RNA helicase DbpA and expressing the helicase inactive R331A DbpA construct. It was shown that the majority of the investigated modifications are incorporated into the 27S large subunit intermediate to similar levels to those in the mature 50S large subunit, indicating that these early modifications or the enzymes that incorporate them play important roles in the initial events of large subunit ribosome assembly (**Table 6**).

Table 6: Summary of the Large Subunit Stages at Which the Modification Enzymes Perform Their Functions in 23S rRNA from (Narayan et al., 2023).

		Ribosome Large Subunit Assembly Stages	
Enzyme[a]	**23S rRNA Modification[b]**	**Early[c]**	**Intermediate/ Late[c]**
RluC	Ψ 957, 2508, 2584		
RluD	Ψ 1919, 1921		
RlmH	$m^3\Psi$ 1919		
RlhA	OH^5C 2505		
RlmN	m^2A 2507		
RluE	Ψ 2461		
RluF	Ψ 2608		
RluB	Ψ 2609		

5. Insight into the Dihydrouridylation reaction catalyzed by Classical Dus Enzymes

5.1 Dihydrouridine: A Modified Base with Unique Structural Properties

Almost all structural and functional properties of dihydrouridine are derived from the study of this modification in tRNAs. It is the most abundant modified base after pseudouridine in many organisms, including prokaryotes like *E. coli*, although it is generally absent in hyperthermophiles.

Dihydrouridine (D) is a pyrimidine nucleoside variant derived from 5,6-dihydrouracil, formed through the hydrogenation of uridine (U) (**Figure 37A**). This process results in a saturated, non-planar nucleobase, a distinctive feature of dihydrouridine (**Figure 37C & 37D**). Although dihydrouracil was chemically synthesized in the late 19th century, it was first biologically detected in 1952 when isolated from beef spleen. Initially demonstrated in vitro, dihydrouridine monophosphate efficiently incorporates into RNA, though its natural presence in cellular RNA was uncertain (Carr & Grisolia, 1964) (Roy-Burman et al., 1965). Later, D was confirmed as a constituent of yeast tRNAAla in the seminal work of Holley and was observed in several tRNA crystal structures (**Figure 37C**) (Holley et al., 1965; Kim et al., 1974; Suddath et al., 1974). Concurrently, Visser et al. proposed enzymatic hydrogenation as a plausible explanation for the non-random distribution of dihydrouridine in nucleic acids (Roy-Burman et al., 1965).
Over the following decades, advancements in X-ray and NMR crystallographic analyses have provided detailed insights into the structural properties of dihydrouridine (**Figure 37C & 37D**) (Dyubankova et al., 2015; Suddath et al., 1974). Structural investigations into the nucleobase and nucleotide of D-containing tRNAs and oligoribonucleotides have yielded several key findings (Lorenz et al., 2017)

- Reduction of the C5-C6 double bond results in carbon 6 (C6) being out of the plane of the nucleobase.
- This non-planarity, due to the loss of aromaticity, diminishes its stacking ability with adjacent nucleobases and positions D outside the stacked bases, rendering it solvent accessible (**Figure 37C**). An important point is that the D base itself cannot engage in canonical hydrogen bonding.

- The ribose moiety adopts a C2'-endo conformation, unlike the usual C3'-endo conformation of canonical ribonucleotides (**Figure 37B**).
- This C2'-endo pucker extends to the 5'-nucleotide, highlighting dihydrouridine's distinctivness (Davis DR, 1998)

These structural characteristics, such as the promotion of the C2'-endo conformation and increased flexibility across the sugar-phosphate backbone (**Figure 37B**), potentially destabilize RNA structure (J. Dalluge, 1996a). Recent NMR investigations into the D loop have further highlighted dihydrouridine's role, showing that its absence leads to unstable stem-loop hairpins that adopt multiple interconverting conformations in solution (Dyubankova et al., 2015). In tRNAs, the increased local flexibility mediated by D may facilitate interactions between neighboring tertiary bases in the critical tRNA elbow region. This region, formed by the interaction between the D and TΨC loops, involves several highly conserved interactions in cytosolic tRNAs, including the Hoogsteen-reverse base pair T54-A58, interloop base pairs G18-Ψ55 and G19-C56, and a stack of four interspersed purine bases, A58-G18-A/G57-G19 (Lorenz et al., 2017) (**Figure 37B**).

The observed decrease in melting temperature of *E. coli* tRNAs in the absence of dihydrouridine could be due to the loss of the necessary flexibility in the D-loop for accommodating these essential tertiary interactions, crucial for maintaining the tRNA's 3D structure (Nomura et al., 2016). In their investigation on siRNA, Sipa et al. evaluated the thermodynamic stability and gene silencing activity of RNA duplexes containing modified bases, including D (Sipa et al., 2007).

Unlike tRNA, RNA duplexes lack tertiary structures and show a decrease in melting temperature when a D unit is present centrally. This effect is attributed to D's destabilizing influence on the C3′-endo sugar conformation and its nonplanar nature, which disrupts stacking interactions with neighboring nucleobases. Hence, it is clear from the published research on dihydrouridine that this modified base plays a crucial role in imparting unique structural properties related to local flexibility.

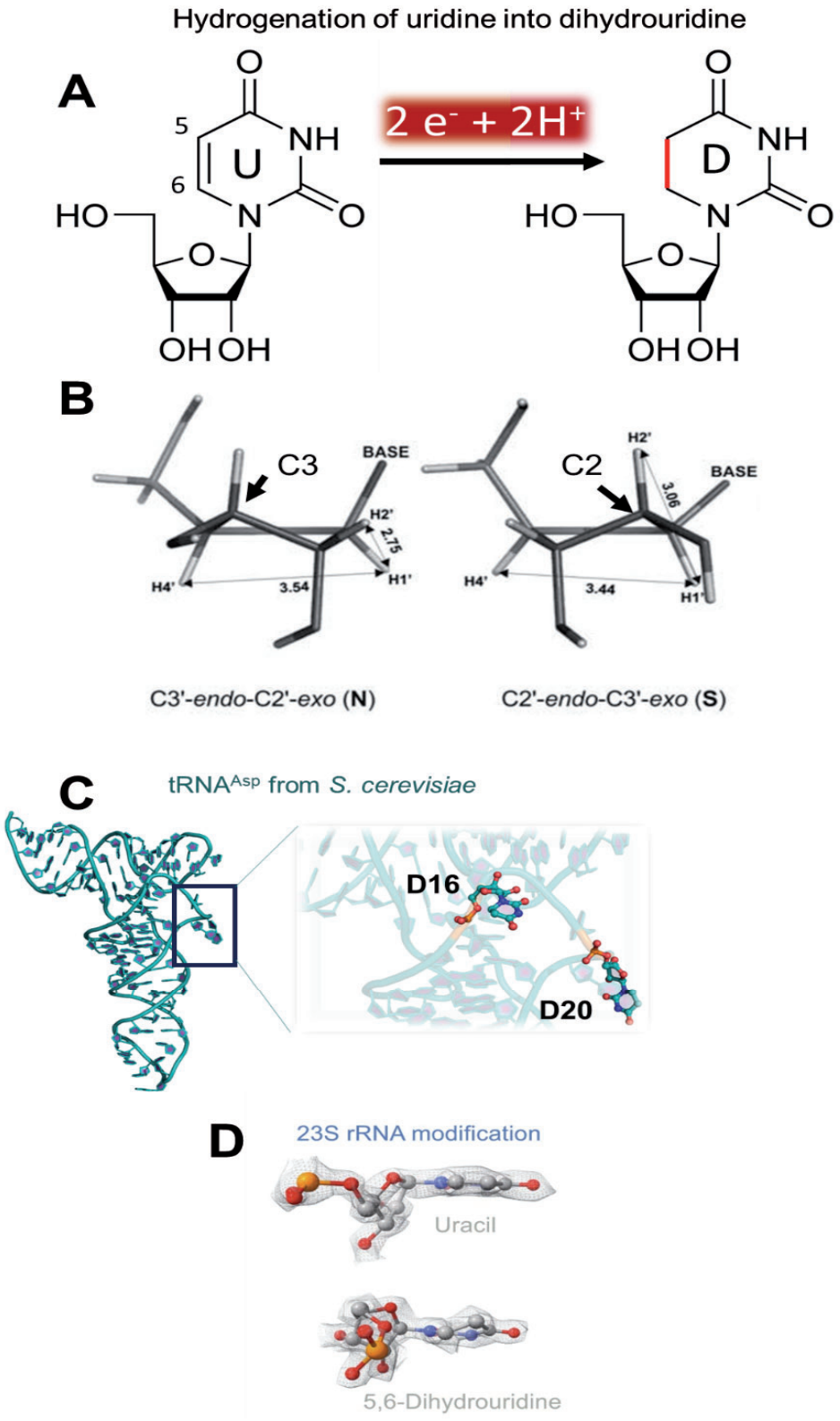


Figure 37: Dihydrouridylation reaction and structural properties of dihydrouridine. (A) Chemical reduction reaction of uridine to dihydrouridine. This process requires 2 electrons and 2 protons, akin to a hydrogenation reaction. (B) Ribose conformation. Dihydrouridine favors the C2'-endo conformation of ribose. (C) Crystallographic structure of yeast aspartate tRNA illustrating the exposure of D16 and D20 to the solvent. (D) Close-up depiction of D2449 modification in *E. coli* 23S rRNA represented as sticks, with overlaid EM densities shown as mesh.

5.2 Occurrence of Dihydrouridine in the RNA world

An analysis of 602 tRNA sequences from viruses, prokaryotes, and eukaryotes shows that dihydrouridine (D) is the second most common tRNA modification, with 925 occurrences, surpassed only by pseudouridine, which has 1,164 occurrences (Boccaletto et al., 2022). Dihydrouridylation is found at canonical sites D16, D17, D20, D20a, D20b, and D47 (**Figure 38**), as well as at rare non-canonical sites D14, D17a, D21, and D48. The most frequently modified positions are D20 and D16, both located in the D-loop, which is crucial for forming secondary and tertiary tRNA structures. All canonical D residues (D16, D17, D20, D20a, D20b) are found in bacteria. D47 is a rare modification, appearing only once in the $tRNA^{Met}$ of *Bacillus subtilis*, despite position 47 being a uridine (U) in nearly 90% of the 134 known bacterial tRNA sequences. D20b is also uncommon and has been found in a cyanobacterial $tRNA^{Glu}$. Dihydrouridine is rare in rRNA. It has been observed at only one location, 2449, in the central loop of domain V in *E. coli* 23S rRNA (Kowalak et al., 1995) and at two positions, 2449 and 2500, in the 23S rRNA of *Clostridium sporogenes* (Kirpekar et al., 2018) (**Figure 38**). Additionally, it has been found at either position 1211 or 1212 in the 16S rRNA of *Clostridium acetobutylicum*, though the exact location is unclear (Emmerechts et al., s. d.). Dihydrouridine has not been detected in other bacterial rRNAs or in any sequenced eukaryotic rRNAs. Interestingly, in the 23S rRNA of *C. sporogenes*, D2449 was found to be methylated at the C5 atom, forming m5D2449 (Kirpekar et al., 2018). It is unknown whether m5D is formed through methylation of m5U followed by reduction, or if dihydrouridine is first formed and then methylated at C5. This latter pathway is unprecedented and warrants further investigation. The enzymes responsible for m5D biosynthesis are still unidentified.

Y RNA, involved in RNA degradation, is another type of non-coding RNA known to undergo dihydrouridylation in the γ-proteobacterium *Salmonella typhimurium* (X. Chen et al., 2014). Recent large-scale transcriptome analyses have identified D in mRNAs of yeasts and humans, but not in bacterial mRNAs(Dai et al., 2021). In eukaryotic mRNAs, D is primarily located in the coding regions of conserved genes, indicating its significant biological role. In *Schizasaccharomyces pombe* transcriptomes, 372 D sites have been identified, with 38% found in mRNA and 61% in tRNA (Finet, Yague-Sanz, Krüger, et al., 2022). Among the 125 mRNAs containing D in *S. pombe*, 87% exhibit a single putative D site, while only two mRNAs (encoding a nonclassical export protein and alanine-tRNA ligase) have at least three distinct D

sites. In human colon cells (HCT116), 112 D sites have been detected within mRNAs. However, the overall prevalence of D in mRNAs is relatively low compared to other modifications.

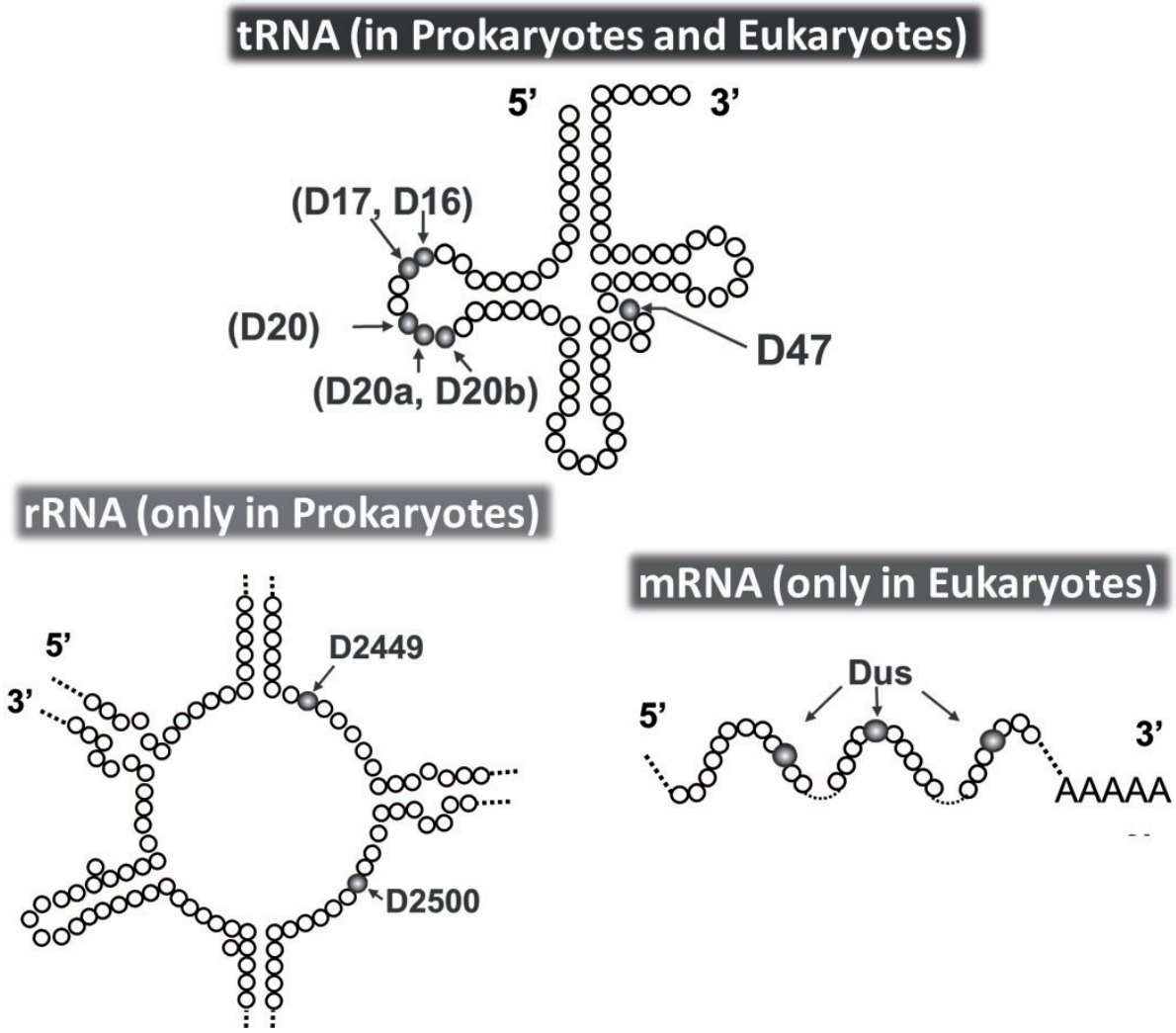


Figure 38 adapted from (Brégeon et al., 2022)**. Distribution of D in the transcriptome.** The figure shows on the top the secondary tRNA structure with the position of D residues observed in prokaryotes and eucaryotes tRNAs. Right Bottom: *E. coli* 23S rRNA sequence showing the position of the unique dihydrouridine (D2449). The 23S subunit of bacterial rRNAs located at the peptidyltransferase center of the ribosome. Uridine 2500 is an unmodified uridine in *E. coli* whereas in *C. sporogenes* this uridine is modified to dihydrouridine, D2500. Left bottom: schematic representation of dihydrouridylated mRNAs sites of eucaryotes.

5.3 Methods for Dihydrouridine labelling and mapping across the transcriptome

Detecting dihydrouridine (D) modifications in RNA involves various techniques, each offering distinct advantages and levels of sensitivity. Key methods include radioactivity-based thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), and advanced RNA sequencing techniques such as RhoSeq and AlkalineSeq. Here is an overview of these methods:

5.3.1 Radioactivity-based Thin-Layer Chromatography (TLC)

Radioactivity-based TLC is one of the traditional methods for detecting RNA modifications, including dihydrouridine. The process involves: Labeling RNA: RNA samples are labeled with radioactive isotopes, typically phosphorus-32 (^{32}P). Hydrolysis: The labeled RNA is then enzymatically hydrolyzed to release nucleoside monophosphates. Chromatography: The hydrolysate is applied to a TLC plate and separated based on the physicochemical properties of the nucleosides. Detection: The radioactive signals corresponding to different nucleosides, including dihydrouridine, are detected using autoradiography.
This method allows for the precise identification and quantification of dihydrouridine by comparing the migration patterns of known standards.

5.3.2 High-Performance Liquid Chromatography (HPLC)

HPLC is a widely used analytical technique for separating and quantifying nucleosides in RNA. For detecting dihydrouridine: Sample Preparation: RNA is enzymatically digested into nucleosides. Separation: The digested sample is injected into an HPLC system, where nucleosides are separated based on their interactions with the column matrix under high pressure. Detection: Diode array detectors (DAD) or mass spectrometers (MS) are typically used to identify and quantify the nucleosides based on their unique retention times and spectral properties. HPLC offers high sensitivity and specificity, making it suitable for detecting low-abundance modifications like dihydrouridine. This technique has been used during my thesis in collaboration with group of Pr Mark Helm, in university of Mainz, Germany. However, the radioactivity technique and the HPLC does not provide any information regarding the sequence, which is provided through RNAseq techniques.

5.3.3 RNA Sequencing Techniques

5.3.3.1 Rho-Seq

Early research demonstrated that D can be specifically labeled with the rhodamine fluorophore (Betteridge et al., 2007). Recently, the research group of Pr Hermand at the University of Namur, hypothesized that incorporating rhodamine, aside from its fluorescent properties, would disrupt reverse transcription (RT), leaving a detectable mark indicative of D presence. This concept of RT stop footprinting had previously proven effective in epitranscriptomic studies for identifying pseudouridine (Schwartz et al., 2014) inosine (Suzuki et al., 2015), and 1-methyladenosine (Safra et al., 2017). Expanding on this approach, Herman's group developed Rho-seq, a method to comprehensively identify D modifications across the transcriptome. Rho-seq combines rhodamine labeling with high-throughput sequencing (**Figure 39**) (Finet et al., 2022). This technique involves selectively reducing D using sodium borohydride (NaBH4) (Cerutti & Miller, 1967) followed by rhodamine incorporation. The labeled D arrests RT one nucleotide downstream of the modified site (Kaur et al., 2011). By comparing RT stop patterns between NaBH4-treated (R+) and mock-treated (R-) samples, Rho-seq effectively identifies potential D sites (**Figure 39C**). However, since some RNA modifications react to NaBH4 (Cerutti & Miller, 1967), including N4-acetylcytidine (ac4C), N6-formyladenosine (f6A), m1A, m3C, m7G, S4U, and yW, the authors also analyze dihydrouridine-free RNA extracts from a D-synthase-deficient strain tested with R+ and R- treatments, and compared to wild-type samples (**Figure 39C**). This control ensures that observed RT stops specifically reflect D modifications.

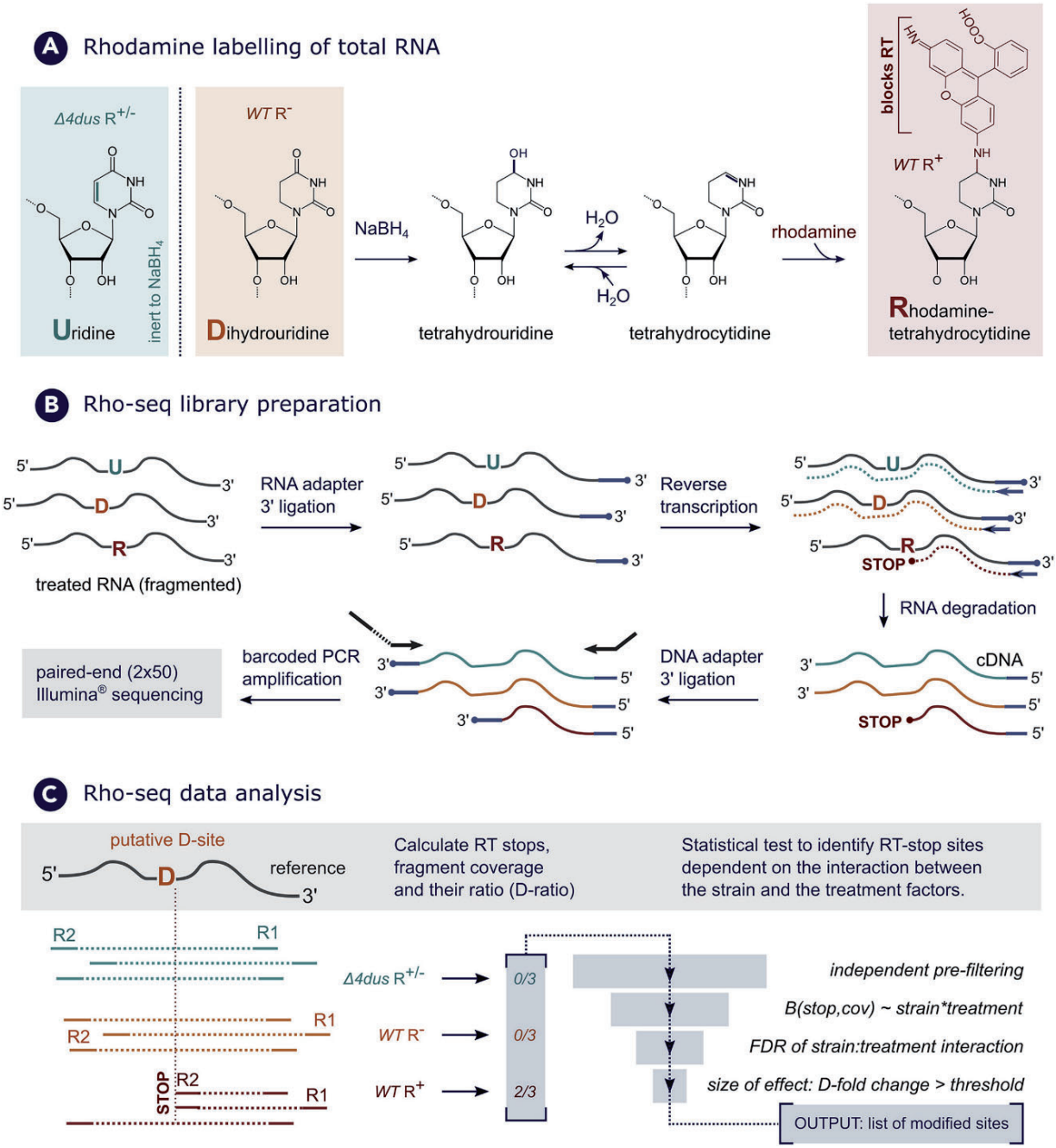


Figure 39: from (Finet, Yague-Sanz, & Hermand, 2022). (A) Total RNA extracts are chemically treated to label dihydrouridine (D) with rhodamine. Sodium borohydride (NaBH4) reduces D, and a nucleophile like rhodamine forms a covalent bond with the intermediate tetrahydrocitidine (Kaur et al., 2011). Effective labeling (*WT R+*, in red) is compared to controls: mock labeling (*WT R-*, using KOH, in orange) and RNA from a strain lacking dihydrouridine synthases (*Δ4dus R+* and *R-,* in cyan blue). This method can also label other RNA modifications, such as m7G and s4U. (B) RNAs from the *R+* and *R-* conditions are prepared for high-throughput sequencing. Fragmented RNA is ligated to a 3′-blocked RNA adapter (blue line with filled circle), followed by reverse transcription. Reverse transcription either proceeds to the RNA end (cyan and orange) or stops prematurely due to rhodamine (red). After ligating a 3′-blocked DNA adaptor (cyan) to the cDNA, the library is amplified and sequenced using Illumina® paired-end chemistry. (C) Sequenced reads are trimmed, mapped, and analyzed to calculate the D-ratio; the ratio of reverse transcription stops (R2 starts) to fragment coverage. To identify D-sites, a three-stage approach is used: filtering sites by

unsupervised criteria, modeling D-ratios with a generalized linear model (considering treatment, strain, and their interaction), and correcting p-values with the Benjamini-Hochberg procedure (FDR). The effect size (D-fold change) is the ratio of the average D-ratio of WT R+ to the average D-ratio of all controls.

5.3.3.2 AlkAniline-Seq

Professor Yuri Motorin's team at the University of Lorraine developed an effective method to pinpoint D sites within tRNAs using deep sequencing (Marchand et al., 2018). They took the advantage that the 5′ phosphate generated during aniline cleavage could aid in library preparation, making its presence crucial (**Figure 40**). This innovative approach not only yielded clear signals. Named AlkAniline-Seq, this pioneering method identifies abasic sites and the modified nucleotides that lead to their formation during processing. AlkAniline-Seq involves three steps: i) alkaline hydrolysis of RNA, ii) thorough dephosphorylation of 5′- and 3′-ends, and iii) aniline cleavage (**Figure 40**). It is known that D residues undergo nucleophilic attack by hydroxide ions under alkaline conditions, resulting in ring opening, base elimination, and abasic site formation. In the second step, RNA fragments undergo terminal repair through treatment with alkaline phosphatase, removing both existing 5′-phosphates in RNA and any forms of 3′-phosphates resulting from alkaline hydrolysis (**Figure 40**). This ensures that all RNA fragments possess 5′- and 3′-OH ends suitable for 3′-end ligation but not for 5′-adapter ligation, which requires a 5′ phosphate. The third step involves aniline treatment, cleaving abasic sites through β- and δ-elimination (**Figure 40**), thereby exposing the 5′ phosphate at the N+1 nucleotide. These RNA fragments, containing a 5′ phosphate, are selectively released from modified nucleosides like D. Positive enrichment of aniline-cleaved fragments is integrated into this selective ligation protocol, highlighting the method's high sensitivity and specificity. Unlike conventional RNA-seq methods, AlkAniline-Seq provides reads where information about modified sites resides at the 5′ end of each read. Double-stranded DNA amplicons resulting from library preparation can be sequenced using Illumina technology, where the start of Read1 corresponds to the N+1 nucleotide unprotected by abasic site cleavage. Consequently, reads are mapped end-to-end onto the reference sequence, and the count of 5′ ends of the reads indicates the position and intensity of cleavage. AlkAniline-Seq have been used for the analysis and mapping of D during my studies.

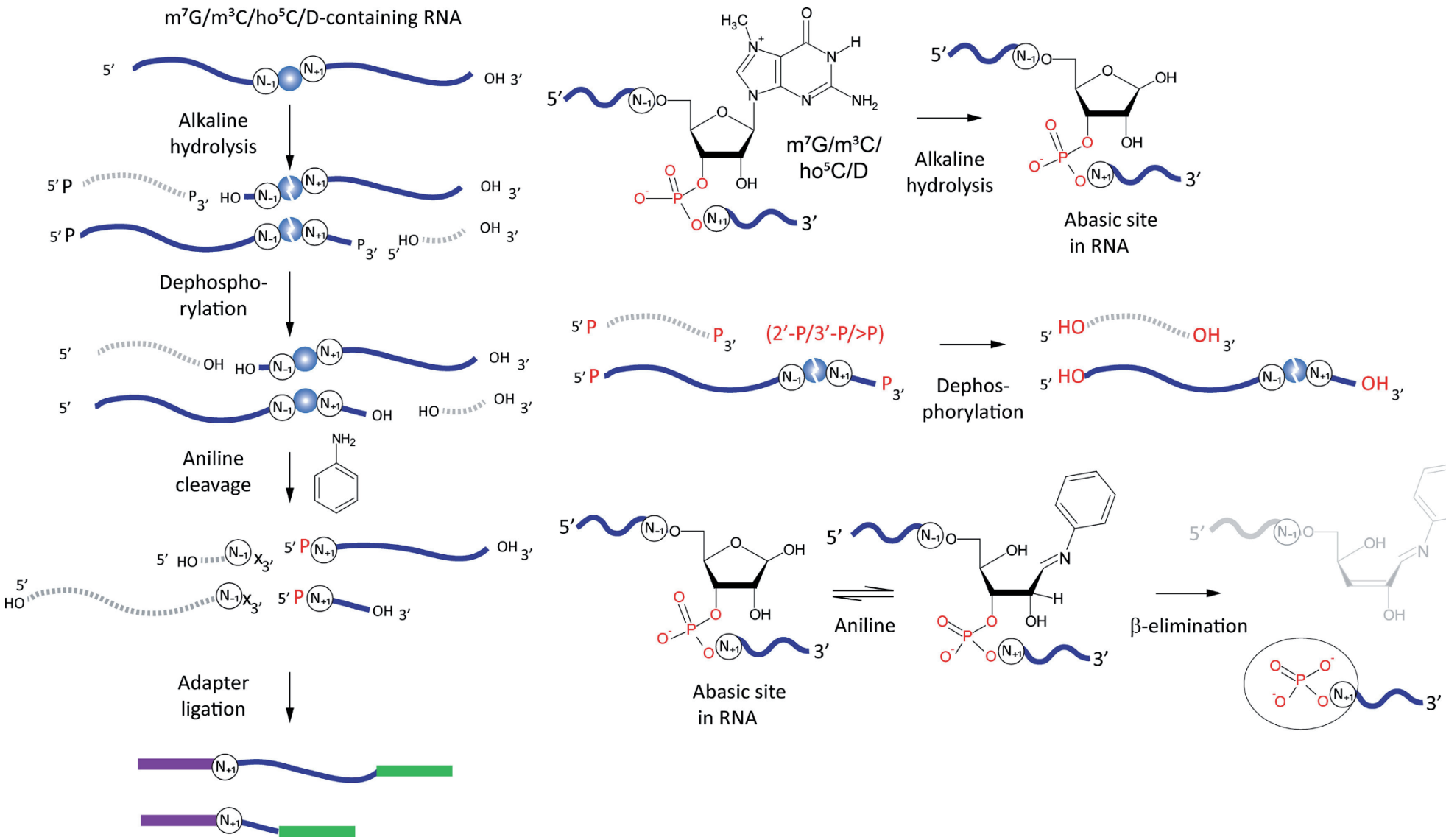

m7G/m3C/ho5C/D-containing RNA
Alkaline hydrolysis
Dephospho-rylation
Aniline cleavage
Adapter ligation
m7G/m3C/ho5C/D
Alkaline hydrolysis
Abasic site in RNA
(2'-P/3'-P/>P)
Dephos-phorylation
Aniline
β-elimination
Abasic site in RNA

Figure 40: Overview of AlkAniline-Seq Technology (Marchand et al., 2018). Schematic Representation: This illustrates the RNA chain cleavage and the subsequent ligation of primers to unique 5′-phosphates formed from abasic site decomposition. The enrichment of sequencing reads starting at the nucleotide N+1 is highlighted. Modified residues such as m7G, m3C, or D are represented as blue dots, with broken blue dots indicating RNA abasic sites. Chemical Reactions: The process involves two main steps. First, alkaline hydrolysis induces RNA fragmentation. Second, aniline cleavage followed by β-elimination occurs, forming a ribose-aniline adduct and creating abasic sites at m7G, m3C, or D residues in RNA. Normalized Cleavage: This is determined by counting the 5′-ends of reads and calculating the normalized cleavage, which is expressed in units. The formula is (reads starting at a given RNA position × 1000 / total reads aligned to this RNA), ranging from 5–25 units for background to a maximum of 1000 units for a single positive hit in RNA. Stop-Ratio: This is calculated as the proportion of reads starting at a specific position relative to all reads overlapping that position. The "Stop-ratio" is extensively used for transcriptome-wide mapping of RNA modifications, such as in Ψ-seq.

5.4 Enzymology of dihydrouridylation reaction

As noted in Section 1, dihydrouridylation is a redox reaction that necessitates two electrons and two protons to reduce the C5=C6 double bond of uracil. Essentially, this reaction mirrors hydrogenation (H2 ↔ 2e- + 2H+). In nature, D biosynthesis is mediated by flavoenzymes utilizing FMN when targeting substrates like tRNA, mRNA, or lncRNA, and by FAD for rRNA, as evidenced in my thesis research. Therefore, it is pertinent to introduce the biochemistry of FMN or FAD—two redox coenzymes intricately bound with apoproteins, whether through covalent or non-covalent associations.

5.4.1 Flavin Coenzymes and Flavoenzymes

In this section of my thesis, I will explore the crucial role of flavin coenzymes in redox biochemistry, emphasizing their impact on various cellular processes. The biosynthesis of dihydrouridine, for instance, strictly relies on redox chemistry involving flavins. Understanding the functioning of flavins is therefore essential to grasp the underlying mechanisms of this biochemical reaction. We will delve into how flavin coenzymes, such as flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), actively participate in critical redox reactions, influencing metabolic processes. This understanding will enable us to fully appreciate the specific role of flavins in dihydrouridine biosynthesis and explore their broader biological implications.

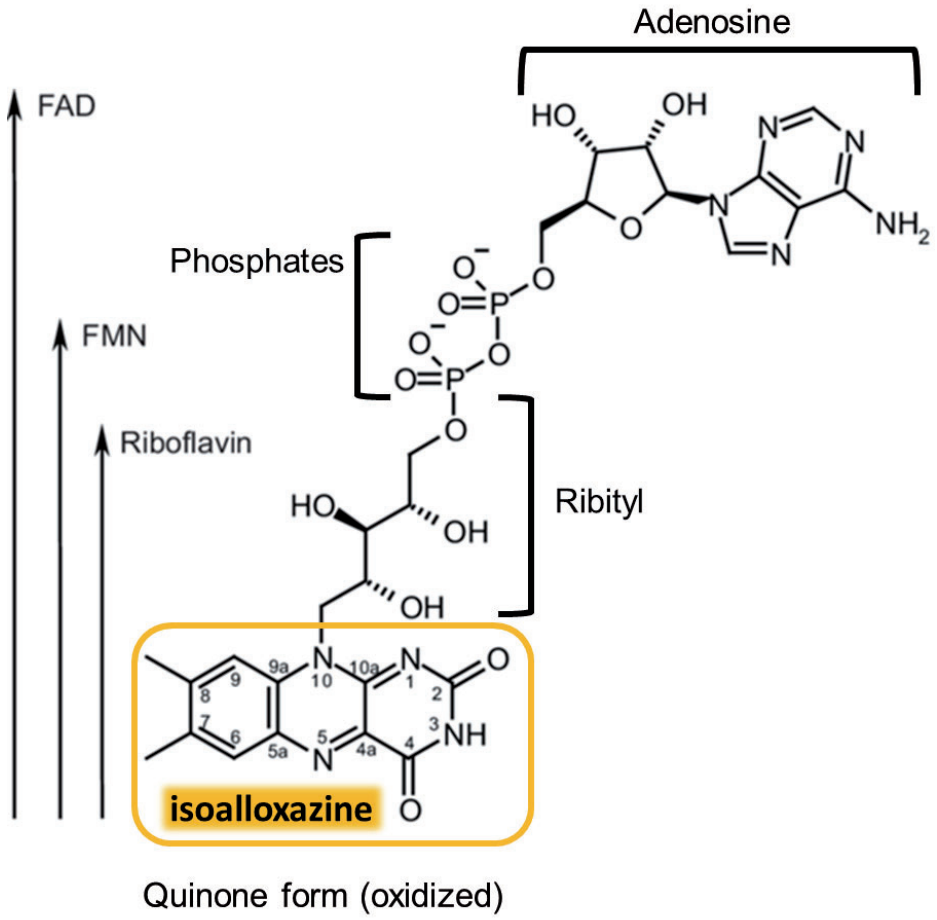


Figure 41: Chemical structures of the flavin coenzymes in the oxidized state.

5.4.1.1 Overview

Flavins are biochemical compounds derived from isoalloxazine (yellow color), playing a crucial role as coenzymes in numerous redox enzymatic reactions (Massey, 2000; F. Müller, 1987)The two primary forms of flavins utilized by enzymes are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (**Figure 41**). These coenzymes, derived from riboflavin (vitamin B2) (**Figure 41**), participate in various metabolic reactions, including energy production, detoxification of free radicals, and biosynthesis of cellular metabolites. Understanding the redox properties of these coenzymes is essential for comprehending their role in cellular biochemistry.

In bacteria such as *E. coli,* FMN and FAD are synthesized from riboflavin (Bacher et al., 2000) In contrast, humans and other animals must obtain riboflavin from the diet, as it is considered an essential vitamin. Riboflavin is synthesized in *E. coli* through a series of well-orchestrated enzymatic reactions from simple precursors like GTP and ribulose-5-phosphate (Fischer & Bacher, 2011). The initial step involves the enzyme GTP cyclohydrolase II, which converts

GTP into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate (DARPP). Subsequently, 3,4-dihydroxy-2-butanone-4-phosphate synthase acts on ribulose-5-phosphate to produce 3,4-dihydroxy-2-butanone-4-phosphate (DHBP). These products, DARPP and DHBP, are then combined by 6,7-dimethyl-8-ribityllumazine synthase to form 6,7-dimethyl-8-ribityllumazine. Finally, riboflavin synthase converts 6,7-dimethyl-8-ribityllumazine into riboflavin. Once ingested, riboflavin is converted into FMN and FAD within the organism. FMN is a phosphorylated form of riboflavin consisting of a riboflavin molecule linked to a phosphate group. In *E. coli*, FMN is synthesized through riboflavin phosphorylation by riboflavin kinase enzyme. FAD, or flavin adenine dinucleotide, is a more complex derivative where FMN is linked to adenosine monophosphate by a pyrophosphate bond. This additional structure enables FAD to interact with a broader range of proteins and enzymes compared to FMN. Then, FAD is formed through adenylation of FMN by the FAD synthetase enzyme (F. Müller, 1987)

5.4.1.2 Flavins Serving a Diverse and Varied Redox Biochemistry

Flavine, in the form of FMN and FAD, is an extremely versatile coenzyme that participates in a multitude of crucial biochemical reactions(Joosten & Van Berkel, 2007; Macheroux et al., 2011). Its versatility stems from its unique ability to accept and donate electrons, enabling a variety of redox transformations essential for cellular metabolism (Massey, 2000) Due to its redox properties, flavine plays a central role in cellular respiration, fatty acid biosynthesis, detoxification of free radicals, and DNA photorepair. For instance, in mitochondria, flavoproteins containing FAD are key players in the Krebs cycle and the electron transport chain, contributing to ATP production. Additionally, flavins are involved in bacterial bioluminescence, degradation of aromatic compounds, and regulation of cellular responses to oxidative stress. Beyond their redox capabilities, flavins are also notable for their ability to perform covalent catalysis, particularly through the N5 position of reduced flavine (Piano et al., 2017). This property allows flavins to form transient covalent bonds with substrates during enzymatic reactions, thereby facilitating complex chemical transformations. For example, in flavin monooxygenases, the N5 of reduced flavine forms a covalent adduct with molecular oxygen, creating a peroxyflavin intermediate that can subsequently hydroxylate an organic substrate (Teufel, 2024). This ability of flavin to stabilize reactive intermediates via covalent interactions significantly expands the range of chemical reactions it can catalyze, including oxidations, reductions, hydroxylations, and C-C bond cleavages.

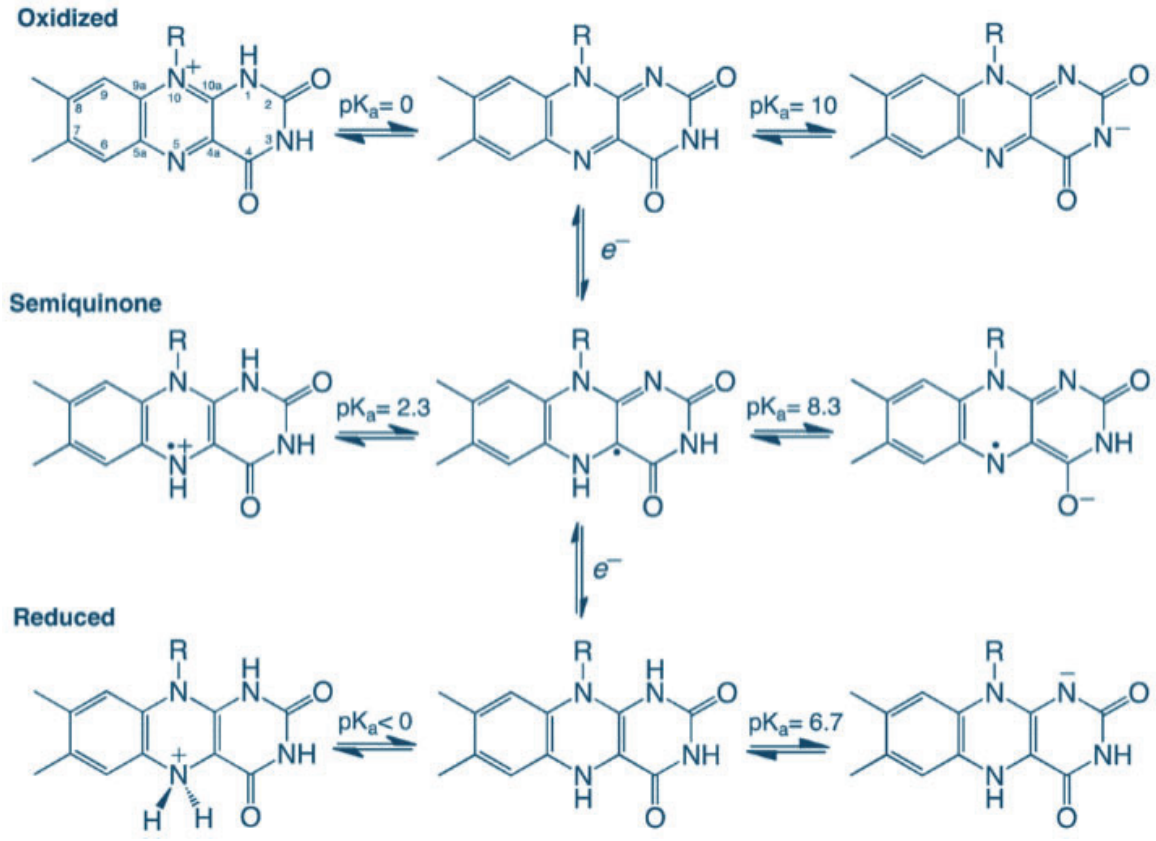


Figure 42: Chemical structure of flavin cofactors and their different redox and protonation states.

Indeed, the coenzymes FMN and FAD can exist in several distinct redox states, enabling their participation in a variety of redox reactions (**Figure 42**) (Massey, 2000). The main forms include the oxidized form (FMN/FAD), the semireduced form (FMNH•/FADH•), also known as the flavin radical, and the reduced form (FMNH2/FADH2). Flavine reduction typically occurs through electron transfer during enzymatic reactions. For instance, FMN or FAD can receive an electron to become FMNH• or FADH•, respectively, and then another electron and a proton to become FMNH2 or FADH2.

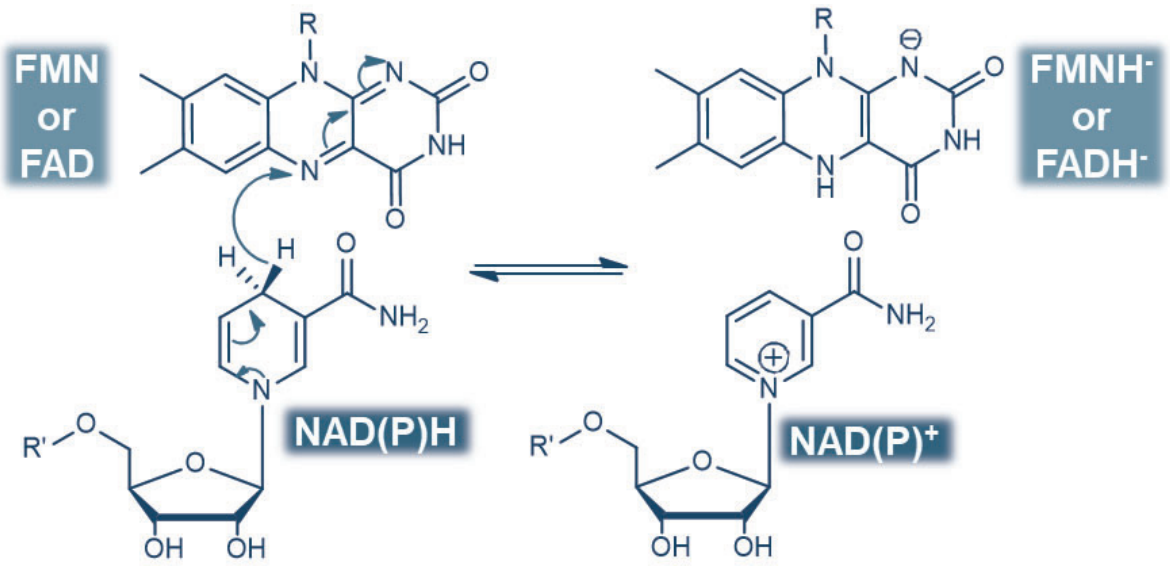


Figure 43: Chemical mechanism of hydride transfer from reduced nicotinamide derivatives (NADH or NADPH) to oxidized flavin.

The reducing chemistry of flavine can also occur directly through reaction of flavoenzymes with NADH or NADPH (**Figure 43**). In most cases, enzymes utilizing FMN and FAD receive hydride from various reducing sources, among which the most important are NADH and NADPH (Massey, 2000). NADH, or nicotinamide adenine dinucleotide, is the primary electron donor in catabolic processes such as cellular respiration and the Krebs cycle. For example, in respiratory chain complex I, NADH-ubiquinone oxidoreductase uses NADH to reduce FMN, which then transfers electrons along the chain. Similarly, in lactate dehydrogenase, NADH is used to reduce pyruvate to lactate, involving flavines as cofactors.

NADPH, or nicotinamide adenine dinucleotide phosphate, serves as the primary electron donor in anabolic processes such as fatty acid biosynthesis and photosynthesis. For instance, glutaredoxin reductase utilizes NADPH to reduce glutathione, involving FAD as a cofactor. Similarly, nitroreductase, a bacterial enzyme, uses NADPH to reduce nitro compounds, with FMN or FAD as cofactors.

In addition to NADH and NADPH, other reducing sources can include ferredoxin, an iron-sulfur protein that transfers electrons to specific flavin enzymes in bacteria and plants, as well as quinones, which participate in electron transport chains and transfer electrons to flavin enzymes.

5.4.1.3 Enzymes Using FMN or FAD, typical structure

Enzymes utilizing the coenzymes FMN and FAD adopt various folding conformations to facilitate the incorporation of these coenzymes and catalyze reactions. A typical structure found in many FAD-utilizing enzymes is the Rossmann fold, a motif composed of parallel beta strands alternating with alpha helices, forming a sandwich-like structure (Hanukoglu, 2015) This folding pattern is characteristic of enzymes involved in fatty acid metabolism and the Krebs cycle. Similarly, a common architecture in FMN-utilizing enzymes is the TIM barrel (Alpha/Beta barrel), consisting of eight beta strands forming a barrel surrounded by eight alpha helices (Wierenga, 2001). This structure is frequently found in oxidoreductases and hydratases that require high stability and flexibility for enzymatic interactions. As we will see these two types of folding are encountered in the enzymology of dihydrouridylation (**see articles 3 in chapter II**).

5.4.2 The classical Dihydrouridine synthases: FMN-dependent Flavoenzymes

A family of enzymes known as dihydrouridine synthases (Dus), utilizing FMN as a coenzyme, catalyzes the synthesis of D in tRNAs, mRNAs, and the bacterial YrlA lncRNA. These flavoenzymes are categorized into three main groups and eight subfamilies, all evolving independently from an ancestral gene. Prokaryotes possess three Dus enzymes (DusA, DusB, and DusC), eukaryotes have four (Dus1 to Dus4), and Archaea have one (Kasprzak et al., 2012) (**Figure 45**). Among these, bacterial DusB is considered the oldest, while Dus3 is possibly the ancestral eukaryotic enzyme from which others evolved via gene duplication. The identity of the rRNA dihydrouridine-synthesizing D-synthase remained unknown until our recent discovery, detailed in upcoming sections.

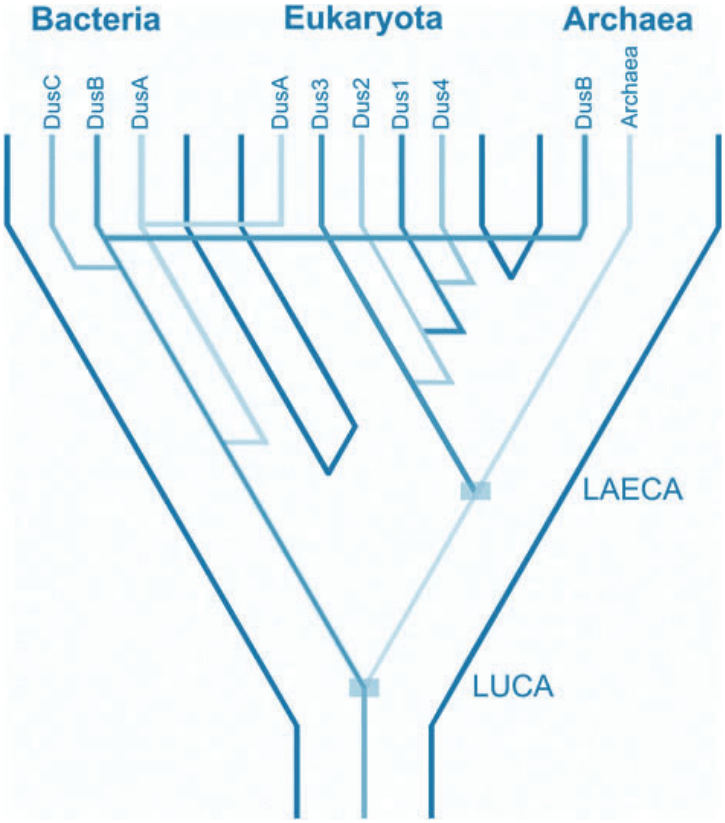


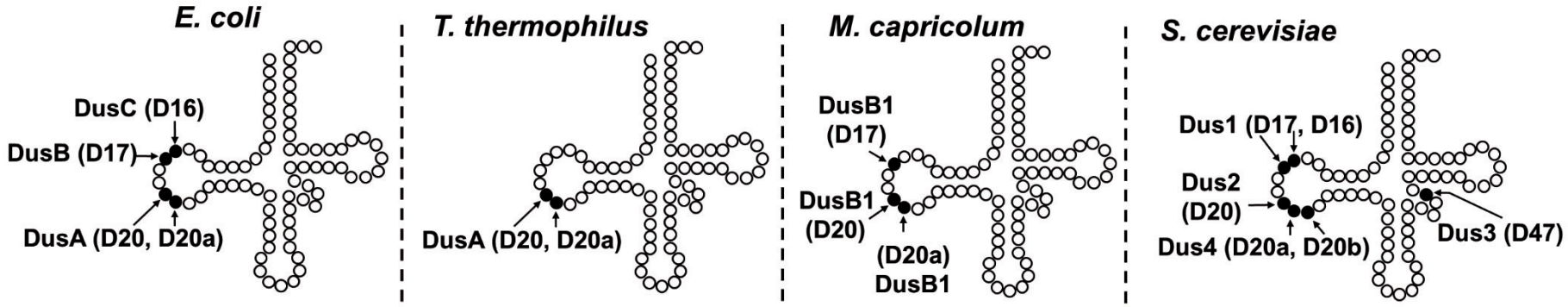
Figure 45: adapted from (Kasprzak et al., 2012) . Speculative scenario for the evolutionary history of the DUS family is based on the following assumptions: Bacteria, Archaea, and Eukaryota are monophyletic; Archaea and Eukaryota are sister lineages; and the root, corresponding to the Last Universal Common Ancestor (LUCA), is between Bacteria and the Last Archaeal and Eukaryotic Common Ancestor (LAECA). The evolutionary tree has three main branches for the three domains of life. Bifurcations represent gene duplications leading to paralogous families, and horizontal lines indicate horizontal gene transfers.

5.4.2.1 Site specificities of Dus enzymes

The site specificity of *S. cerevisiae's* four Dus enzymes Dus1, Dus2, Dus3, and Dus4 was comprehensively characterized using three techniques: determining the molar ratios of dihydrouridine in tRNAs from *dus* mutants, microarray analysis of tRNAs, and primer extension (Xing et al., 2004). These studies revealed that Dus1, Dus2, Dus3, and Dus4 synthesize D16-D17, D20, D47, and D20a-D20b, respectively, with each enzyme exhibiting specific and nonredundant activities (**Figure 46**). Bishop et al. identified COG0042 genes in *E. coli* as encoding Dus enzymes responsible for introducing D into tRNAs (Bishop et al., 2002) Recent research has further elucidated the specificities of Dus enzymes across various organisms, including *E. coli*, *Mycoplasma. capricolum* and *B. subtilis* (Bou-Nader et al., 2018) (Faivre et al., 2021) (Sudol et al., 2024). In *E. coli* DusA, DusB, and DusC catalyze D20-D20a, D17, and D16, respectively (**Figure 46**). Eukaryotic Dus but not bacterial enzymes also modify mRNA substrates (Dai et al., 2021; Draycott et al., 2022; Finet, Yague-Sanz, Krüger, et al.,

2022). The involvement of Dus enzymes in both tRNA and mRNA dihydrouridine synthesis underscores their critical role in RNA modification. This broader trend is reflected in other enzymes like NSUN2 and Trm4 (Squires et al., 2012) which modify both tRNA and mRNA substrates, illustrating a conserved function across diverse RNA types.

The lack of knowledge in prokaryotic dihydrouridine (D) biosynthesis notably Gram+ bacteria prompted us to focus on Gram-positive bacteria. Phylogenetic analysis done in the lab before my arrival identified three DusB protein subgroups in Firmicutes: DusB1, DusB2, and DusB3 (Faivre et al., 2021). While most organisms carry either DusB1 or DusB2, approximately 40% have both. *Bacillus* species generally retain both DusB1 and DusB2, whereas Mollicutes have DusB1 but lost DusB2. In contrast, most *Staphylococcus* species kept DusB2 and lost DusB1. Both DusB subgroups originated from an ancestral DusB duplication in the common ancestor of Firmicutes. The low abundance of DusB3, primarily found in *Clostridiae*, suggests it arose more recently. Recent biochemical characterization from our lab revealed that *M. capricolum* DusB1 (McDusB1) is multisite specific, modifying all expected U residues (U17, U20, and U20a) (**Figure 46**) (Faivre et al., 2021). This multisite specificity is supported by tRNA modification profiles of *Lactococcus lactis* and *Streptomyces griseus*, which have only a single DusB1 homolog but display D17, D20, and D20a in their tRNAs(Boccaletto et al., 2022). These findings suggest that DusB1's multisite specificity is a common trait in Gram-positive bacteria. DusB2 likely exhibits similar behavior in organisms that possess only this homolog, such as *S. aureus*, which shows D17, D20, and D20a modifications in its tRNA. **In article N°2**, we thoroughly investigated dihydrouridine (D) biosynthesis in *B. subtilis* through genetic, biochemical, and epitranscriptomic approaches. Our findings revealed that the two DusB1 and DusB2 homologues introduce all D residues into its tRNAs. DusB1 demonstrated multisite enzyme activity, forming D at positions 17, 20, 20a, and 47, while DusB2 specifically catalyzed D at positions 20 and 20a, indicating functional redundancy among these modification enzymes (**Figure 46**). Extensive tRNA-wide D-mapping using AlkAniline-Seq showed that this redundancy affects most tRNAs, with DusB2 displaying higher dihydrouridylation efficiency than DusB1. Notably, BsDusB2 can function similarly to BsDusB1 when overexpressed in vivo and at higher enzyme concentrations in vitro (**Figure 46**). This property is unique in RNA modification enzymology.



The particular case *of B. subtilis*

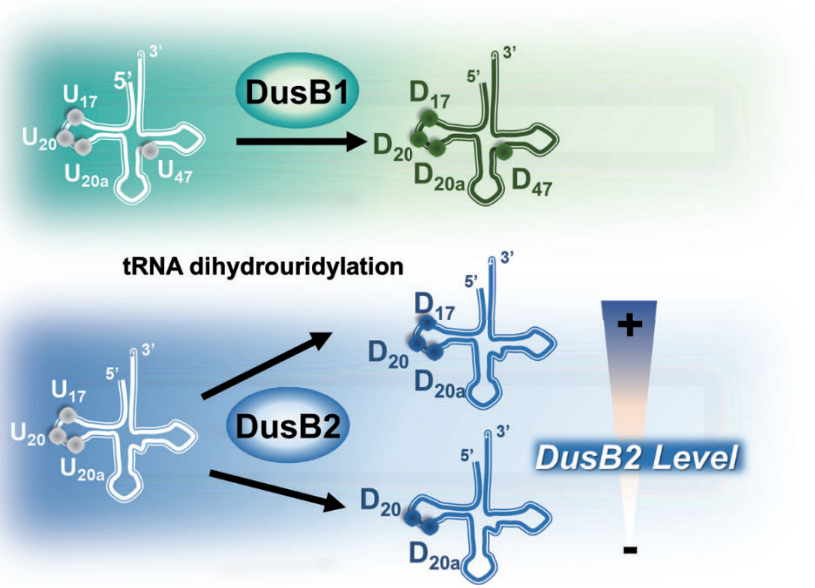


Figure 46: adapted from (Sudol et al., 2024). Location of D-sites in tRNA and the corresponding enzyme involved in site dihydrouridylation determined experimentally. Schematic representation of the secondary structure of tRNA, showing the location of D residues and the corresponding Dus enzyme responsible for their synthesis in *E. coli*, *T. thermophilus* and *M. capricolum* for eubacteria and *S. cerevisiae* for eukaryotes. In the lower panel is shown the case of B. subtilis in which the specificities of the two Dus homologues can be modulated by enzyme concentration in vivo and in vitro.

5.4.2.2 Structural organization of Dus enzymes

The X-ray crystallographic structures of several Dus enzymes have established a canonical fold for this enzyme family (Lombard & Hamdane, 2017) (**Figure 47**). To date, structures have been resolved for three bacterial Dus homologues: DusA from *Thermus. thermophilus*, and DusB and DusC from *E. coli*(Bou-Nader et al., 2018; Byrne et al., 2015; Yu et al., 2011). However, structural data for eukaryotic enzymes are limited, with only isolated domain structures available for human Dus2 (hDus2), recently solved by our group (Bou-Nader et al., 2015, 2019; Lombard et al., 2022). All Dus enzymes feature a shared catalytic N-terminal domain that adopts a TIM-barrel fold (**Figure 47**). This domain hosts an active site within a solvent-accessible crevice that accommodates the FMN prosthetic group and uridine substrate. Following this catalytic domain is a helical domain (HD), composed of four parallel helices organized into a bundle (**Figure 47**). Additionally, the canonical fold can accommodate extra domains, which can be positioned at the N-terminal or C-terminal ends, or both—a feature primarily observed in eukaryotic Dus enzymes (**See Article 1**) (**See review**), involved in recognizing RNA substrates.

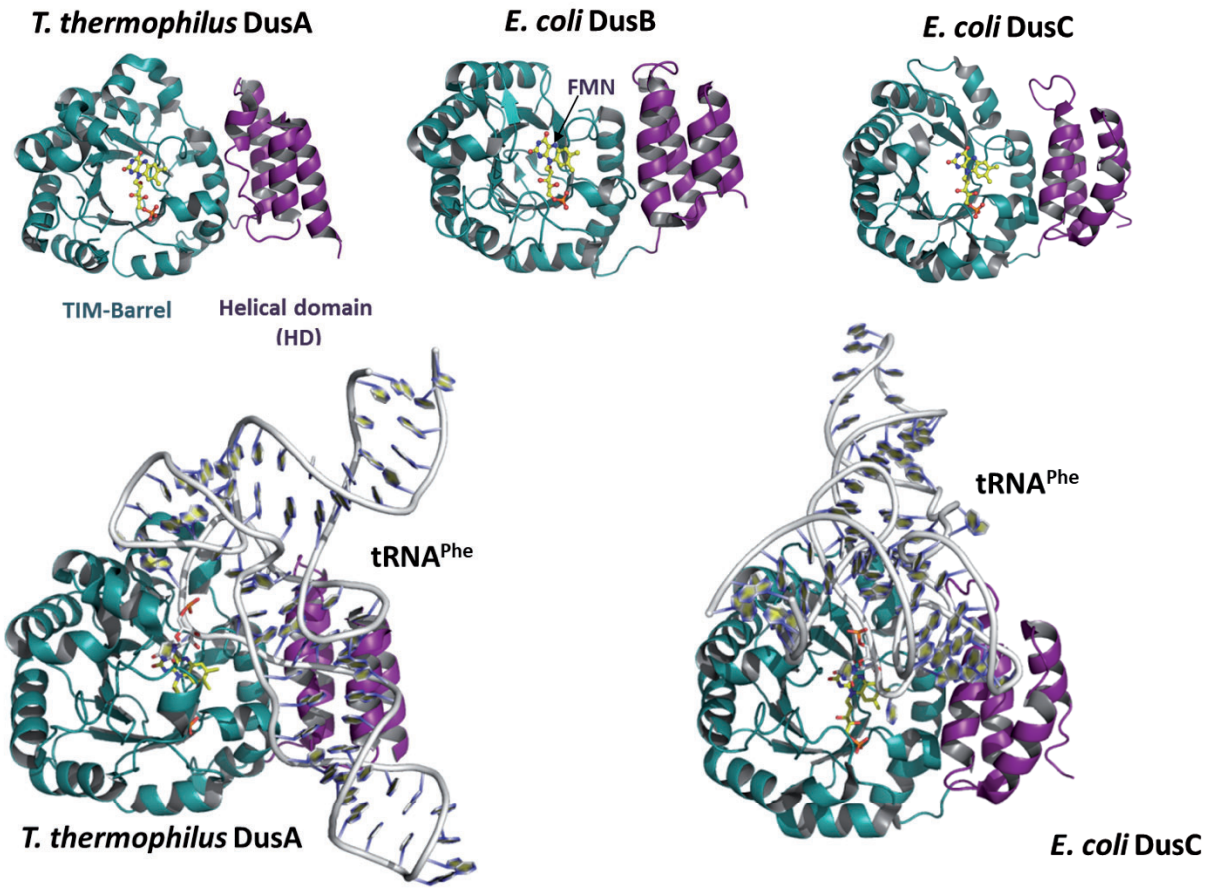


Figure 47: Crystallographic structures of bacterial Dus alone or in complex with tRNA adapted from (Brégeon et al., 2022)**.** The TIM-barrel, HD, and dsRBD domains are colored in deep-teal, purple, and olive, respectively. The prosthetic group, FMN, is in yellow stick. The pdb codes used are 3B0P, 6EI9, 4BFA for *T. thermophilus* DusA, *E. coli* DusB and *E. coli* DusC, respectively. Below the crystallographic structure of the protein/RNA complexes, namely *T. thermophilus* DusA:tRNA^{phe} and *E. coli* DusC:tRNA^{Phe} whose PDB codes are 3B0V and 4YCO, respectively.

The structures of *T. thermophilus* DusA and *E. coli* DusC in complex with tRNAs have elucidated the molecular basis of tRNA substrate recognition by bacterial enzymes(Byrne et al., 2015; Yu et al., 2011) (**Figure 47**). These structures reveal that the two canonical domains provide a platform for RNA recognition, involving numerous ionic interactions, particularly between positively charged residues (e.g., Lys, Arg) and the bases, ribose, and phosphate backbone of the substrate. The recognition mode differs depending on whether the enzyme is specific for U16 or U20. Although the overall structures of DusA and DusC are conserved, they access the target uridine by positioning their substrate tRNAs differently on their surfaces, involving a 160° rotation between the tRNA on DusC and that on DusA (**Figure 47**). DusA recognizes a larger portion of the tRNA than DusC, particularly the anticodon stem, which

DusC does not bind. Both enzymes recognize the elbow region, the D-stem-loop, and the T-loop. For DusA, the D and T-loops are recognized solely by the TIM-barrel domain, while the D and anticodon stems are recognized by the HD. Conversely, for DusC, the D-loop is recognized by both canonical domains, the D-stem is recognized only by the TIM-barrel domain, and the T-loop is recognized by the HD. In these complexes, the Dus enzymes bind to tRNAs without disrupting the crucial interactions that maintain their tertiary structure. Thus, the tRNA elbow likely serves as a quality control checkpoint that Dus scrutinizes before dihydrouridylation. Finally, the two enzymes flip their uridine substrate and stack it on the isoalloxazine ring of the FMN to proceed with its reduction (see the mechanism of Dus in the next section).

5.4.3 Chemical mechanism of tRNA dihydrouridylation

The chemical mechanism of dihydrouridylation catalyzed by dihydrouridine synthases (Dus) has been elucidated mainly through the study of yeast Dus2 (Lombard & Hamdane, 2017) and crystallographic structures of various Dus active sites, particularly in complex with tRNA(Byrne et al., 2015; Yu et al., 2011). Dus enzymes follow a flavoenzyme catalytic cycle: FMN is reduced by NADPH, followed by oxidation of the reduced flavin by the uridine substrate (**Figure 48**). Structural studies of bacterial Dus show spatial constraints preventing simultaneous binding of NADPH and uridine, suggesting a ping-pong mechanism where NADP+ exits before uridine binds. FMNH– then transfers a hydride to the C6 position of uridine, followed by protonation at C5, altering the base's saturation (**See review**) (**Figure 48**). A conserved cysteine (C93 in T. thermophilus DusA, and C98 in E. coli DusC) acts as a general acid in this step (**Figure 48**). The validity of this redox mechanism is supported by structures of *T. thermophilus* DusA and *E. coli* DusC, where uridine is positioned near the N5 of FMN, stabilized by hydrogen bonds with conserved residues (**Figure 48**). Similar redox mechanisms are seen in related enzymes like dihydropyridine and dihydroorotate dehydrogenases, which share a homologous catalytic domain with Dus enzymes (Bishop et al., 2002; Dobritzsch, 2001).

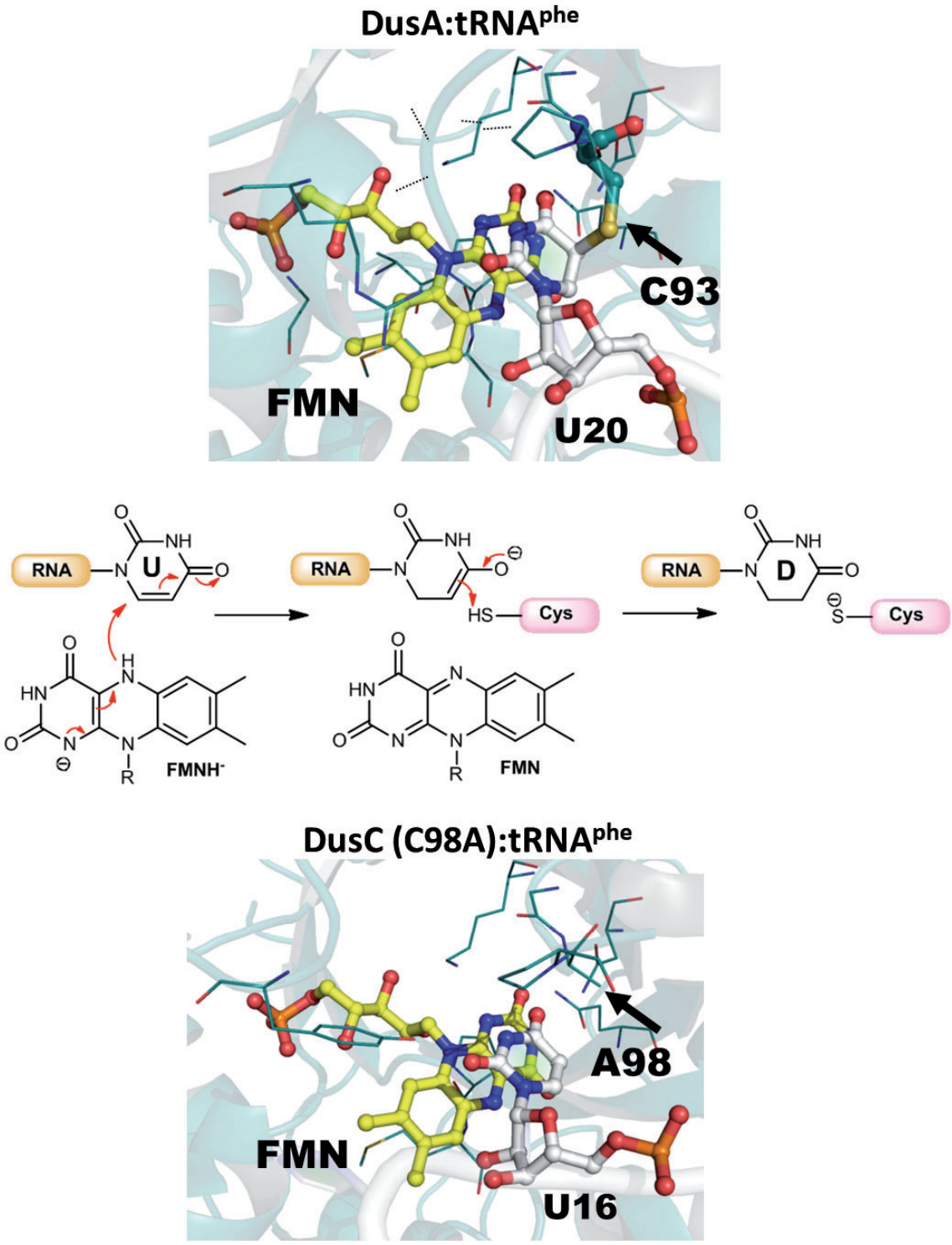


Figure 48: adapted from (Brégeon et al., 2022)**.** Proposed chemical mechanism of D biosynthesis and activity-based inhibition of Dus. (A) Focus on the active site of *T. thermophilus* DusA and *E. coli* DusC in complex with their respective tRNA substrates. The FMN is represented in yellow ball–sticks, while the uridines are in white ball–sticks. The figure shows the postulated chemical mechanism of Dus enzymes between the two panels of Dus active sites.

5.5 Biological D-Relevance in Prokaryotes

The relevance of D in the conformational dynamics of RNAs was established early on by structural approaches (see previous sections). On the other hand, this function took on its full biological meaning when MacCloskey analyzed the quantitative composition of post-transcriptional modifications in tRNAs isolated from psychrophilic organisms having the capacity to grow under extremely low temperatures ranging from −5 to 12 °C(J. J. Dalluge et al., 1997; Noon et al., 2003). These organisms have implemented evolutionary strategies to counteract the restriction of molecular mobility and to maintain a form of resilience in the face of low temperature by incorporating into their biomolecule, biochemical components that have the capacity to maintain molecular flexibility. Similarly, these organisms show much less post-transcriptional modifications in their tRNAs, although they retain some of them such as pseudouridine, m5U, and m7G at normal levels. In dramatic contrast, D levels are found to be between 40 and 70% higher than those found in mesophilic organisms, such as *E. coli.* Thus, these biological data together with structural information corroborate the role of D in promoting the local fluctuation and mobility of nucleic acids. Depletion of D by deletion of dus genes does not cause significant defects in growth phenotype of model organisms such as eubacteria *T. thermophilus*, (Kusuba et al., 2015), *E. coli*, (Bishop et al., 2002) or *S. cerevisiae*, (Xing et al., 2004). Similarly, in the absence of identified enzyme-catalyzing rRNA dihydrouridine synthesis, O'Connor et al. replaced U2449 target of dihydrouridylation C2449 and evaluated its effect on *E. coli* physiology and ribosome function by direct mutation of U2449 to C2449. The results obtained in this study pointed out that ribosomal D2449 is dispensable to the cell, but mutation to A2449 or G2449 was not viable (O'Connor, 2001). These results are not surprising by themselves because many so-called "nonessential" modifications, which are located mainly in the body of the RNA molecule, produce only minor phenotypic impact following their removal. On the contrary, this phenomenon makes sense given the fact that these modifications are part of an interconnected network where compensation phenomena, or functional redundancy, may occur. The biological relay of these modifications becomes relevant when this network is disturbed under particular stressing events or beyond the simple loss of a single modification. It is in this context that Phizicky uncovers the importance of D in combination with m7G46 in yeast $tRNA^{Val}_{AAC}$ (Alexandrov et al., 2006). Indeed, the double mutants *dus/trm8* and particularly *dus3/trm8* produce severe growth defects. Molecular analysis has revealed that this growth defect coincides with a rapid decrease at the steady-state level of the pool of this tRNA via its rapid intracellular degradation, which approaches the degradation

rates of mRNAs. Hence, by maintaining the functional folding of RNAs in cooperation with its relatives, D acts as a kind of quality control mark for RNAs.

Our work on *B. subtilis* seems to comfort the idea of the need of D to cope with suboptimal temperature (**see article N°2**). *B. Subtilis* is a fast-growing, Gram-positive, aerobic bacterium with rod-shaped cells that are typically 2–6 µm long and just less than 1 µm in diameter. The optimal growth temperature is about 30–37 C. The influence of the lack of BsDusB and by extrapolation of D on *B. subtilis* growth was investigated in LB medium at 23, 30 and 37°C (**Figure 49**). At 37°C, the generation time of *B. subtilis* W168 is 21 min while in the case of the three strains deleted in dus it is very slightly increased by 4 minutes. The effect is slightly more visible when the growth curves were conducted at 30°C, where the generation time was raised from 32 min for the wild type to 40 min for the two single mutants, and 42 min for the double mutant strain. Remarkably, this difference becomes significantly larger when the temperature is lowered to 23°C, with generation time increasing from 49 min for the WT strain to 87 min for the three mutant strains due to the absence of D. Thus, the absence of D does not seem to have too great an impact on B. subtilis at physiological growth temperatures but becomes significant during cell growth at 23°C that is consistent with the role of this modified base in promoting structural flexibility at the tRNA level rather needed at lower temperatures than at higher ones.

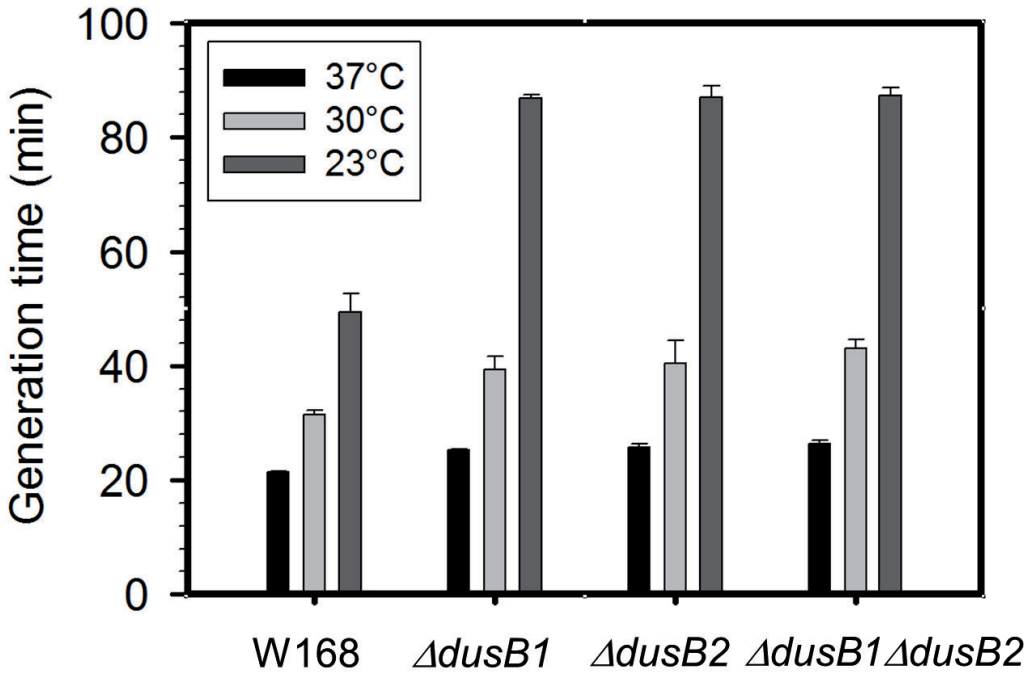


Figure 49: Role of BsDusB in the growth phenotype of *B. subtilis.*

See review: Dihydrouridine in the Transcriptome: New Life for This Ancient RNA Chemical Modification.

pubs.acs.org/acschemicalbiology

Reviews

Dihydrouridine in the Transcriptome: New Life for This Ancient RNA Chemical Modification

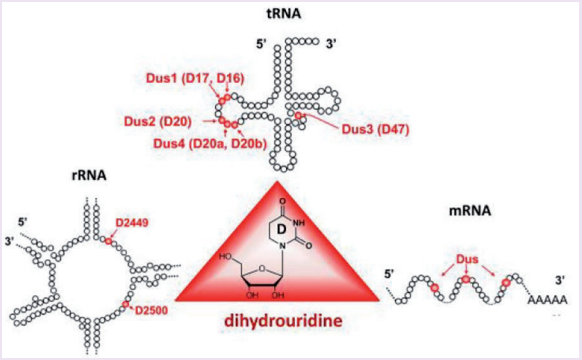
Damien Brégeon, Ludovic Pecqueur, Sabrine Toubdji, Claudia Sudol, Murielle Lombard, Marc Fontecave, Valérie de Crécy-Lagard, Yuri Motorin, Mark Helm, and Djemel Hamdane*

Cite This: *ACS Chem. Biol.* 2022, 17, 1638−1657

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Until recently, post-transcriptional modifications of RNA were largely restricted to noncoding RNA species. However, this belief seems to have quickly dissipated with the growing number of new modifications found in mRNA that were originally thought to be primarily tRNA-specific, such as dihydrouridine. Recently, transcriptomic profiling, metabolic labeling, and proteomics have identified unexpected dihydrouridylation of mRNAs, greatly expanding the catalog of novel mRNA modifications. These data also implicated dihydrouridylation in meiotic chromosome segregation, protein translation rates, and cell proliferation. Dihydrouridylation of tRNAs and mRNAs are introduced by flavin-dependent dihydrouridine synthases. In this review, we will briefly outline the current knowledge on the distribution of dihydrouridines in the transcriptome, their chemical labeling, and highlight structural and mechanistic aspects regarding the dihydrouridine synthases enzyme family. A special emphasis on important research directions to be addressed will also be discussed. This new entry of dihydrouridine into mRNA modifications has definitely added a new layer of information that controls protein synthesis.



INTRODUCTION

Many canonical nucleotide residues found in RNA polymers undergo extensive chemical modifications after the transcriptional step catalyzed by a wide variety of enzymes called "RNA modifying enzymes".[1] These post-transcriptional modifications are part of the complex maturation processes that eventually generate functional RNA molecules. Seventy years ago, pseudouridine (Ψ), also called the "fifth nucleotide" of RNA, was discovered.[2] This modified base has since proven to be one of the most abundant modifications in the transcriptome. Today, more than 170 distinct chemical modifications have been identified, and their number is still steadily climbing. Modifications are widely distributed among different types of RNA, including transfer RNA (tRNA), ribosomal RNA (rRNA), long noncoding RNA (lncRNA), and messenger RNA (mRNA).[1] After having been the realm of specialists for many years, enthusiasm for this field of research has been rekindled as more RNA modifications are shown to have essential biological roles, particularly in regulating gene expression.[3] The understanding that some of RNA modifications (at least m^6A,[4] and, probably m^1A, m^3C, m^1G and m^2_2G[5,6,78]) may be removed by carbon oxidation and thus reversible and respond to environmental changes have revived a renewed interest in the discovery of their biological functions.[9] Most of the modified bases identified to date have been in tRNA and rRNA; however, rapid advances in the field notably with new developments in the accurate detection and quantification of these epitranscriptomic marks, combined with the ability to detect low-content RNA modifications, have expanded the list of such modifications in mRNA.[10] Although, detecting RNA modifications remains a challenging task, these technical advances, particularly with the availability of much more specific labeling, have allowed to establish the patterns of distribution of these modifications at the level of the whole transcriptome. On the other hand, many aspects of the enzymology and structural biology of RNA-modifying enzymes, which provide the molecular basis underlying the biogenesis of these modifications, are still lagging far behind.

Because tRNAs are the most heavily modified RNA molecules, their study has provided the core of our understanding of the RNA modification machinery and progressively revealed the functions of many RNA modifications in recent years.[11−16] Most RNA modifications have a

Received: April 11, 2022
Accepted: June 7, 2022
Published: June 23, 2022

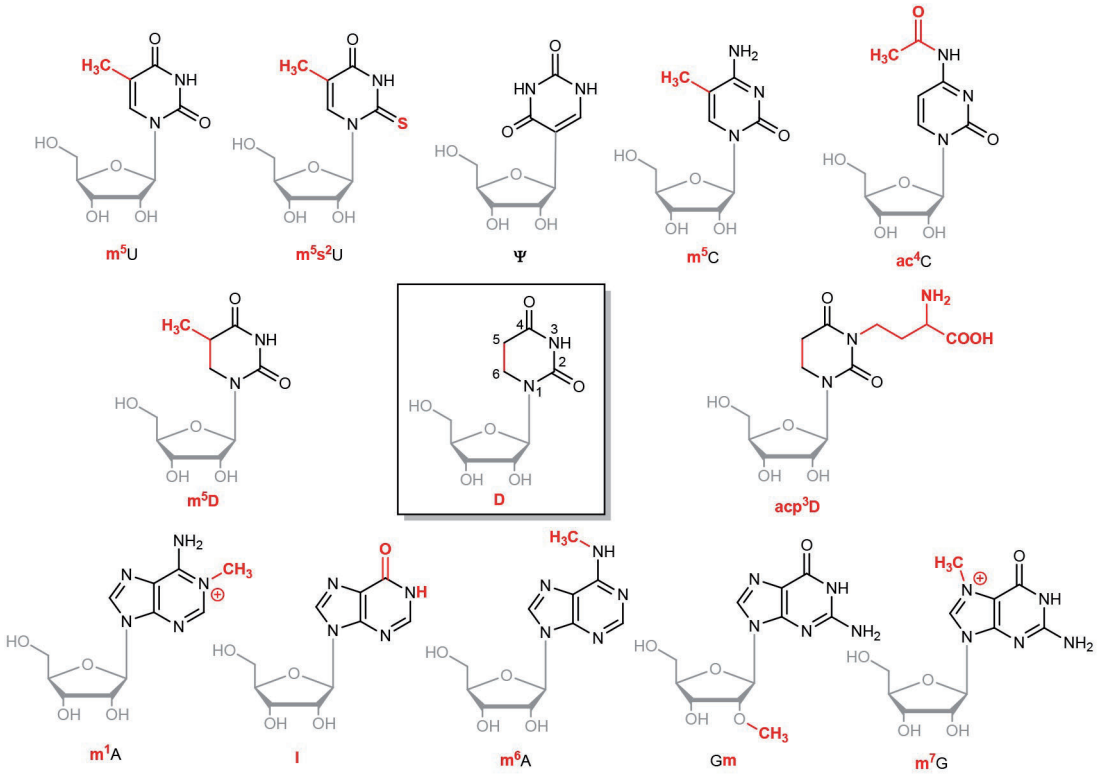


Figure 1. Chemical structure of some modified nucleosides found in different types of RNAs. The modification made to the canonical base or ribose is indicated in red. The chemical structure of dihydrouridine (D) is boxed, and the atom numbering of the base is also indicated. The acronym of the modifications is indicated under the corresponding modified nucleoside.

fairly simple chemical nature and often involve methylation, isomerization, reduction, or deamination reactions. For example, 5-methyluridine (m^5U), pseudouridine (Ψ), 1-methyladenosine (m^1A), dihydrouridine (D), or 2′-O-methylguanosine (G_m) (Figure 1) are often observed in loop regions of tRNAs and rRNAs. Evolutionarily conserved enzymes typically catalyze the biosynthesis of these classes of modified nucleosides, whose position and identity are conserved in the majority of these RNA species. Others may have a very complex chemical nature, but they are usually found in the anticodon loop of tRNAs. They facilitate or prevent interactions with amino-acyl-tRNA synthetases and translation factors, and they allow a precise decoding of mRNAs on the ribosome via accurate codon/anticodon interactions.[17] This second category represents hypermodified bases, and their biosynthesis often involves many enzymatic steps, sometimes within protein complexes. In mRNA, only simple modifications have been identified to date (Figure 1), and their functions are slowly being deciphered. Modifications in mRNA (N6-methyladenosine (m^6A), 5-methylcytidine (m^5C), pseudouridine (Ψ), inosine (I), N1-methyladenosine (m^1A) and 7-methylguanosine (m^7G), N4-acetylcytidine (ac^4C)) play important structural roles and are involved in RNA stability (directly or indirectly) and also found to regulate several mRNA cell cycle processes like mRNA export, splicing and translation.[4,16,18−21] The role of m^6A on RNA stability depends on the m^6A binding proteins. The general belief is that m^6A on mRNA promotes RNA degradation. However, when m^6A is recognized by IGF2BP proteins, the mRNA is stabilized.[22] Recently, D, one of the most abundant modified bases in tRNAs, that gave its name to tRNA's D-loop structure, has just entered the world of mRNA modifications with important physiological roles in cell growth.[23,24] In this review, we attempt to summarize the current knowledge about the distribution/frequencies of D modification in the transcriptome, the chemical labeling tools used to detect D, and the enzymology of D formation as well as the relevance of these modifications in translation and cancer biology. In addition to the already existing structural information on Dus proteins, we also present the first accurate structural models of human enzymes catalyzing RNA dihydrouridylation obtained by AlphaFold. The accuracy of the 3D models can be found at https://alphafold.ebi.ac.uk/. All these structural elements serve as a solid basis to discuss some important aspects related to the functional involvement of these proteins in epitranscriptomic under normal and pathological cellular states.

STRUCTURAL VERSUS CHEMICAL PROPERTIES OF DIHYDROURIDINE AND QUANTITATIVE APPROACHES

D is one of the most conserved and abundant modified bases in the transcriptome, second only after pseudouridine, primarily due to its high presence in tRNAs. The D content in tRNAs seems to correlate with the growth temperature, since at high growth temperatures, D ring undergoes hydrolysis.[25] Thus, the highest tRNA D-content is observed in psychrophilic organisms, where 40−70% more of D was found, compared to mesophilic bacteria, while much lower D content was found in hyperthermophiles.[26−28] D is formed by the reduction of the C5−C6 double bond of the pyrimidine ring of uridine leading to a saturated base (Figure 1), a unique feature found to date in nucleic acids. We will see below that

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

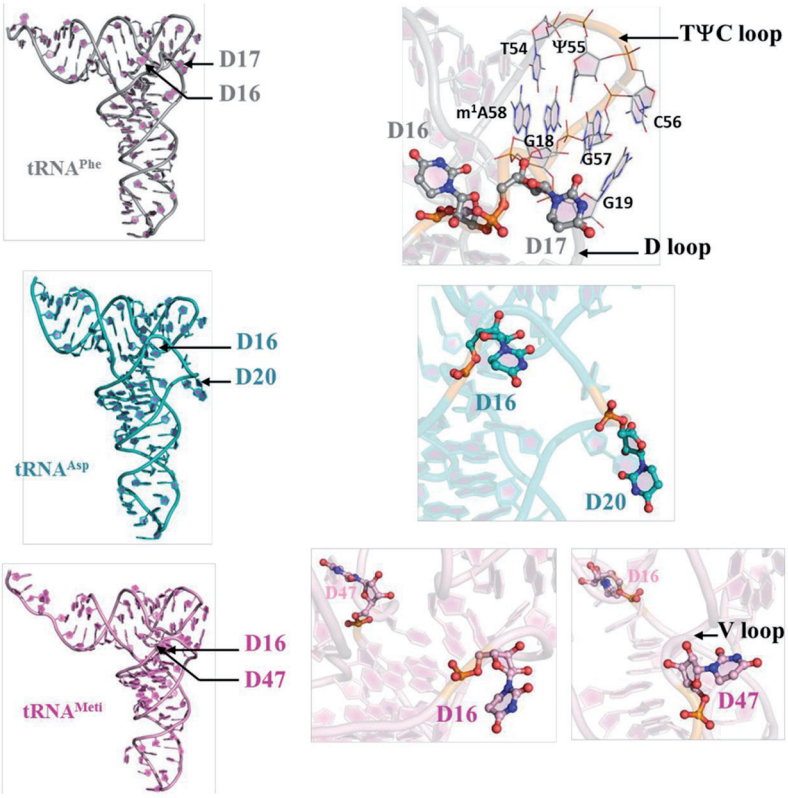


Figure 2. X-ray structures of three matured yeast tRNAs. The full-length tRNA structures Phe, Asp, and Met initiator are shown on the left side in gray, cyan, and pink colored cartoons, respectively. The pdb codes used are 1EVV, 3TRA, and 1YFG for $tRNA^{Phe}$, $tRNA^{Asp}$, and $tRNA^{Met}_{i}$, respectively. On the right side, the D-residues found in each of the tRNAs are represented in stick in its corresponding color. Nucleotide involved in tertiary interactions that are highly conserved in cytosolic tRNAs, including the Hoogsteen-reverse base pair T54-A58, the interloop base pairs G18-Ψ55 and G19-C56, and the stack of four mutually interspersed purine bases A58-G18-A/G57-G19 are shown in $tRNA^{Phe}$.

this property, which distinguishes it from other modifications, gives it unique physicochemical and structural properties, which can be judiciously exploited for analytical purposes. Initially discovered in tRNAs by Holley's group in the mid-1960s,[29] this modified base was later observed in much lower abundance in rRNAs and some long noncoding RNAs. Early studies by Bonner had shown that D can also be found in histone-associated RNAs in plants and mammals,[30] albeit without localization data. The very recent and independent studies of Dai et al.[23] and Finet et al.[24] using two different approaches came to the same conclusion, namely the presence of D in mRNAs. We now know that it is more than likely that the distribution of D in the transcriptome is broader than expected.

Structural Properties of Dihydrouridine. The peculiar structural properties of D nucleoside were defined early on from the crystallographic structure of free D hemihydrate[31] and the first structures of mature tRNAs isolated from *Saccharomyces cerevisiae*[32−35] (Figure 2), while NMR studies on D-containing oligonucleotides have provided important information regarding the dynamical attributes that this modified base plays on RNA.[36−38] These structures reveal the lack of planarity of the D heterocycle and a shift of carbons C5 and C6 to opposite sides of the plane formed by positions N1, C2, N3 and C4, resulting in a half-chair conformation (Figure 2, zoom on D sites). This has notable structural consequences, among them D cannot participate in stacking interactions with neighboring aromatic bases. Moreover, the presence of a CH_2 in place of a CH increases the electronic volume of the base, thus promoting the inherently flexible C2′-endo conformation of its ribose. To date, this modification is the only one known to favor such a conformation, whereas others RNA modifications such as 2′-O-ribose methylation,[39] 2-thiolation of ribothymidine,[40] and pseudouridylation[41] rather favor the C3′-endo conformation observed in A-type RNA helix, conferring stability to the RNA by reinforcing base stacking.[14] Thus, these structural properties of D, and in particular the C2′-endo conformation of the sugar, locally affect the RNA structure by introducing functionally important local flexibility. This dynamic property will have different consequences depending on whether the RNA has a particular structure or not. Indeed, this increased local flexibility mediated by D may facilitate formation of interactions between neighboring tertiary bases in the critical tRNA elbow region. In fact, this region, formed by the kissing between the D and TΨC loops, involves several critical interactions that are highly conserved in cytosolic tRNAs, including the Hoogsteen-reverse base pair T54-A58, the interloop base pairs G18-Ψ55, G19-C56, and the stack of four mutually interspersed purine bases, A58-G18-A/G57-G19 (Figure 2). The decrease in melting temperature of *Escherichia coli* tRNAs observed in the absence of dihydrouridine[42] could well be explained by the loss of flexibility of the D-loop required to accommodate these tertiary interactions essential for the maintenance of the tRNA 3D structure. In the course of their investigation on siRNA, Sipa et al. evaluated the thermodynamic stability and gene silencing activity of a series of nucleobase-modified RNA duplexes containing modified bases nucleosides, including D.[43]

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

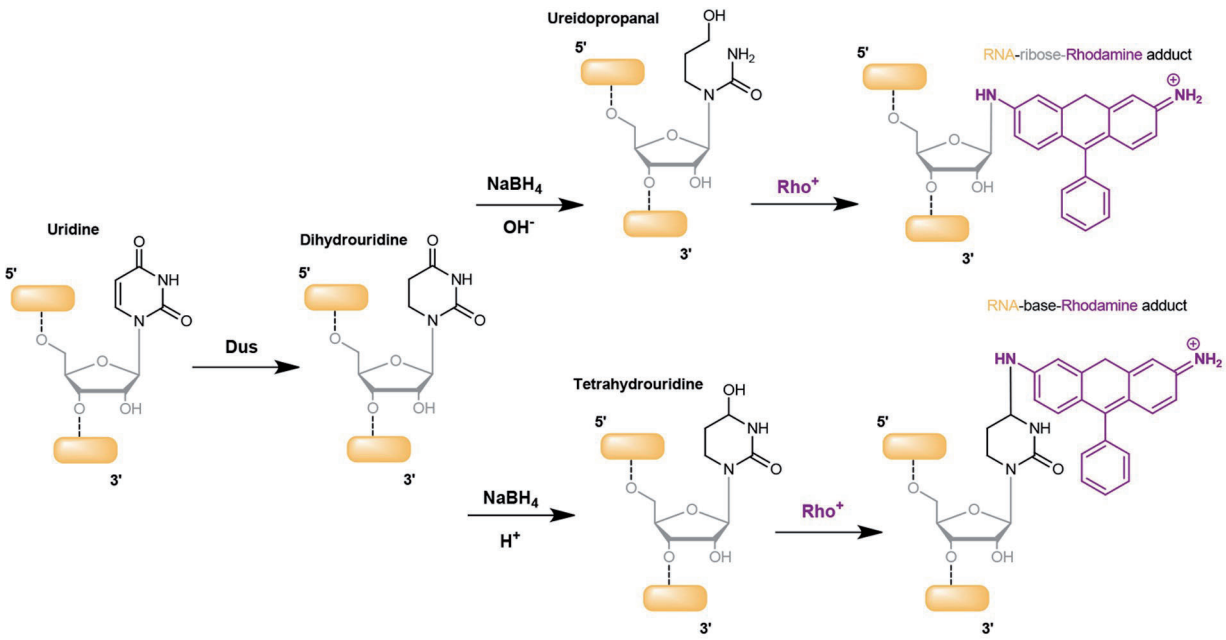


Figure 3. Strategy of chemical labeling of D sites with rhodamine. Reduction of D by $NaBH_4$ in basic conditions leads to the opening of the pyrimidine ring and to a labeling of the ribose by elimination of the ureidopropanal moiety. On the other hand, reduction under acidic conditions produces the tetrahydrouridine which is directly labeled by replacement of its C4-hydroxyl group.

Unlike tRNA, RNA duplexes lack tertiary structures and rather show a decrease in melting temperature when a D unit is present in their central part. Again, this effect can obviously be attributed to the destabilizing effect of a D nucleobase on the C3′-endo sugar conformation and its nonplanar nature that disrupts stacking interactions with neighboring nucleobases. These negative effects of D on duplex formation are likely the driving force for the modified D arm to exclusively adopt the hairpin conformation in solution. Moving in that direction, Dyubankova et al. performed an NMR study to understand the role of dihydrouridine modification on a 15-nucleotides long D-arm of a *Schizosaccharomyces pombe* $tRNA_i^{Met}$ mimetic. While the unmodified oligonucleotide adopts several undefined conformations that interchange rapidly in solution, the presence of D triggers a hairpin folding with a stable stem and a flexible loop region.[37] Although the structural role of D in mRNAs has not yet been investigated, based on the studies we have cited above, it is likely that this base may also serve some regions to adopt a hairpin structure or perhaps to prevent alternative RNA structures in the cell.

Dihydrouridine Labeling and Detection. The study of modifications and their biosynthesis requires the development of tools allowing their accurate detection and quantification. Initially, modified nucleosides including D were identified solely on the basis of their chromatographic mobility involving ^{32}P and/or ^{14}C labeling, 2D electrophoresis combined with thin-layer chromatography or anion exchange chromatography and HPLC analysis.[44] However, these methods suffer from low specificity and reproducibility, and identification becomes problematic as the number of modifications or the length of the RNA chain increases. In contrast, analysis of oligonucleoside fragments by mass spectrometry generated by treatment of the RNA polymer with RNase has proven to be a much better technique for analysis of post-transcriptional modifications, as almost all modifications produce a change in the mass of canonical nucleosides.[10,45−48] In the case of dihydrouridine an m/z + 2 is expected as compared to an unmodified uridine. In that respect, we have successfully applied such a methodology to determine the specificity of not only Dus enzymes from *E. coli* and that of *Mycoplasma capricolum* (see below) but also other RNA-modifying enzymes including m^5U methyltransferases[49,50] and TrmI from *Pyrococcus abyssi*, which catalyzes the sequential double methylation of A57-A58 to m^1A57-m^1A58.[51−53] However, although this technique allows accurate identification of modifications, it does not have the ability to map all D sites in the transcriptome of a given organism on a large scale and in a single shot. Newer detection and quantitative assessment methods that examine RNA modifications on a larger scale, such as in tissues or whole cells, are sequencing methods that target a subset of RNA modifications open to detection by reverse transcription (generally coupled with selective chemical treatment), yet with many limitations.

Beyond its structural properties, the saturation of the C5−C6 bond of D offers interesting consequences on its chemical reactivity, which have been judiciously exploited for RNA labeling and D-site mapping. Earlier works showed that under moderate alkaline hydrolysis, the dihydrouracil undergoes a ring opening via hydrolysis of the N3−C4 bond.[54,55] This reaction produces a β-ureidopropionic acid adduct, which is thought to lose base-pairing ability and generates primer extension arrest at D sites. Xing et al. successfully applied this procedure to map dihydrouridine modification sites of several cytoplasmic tRNAs from yeast with the 5′ end ^{32}P radiolabeled primer extension technique using a primer complementary to the 3′-end region of tRNAs.[56] Alternatively, the heterocyclic ring of D can also be subjected to reductive cleavage by sodium borohydride ($NaBH_4$) under alkaline conditions, yielding in this case a 3-ureidopropanol bound to the ribose C-1′ position[57] (Figure 3). Interestingly, Zachau and his co-workers and others demonstrated that several R-NH_2 compounds (e.g., amines, hydrazines, hydrazides) could, in principle, be used to replace 3-ureidopropanol within

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

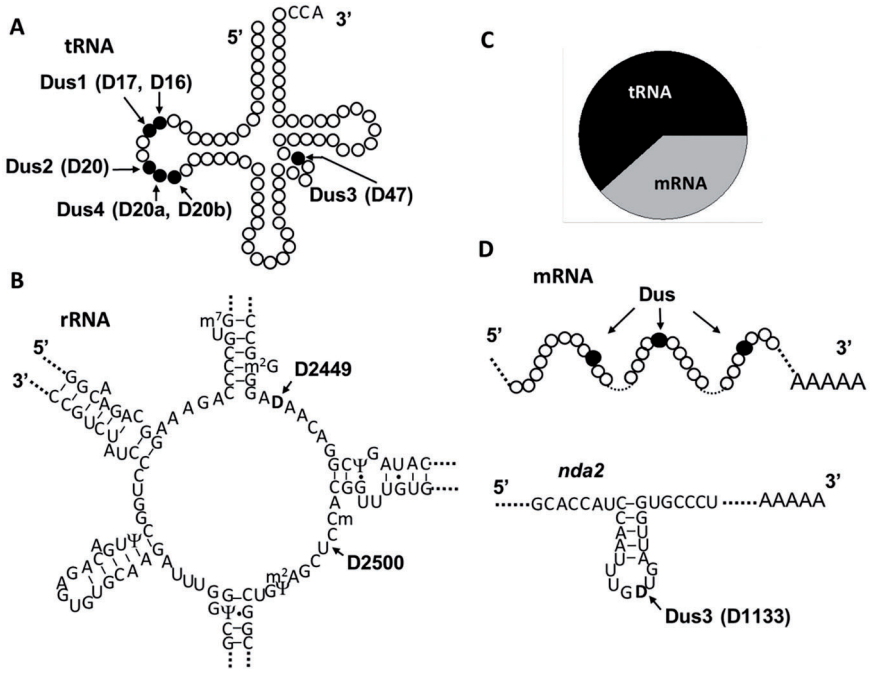


Figure 4. Distribution of D in the transcriptome. (A) Secondary tRNA structure showing the position of D residues observed in *S. cerevisiae* tRNAs. The Dus enzymes responsible for biosynthesis are also shown next to the D residues in parentheses. (B) *E. coli* 23S rRNA sequence showing the position of the unique dihydrouridine (D2449). The 23S subunit of bacterial rRNAs located at the peptidyltransferase center of the ribosome. Uridine 2500 is an unmodified uridine in *E. coli* whereas in *Clostridium sporogenes* this uridine is modified to dihydrouridine, D2500. (C) Pie chart showing the distribution of D in the *S. pombe* transcriptome within tRNAs and mRNAs. (D) Putative Dus-catalyzed dihydrouridylated mRNAs carrying either 3 D-sites as *ala1* and *fhn1* or one as *nda2*. In the case of *nda2*, the sequence part of the mRNA shows the putative stem-loop structure containing the D1133 catalyzed by the enzyme Dus3.

tRNA,[58−60] which opened up an efficient pathway for tRNA labeling as well as applications for sequencing purposes (Figure 3). For instance, in studies of tRNA-ribosome interactions, this approach has been used to label tRNA at the D-site by fluorophores bearing a primary amino group such as proflavin or rhodamine 110.[61] Fluorescent-labeled tRNAs have been used extensively to provide mechanistic insights into *in vitro* protein synthesis and protein synthesis within intact cells.

A conceptually similar approach consisting in producing an abasic site but this time for broad sequencing purposes and detection of certain modifications including D in RNA fragments from tRNAs or rRNAs has been recently developed.[62] This new approach, termed AlkAniline-Seq, can map abasic sites and modified nucleotides in successive treatments that combine: (i) alkaline hydrolysis of RNA at high temperatures, (ii) extensive 5′- and 3′-dephosphorylation, and (iii) aniline-dependent cleavage of the sugar moiety and subsequent formation of RNA abasic sites. This methodology enables simultaneous detection of 7-methylguanosine (m^7G), 3-methylcytidine (m^3C), and D in RNA at single nucleotide resolution. It should be noted, however, that the AlkAniline-Seq signal intensity is considerably lower for D than for m^7G and m^3C, which is most likely due to the fact that reduction of uridine at many potential D-sites is often partial.

In their quest to understand the mechanisms of fluorescent labeling of D in tRNA, Kaur et al. showed that D is converted to tetrahydrouridine in the presence of a large excess of $NaBH_4$ under acidic conditions, and in a second step, its C4 hydroxyl group is replaced by benzohydrazide via nucleophilic substitution, ultimately producing a covalent adduct with the base.[63] Finet et al. have recently developed a Rho-seq integrated pipeline, based on a concept similar to that described by Kaur, but with a variation involving the replacement of the hydrazide nucleophile with rhodamine (Rho), for the purpose of transcription-wide mapping of D.[24] This approach is expected to produce an irreversible RNA-base-Rho adduct primarily at D sites, providing a specific labeling for this modification (Figure 3). The presence of a bulky rhodamine moiety produces a specific and clear stop during reverse transcription at the predicted D-sites in tRNAs validating the specificity of this experimental procedure. Application of this Rho-seq to the *E. coli*, *S. pombe*, and human transcriptome showed (i) the absence of D in *E. coli* mRNAs while the expected D2449 in 23S rRNA as well as D in tRNAs were detected, (ii) the absence of D in *S. pombe* yeast rRNA as expected, (iii) established of D distribution in tRNAs and, (iv) led to the discovery of D in mRNAs (see below).

DISTRIBUTION OF DIHYDROURIDINE IN THE TRANSCRIPTOME

Location and Frequencies of Dihydrouridine Residues in Noncoding RNAs. The study of tRNAs has revealed some interesting paradigm by which D influences the RNA structure/function relationship, although this modification was long neglected compared to other modifications such as pseudouridine for example. D is often found in multiple locations in bacterial and eukaryotic tRNAs, and its abundance varies with both the organism and the type of tRNA.[1] For example, there are up to five positions where D can be found in prokaryotes (Figure 4A), most frequently at positions 16, 17, 20, and 20a, all of which are located in the "D loop", and at position 47 in the variable loop (V loop); however, the latter has so far been observed in only one *Bacillus subtilis* tRNA,

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

$tRNA^{Met}_{CAU}$. In eukaryotes, D is in general more abundant than in prokaryotes, since D is observed in no less than six different sites, five of which are in the D-loop (D16, D17, D20, D20a, and D20b), and one is in the variable loop (D47) (Figure 4A). This persistent presence of D at these positions in so many different organisms reveals the evolutionary importance of the modification and alteration sites. Beyond these "D-canonical positions", D can also locate at other positions such as 14, 17a, and 21 in D loops and at position 48 in the variable arm, but all of these cases are exceptional (see below). An accurate picture of D-distribution in tRNAs can be obtained from *S. cerevisiae* since they have all been sequenced. These can be viewed in MODOMICS (http://genesilico.pl/modomics), a database of RNA modifications that provides comprehensive information concerning the location of modified residues in RNA sequences.[1] Interestingly, all uridine residues in yeast tRNAs at positions 16, 17, 20, 20a and 20b in the D-loop can be converted to dihydrouridine, which may be a consequence of the high solvent accessibility observed for all dihydrouridine-modified positions as evidenced from the crystal structures of three well-known yeast tRNAs, namely $tRNA^{Phe}$, $tRNA^{Asp}$, and $tRNA^{Met}_{i}$[32,33,35] (see Figure 2). To take this analysis a step further, we examined the modification pattern of D sites in all cytosolic tRNA sequences from single cells of fungi and metazoa that are available in RNA databases, i.e., 173 sequences from 22 species. The observation is again consistent with the fact that uridines at positions 16, 17, 20, 20a, 20b, and 47 are predominantly modified to D (Table 1).

Table 1. Dihydrouridine Modification Profile for tRNA Sequences from Cytosol and Mitochondria[a]

positions	modifications	frequency of occurrence
	Cytosolic tRNAs	
16	D/U	123/25
17	D/U	39/0
20	acp^3U/D/U	4/118/11
20a	acp^3U/D/ψ/xU/U	6/64/2/2/2
20b	acp^3U/D/ψ/U	8/6/2
47	D/xU/U	83/1/16
	Mitochondrial tRNAs	
14	D/U	1/3
16	D/U	27/21
17	D/U	8/19
17a	D/U	2/5
20	D/U	49/9
20a	D/U	7/9
47	acp^3U/D/U	3/3/24
48	D/U	1/59

[a]x before U means that the uridine is modified by an unidentified chemical group.

Dihydrouridine is also present in mitochondrial and plastid tRNAs, but is less frequent there, leaving many uridines unmodified (124 mt-tRNA sequences from 18 species, see Table 1). A comprehensive analysis performed on the bovine mt-tRNAs identified 15 types of modified nucleosides distributed over 7.5% of all mt-tRNA bases.[64] Although post-transcriptional modifications in mt-tRNAs are less abundant than in cytosolic tRNAs, their biogenesis requires a large panel of specialized enzymes, some exclusively dedicated to function in mitochondria.[65] Among the D residues, those at positions 16 and 20 seem to be the most widespread modifications in eukaryotic tRNAs, whether cytosolic or mitochondrial. Indeed, a recent transcriptome-wide D mapping using the Rho-seq method was able to confirm this in *S. pombe*.[24] Among the 228 modified positions identified on 141 *S. pombe* tRNAs, 98.7% of the identified D-sites were located within the D-loop, with the most prevalent positions being the expected D16 and D20.

The tRNAs can often harbor doubly modified bases such as m^2_2G26, $cmnm^5s^2U54$, m^1I, or even ms^2i^6A, for example, and Table 1 shows that D is no exception to this rule. Starting with a simple modification, additional modifications can quickly lead to a hypermodified base. In search of new modifications, Krog et al. identified, by MS analysis of *Trypanosoma brucei* $tRNA^{Lys}_{UUU}$ digestion fragments, a new D-derivative, namely 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine (acp^3D) at position 47[66] (Figure 1). Interestingly, a pattern, with dihydrouridylation and/or acp^3 modification of various tRNA species seems to exist; however, how these modifications are introduced remains unclear. The fact that we find uridines bearing only an acp^3 group such as acp^3U47 may suggest that the biosynthetic reaction-giving rise to acp^3D takes place in a particular order where reduction would occur as the last step even if we cannot rule out that the reduction occurs first. In fact, both types of scenarios can be encountered in the biosynthesis of modified bases. For instance, a sequential synthesis of ms^2i^6A37 is observed where the initial grafting of an isopentenyl by MiaA is essential to allow the second enzyme MiaB to introduce a thiomethyl group on the C2 of A37.[67,68] Stepwise modification also extends to the case of m^1I biosynthesis. After the conversion of A37 to I37 by ADAT1, SAM-dependent Trm5, also known for its ability to synthesize m^1G37, methylates directly the inosine N1 nitrogen.[69−72] In archaea, a SAM-dependent TrmI enzyme first methylates A57 before it becomes a substrate for deamination to inosine[73] by a yet unidentified deaminase. The C32 in the anticodon loop of *T. brucei* $tRNA^{Thr}$ is methylated to 3-methylcytosine (m^3C) by Trm140 as a prerequisite for C-to-U deamination by the deaminase ADAT2/3.[74] The introduction of stepwise modifications at positions 34 and 37 in the anticodon loop is frequently observed and is attributed to the fact that the first modification acts as an additional recognition determinant for subsequent modifying enzymes.[75] Cases where the order does not matter are also found, such as in $cmnm^5s^2U54$ biosynthesis, where the C2 thiolation can occur before or after C5−U54 carboxyamino-methylation.[76]

D is not abundant in rRNA, however, unlike tRNAs, the number of sequenced rRNAs remains low. D has so far only been observed at a single location, 2449, in the central loop of domain V in *E. coli* 23 S rRNA,[77] two positions, 2449 and 2500, in 23S rRNA of *C. sporogenes*[78] (Figure 4B), and one position, 1211 or 1212, in 16S rRNA of *Clostridium acetobutylicum*, but the exact location remains unclear.[79] D has not been detected in other bacterial rRNAs nor in any of the eukaryotic rRNAs sequenced to date. Interestingly, in 23S rRNA of *C. sporogenes*, D2449 was found to be methylated at the C5 atom to give m^5D2449.[78] Again, it is not known if there is a particular order in the m^5D biosynthetic pathway. Is m^5U formed first and then reduced to m^5D or is dihydrouridine formed first and then its C5 is methylated? This last scenario is unprecedented so far but deserves further investigation. The enzyme(s) responsible for m^5D biosynthesis remain presently unknown.

Presence of Dihydrouridine in Eukaryotic mRNAs. The first reports on chemical modifications in mRNAs date

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

back to the 1970s, when development of poly(A) tail purification techniques made possible preparation of sufficiently enriched mRNAs, reducing eventual contamination by other abundant RNA species such as tRNA or rRNA. These studies revealed the presence of three in situ modifications, namely m^6A, m^5C, and G_m (Figure 1).[16] In recent years, many other modifications have been added to the mRNA repertoire such as pseudouridine,[80,81] m^1A,[82] and ac^4C[83,84] (Figure 1). Very recently, large transcriptome analysis discovered D in *S. pombe* and human mRNAs; however, no trace of this modification was detected in bacterial mRNAs.[23,24] D is distributed predominantly in the coding regions of conserved genes, supporting a biologically relevant function of dihydrouridylation, which may as well be a general feature of eukaryotic mRNAs. The methodologies and strategies that led to mapping the D sites in the transcriptomes have been discussed in the preceding sections. Of the total D sites detected (372 sites) in the *S. pombe* transcriptome, 38% and 61% are found in mRNA and tRNA, respectively (Figure 4C). Of the 125 D-containing mRNAs in *S. pombe*, 87% have a single putative D, whereas only two mRNAs (encoding nonclassical export protein and alanine-tRNA ligase) carry at least three distinct D sites (Figure 4D). In colon human cells (HCT116), 112 D-sites within mRNAs were also identified. The total D content in mRNAs, however, appears to be relatively low compared with that observed for other modifications. The exact numbers of modified residues vary between studies. The numbers of reproducible peaks seem to converge on approximately 13,000 sites in 5000−7000 mRNAs for m^6A.[8586,87] Regarding pseudouridine, approximately 250−300 sites have been mapped in yeast mRNA at single base resolution;[80,81] however, overlap between these sites is only modest. In humans, the reported number of mRNA pseudouridine sites varies widely from 96 in one study, 23−353 in a second, 22 and up to 2084 in a third study.[80,88] For m^1A, the first study identified 7154 peaks in 4151 coding genes, while the other identified 887 peaks in 600 genes,[89−92] while another study reported only a handful of sites.[93] It is therefore important to keep in mind that the number of identified Ds in mRNAs can vary depending on the methods used for mapping. Dihydrouridylated mRNAs are derived from genes with conserved functions, as it was estimated that 73% of these are conserved in vertebrates.[24] Notably, several mRNAs encoding cytoskeleton-related proteins have been identified as D-containing transcripts. For instance, in *S. pombe nda2* and *nda3* encoding the α- and β-tubulin are found to be a D-containing mRNAs presenting a single D at position 1133 in GDU valine codon and 586 in DCU serine codon, respectively (Figure 4D).

DIHYDROURIDINE SYNTHASE ENZYMES: STRUCTURE−FUNCTION RELATIONSHIP

A large family of enzymes called dihydrouridine synthases (Dus) that uses flavin mononucleotide (FMN) as coenzyme catalyzes the synthesis of dihydrouridine in tRNAs,[94] mRNAs, and bacterial YrlA lncRNA. Nicotinamide adenine dinucleotide phosphate (NADPH) serves as a reducing agent for flavins. This family of flavoenzymes is classified into three major groups and eight subfamilies, all of which evolved by independent duplications of an ancestral gene.[95] The first group found in prokaryotes includes three Dus (Dus A, B and C), while the second is eukaryotic and contains four different enzymes (Dus 1 to 4). Archaea has a single Dus, and it characterizes the last group. The bacterial enzyme, DusB, is considered as the oldest enzyme, while Dus3 could be the ancestral eukaryotic enzyme, from which the other Dus were derived by gene duplication starting with Dus2, then Dus1 and finally Dus4.

Dihydrouridine Site Specificities. An experimental evidence for the enzymatic introduction of D into RNAs comes from the work of Grosjean's lab which systematically tested all enzymatic activities in cell-free yeast extract and showed formation of 11 of the 14 naturally occurring modifications present in mature yeast $tRNA^{Phe}_{GAA}$, including D17, on a radioactive tRNA transcript.[96] It was shown that D17 biosynthesis is insensitive to the presence of the intron since the modification reaction proceeds with the same efficiency in both the intron-free and intron-containing $tRNA^{Phe}$ precursor. Five years later, the first genes encoding yeast Dus were discovered by screening the genomic library of *S. cerevisiae* GST-ORF proteins with pre-$tRNA^{Phe}$ substrate whose dihydrouridylated form carries both D16 and D17 and showed both *in vivo* and *in vitro* that Dus1 catalyzes D17 synthesis in several tRNAs.[97] The complete characterization of the site specificity of the four existing Dus in *S. cerevisiae* could eventually be achieved by the same group using three complementary techniques, namely (i) determination of the molar ratio of dihydrouridine in purified tRNAs from different dus mutants; (ii) microarray analysis of a large number of tRNAs based on differential hybridization of uridine- and dihydrouridine-containing tRNAs to complementary oligonucleotides; and (iii) the primer extension analysis.[56] These *in vivo* studies established that Dus1, Dus2, Dus3, and Dus4 are responsible for the synthesis of D16-D17, D20, D47, and D20a-D20b, respectively (Figure 4A and Table 2). Furthermore, the Dus proteins faithfully retain their specificity in the absence of the other Dus, indicating that they have nonredundant activities. In parallel, Bishop et al. used comparative genomics and computational methods to identify members of the orthologous gene cluster, COG0042, annotated in protein sequence databases as 'predicted TIM-barrel enzymes, possibly dehydrogenases, nifR3 family' as genes encoding dihydrouridine synthase and identified three

Table 2. Experimentally Established Specificities of Dihydrouridine Synthases

enzymes	organisms	substrate(s)	products in tRNA	refs
DusA	*E. coli*	tRNA	D20, D20a	52, 98
	T. thermophilus	tRNA	D20, D20a	101
DusB	*E. coli*	tRNA	D17	52
	M. capricolum	tRNA	D17, D20, D20a	53
DusC	*E. coli*	tRNA	D16	52, 99
	S. cerevisiae	tRNA	D16, D17	56, 97
Dus1	*S. pombe*	tRNA, mRNA	D16, D17	24
	H. sapiens	tRNA, mRNA	D16	24
	S. cerevisiae	tRNA	D20	56, 97
Dus2	*S. pombe*	tRNA, mRNA	D20	24
	H. sapiens	tRNA	D20	24, 103
	S. cerevisiae	tRNA	D47	56
Dus3	*S. pombe*	tRNA, mRNA	D47	24
	H. sapiens	tRNA, mRNA	D47	23, 24
	S. cerevisiae	tRNA	D20a, D20b	56
Dus4	*S. pombe*	tRNA, mRNA	D20a, D20b	24
	H. sapiens	tRNA	D20a	24

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

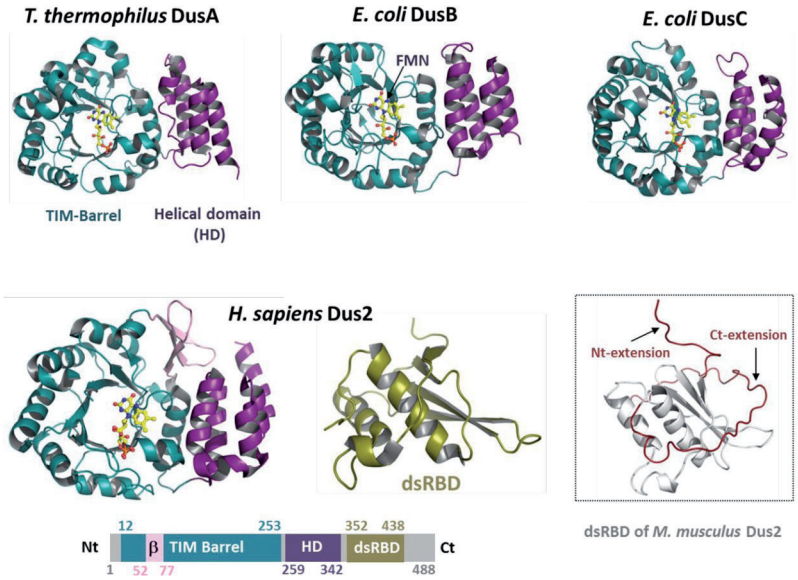


Figure 5. Crystallographic structures of bacterial and human Dus. The TIM-barrel, HD, and dsRBD domains are colored in deep-teal, purple, and olive, respectively. The prosthetic group, FMN, is in yellow stick. In pink is the inserted domain found in hDus2. Inset is the putative mouse Dus2 dsRBD whose structure was resolved by NMR and annotated in the PDB as dsRBD from hypothetical protein BAB26260. This solution structure shows extensions in N-terminal (Nt) and C-terminal (Ct) regions. The pdb codes used are 3B0P, 6EI9, 4BFA, 4WFS, 4WFT, and 1WFN for *T. thermophilus* DusA, *E. coli* DusB, *E. coli* DusC, the canonical domains of hDus2, the dsRBD of hDus2, and the dsRBD of BAB26260 protein, respectively. Below the crystallographic structure of hDus2 is shown the diagram of the modular organization of this enzyme as well as the delimitation of the respective domains.

members of the COG0042 family in *E. coli*, named DusA, DusB, and DusC, which are responsible for introducing all D content into tRNAs and that DusA is likely the D20-catalyzing enzyme in $tRNA^{fMet}_2$.[98] Recently, we and others have contributed to unraveling the specificity of Dus enzymes from several model organisms, including those in humans. In *E. coli*, DusA, DusB and DusC catalyze the formation of D20-D20a, D17, and D16, respectively.[52,98,99] *Thermus thermophilus* harbors a single Dus, which is a homologue of DusA and has been shown to catalyze the formation of D20-D20a.[100,101] DusA is also responsible for the biosynthesis of D76 in the lncRNA YrlA of *Salmonella typhimurium*.[102] Indeed, D is observed at position 76 in YrlA from *S. typhimurium*, which is a long noncoding RNA that attaches the Ro60 protein to the polynucleotide phosphorylase, thus targeting this exoribonuclease for degradation of structured RNAs. In *M. capricolum*, the single DusB introduces all D content present in tRNAs, namely D17, D20, and D20a.[53] In *S. pombe* and *Homo sapiens*, the Dus enzymes share the same site specificities for tRNAs as those of *S. cerevisiae*.[23,24,103] However, it was shown that all four Dus enzymes are involved in modifying both tRNA and mRNA, and in the case of Dus3, mRNAs seem to be its predominant target.[23,24] It is plausible that Dus-catalyzed mRNA dihydrouridylation is a conserved functional feature in eukaryotes; however, no D residues were reported in *S. cerevisiae* mRNA.[84] The involvement of Dus in the synthesis of D in both tRNA and mRNA is not unique, as it has now been reported that most enzyme systems catalyzing mRNA marks are those involved in the modification of other RNAs, mainly tRNAs. The only known exception to date regards the biosynthesis of m^6A in mRNA that is catalyzed by the dedicated METTL3/METTL4 complex.[9] For example, NSUN2, the human SAM-dependent tRNA methyltransferase involved in the formation of m^5C at positions 34, 48, 49, and 50 of tRNAs[104−106] also mediates m^5C synthesis in nearly 300 mRNAs.[107] For comparison, the orthologous yeast enzyme Trm4 catalyzes m^5C formation at positions 34, 40, 48, or 49 depending on the tRNA.[108] The *S. cerevisiae* several pseudouridine synthases demonstrate multisite substrate specificity, Pus1 modifies tRNAs at multiple locations[109] as well as U2 snRNA,[110] Pus4 forms a universally conserved ψ55 in tRNAs,[111] but was also reported to modify mRNAs,[112,113] while Pus7 modifies U13 in several tRNAs, U35 in pre-$tRNA^{Tyr}$, U35 in the small nuclear RNA U2, U50 in 5S rRNA,[114,115] and several U residues in mRNA.[116] In human cells, the majority of cytoplasmic tRNAs carry the m^1A58 modification catalyzed by the heterotetrameric TRMT61A/TRMT6 enzyme complex, which is also responsible for the m^1A modification in mRNAs.[93]

Today, the enzymes responsible of D incorporation in rRNAs remain to be identified since in the case of D2449 present at the peptidyl transferase site in *E. coli* 23S, the deletion of the three bacterial *Dus* genes does not abolish its formation,[24] suggesting that another class of dihydrouridine synthase specific to rRNAs must exist in this bacteria. Finally, the enzymes that introduce D into mitochondrial tRNAs also remain to be identified. However, it might be possible that the cytosolic Dus are also responsible for the biosynthesis of the corresponding D in mitochondria. This dual specificity is common for other tRNA modification enzymes such as pseudouridines synthases[111,117] or m^5U54 tRNA methyltransferases.[118] The mammalian Dus2 was indeed detected in mitochondria in a study aiming to create a mitochondrial compendium of 1098 genes and their protein expression across 14 mouse tissues.[119]

Structural Analysis of Dihydrouridine Synthases. The X-ray crystallographic structures of some Dus have defined a canonical fold for this family of enzymes (Figure 5). To date, the structures of the three bacterial Dus homologues, namely DusA from *T. thermophilus*, DusB and DusC from *E. coli* were

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

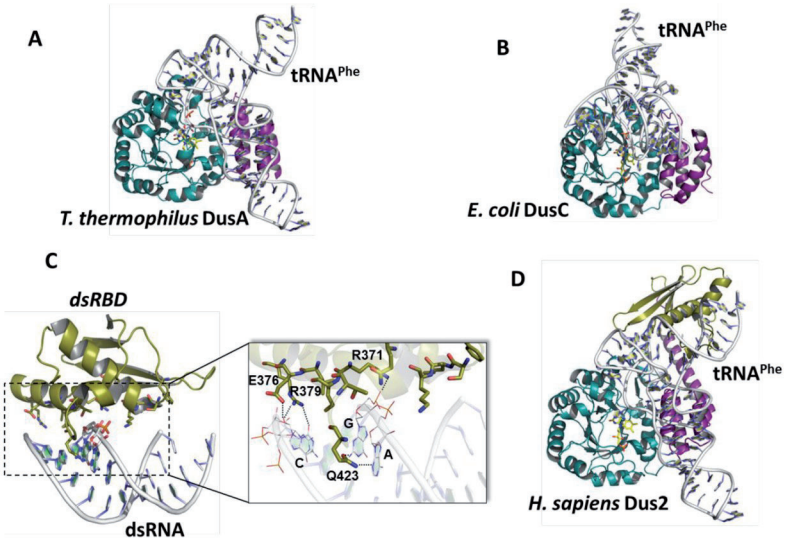


Figure 6. Molecular basis of tRNA substrate recognition by *T. thermophilus* DusA, *E. coli* DusC, and human Dus2. The protein/RNA complexes, namely *T. thermophilus* DusA:tRNAphe (A), *E. coli* DusC:tRNAPhe (B), and hDus2 dsRBD:dsRNA (C) are crystallographic structures whose PDB codes are 3B0V, 4YCO, and 5OC6, respectively. (D) Molecular model of the hDus2:tRNA complex generated as described in Bou-Nader et al.[122]

solved.[52,99,100,120] However, the data are scarce for eukaryotic enzymes since only the structures of the isolated domains of human Dus2 (hDus2) recently solved by our group and that of Antson are available (Figure 5), seriously impeding our understanding on this subfamily of enzymes.[103,121] All Dus enzymes share a common catalytic N-terminal domain, which adopts a TIM-barrel type fold. The latter carries an active site circumscribed within a solvent-accessible crevice providing a binding site for the FMN prosthetic group and uridine substrate. The catalytic domain is followed by a domain formed by four parallel helices organized in a bundle and which constitutes the helical domain (HD). Beyond these domains, the canonical fold can accommodate additional domains that are added either on the N-terminal end or on the C-terminal end or even both, but this concerns essentially eukaryotic Dus.[95,98,103] Obviously, this complex modularity could serve new biological purposes that do not exist in prokaryotic organisms, such as RNA substrate transport and localization, or specialization in the recognition of various RNA substrates, as is the case in yeast and human Dus involved in tRNA and mRNA modification.

The structures of *T. thermophilus* DusA and *E. coli* DusC in complex with tRNAs have elucidated the molecular basis of tRNA substrate recognition by bacterial enzymes (Figure 6A,B).[99,100] Here, the two canonical domains provide the platform for RNA recognition and involve numerous ionic interactions, including positively charged residues (i.e., Lys, Arg) that interact specifically with bases, ribose and phosphate constituting the substrate backbone. The recognition mode actually differs depending on whether the enzyme is specific for U16 or U20. While the structure of both proteins is globally conserved, these enzymes access the target uridine by positioning their substrate tRNAs differently on their respective surfaces involving a 160° rotation from tRNA on DusC to that of DusA (Figure 6A,B). DusA recognizes a more important portion of the tRNA than DusC does, in particular the anticodon stem that DusC does not bind. Both enzymes recognize the elbow region, the D-stem-loop and the T-loop. In the case of DusA, the D and T-loops are recognized only by the TIM-barrel domain, while the D and anticodon stems are recognized by the HD. In contrast, in the case of DusC, the D-loop is recognized by both canonical domains, while the D-stem is recognized only by the TIM-barrel domain and the HD recognizes the T-loop. In these complexes, the Dus enzymes appear to bind to tRNAs without disrupting the crucial interactions that maintain their tertiary structure. Thus, the tRNA elbow must be a quality control checkpoint that Dus scrutinizes before dihydrouridylation. Finally, the two enzymes flip their uridine substrate and stack it on the isoalloxazine to proceed with its reduction (see the mechanism of Dus in the next section).

We have recently shown that the recognition mechanism of the tRNA substrate by *Homo sapiens* Dus2 (hDus2 or Dus2L) is much more complex than that observed in bacterial enzymes (Figure 6).[122] Indeed, hDus2 has a structural insertion within the TIM-barrel and an additional double-stranded RNA binding domain (dsRBD) that is appended to the polypeptide just after the HD, both playing a role in tRNA recognition (Figures 5 and 6).[103,122,123] The dsRBD is a double-stranded RNA (dsRNA) recognition module and is mainly found in proteins involved in mRNA transport, processing or editing.[124−127] Our structures of the hDus2 dsRBD in complex with a dsRNA (Figure 6C), as well as in-depth investigations by site-directed mutagenesis, nuclear magnetic resonance (NMR), and small-angle X-ray scattering (SAXS) of the interaction of the dsRBD with human $tRNA^{Lys}_3$ revealed how this domain binds to tRNA. Indeed, this domain has a particular mode of tRNA recognition involving, in addition to the well-known canonical interactions between dsRBDs and dsRNAs (ribose and phosphate recognition), specific interactions with RNA, i.e., hydrogen bonds between some residues of dsRBD and RNA bases. The model of the hDus2/tRNA complex showed that in addition to engaging the TIM-barrel and to a lesser extent the HD, the dsRBD provides an important substrate recognition platform by binding to the long double-stranded region formed by the junction between the acceptor and T stems of tRNA (Figure 6D). Interestingly, we can infer that mammalian Dus2 encompasses almost the entire tRNA, where only the anticodon loop seems to be left out. Although we have no experimental data at this stage, we

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

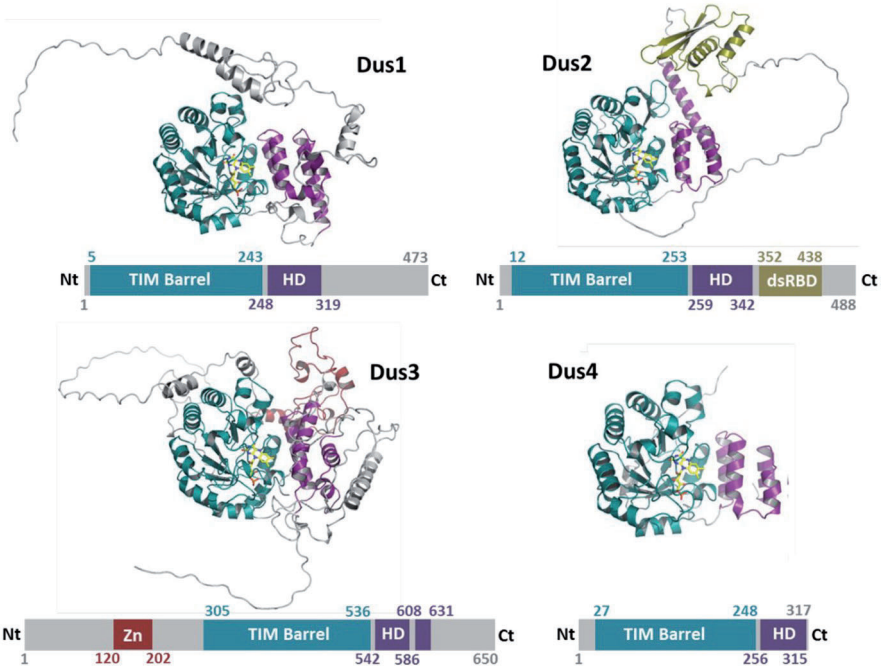


Figure 7. 3D models of human Dus generated by AlphaFold. The architectural organization and the delimitation of the domains are indicated under each corresponding Dus model.

believe that the dsRBD may also be important for mRNA with dsRNAs structures recognition but this remains to be tested. For instance, the dsRBDs of adenosine deaminase ADAR2, which acts on mRNA to recode genomic information by the site-selective deamination of adenosine, binds to a stem-loop pre-mRNA encoding the R/G editing site of GluR-2 by recognizing the shape and sequence of the dsRNA.[128]

3D Models of Human Dihydrouridine Synthases. Recently, a revolution in the world of structural biology has occurred with the arrival of artificial intelligence in the high precision prediction of 3D protein models generated by AlphaFold (https://alphafold.ebi.ac.uk/),[129] opening a new area in protein modeling.[129−131] This accuracy applies not only to the prediction of folding but also to the positioning of residues within active sites for example. By applying AlphaFold2, Tunyasuvunakool et al. markedly expand the structural coverage of the proteome at a scale that covers almost the entire human proteome (98.5% of human proteins).[132] They also provide some case studies to illustrate how high-quality predictions could be used to generate biological hypotheses. We have seized this unique opportunity to obtain accurate 3D models of the four human Dus as shown in Figure 7. The resulting models predict, as expected, the conservation of the canonical domains in all four human Dus with, however, some peculiarities that can be noted. For example, the β-sheet inserted into the TIM-barrel of hDus2 is absent in the other three enzymes, making it a unique feature of this enzyme subfamily. While the TIM-barrel structure is largely maintained in all four Dus, Dus3 has an HD that is different from the others (see below for details). With the exception of Dus4, which has no additional domain, all the others have additional structural elements added to their respective canonical structures. Dus1 has an extension of 154 residues on the C-ter side, part of which forms three independent helices (368−380, 412−423, and 432−448), while the remaining sequences are organized into loops and unstructured regions. Regarding Dus2, the model is in perfect agreement with our structures, wherein as indicated above, the HD is followed by the dsRBD. The protein ends with a predicted unfolded 50-amino acid extension, probably an intrinsically disordered region as we recently showed whose truncation does not abrogate tRNA dihydrouridylation.[103] It should be noted that a structural genomics group has published in the PDB the solution structure of the isolated dsRBD domain of mouse Dus2 under the code 1WHN and annotated "Solution structure of the dsRBD from hypothetical protein BAB26260". In this NMR models, the long C-terminal extension is disordered likely due to the lack of restraints, a consequence of the intrinsic flexibility of this region (Figure 5). Dus3 is the enzyme that is clearly distinguished from the three others by the presence of several features that we list as follows: (i) It is the Dus orthologue with the largest size, with a polypeptide of 650 amino acids, i.e., twice as large as Dus4 for example; and (ii) the enzyme has a much more complex modularity with the addition of several structural elements that are appended to the N-terminal side of the sequence, thereby extending several regions of the canonical domains in the 3D space. Specifically, residues 36−60 form a helix that flanks the HD, extending that domain to the right, while residues 211−236 adopt two helices that cover the TIM-barrel. An additional zinc finger domain (ZnD) is inserted between these structural elements and is positioned above the HD. A zinc finger is a small protein structural motif that is characterized by the coordination of one or more zinc ions (Zn^{2+}) to stabilize the fold. However, despite the wide variety of these motifs, the vast majority typically functions as interaction modules that bind DNA or RNA and structural variations serve primarily to alter the binding specificity of a particular protein. In the case of hDus3, the ZnD carries a C161-X8-C171-X5-C178-X5-H183 motif, with perfect preorganization of the Zn^{2+} binding site; as if the cysteines and histidine ligands were preoriented to readily coordinate the metal. (iii) Finally, the four helices of

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

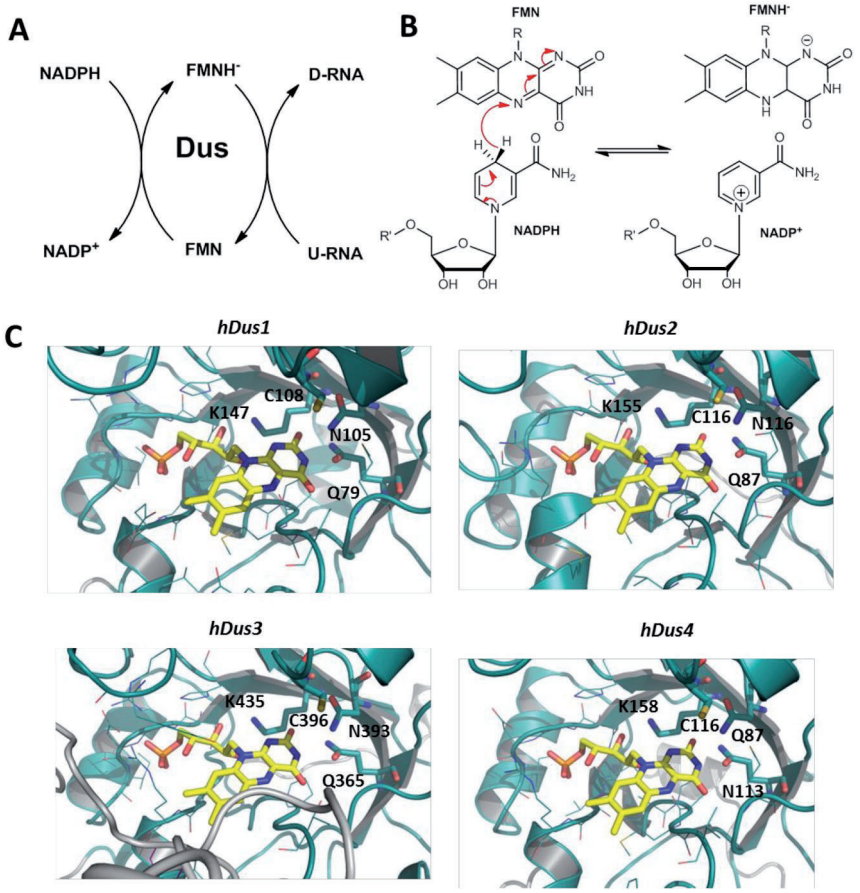


Figure 8. Chemical mechanism of D synthesis catalyzed by Dus. (A) General reaction scheme of the Dus catalytic cycle. (B) Proposed stereochemical mechanism of flavin reduction of Dus by NADPH. (C) Sections of the active sites of human Dus obtained from AlphaFold protein models. The FMN coenzyme is represented in yellow ball−sticks.

the HD no longer form the bundle that characterizes this domain, while it is interspersed with an insertion of about 20 amino acids that form a large loop.

All these structural features of eukaryotic Dus raise an important question: What exactly is the utility of having these extensions knowing that, for instance, Dus4 can obviously function only with the canonical architecture and counts tRNAs and mRNAs as well among its known substrates? These structural elements may play a functional role, such as cooperative participation with the canonical domains in substrate recognition, as we recently demonstrated for the dsRBD of hDus2.[103,122] In the case of the zinc finger of Dus3, the region of this domain that faces the HD bears a positive electrostatic surface, suggesting its likely involvement in RNA binding. Thus, all these models offer interesting perspectives to evaluate in the future the role of these extensions and domains in the dihydrouridylation reactions of RNA substrates.

MOLECULAR MECHANISM OF DIHYDROURIDINE BIOSYNTHESIS AND INHIBITION

Chemical Mechanism of the Dus-Catalyzed Dihydrouridylation. The chemical mechanism of Dus has been primarily elucidated through the study of yeast Dus2,[133] subsequently corroborated by crystallographic structures of several Dus active sites, including those obtained in complex with tRNA. Overall, the catalytic cycle of these enzymes is composed, like for most flavoenzymes, of a first reductive step that involves the reduction of FMN by the natural reductant followed by a second step that involves the oxidation of the reduced flavin by the second substrate (Figure 8A). The fast kinetic stopped-flow approach by Rider et al. showed that NADPH rapidly reduces FMN (2.5 s^{-1} at 4 °C) to give rise to a flavin hydroquinone.[133] We also showed, by monitoring the oxidase activity of hDus2, DusB from *E. coli* and *M. capricolum*, that these enzymes prefer NADPH to NADH. Although the latter may be a potential reductant, kinetic evidence argues against its utilization while strengthening general use of NADPH as the physiological substrate by all Dus enzymes.[52,53,103] Only a structure in complex with NADPH will eventually identify potential residues involved in the physiological discrimination of the flavin reducing agent. The reduction of FMN to hydroquinone by NADPH occurs via a hydride-transfer reaction, and it was shown that Dus2 is specific for the proR hydrogen of NADPH[133] (Figure 8B). We propose that the form of hydroquinone produced in this reaction is $FMNH^-$ and not $FMNH_2$ as previously proposed.[100,133] This is supported by the fact that all the Dus structures and models strictly conserve a lysine residue (K147, K155, K435, K158, K132, K139 in hDus1, hDus2, hDus3, hDus4, *T. thermophilus* DusA, *E. coli* DusC, respectively) in their active sites that faces the N1-isoalloxazine and that is perfectly poised to stabilize the eventual negative

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

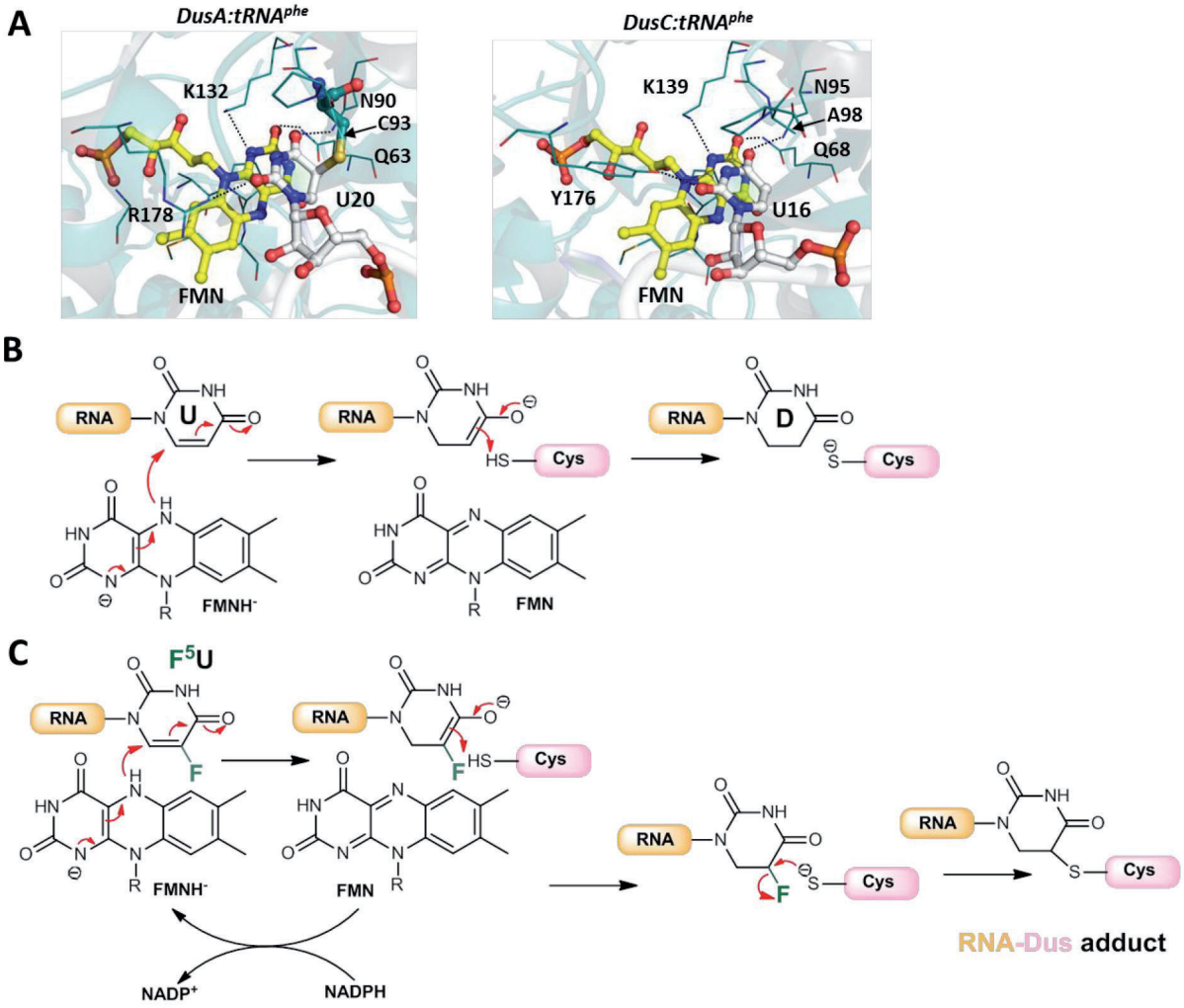


Figure 9. Proposed chemical mechanism of D biosynthesis and activity-based inhibition of Dus. (A) Focus on the active site of *T. thermophilus* DusA and *E. coli* DusC in complex with their respective tRNA substrates. The FMN is represented in yellow ball−sticks, while the uridines are in white ball−sticks. (B) Postulated chemical mechanism of Dus enzymes. (C) Postulated chemical mechanism for the activity-based inhibition of Dus via formation of a covalent Dus/RNA complex in the presence of fluorinated uridine.

charge built on this flavin nitrogen via hydrogen bonding (Figures 8C and 9A). Confirmation of its importance in the biosynthesis of D was evidenced by its replacement by an alanine in *E. coli* DusA causing a collapse of D-level in a triple-Dus *E. coli* strain complemented with *dusA-K153A* mutant compared to the same strain complemented with a wild-type *dusA*.[134] Another strictly conserved active site residue, namely a glutamine (Q79, Q87, Q365, Q87, Q63, Q68 in hDus1, hDus2, hDus3, hDus4, *T. thermophilus* DusA, *E. coli* DusC, respectively), engages two H-bond with C2═O and N3H of FMN, to assist in the stabilization of $FMNH^-$ (Figures 8C and 9A). Thus, this structural information implies that $FMNH^-$ is likely to be the RNA reducing entity. The structures of bacterial Dus in complex with tRNA and that of hDus2 also reveal that there is not enough space to accommodate both NADPH and the target uridine at the same time. Hence, after the reduction of FMN by NADPH, the first product of the reaction, namely $NADP^+$, will have to leave the active site to allow the accommodation of the second substrate, which is consistent with a ping-pong-type enzymatic mechanism. In a second step, $FMNH^-$ transfers its hydride to the pyrimidine substrate uridine at C6 followed by a subsequent protonation step at C5 breaking eventually the unsaturation character of the base (Figure 9B).[133] A conserved cysteine in Dus acts as a general acid in the protonation step (C108, C116, C396, C116, C93, and C98 in hDus1, hDus2, hDus3, hDus4, *T. thermophilus* DusA, *E. coli* DusC, respectively) (Figures 8C and 9A,B). This redox mechanism is validated by the structures of *T. thermophilus* DusA:tRNAphe and *E. coli* DusC:tRNAphe complexes,[99,100] where we clearly distinguish the substrate uridines stacked against the *si*-face of the isoalloxazine with their C5 pointing within 3.5 Å of the hydride donor, i.e., the N5-FMN (Figure 9A). This productive orientation of uracil is firmly maintained by two hydrogen bonds between, on the one hand the C2═O of the pyrimidine with the side chain of a polar amino acid (R178 and Y176 for DusA and DusC, respectively), and on the other hand between the C4═O of uracil and the side chain of an asparagine strictly conserved in Dus (Figure 9A). It is worth mentioning that similar mechanisms are also observed in dihydropyrimidine and dihydroorotate dehydrogenases, both of which share a TIM-barrel catalytic domain homologous to that of Dus enzymes.[98,135,136]

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

Dihydrouridylation and Modifications Interdependence. Remarkably, the study by Rider et al. showed that the rate of U20 reduction by Dus2 is dependent on the presence of other modifications.[133] Indeed, U20 of yeast $tRNA^{Leu}$ is reduced 600 times faster when the substrate is matured, i.e., it bears all its modifications. Similarly, we have shown that the dihydrouridylation activity of hDus2 is more efficient on modified bulk tRNAs than on naked transcripts.[103] This functional behavior does not seem to be a peculiarity of the fungal and mammalian enzymes because *T. thermophilus* DusA also behaves in the same way. These studies raise the possibility of modification interdependencies that may translate into a precise order in the biosynthesis of certain modifications and D20 could be among those that appear late in the maturation pathway. In this context, Barraud et al. have ingeniously developed a recent methodology allowing them to follow the maturation of labeled $tRNA^{Phe}$ in yeast cell extracts using time-resolved NMR.[137] Remarkably, they observed that modifications are introduced in a predefined sequential order and that this timing seems to be controlled by interdependencies between modification elements. Among these events, the introduction of D by Dus1 occurs after the biosynthesis of Ψ55 and m^5U54 in the T-arm. While D has a beneficial effect on the biosynthesis of m^1A58 in the T-arm, it is shown to have a negative effect on the introduction of m^2G10 in the D-arm, revealing a complex connected circuit that could be specific to each tRNA or even each isoacceptor. It is important to mention that such hierarchical circuits not only concern tRNA core modifications but also are quite frequent for modifications affecting the anticodon loop region. In these cases, the modification circuits can enhance the specificity of the modification enzymes by using the first modification in the circuit as an additional recognition identity factor for the following modification.[75] Nevertheless, beyond this concept, the structures of Dus/tRNA complexes may indeed provide convincing rationale for these phenomena at least for D biosynthesis. As previously noted, dihydrouridylation appears to require the architectural integrity of the tRNA elbow, so it is quite logical to expect that any modification that stabilizes tRNA tertiary interactions would have a direct positive impact on the efficiency of D biosynthesis. Interestingly, Cavaillé et al. observed that D20 biosynthesis in tRNA incubated in yeast cell-free extracts is completely abolished in tRNA monomutants affecting substrate architecture (G18C, G19C, C56G).[138] Consistent with this analysis, D is introduced after isomerization of U55 and methylation of U54, both of which are known to stabilize the elbow structure. It will therefore be interesting to validate this hypothesis by determining the dihydrouridylation activity on tRNAs lacking these two modifications.

Activity-Based Inhibition of Dihydrouridine Synthases. Dai et *al.* recently developed a chemo-proteomic strategy based on an RNA-mediated protein profiling approach to map the *in vivo* interactions existing in human cells between C5-pyrimidine-RNA methyltransferases and mRNAs by directly feeding the living cells with 5-fluorocytidine (5FC) or 5-fluorouridine (5FU), both analogues of C or U, respectively.[23] The principle is to use these fluorinated nucleotides as baits, thanks to their potential incorporation into RNA especially at the sites targeted by these enzymes, and to generate dead-end enzyme/substrate covalent complexes formed during catalysis. This strategy has been widely used *in vitro* to trap covalent complexes between C5-pyrimidine-RNA methyltransferases and miniRNAs containing fluorinated substrate analogues with a fluorine atom at the C5 of the target pyrimidines during reaction with the natural carbon donor (*S*-adenosylmethionine for SAM-dependent m^5C or m^5U methyltransferases and 5-methylenetetrahydrofolate for folate-dependent m^5U methyltransferases).[139−142] The mechanistic principle was established a long time ago and consists of an activity-based inhibition. Like all C5-pyrimidine methyltransferases, the catalytic mechanism requires, as a first step, preactivation of the base by addition of a cysteine present in the active site to the C6 carbon via a nucleophilic attack (Michael addition type of reaction). This leads to a transient enolate with an activated C5. The latter then attacks the methyl donor, allowing the transfer of the methyl group onto the RNA. Finally, labile H5 proton abstraction by a base residue triggers β-elimination and dissociation of the RNA from the protein. However, when F, Br, or I replaces the H5, as the general base cannot abstract the halogen atom, the covalent RNA-enzyme species is the final stable product. This method has unexpectedly led to the capture of an mRNA-hDus3 covalent complex.[23] This cross-linking was not observed when the potential general acid in the dihydrouridylation mechanism, namely C396, was replaced by an alanine suggesting that in the presence of a 5-halodihydrouridine; this residue likely becomes a reactive nucleophile. An activity-based mechanism for this Dus-RNA trapping can be formulated as proposed in Figure 9C, which is in theory a strategy applicable to all Dus: (i) reduction of FMN to $FMNH^-$ by NADPH, (ii) hydride transfer from FMNH- to C6 of 5FU, (iii) protonation of C5 by the conserved cysteine and formation of 5-fluorodihydrouridine, and (iv) formation of the covalent bond between cysteine and dihydrouridine by nucleophilic substitution of fluorine leading to an RNA-Dus adduct.

DIHYDROURIDINE AND BIOLOGICAL IMPLICATIONS

Biological D-Relevance in RNA Folding and Architectural Stability. The relevance of D in the conformational dynamics of RNAs was established early on by structural approaches (see previous sections). On the other hand, this function took on its full biological meaning when MacCloskey analyzed the quantitative composition of post-transcriptional modifications in tRNAs isolated from psychrophilic organisms having the capacity to grow under extremely low temperatures ranging from −5 to 12 °C.[26,27] These organisms have implemented evolutionary strategies to counteract the restriction of molecular mobility and to maintain a form of resilience in the face of low temperature by incorporating into their biomolecule, biochemical components that have the capacity to maintain molecular flexibility. Among these compounds, the observed high incorporation of unsaturated fatty acids in the lipid bilayers tends to fluidity the cell membranes while the limitation of hydrophobic clusters or salt bridges, known for their stabilizing effect on proteins. Similarly, these organisms show much less post-transcriptional modifications in their tRNAs, although they retain some of them such as pseudouridine, m^5U, and m^7G at normal levels. In dramatic contrast, D levels are found to be between 40 and 70% higher than those found in mesophilic organisms, such as *E. coli*. Thus, these biological data together with structural information corroborate the role of D in promoting the local fluctuation and mobility of nucleic acids.

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

Depletion of D by deletion of *dus* genes does not cause significant defects in growth phenotype of model organisms such as eubacteria *T. thermophilus*,[101] *E. coli*,[98] or *S. cerevisiae*.[56] Similarly, in the absence of identified enzyme-catalyzing rRNA dihydrouridine synthesis, O'Connor et al. removed D2449 and evaluated its effect on *E. coli* physiology and ribosome function by direct mutation of U2449 to C2449. The results obtained in this study pointed out that ribosomal D2449 is dispensable to the cell.[143] These results are not surprising by themselves because many so-called "nonessential" modifications, which are located mainly in the body of the RNA molecule, produce only minor phenotypic impact following their removal. On the contrary, this phenomenon makes sense given the fact that these modifications are part of an interconnected network where compensation phenomena, or functional redundancy, may occur. The biological relay of these modifications becomes relevant when this network is disturbed under particular stressing events or beyond the simple loss of a single modification. It is in this context that Phizicky uncovers the importance of D in combination with m^7G46 in yeast $tRNA^{Val}_{AAC}$.[144] Indeed, the double mutants *dus/trm8* and particularly *dus3/trm8* produce severe growth defects. Molecular analysis has revealed that this growth defect coincides with a rapid decrease at the steady-state level of the pool of this tRNA via its rapid intracellular degradation, which approaches the degradation rates of mRNAs. Hence, by maintaining the functional folding of RNAs in cooperation with its relatives, D acts as a kind of quality control mark for RNAs.

Dihydrouridine and Protein Translation. Finet et al. recently showed that D has a critical role in the control of cell cytoskeleton dynamics in *S. pombe* via its presence in a codon of the *nda2* and *nda3* mRNAs encoding α- and β-tubulin, respectively.[24] The absence of D in these mRNAs, obtained either by deletion of Dus3 or by replacement of D by C using mutagenesis, leads to the same result, namely a cell growth defect observed in the presence of the depolymerized tubulin drug 2-(4-thiazolyl)benzimidazole. This growth defect appears to be attributed to a meiosis problem caused by an excess of tubulin. Here, the function of D is to slow down the translation rate of the α- and β-tubulin genes to allow a controlled accumulation of the α/β-tubulin pool compatible with functional concentrations, preventing any imbalance of these components, which is known to be detrimental to cell's life. Conversely, the absence of D provokes overexpression of these cytoskeleton proteins, which ultimately leads to altered chromosome segregation and reduced gamete viability. D in the human transcriptome seems to have an opposite behavior on translation since it has been shown that its absence obtained via hDus3 deletion impairs translational efficiency, leading to a strong decrease in cell viability.[23] We speculate that D could therefore be considered as a key mark in the control of the translational homeostasis of certain genes.

Implication of Dihydrouridine and Dus Enzymes in Cancers. Because of their importance in translation, defects in post-transcriptional RNA modifications and in enzymes that catalyze them are often associated with severe human diseases.[145−151] Situations where an over-representation of certain modifications can also occur in some cancers; however, the molecular mechanisms that link aberrant RNA modifications to human diseases are largely unknown. The case of overexpression of modifying enzymes in some cancers can also be observed.[145] Both types of situations are observed in some cancers for the case of dihydrouridine, i.e., (i) increased D content, (ii) overexpression of Dus (see below), and (iii) Dus activity (specific or not). However, the link between (i), (ii), and (iii) has never been established, and therefore, it cannot be concluded whether there is a correlation between overexpression of the enzyme and increased tRNA dihydrouridylation activity.

Increase of D Levels. Kuchino and Borek reported in the late 1970s excessively abnormal levels of D, together with m^5C, in tumor-specific phenylalanine tRNA isolated from Novikoff hepatoma and Ehrlich ascites cancerous tissues.[152] These pathogenic tissues showed an increase in D content of up to 50% compared to healthy tissues. Most of the subsequent work focused only on the observed increased methylation, yet the increase in D levels raises interesting questions that are yet to be addressed. Of note, human $tRNA^{Phe}_{GmAA}$ has three D residues at positions 16, 17, and 47 on its sequence.[1] Given this information, one can ask whether U16, U17, and U47 are fully modified in tRNAs from healthy cells, and if not, this could possibly suggest that the increase in D content observed in tumorigenic tissue is the result of complete dihydrouridylation of vacant D-sites, as it was known that D modification is rarely stoichiometric.[153]

Dus Overexpression. By using screening for up-regulated genes in cancer cells based on genetic information obtained on cDNA microarrays, combined with high-throughput screening of their effects on cell growth, Kato et al. found that hDus2 is frequently overexpressed in clinical lung cancer samples and nonsmall lung cancer cell lines and that the overexpressed hDus2 is necessary for survival/growth of lung cancer cells.[154] The contribution of hDus2 to lung carcinogenesis was revealed using siRNA to suppress hDus2 expression, showing a reduced dihydrouridylation of total RNA and a growth suppression of these pathogenic cells. However, it is important to note that this observation does not mean that Dus2 activity is *per se* responsible for tumor activity since it is expected that the decrease in hDus2 biosynthesis by siRNA would decrease its physiological activity, i.e., D20 biosynthesis (see below). Specific interaction of hDus2 with cellular partner has also been documented. Using immunoprecipitations assays, a multifunctional glutamyl-prolyl-tRNA synthetase, which catalyzes the aminoacylation of glutamic acid and proline tRNA species, has been found to interact with hDus2,[154] but the exact function of such an interaction is still obscure. Mittelstadt et al. also reported the capacity of hDus2 to interact with other cellular partner, namely protein kinase R (PKR) and interferon (IFN)-inducible double-stranded RNA-dependent protein kinase activator A (PACT), using yeast two-hybrid screen and immunoprecipitation assays.[155] The IFN-induced, dsRNA-activated protein kinase PKR, a serine/threonine kinase, is a major mediator of the antiproliferative and antiviral actions of IFN.[156−158] Although induced at the transcriptional level by IFNs, PKR is present at a low, basal level in most cell types until it binds to its activators, including the protein activator PACT. Once activated, PKR phosphorylates eukaryotic translation initiation factor 2 (eIF2α), leading to an inhibition of protein synthesis. Binding of hDus2 to PKR resulted in an inhibition of PKR activity both *in vitro* as well as in mammalian cells. Moreover, overexpression of hDus2 seems to inhibit stress-induced apoptosis, indicating that it acts as an important negative regulator of PKR activity in cells.[155] How hDus2 enhances the rate of translation is not clear, but inhibition of

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

eIF2α phosphorylation brought about by activated PKR could be a possibility for efficient translation.

Recent advances in transcriptome sequencing have facilitated identification of novel fusion transcripts in human gastric cancer, including a Dus4-Bcap29 fusion transcript present in most of the analyzed gastric cancer tissues,[159] Bcap29 being a B-cell receptor associated protein.[160] Dus4-Bcap29 fusion transcript exists also in a variety of normal tissues, notably in noncancerous epithelial and fibroblast cell lines.[161,162] Suppression of the highly expressed Dus4-Bcap29 transcript without affecting expression of Dus4 stopped cell proliferation, while a siRNA specifically targeting Dus4 did not. Induced overexpression of this transcript in SNU-638 cell lines promotes cell proliferation in a time-dependent manner, suggesting that the Dus4-Bcap29 is tumorigenic in gastric cancer.

Dihydrouridylation. As mentioned above, there has never been a clear link between RNA dihydrouridylation activity and the possibility of leading to tumor processes. However, we would like to discuss this aspect, even if it is more speculation than established evidence. Human Dus2 protein appears to impact translation efficiency since the *in vitro* rabbit-reticulocyte lysate system increased protein production in the presence of hDus2, reminiscent of the recent results by Dai et al. showing that hDus3 is important for translational efficiency.[22] This impact on translation has not been rationalized. In that respect, an action on mRNA is not excluded given the presence of D in mRNAs and their effect on the regulation of certain genes, including those involved in the cytoskeleton have been confirmed. Hence, the increase of dihydrouridylation activity in the transcriptome could also target mRNAs, providing additional tricks used by cancer cells to stimulate protein translation and thereby their cellular metabolism. Another reasonable explanation that we propose here is that additional D sites could appear at noncanonical positions due to, for example, nonspecific Dus activity. This nonspecific activity could occur during protein overexpression events, as is often the case in cancer cells (see below). On closer inspection, the sequence of the Dus4 transcript includes residues 1−237 of the protein, which corresponds to the majority of the TIM-barrel deleted from its last β-strand based on our hDus4 3D-model (Figure 7). In our opinion, this form of Dus4 is certainly nonfunctional, especially in the absence of its HD domain. We can therefore legitimately think that Dus4-Bcpa29 is devoid of RNA dihydrouridylation function, thus excluding in that specific case the involvement of Dus activity in the tumorigenic process. However, the dihydrouridylation activity of Dus in general and of Dus4 in particular could obviously be relevant in other types of cancer, in particular in lung cancer. Indeed, a recent study has proven the tumorigenic role of hDus4 in lung cancers.[163] Overall, special attention should be given to a more detailed examination of the relevance of D in cancer biology in the future.

CONCLUSION

This is the first comprehensive review devoted to dihydrouridine in the transcriptome in which we have addressed all aspects related to its physicochemical and structural properties and its distribution in the transcriptome as well as its biosynthesis and functional and pathological implications. Although, in our opinion, we are just beginning to glimpse its potential biological role in protein translation, there are still many unknowns that will need to be addressed in the future. It is undeniable that this modified base carries a structural attribute by preventing and/or promoting the formation of certain RNA folds. The consequences are specific to the nature of the substrate since D seems to stabilize tRNAs, whereas it destabilizes dsRNAs. This point needs to be further investigated, in particular by trying to apply it to specific mRNA sequences that have been identified as carrying this modification. Beyond its structuring role and its impact on translation, it is quite legitimate to wonder whether D could fulfill other potential functions as a local identity element for yet-to-be-discovered cellular partners, as is already the case for other types of modifications. But to better appreciate the importance of this modified base, more precise mapping of D in the transcriptome wide is required, which will require breaking through common roadblocks imposed by large sequencing techniques. Although the recent Rho-seq technique is an obvious advance in the field, since it does not rely on the generation of abasic sites but on a specific tagging of the D-base, there is still room to improve this method. For example, demethylation steps of the RNA sample by treatment with AlkB demethylases could be introduced to remove m^6A, m^1A, and m^5C. Finally, we would like to emphasize that this review has highlighted the extent of what remains to be accomplished to fully understand the enzymology of D synthesis. Certainly, the enzymes of Dus have been identified and their site specificity determined, the full understanding of nucleotide level determinants beyond elbow integrity for tRNA substrates at least is yet to be determined, particularly for eukaryotic enzymes that target both tRNAs and mRNAs. Besides, the fact that mRNA-modifying enzymes usually also have tRNA as a substrate, as is the case for Dus, seriously complicates the interpretation of genetic targeting, encouraging us to elucidate the molecular basis behind the substrate discrimination of these enzymes. Likewise, the complex modularity of eukaryotic Dus enzymes needs to be better characterized, and AlphaFold can help us considerably in that endeavor. Have these eukaryotic specific domains been acquired during evolution for functional purposes of recognition, substrate discrimination, transport, cellular localization, are questions begging for answers. These points will certainly be substantial to advance our understanding of the role of D and its enzymes in carcinogenic mechanisms. Finally, the abundance of D in tRNAs and more recently in mRNAs should not make us forget rRNAs, certainly the least understood dihydrouridylation substrates. The fact that *bona fide* Dus enzymes do not appear to be involved in D biosynthesis in rRNAs suggests that a fascinating new enzyme system for D biosynthesis exists in nature and is just waiting to be brought to light.

AUTHOR INFORMATION

Corresponding Author

Djemel Hamdane − *Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France;* orcid.org/0000-0002-1737-8320; Phone: +33-(0)1-44271645; Email: djemel.hamdane@college-de-france.fr

Authors

Damien Brégeon − *IBPS, Biology of Aging and Adaptation, Sorbonne Université, Paris 75252, France*

Ludovic Pecqueur − *Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France*

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

Sabrine Toubdji − *IBPS, Biology of Aging and Adaptation, Sorbonne Université, Paris 75252, France; Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France*

Claudia Sudol − *IBPS, Biology of Aging and Adaptation, Sorbonne Université, Paris 75252, France; Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France*

Murielle Lombard − *Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France*

Marc Fontecave − *Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège De France, Université Pierre et Marie Curie, 75231 Paris, France;* orcid.org/0000-0002-8016-4747

Valérie de Crécy-Lagard − *Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611, United States; Genetics Institute, University of Florida, Gainesville, Florida 32610, United States;* orcid.org/0000-0002-9955-3785

Yuri Motorin − *Université de Lorraine, CNRS, INSERM, UMS2008/US40 IBSLor, EpiRNA-Seq Core Facility, Nancy F-54000, France; Université de Lorraine, CNRS, UMR7365 IMoPA, Nancy F-54000, France*

Mark Helm − *Institut für pharmazeutische und biomedizinische Wissenschaften (IPBW), Johannes Gutenberg-Universität, Mainz 55128, Germany;* orcid.org/0000-0002-0154-0928

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.2c00307

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank all past and present members of our laboratories who participated in the dihydrouridine project. This research is funded by ANR/DFG grant DERASE (no. 20-CE92-0030), LABEX DYNAMO, ANR-11-LABX-0011, by grant GM132254 to V.d.C.L. M.H. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR-319-TP A05, SPP1784, HE 3397/13-2, and HE 3397/21-1.

■ KEYWORDS

tRNA, tRNA is an adaptor molecule typically composed of around 76 nucleotides that serves as a physical link between mRNA and the amino acid sequence of proteins. The tRNA performs this function by transporting an amino acid to the ribosome where it pairs via its anticodon to the codon of an mRNA; mRNA, messenger ribonucleic acid is a single-stranded molecule of RNA that corresponds to the genetic sequence of a gene and is read by the ribosome machinery during the translation process; rRNA, ribosomal ribonucleic acid is noncoding RNA found as the primary component of ribosomes. rRNA allows tRNA and mRNA to process and translate the latter into proteins; RNA modifications, chemically modified nucleotides found in mature RNA species. The modifications are specifically catalyzed by enzymes after transcription and are one of the maturation steps leading to functional RNAs; epitranscriptome, all functionally relevant biochemical modifications of the RNAs (the transcriptome) within a cell; dihydrouridine, one of the most abundant modifications of the transcriptome. This base is formed by the reduction of uridine and is therefore nonaromatic; dihydrouridine synthases, a family of flavoenzymes using flavin as a coenzyme and NADPH as a reductant to catalyze the dihydrouridylation (dihydrouridine synthesis) of tRNAs and mRNAs; flavin, flavin is a redox-active organic coenzyme associated with various enzymes (flavoenzymes)

■ REFERENCES

(1) Boccaletto, P.; Baginski, B. MODOMICS: An Operational Guide to the Use of the RNA Modification Pathways Database. *Methods Mol. Biol.* **2021**, *2284*, 481−505.

(2) Spenkuch, F.; Motorin, Y.; Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA biology* **2014**, *11*, 1540−1554.

(3) Frye, M.; Jaffrey, S. R.; Pan, T.; Rechavi, G.; Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nature reviews. Genetics* **2016**, *17*, 365−372.

(4) Jia, G.; Fu, Y.; He, C. Reversible RNA adenosine methylation in biological regulation. *Trends in genetics: TIG* **2013**, *29*, 108−115.

(5) Cozen, A. E.; Quartley, E.; Holmes, A. D.; Hrabeta-Robinson, E.; Phizicky, E. M.; Lowe, T. M. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **2015**, *12*, 879−884.

(6) Ma, C. J.; Ding, J. H.; Ye, T. T.; Yuan, B. F.; Feng, Y. Q. AlkB Homologue 1 Demethylates N(3)-Methylcytidine in mRNA of Mammals. *ACS Chem. Biol.* **2019**, *14*, 1418−1425.

(7) Hrabeta-Robinson, E.; Marcus, E.; Cozen, A. E.; Phizicky, E. M.; Lowe, T. M. High-Throughput Small RNA Sequencing Enhanced by AlkB-Facilitated RNA de-Methylation (ARM-Seq). *Methods Mol. Biol.* **2017**, *1562*, 231−243.

(8) Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T. Selective Enzymatic Demethylation of N(2),N(2) -Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. *Angew. Chem., Int. Ed. Engl.* **2017**, *56*, 5017−5020.

(9) Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017**, *169*, 1187−1200.

(10) Helm, M.; Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nature reviews. Genetics* **2017**, *18*, 275−291.

(11) El Yacoubi, B.; Bailly, M.; de Crécy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual review of genetics* **2012**, *46*, 69−95.

(12) Jackman, J. E.; Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley interdisciplinary reviews. RNA* **2013**, *4*, 35−48.

(13) Helm, M.; Alfonzo, J. D. Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. *Chemistry & biology* **2014**, *21*, 174−185.

(14) Lorenz, C.; Lunse, C. E.; Morl, M. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **2017**, *7*, 35.

(15) Barraud, P.; Tisne, C. To be or not to be modified: Miscellaneous aspects influencing nucleotide modifications in tRNAs. *IUBMB life* **2019**, *71*, 1126−1140.

(16) Motorin, Y.; Helm, M. RNA nucleotide methylation: 2021 update. *Wiley interdisciplinary reviews. RNA* **2022**, *13*, No. e1691.

(17) Grosjean, H.; de Crécy-Lagard, V.; Marck, C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS letters* **2010**, *584*, 252−264.

(18) Pan, T. N6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends in biochemical sciences* **2013**, *38*, 204−209.

(19) Liu, J.; Jia, G. Methylation modifications in eukaryotic messenger RNA. *Journal of genetics and genomics* **2014**, *41*, 21−33.
(20) Zhao, B. S.; Roundtree, I. A.; He, C. Post-transcriptional gene regulation by mRNA modifications. *Nature reviews. Molecular cell biology* **2017**, *18*, 31−42.
(21) Franco, M. K.; Koutmou, K. S. Chemical modifications to mRNA nucleobases impact translation elongation and termination. *Biophys. Chem.* **2022**, *285*, 106780.
(22) Huang, H.; Weng, H.; Sun, W.; Qin, X.; Shi, H.; Wu, H.; Zhao, B. S.; Mesquita, A.; Liu, C.; Yuan, C. L.; et al. Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature cell biology* **2018**, *20*, 285−295.
(23) Dai, W.; Li, A.; Yu, N. J.; Nguyen, T.; Leach, R. W.; Wuhr, M.; Kleiner, R. E. Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat. Chem. Biol.* **2021**, *17*, 1178−1187.
(24) Finet, O.; Yague-Sanz, C.; Kruger, L. K.; Tran, P.; Migeot, V.; Louski, M.; Nevers, A.; Rougemaille, M.; Sun, J.; Ernst, F. G. M.; et al. Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Molecular cell* **2022**, *82*, 404−419.
(25) House, C. H.; Miller, S. L. Hydrolysis of dihydrouridine and related compounds. *Biochemistry* **1996**, *35*, 315−320.
(26) Dalluge, J. J.; Hamamoto, T.; Horikoshi, K.; Morita, R. Y.; Stetter, K. O.; McCloskey, J. A. Posttranscriptional modification of tRNA in psychrophilic bacteria. *Journal of bacteriology* **1997**, *179*, 1918−1923.
(27) Noon, K. R.; Guymon, R.; Crain, P. F.; McCloskey, J. A.; Thomm, M.; Lim, J.; Cavicchioli, R. Influence of temperature on tRNA modification in archaea: *Methanococcoides burtonii* (optimum growth temperature [Topt], 23 °C) and *Stetteria hydrogenophila* (Topt, 95 degrees C). *Journal of bacteriology* **2003**, *185*, 5483−5490.
(28) Best, A. N. Composition and Characterization of tRNA from *Methanococcus vannielii*. *Journal of bacteriology* **1978**, *133*, 240−250.
(29) Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.; Penswick, J. R.; Zamir, A. Structure of a Ribonucleic Acid. *Science* **1965**, *147*, 1462−1465.
(30) Huang, R. C.; Bonner, J. Histone-bound RNA, a component of native nucleohistone. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *54*, 960−967.
(31) Sundaralingam, M.; Rao, S. T.; Abola, J. Molecular conformation of dihydrouridine: puckered base nucleoside of transfer RNA. *Science* **1971**, *172*, 725−727.
(32) Kim, S. H.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Sussman, J. L.; Wang, A. H.; Seeman, N. C.; Rich, A. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **1974**, *185*, 435−440.
(33) Suddath, F. L.; Quigley, G. J.; McPherson, A.; Sneden, D.; Kim, J. J.; Kim, S. H.; Rich, A. Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0angstroms resolution. *Nature* **1974**, *248*, 20−24.
(34) Westhof, E.; Dumas, P.; Moras, D. Crystallographic refinement of yeast aspartic acid transfer RNA. *Journal of molecular biology* **1985**, *184*, 119−145.
(35) Basavappa, R.; Sigler, P. B. The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO journal* **1991**, *10*, 3105−3111.
(36) Dalluge, J. J.; Hashizume, T.; Sopchik, A. E.; McCloskey, J. A.; Davis, D. R. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic acids research* **1996**, *24*, 1073−1079.
(37) Dyubankova, N.; Sochacka, E.; Kraszewska, K.; Nawrot, B.; Herdewijn, P.; Lescrinier, E. Contribution of dihydrouridine in folding of the D-arm in tRNA. *Organic & biomolecular chemistry* **2015**, *13*, 4960−4966.
(38) Davis, D. R.; Griffey, R. H.; Yamaizumi, Z.; Nishimura, S.; Poulter, C. D. 15N-labeled tRNA. Identification of dihydrouridine in *Escherichia coli* tRNAfMet, tRNALys, and tRNAPhe by 1H-15N two-dimensional NMR. *J. Biol. Chem.* **1986**, *261*, 3584−3587.
(39) Kawai, G.; Yamamoto, Y.; Kamimura, T.; Masegi, T.; Sekine, M.; Hata, T.; Iimori, T.; Watanabe, T.; Miyazawa, T.; Yokoyama, S. Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **1992**, *31*, 1040−1046.
(40) Shigi, N.; Suzuki, T.; Tamakoshi, M.; Oshima, T.; Watanabe, K. Conserved bases in the TPsi C loop of tRNA are determinants for thermophile-specific 2-thiouridylation at position 54. *J. Biol. Chem.* **2002**, *277*, 39128−39135.
(41) Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic acids research* **1995**, *23*, 5020−5026.
(42) Nomura, Y.; Ohno, S.; Nishikawa, K.; Yokogawa, T. Correlation between the stability of tRNA tertiary structure and the catalytic efficiency of a tRNA-modifying enzyme, archaeal tRNA-guanine transglycosylase. *Genes Cells* **2016**, *21*, 41−52.
(43) Sipa, K.; Sochacka, E.; Kazmierczak-Baranska, J.; Maszewska, M.; Janicka, M.; Nowak, G.; Nawrot, B. Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA. *RNA* **2007**, *13*, 1301−1316.
(44) Grosjean, H.; Keith, G.; Droogmans, L. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. *Methods Mol. Biol.* **2004**, *265*, 357−391.
(45) Hossain, M.; Limbach, P. A. Mass spectrometry-based detection of transfer RNAs by their signature endonuclease digestion products. *RNA* **2007**, *13*, 295−303.
(46) Su, D.; Chan, C. T. Y.; Gu, C.; Lim, K. S.; Chionh, Y. H.; McBee, M. E.; Russell, B. S.; Babu, I. R.; Begley, T. J.; Dedon, P. C. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Protoc* **2014**, *9*, 828−841.
(47) Solivio, B.; Yu, N. X.; Addepalli, B.; Limbach, P. A. Improving RNA modification mapping sequence coverage by LC-MS through a nonspecific RNase U2-E49A mutant. *Analytica chimica acta* **2018**, *1036*, 73−79.
(48) Douthwaite, S.; Kirpekar, F. Identifying modifications in RNA by MALDI mass spectrometry. *Methods Enzymol.* **2007**, *425*, 3−20.
(49) Hamdane, D.; Guerineau, V.; Un, S.; Golinelli-Pimpaneau, B. A catalytic intermediate and several flavin redox states stabilized by folate-dependent tRNA methyltransferase from *Bacillus subtilis*. *Biochemistry* **2011**, *50*, 5208−5219.
(50) Sirand-Pugnet, P.; Bregeon, D.; Beven, L.; Goyenvalle, C.; Blanchard, A.; Rose, S.; Grosjean, H.; Douthwaite, S.; Hamdane, D.; Crecy-Lagard, V. Reductive Evolution and Diversification of C5-Uracil Methylation in the Nucleic Acids of Mollicutes. *Biomolecules* **2020**, *10*, 587.
(51) Hamdane, D.; Guelorget, A.; Guerineau, V.; Golinelli-Pimpaneau, B. Dynamics of RNA modification by a multi-site-specific tRNA methyltransferase. *Nucleic acids research* **2014**, *42*, 11697−11706.
(52) Bou-Nader, C.; Montemont, H.; Guerineau, V.; Jean-Jean, O.; Bregeon, D.; Hamdane, D. Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: perspectives on the evolution scenario. *Nucleic acids research* **2018**, *46*, 1386−1394.
(53) Faivre, B.; Lombard, M.; Fakroun, S.; Vo, C. D.; Goyenvalle, C.; Guerineau, V.; Pecqueur, L.; Fontecave, M.; de Crécy-Lagard, V.; Bregeon, D.; et al. Dihydrouridine synthesis in tRNAs is under reductive evolution in Mollicutes. *RNA biology* **2021**, *18*, 2278−2289.
(54) Magrath, D. I.; Shaw, D. C. The occurrence and source of beta-alanine in alkaline hydrolysates of sRNA: a sensitive method for the detection and assay of 5,6-dihydrouracil residues in RNA. *Biochemical and biophysical research communications* **1967**, *26*, 32−37.
(55) Molinaro, M.; Sheiner, L. B.; Neelon, F. A.; Cantoni, G. L. Effect of chemical modification of dihydrouridine in yeast transfer ribonucleic acid on amino acid acceptor activity and ribosomal binding. *J. Biol. Chem.* **1968**, *243*, 1277−1282.
(56) Xing, F.; Hiley, S. L.; Hughes, T. R.; Phizicky, E. M. The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.* **2004**, *279*, 17850−17860.
(57) Cerutti, P.; Miller, N. Selective reduction of yeast transfer ribonucleic acid with sodium borohydride. *Journal of molecular biology* **1967**, *26*, 55−66.

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

(58) Wintermeyer, W.; Zachau, H. G. Replacement of odd bases in tRNA by fluorescent dyes. *Methods Enzymol.* **1974**, *29*, 667−673.
(59) Wintermeyer, W.; Zachau, H. G. Replacement of Y base, dihydrouracil, and 7-methylguanine in tRNA by artificial odd bases. *FEBS letters* **1971**, *18*, 214−218.
(60) Yang, C. H.; Soll, D. Studies of transfer RNA tertiary structure of singlet-singlet energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2838−2842.
(61) Betteridge, T.; Liu, H.; Gamper, H.; Kirillov, S.; Cooperman, B. S.; Hou, Y. M. Fluorescent labeling of tRNAs for dynamics experiments. *RNA* **2007**, *13*, 1594−1601.
(62) Marchand, V.; Ayadi, L.; Ernst, F. G. M.; Hertler, J.; Bourguignon-Igel, V.; Galvanin, A.; Kotter, A.; Helm, M.; Lafontaine, D. L. J.; Motorin, Y. AlkAniline-Seq: Profiling of m(7) G and m(3) C RNA Modifications at Single Nucleotide Resolution. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 16785−16790.
(63) Kaur, J.; Raj, M.; Cooperman, B. S. Fluorescent labeling of tRNA dihydrouridine residues: Mechanism and distribution. *RNA* **2011**, *17*, 1393−1400.
(64) Suzuki, T. A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic acids research* **2014**, *42*, 7346−7357.
(65) Bohnsack, M. T.; Sloan, K. E. The mitochondrial epitranscriptome: the roles of RNA modifications in mitochondrial translation and human disease. *Cellular and molecular life sciences: CMLS* **2018**, *75*, 241−260.
(66) Krog, J. S.; Espanol, Y.; Giessing, A. M.; Dziergowska, A.; Malkiewicz, A.; Ribas de Pouplana, L.; Kirpekar, F. 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine is one of two novel post-transcriptional modifications in tRNALys(UUU) from *Trypanosoma brucei*. *FEBS journal* **2011**, *278*, 4782−4796.
(67) Esberg, B.; Leung, H. C.; Tsui, H. C.; Bjork, G. R.; Winkler, M. E. Identification of the miaB gene, involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. *Journal of bacteriology* **1999**, *181*, 7256−7265.
(68) Pierrel, F.; Douki, T.; Fontecave, M.; Atta, M. MiaB protein is a bifunctional radical-S-adenosylmethionine enzyme involved in thiolation and methylation of tRNA. *J. Biol. Chem.* **2004**, *279*, 47555−47563.
(69) Gerber, A.; Grosjean, H.; Melcher, T.; Keller, W. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J.* **1998**, *17*, 4780−4789.
(70) Maas, S.; Gerber, A. P.; Rich, A. Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8895−8900.
(71) Brule, H.; Elliott, M.; Redlak, M.; Zehner, Z. E.; Holmes, W. M. Isolation and characterization of the human tRNA-(N1G37) methyltransferase (TRM5) and comparison to the *Escherichia coli* TrmD protein. *Biochemistry* **2004**, *43*, 9243−9255.
(72) Macbeth, M. R.; Schubert, H. L.; VanDemark, A. P.; Lingam, A. T.; Hill, C. P.; Bass, B. L. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **2005**, *309*, 1534−1539.
(73) Grosjean, H.; Constantinesco, F.; Foiret, D.; Benachenhou, N. A novel enzymatic pathway leading to 1-methylinosine modification in *Haloferax volcanii* tRNA. *Nucleic acids research* **1995**, *23*, 4312−4319.
(74) Rubio, M. A. T.; Gaston, K. W.; McKenney, K. M.; Fleming, I. M. C.; Paris, Z.; Limbach, P. A.; Alfonzo, J. D. Editing and methylation at a single site by functionally interdependent activities. *Nature* **2017**, *542*, 494−497.
(75) Han, L.; Phizicky, E. M. A rationale for tRNA modification circuits in the anticodon loop. *RNA* **2018**, *24*, 1277−1284.
(76) Armengod, M. E.; Moukadiri, I.; Prado, S.; Ruiz-Partida, R.; Benitez-Paez, A.; Villarroya, M.; Lomas, R.; Garzon, M. J.; Martinez-Zamora, A.; Meseguer, S.; et al. Enzymology of tRNA modification in the bacterial MnmEG pathway. *Biochimie* **2012**, *94*, 1510−1520.
(77) Kowalak, J. A.; Bruenger, E.; McCloskey, J. A. Posttranscriptional modification of the central loop of domain V in *Escherichia coli* 23 S ribosomal RNA. *J. Biol. Chem.* **1995**, *270*, 17758−17764.
(78) Kirpekar, F.; Hansen, L. H.; Mundus, J.; Tryggedsson, S.; Teixeira Dos Santos, P.; Ntokou, E.; Vester, B. Mapping of ribosomal 23S ribosomal RNA modifications in Clostridium sporogenes. *RNA biology* **2018**, *15*, 1060−1070.
(79) Emmerechts, G.; Barbe, S.; Herdewijn, P.; Anne, J.; Rozenski, J. Post-transcriptional modification mapping in the *Clostridium acetobutylicum* 16S rRNA by mass spectrometry and reverse transcriptase assays. *Nucleic acids research* **2007**, *35*, 3494−3503.
(80) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014**, *515*, 143−146.
(81) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014**, *159*, 148−162.
(82) Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim, M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.; et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **2016**, *530*, 441−446.
(83) Arango, D.; Sturgill, D.; Alhusaini, N.; Dillman, A. A.; Sweet, T. J.; Hanson, G.; Hosogane, M.; Sinclair, W. R.; Nanan, K. K.; Mandler, M. D.; et al. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **2018**, *175*, 1872−1886.
(84) Tardu, M.; Jones, J. D.; Kennedy, R. T.; Lin, Q.; Koutmou, K. S. Identification and Quantification of Modified Nucleosides in *Saccharomyces cerevisiae* mRNAs. *ACS Chem. Biol.* **2019**, *14*, 1403−1409.
(85) Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y. G.; et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **2011**, *7*, 885−887.
(86) Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **2012**, *485*, 201−206.
(87) Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **2012**, *149*, 1635−1646.
(88) Li, X. Y.; Zhu, P.; Ma, S. Q.; Song, J. H.; Bai, J. Y.; Sun, F. F.; Yi, C. Q. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **2015**, *11*, 592−U593.
(89) Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S.; Peer, E.; Kol, N.; Ben-Haim, M. S.; Dai, Q.; Di Segni, A.; Salmon-Divon, M.; Clark, W. C.; et al. The dynamic N-1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **2016**, *530*, 441−446.
(90) Li, X. Y.; Xiong, X. S.; Wang, K.; Wang, L. X.; Shu, X. T.; Ma, S. Q.; Yi, C. Q. Transcriptome-wide mapping reveals reversible and dynamic N-1-methyladenosine methylome. *Nat. Chem. Biol.* **2016**, *12*, 311−316.
(91) Li, X.; Xiong, X.; Zhang, M.; Wang, K.; Chen, Y.; Zhou, J.; Mao, Y.; Lv, J.; Yi, D.; Chen, X. W.; et al. Base-Resolution Mapping Reveals Distinct m^1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Molecular cell* **2017**, *68*, 993−1005.
(92) Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014**, *505*, 117−120.
(93) Safra, M.; Sas-Chen, A.; Nir, R.; Winkler, R.; Nachshon, A.; Bar-Yaacov, D.; Erlacher, M.; Rossmanith, W.; Stern-Ginossar, N.;

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

Schwartz, S. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017**, *551*, 251−255.
(94) Lombard, M.; Hamdane, D. Flavin-dependent epitranscriptomic world. *Archives of biochemistry and biophysics* **2017**, *632*, 28−40.
(95) Kasprzak, J. M.; Czerwoniec, A.; Bujnicki, J. M. Molecular evolution of dihydrouridine synthases. *BMC bioinformatics* **2012**, *13*, 153.
(96) Jiang, H. Q.; Motorin, Y.; Jin, Y. X.; Grosjean, H. Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. *Nucleic acids research* **1997**, *25*, 2694−2701.
(97) Xing, F.; Martzen, M. R.; Phizicky, E. M. A conserved family of *Saccharomyces cerevisiae* synthases effects dihydrouridine modification of tRNA. *RNA* **2002**, *8*, 370−381.
(98) Bishop, A. C.; Xu, J.; Johnson, R. C.; Schimmel, P.; de Crécy-Lagard, V. Identification of the tRNA-dihydrouridine synthase family. *J. Biol. Chem.* **2002**, *277*, 25090−25095.
(99) Byrne, R. T.; Jenkins, H. T.; Peters, D. T.; Whelan, F.; Stowell, J.; Aziz, N.; Kasatsky, P.; Rodnina, M. V.; Koonin, E. V.; Konevega, A. L.; et al. Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6033−6037.
(100) Yu, F.; Tanaka, Y.; Yamashita, K.; Suzuki, T.; Nakamura, A.; Hirano, N.; Yao, M.; Tanaka, I. Molecular basis of dihydrouridine formation on tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 19593−19598.
(101) Kusuba, H.; Yoshida, T.; Iwasaki, E.; Awai, T.; Kazayama, A.; Hirata, A.; Tomikawa, C.; Yamagami, R.; Hori, H. In vitro dihydrouridine formation by tRNA dihydrouridine synthase from *Thermus thermophilus*, an extreme-thermophilic eubacterium. *Journal of biochemistry* **2015**, *158*, 513−521.
(102) Chen, X.; Sim, S.; Wurtmann, E. J.; Feke, A.; Wolin, S. L. Bacterial noncoding Y RNAs are widespread and mimic tRNAs. *RNA* **2014**, *20*, 1715−1724.
(103) Bou-Nader, C.; Pecqueur, L.; Bregeon, D.; Kamah, A.; Guerineau, V.; Golinelli-Pimpaneau, B.; Guimaraes, B. G.; Fontecave, M.; Hamdane, D. An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic acids research* **2015**, *43*, 9446−9456.
(104) Brzezicha, B.; Schmidt, M.; Makalowska, I.; Jarmolowski, A.; Pienkowska, J.; Szweykowska-Kulinska, Z. 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* **2006**, *34*, 6034−6043.
(105) Squires, J. E.; Patel, H. R.; Nousch, M.; Sibbritt, T.; Humphreys, D. T.; Parker, B. J.; Suter, C. M.; Preiss, T. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic acids research* **2012**, *40*, 5023−5033.
(106) Auxilien, S.; Guerineau, V.; Szweykowska-Kulinska, Z.; Golinelli-Pimpaneau, B. The Human tRNA m(5)C methyltransferase Misu is multisite-specific. *RNA biology* **2012**, *9*, 1331−1338.
(107) Huang, T.; Chen, W. Y.; Liu, J. H.; Gu, N. N.; Zhang, R. Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nature structural & molecular biology* **2019**, *26*, 380−388.
(108) Motorin, Y.; Grosjean, H. Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast *Saccharomyces cerevisiae*: identification of the gene and substrate specificity of the enzyme. *RNA* **1999**, *5*, 1105−1118.
(109) Motorin, Y.; Keith, G.; Simon, C.; Foiret, D.; Simos, G.; Hurt, E.; Grosjean, H. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA* **1998**, *4*, 856−869.
(110) Massenet, S.; Motorin, Y.; Lafontaine, D. L.; Hurt, E. C.; Grosjean, H.; Branlant, C. Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Molecular and cellular biology* **1999**, *19*, 2142−2154.
(111) Becker, H. F.; Motorin, Y.; Planta, R. J.; Grosjean, H. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic acids research* **1997**, *25*, 4493−4499.
(112) Lovejoy, A. F.; Riordan, D. P.; Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PloS one* **2014**, *9*, No. e110799.
(113) Begik, O.; Lucas, M. C.; Pryszcz, L. P.; Ramirez, J. M.; Medina, R.; Milenkovic, I.; Cruciani, S.; Liu, H.; Vieira, H. G. S.; Sas-Chen, A.; et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nature biotechnology* **2021**, *39*, 1278−1291.
(114) Urban, A.; Behm-Ansmant, I.; Branlant, C.; Motorin, Y. RNA sequence and two-dimensional structure features required for efficient substrate modification by the *Saccharomyces cerevisiae* RNA:{Psi}-synthase Pus7p. *J. Biol. Chem.* **2009**, *284*, 5845−5858.
(115) Decatur, W. A.; Schnare, M. N. Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. *Molecular and cellular biology* **2008**, *28*, 3089−3100.
(116) Safra, M.; Nir, R.; Farouq, D.; Vainberg Slutskin, I.; Schwartz, S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome research* **2017**, *27*, 393−406.
(117) Lecointe, F.; Simos, G.; Sauer, A.; Hurt, E. C.; Motorin, Y.; Grosjean, H. Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. *J. Biol. Chem.* **1998**, *273*, 1316−1323.
(118) Hopper, A. K.; Furukawa, A. H.; Pham, H. D.; Martin, N. C. Defects in modification of cytoplasmic and mitochondrial transfer RNAs are caused by single nuclear mutations. *Cell* **1982**, *28*, 543−550.
(119) Pagliarini, D. J.; Calvo, S. E.; Chang, B.; Sheth, S. A.; Vafai, S. B.; Ong, S. E.; Walford, G. A.; Sugiana, C.; Boneh, A.; Chen, W. K.; et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **2008**, *134*, 112−123.
(120) Chen, M.; Yu, J.; Tanaka, Y.; Tanaka, M.; Tanaka, I.; Yao, M. Structure of dihydrouridine synthase C (DusC) from *Escherichia coli*. *Acta crystallographica. Section F, Structural biology and crystallization communications* **2013**, *69*, 834−838.
(121) Whelan, F.; Jenkins, H. T.; Griffiths, S. C.; Byrne, R. T.; Dodson, E. J.; Antson, A. A. From bacterial to human dihydrouridine synthase: automated structure determination. *Acta crystallographica. Section D, Biological crystallography* **2015**, *71*, 1564−1571.
(122) Bou-Nader, C.; Barraud, P.; Pecqueur, L.; Perez, J.; Velours, C.; Shepard, W.; Fontecave, M.; Tisne, C.; Hamdane, D. Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic acids research* **2019**, *47*, 3117−3126.
(123) Bou-Nader, C.; Pecqueur, L.; Barraud, P.; Fontecave, M.; Tisne, C.; Sacquin-Mora, S.; Hamdane, D. Conformational Stability Adaptation of a Double-Stranded RNA-Binding Domain to Transfer RNA Ligand. *Biochemistry* **2019**, *58*, 2463−2473.
(124) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. The double-stranded-RNA-binding motif: interference and much more. *Nature reviews. Molecular cell biology* **2004**, *5*, 1013−1023.
(125) Masliah, G.; Barraud, P.; Allain, F. H. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol. Life Sci.* **2013**, *70*, 1875−1895.
(126) Barraud, P.; Emmerth, S.; Shimada, Y.; Hotz, H. R.; Allain, F. H.; Buhler, M. An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO journal* **2011**, *30*, 4223−4235.
(127) Barraud, P.; Banerjee, S.; Mohamed, W. I.; Jantsch, M. F.; Allain, F. H. A bimodular nuclear localization signal assembled via an extended double-stranded RNA-binding domain acts as an RNA-sensing signal for transportin 1. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1852−1861.
(128) Stefl, R.; Oberstrass, F. C.; Hood, J. L.; Jourdan, M.; Zimmermann, M.; Skrisovska, L.; Maris, C.; Peng, L.; Hofr, C.; Emeson, R. B.; et al. The solution structure of the ADAR2 dsRBM-

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **2010**, *143*, 225−237.
(129) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(130) Cramer, P. AlphaFold2 and the future of structural biology. *Nature structural & molecular biology* **2021**, *28*, 704−705.
(131) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **2022**, *50*, D439−D444.
(132) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590−596.
(133) Rider, L. W.; Ottosen, M. B.; Gattis, S. G.; Palfey, B. A. Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *J. Biol. Chem.* **2009**, *284*, 10324−10333.
(134) Savage, D. F.; de Crécy-Lagard, V.; Bishop, A. C. Molecular determinants of dihydrouridine synthase activity. *FEBS letters* **2006**, *580*, 5198−5202.
(135) Dobritzsch, D.; Schneider, G.; Schnackerz, K. D.; Lindqvist, Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *EMBO journal* **2001**, *20*, 650−660.
(136) Dobritzsch, D.; Ricagno, S.; Schneider, G.; Schnackerz, K. D.; Lindqvist, Y. Crystal structure of the productive ternary complex of dihydropyrimidine dehydrogenase with NADPH and 5-iodouracil. Implications for mechanism of inhibition and electron transfer. *J. Biol. Chem.* **2002**, *277*, 13155−13166.
(137) Barraud, P.; Gato, A.; Heiss, M.; Catala, M.; Kellner, S.; Tisne, C. Time-resolved NMR monitoring of tRNA maturation. *Nat. Commun.* **2019**, *10*, 3373.
(138) Cavaille, J.; Chetouani, F.; Bachellerie, J. P. The yeast Saccharomyces cerevisiae YDL112w ORF encodes the putative 2'-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs. *RNA* **1999**, *5*, 66−81.
(139) Kealey, J. T.; Santi, D. V. Identification of the catalytic nucleophile of tRNA (m5U54)methyltransferase. *Biochemistry* **1991**, *30*, 9724−9728.
(140) Liu, Y.; Santi, D. V. m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8263−8265.
(141) Lee, T. T.; Agarwalla, S.; Stroud, R. M. A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **2005**, *120*, 599−611.
(142) Hamdane, D.; Argentini, M.; Cornu, D.; Myllykallio, H.; Skouloubris, S.; Hui-Bon-Hoa, G.; Golinelli-Pimpaneau, B. Insights into folate/FAD-dependent tRNA methyltransferase mechanism: role of two highly conserved cysteines in catalysis. *J. Biol. Chem.* **2011**, *286*, 36268−36280.
(143) O'Connor, M.; Lee, W. M.; Mankad, A.; Squires, C. L.; Dahlberg, A. E. Mutagenesis of the peptidyltransferase center of 23S rRNA: the invariant U2449 is dispensable. *Nucleic acids research* **2001**, *29*, 710−715.
(144) Alexandrov, A.; Chernyakov, I.; Gu, W.; Hiley, S. L.; Hughes, T. R.; Grayhack, E. J.; Phizicky, E. M. Rapid tRNA decay can result from lack of nonessential modifications. *Molecular cell* **2006**, *21*, 87−96.
(145) Boriack-Sjodin, P. A.; Ribich, S.; Copeland, R. A. RNA-modifying proteins as anticancer drug targets. *Nature reviews. Drug discovery* **2018**, *17*, 435−453.
(146) Delaunay, S.; Frye, M. RNA modifications regulating cell fate in cancer. *Nature cell biology* **2019**, *21*, 552−559.
(147) Barbieri, I.; Kouzarides, T. Role of RNA modifications in cancer. *Nature reviews. Cancer* **2020**, *20*, 303−322.
(148) Haruehanroengra, P.; Zheng, Y. Y.; Zhou, Y.; Huang, Y.; Sheng, J. RNA modifications and cancer. *RNA biology* **2020**, *17*, 1560−1575.
(149) Chujo, T.; Tomizawa, K. Human transfer RNA modopathies: diseases caused by aberrations in transfer RNA modifications. *FEBS journal* **2021**, *288*, 7096−7122.
(150) Zhou, J. B.; Wang, E. D.; Zhou, X. L. Modifications of the human tRNA anticodon loop and their associations with genetic diseases. *Cell. Mol. Life Sci.* **2021**, *78*, 7087−7105.
(151) Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Bio* **2021**, *22*, 375−392.
(152) Kuchino, Y.; Borek, E. Tumour-specific phenylalanine tRNA contains two supernumerary methylated bases. *Nature* **1978**, *271*, 126−129.
(153) Dalluge, J. J.; Hashizume, T.; McCloskey, J. A. Quantitative measurement of dihydrouridine in RNA using isotope dilution liquid chromatography mass spectrometry (LC/MS). *Nucleic acids research* **1996**, *24*, 3242−3245.
(154) Kato, T.; Daigo, Y.; Hayama, S.; Ishikawa, N.; Yamabuki, T.; Ito, T.; Miyamoto, M.; Kondo, S.; Nakamura, Y. A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer research* **2005**, *65*, 5638−5646.
(155) Mittelstadt, M.; Frump, A.; Khuu, T.; Fowlkes, V.; Handy, I.; Patel, C. V.; Patel, R. C. Interaction of human tRNA-dihydrouridine synthase-2 with interferon-induced protein kinase PKR. *Nucleic acids research* **2007**, *36*, 998−1008.
(156) Lee, Y. S.; Kunkeaw, N.; Lee, Y.-S. Protein kinase R and its cellular regulators in cancer: An active player or a surveillant? *Wiley interdisciplinary reviews. RNA* **2020**, *11*, No. e1558.
(157) Bou-Nader, C.; Gordon, J. M.; Henderson, F. E.; Zhang, J. The search for a PKR code-differential regulation of protein kinase R activity by diverse RNA and protein regulators. *RNA* **2019**, *25*, 539−556.
(158) Hull, C. M.; Bevilacqua, P. C. Discriminating Self and Non-Self by RNA: Roles for RNA Structure, Misfolding, and Modification in Regulating the Innate Immune Sensor PKR. *Accounts of chemical research* **2016**, *49*, 1242−1249.
(159) Kim, H. P.; Cho, G. A.; Han, S. W.; Shin, J. Y.; Jeong, E. G.; Song, S. H.; Lee, W. C.; Lee, K. H.; Bang, D.; Seo, J. S.; et al. Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene* **2014**, *33*, 5434−5441.
(160) Kim, K. M.; Adachi, T.; Nielsen, P. J.; Terashima, M.; Lamers, M. C.; Kohler, G.; Reth, M. Two new proteins preferentially associated with membrane immunoglobulin D. *EMBO journal* **1994**, *13*, 3793−3800.
(161) Tang, Y.; Qin, F.; Liu, A.; Li, H. Recurrent fusion RNA DUS4L-BCAP29 in non-cancer human tissues and cells. *Oncotarget* **2017**, *8*, 31415−31423.
(162) Tang, Y.; Guan, F.; Li, H. Case Study: The Recurrent Fusion RNA DUS4L-BCAP29 in Noncancer Human Tissues and Cells. *Methods Mol. Biol.* **2020**, *2079*, 243−258.
(163) Li, Z.; Yin, C.; Li, B.; Yu, Q. Y.; Mao, W. J.; Li, J.; Lin, J. P.; Meng, Y. Q.; Feng, H. M.; Jing, T. DUS4L Silencing Suppresses Cell Proliferation and Promotes Apoptosis in Human Lung Adenocarcinoma Cell Line A549. *Cancer management and research* **2020**, *12*, 9905−9913.

https://doi.org/10.1021/acschembio.2c00307
ACS Chem. Biol. 2022, 17, 1638−1657

See article 1: Evolutionary Diversity of Dus2 Enzymes Reveals Novel Structural and Functional Features among Members of the RNA Dihydrouridine Synthases Family.

biomolecules

MDPI

Article

Evolutionary Diversity of Dus2 Enzymes Reveals Novel Structural and Functional Features among Members of the RNA Dihydrouridine Synthases Family

Murielle Lombard [1], Colbie J. Reed [2], Ludovic Pecqueur [1], Bruno Faivre [1], Sabrine Toubdji [1,3], Claudia Sudol [1,3], Damien Brégeon [3], Valérie de Crécy-Lagard [2,4] and Djemel Hamdane [1,*]

1 Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège de France, Université Pierre et Marie Curie, 11 Place Marcelin Berthelot, CEDEX 05, 75231 Paris, France
2 Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA
3 IBPS, Biology of Aging and Adaptation, Sorbonne Université 7 quai Saint Bernard, CEDEX 05, 75252 Paris, France
4 Genetics Institute, University of Florida, Gainesville, FL 32610, USA
* Correspondence: djemel.hamdane@college-de-france.fr; Tel.: +33-(0)-1-4427-1642

Abstract: Dihydrouridine (D) is an abundant modified base found in the tRNAs of most living organisms and was recently detected in eukaryotic mRNAs. This base confers significant conformational plasticity to RNA molecules. The dihydrouridine biosynthetic reaction is catalyzed by a large family of flavoenzymes, the dihydrouridine synthases (Dus). So far, only bacterial Dus enzymes and their complexes with tRNAs have been structurally characterized. Understanding the structure-function relationships of eukaryotic Dus proteins has been hampered by the paucity of structural data. Here, we combined extensive phylogenetic analysis with high-precision 3D molecular modeling of more than 30 Dus2 enzymes selected along the tree of life to determine the evolutionary molecular basis of D biosynthesis by these enzymes. Dus2 is the eukaryotic enzyme responsible for the synthesis of D20 in tRNAs and is involved in some human cancers and in the detoxification of β-amyloid peptides in Alzheimer's disease. In addition to the domains forming the canonical structure of all Dus, i.e., the catalytic TIM-barrel domain and the helical domain, both participating in RNA recognition in the bacterial Dus, a majority of Dus2 proteins harbor extensions at both ends. While these are mainly unstructured extensions on the N-terminal side, the C-terminal side extensions can adopt well-defined structures such as helices and beta-sheets or even form additional domains such as zinc finger domains. 3D models of Dus2/tRNA complexes were also generated. This study suggests that eukaryotic Dus2 proteins may have an advantage in tRNA recognition over their bacterial counterparts due to their modularity.

Keywords: dihydrouridine; tRNA; dihydrouridine synthase; tRNA binding; phylogeny; AlphaFold; structural-protein-evolution

Citation: Lombard, M.; Reed, C.J.; Pecqueur, L.; Faivre, B.; Toubdji, S.; Sudol, C.; Brégeon, D.; de Crécy-Lagard, V.; Hamdane, D. Evolutionary Diversity of Dus2 Enzymes Reveals Novel Structural and Functional Features among Members of the RNA Dihydrouridine Synthases Family. *Biomolecules* **2022**, *12*, 1760. https://doi.org/10.3390/biom12121760

Academic Editors: Philippe Urban and Denis Pompon

Received: 4 November 2022
Accepted: 24 November 2022
Published: 26 November 2022

1. Introduction

Dihydrouridine (D) is one of the most abundant post-transcriptional modified bases in the transcriptome [1–3]. Present mainly in transfer RNA (tRNA) and occasionally in bacterial ribosomal RNA (rRNA), D has recently entered the messenger RNA (mRNA) world [1,2]. Indeed, this modification was recently detected in fission yeast mRNAs, including those encoding cytoskeleton-related proteins (2), in *Saccharomyces cerevisiae* mRNAs [4], but also in human mRNAs [5,6]. D is formed by the reduction of the C5 = C6 double bond of uridine, resulting in a loss of aromaticity, a unique feature among base modifications (Figure 1A). The lack of aromaticity leads to a pyrimidine that is unable to participate in stacking interactions [7,8].

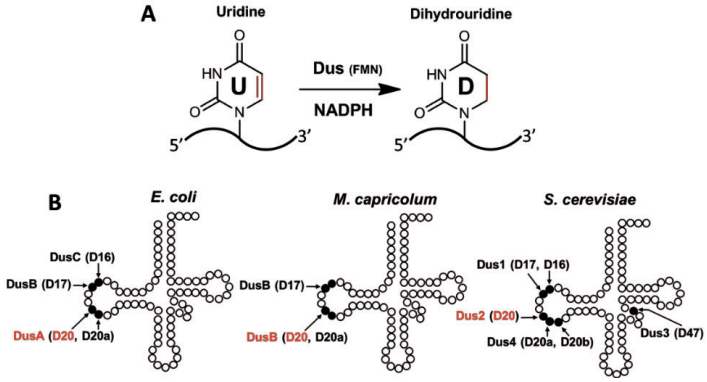


Figure 1. Dihydrouridine biosynthesis and localization of D residues in the tRNA. (**A**) Reduction reaction of uridine to dihydrouridine in tRNAs and mRNAs catalyzed by dihydrouridine synthases (Dus). These proteins are flavoenzymes that use FMN as a redox coenzyme and NADPH as a reductant source. (**B**) Cloverleaf secondary structure of the tRNAs shows the location of the D residues as well as the Dus enzymes that introduce them into *Escherichia coli*, *Mycoplasma capricolum*, and *Saccharomyces cerevisiae*.

The exact physiological role of D remains to be clearly defined, although some studies suggest that it plays an important role in 3D RNA shaping by promoting local flexibility and RNA backbone dynamics [9,10]. Consistent with this structural property, D is found primarily in single-stranded loops and in regions of the RNA that require flexibility, such as the tRNA elbow region, formed by the interaction between the D and TΨC loops and involving important tertiary interactions necessary to maintain its particular L-shaped structure [7,8]. Indeed, tRNAs lacking D in combination with other modifications have been shown to undergo rapid degradation [11], probably due to a defect in conformational flexibility. This could possibly explain how some cancer cells can prevent tRNA turnover by significantly increasing their D level in tRNAs [12] and thus promote cell growth [13]. In mRNAs, the absence of D has been shown to strongly affect meiotic chromosome segregation, leading to low gamete viability in yeast (2). In humans, D plays a role in the efficiency of the translation via a mechanism of action that remains to be established [5]. A relatively high level of D was observed in the tRNAs of cancer cells [12] and in those of psychrophilic organisms, where a greater demand for molecular flexibility is required [14].

D is often present at multiple positions in bacterial and eukaryotic tRNAs, and its abundance varies with both organism and tRNA type. In prokaryotes, D can be present at five positions of the tRNA (Figure 1B), namely positions 16, 17, 20 and 20a, all located in the D loop, but also at position 47 in the variable loop (V loop) [2], which has so far only been observed in *Bacillus subtilis* $tRNA^{Met}$(CAU). In eukaryotes, D is observed in as many as six sites, including five in the D-loop (D16, D17, D20, D20a, and D20b) and one in the variable loop (D47) [3] (Figure 1B), with D20 being the most frequent D-modification in tRNAs [2]. D residues are introduced in tRNAs and mRNAs by a set of conserved dihydrouridine synthases (Dus) that are members of the Cluster of Orthologous Group family COG0042 [2,15–17]. These flavoenzymes catalyze an NADPH-dependent reduction of specific uridines using the flavin mononucleotide (FMN) as a tRNA-reducing coenzyme (Figure 1A) [2,15,16,18,19]. A phylogenetic study classified these enzymes into eight subfamilies, namely DusA, DusB, DusC, Dus1, Dus2, Dus3, Dus4, and archaeal Dus [20]. The first three enzymes are bacterial proteins, while Dus1 through Dus4 are found in eukaryotes, the last one being the unique member of Dus observed in archaea. Since DusB is present in almost all bacteria, a model where DusB is the bacterial ancestor, and DusA and DusC are the products of DusB duplication events that occurred shortly after the divergence of the major Proteobacteria groups was proposed [20]. In eukaryotes, Dus3 is considered the ancestral enzyme from which the other three are derived, starting with Dus2, followed by Dus1, and finally Dus4.

The tRNA-substrate specificity of Dus enzymes has been fully established in several model organisms, including *Mycoplasma capricolum* [21], *Escherichia coli* [15,19], and *Thermus thermophilus* [22,23] for prokaryotes, and *S. cerevisiae* [16,24], *S. pombe* [6], and humans [5,6,25] for eukaryotes. These studies revealed that Dus can generally modify up to three positions in a given tRNA substrate (Figure 1B). DusA, Dus1, and Dus4 are dual-site enzymes catalyzing the formation of D20/D20a, D16/D17, and D20a/D20b, respectively. In contrast, DusC, Dus2, and Dus3 can modify only one position and synthesize D17, D16, D20, and D47, respectively. Concerning DusB, we have recently shown that this enzyme can be either mono-site specific, such as the *E. coli* enzyme that catalyzes the synthesis of D17 [19], or tri-site specific, catalyzing D17, D20, and D20a in *M. capricolum* [21]. From an evolutionary standpoint, the mono-site specificity of a subset of Dus proteins could thus be regarded as a functional feature that has evolved recently, at least in the prokaryotes.

To date, three-dimensional structures, obtained by X-ray crystallography, are available in the PDB for *T. thermophilus* DusA [22], *E. coli* DusB [19], and *E. coli* DusC [26,27] (Table 1). The crystallographic structures of T. *thermophilus* DusA in complex with $tRNA^{Phe}$ and *E. coli* DusC in complex with $tRNA^{Phe}$ or $tRNA^{Trp}$ were also solved [22,27]. These key data elucidated the structural and molecular basis of dihydrouridine biosynthesis in bacteria. In contrast, structural studies are limited for eukaryotic enzymes. The only data available to date are the structures of isolated domains of human Dus2 (hDus2 or Dus2L) [25,28,29] (Table 1). Beyond its physiological role, hDus2 seems to play a role in some cancers [13,30] but also in Alzheimer's disease [31]. This enzyme promotes cell growth through its ability to interact with other enzymes, notably the aminoacyl tRNA synthetase complex EPRS [13] and the protein kinase R [30], by a mechanism that remains to be established. In addition, overexpression of hDus2 in tumorigenic cells appears to be associated with a poor prognosis for lung cancer patients [13].

Table 1. X-ray structures of Dus available in the protein data bank.

Proteins	Products	Domain/Complex	PDB Code	Resolution
T. thermophilus DusA	D20/D20a	full length	3B0P	1.7
T. thermophilus DusA	D20/D20a	full length + RNA fragment	3B0U	1.94
T. thermophilus DusA	D20/D20a	full length + $tRNA^{Phe}$	3B0V	3.51
E. coli DusB	D17	full length	6EI9	2.55
E. coli DusC	D16	full length	3W9Z	2.1
E. coli DusC	D16	full length	4BFA	1.65
E. coli DusC	D16	full length + $tRNA^{Trp}$	4YCP	2.55
E. coli DusC	D16	full length + $tRNA^{Phe}$	4YCO	2.1
Homo sapiens Dus2	D20	TIM Barrel + HD	4XP7	1.9
Homo sapiens Dus2	D20	TIM Barrel + HD	4WFS	2.68
Homo sapiens Dus2	D20	dsRBD	4WFT	1.7
Homo sapiens Dus2	D20	dsRBD + dsRNA	5OC6	3.2

Overall, the structural analyses of all these structures allow for the definition of a canonical Dus fold, which consists of: (i) a TIM-barrel domain on the N-terminal side carrying the active site with the FMN located in the center of the barrel; (ii) a helical domain (HD) formed by 4 helices in a bundle lying on the C-terminal side; and (iii) a linker, which connects these two domains (Figure 2A,B).

In bacterial enzymes both the TIM-barrel and HD participate in tRNA binding [22,27]. In contrast, hDus2 was shown to carry an additional domain after the HD, namely a double-strand binding domain (dsRBD) (Figure 2B) cooperating with the TIM-barrel for tRNA recognition and binding [25]. This gain in architectural modularity is accompanied by a loss of electropositivity on the HD surface of hDus2 compared to its bacterial counterparts, which no longer fully participate in substrate binding [25,32]. The dsRBD of hDus2 is unique among members of this family as it carries a new type of N-terminal extension (NTE) [25,33]. This finding raised the possibility that this new prototype of dsRBD has

evolved to specifically recognize the particular 3D-L-shaped structure of tRNAs. Indeed, compelling structural evidence, namely the crystal structure of this dsRBD in complex with a dsRNA, structural characterization of the dsRBD/tRNA complex by NMR, SAXS, and extensive mutagenesis, provided evidence that this dsRBD is specialized to recognize a tRNA substrate via its NTE [29]. The latter provides specific residues that, in combination with those on the canonical dsRBD structure, expand the RNA-binding interface, allowing the newly evolved domain to bind tRNA. These observations led us to hypothesize that perhaps the modularity acquired by hDus2 may not be an isolated case within the Dus2 family and that the tRNA recognition mechanism may have undergone various evolutionary modifications.

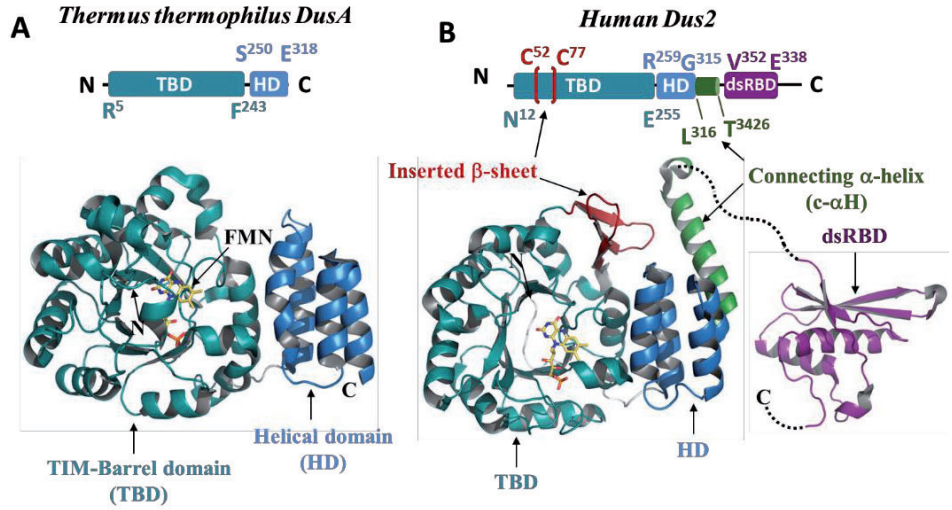


Figure 2. Crystallographic structures of Dus. (**A**,**B**) crystal structures of *T. thermophilus* DusA (PDB: 3B0V), human Dus2 without the dsRBD (PDB: 4XP7), and the dsRBD of hDus2 (PDB: 4WFT), respectively. The TIM-barrel domain (TBD) appears in teal, while the helical domain (HD) is in blue, the inserted beta-sheet in red, the connecting alpha-helix (c-αH) in green, and the dsRBD in purple. The FMN coenzyme is denoted in yellow. Above each of the structures is a scheme of the modular organization of Dus2, in which the delineation of each domain is shown.

Using phylogenetic analysis and accurate 3D protein structure prediction, we investigate here the structural evolution of eukaryotic Dus2 to identify novel modes of tRNA binding along the evolutionary tree of life. We found that Dus2 exhibits significant structural variability beyond the level of their canonical domains. Dus2 enzymes can carry structural extensions primarily on the C-terminal side that range from simple helix acquisition to the addition of a new domain. In addition to the dsRBD, we have identified five new domains that may be present in Dus2, including zinc finger modules. More importantly, analyses of protein surface electrostatics and modeling of Dus2/tRNA complexes suggest that some of these extensions are likely involved in RNA recognition. Our study illustrates how nature opportunistically refines Dus structures by decorating the canonical fold with new structural elements that function as effectors to generate new substrate recognition units.

2. Materials and Methods

2.1. Phylogenetic Distributions across the Dus Superfamily

Using representative sequences of each Dus subfamily (i.e., Q9HGN6, Dus1 of *Schizosaccharomyces pombe*; O74731, Dus2 of *S. pombe*; Q9UTH9, Dus3 of *S. pombe*; O74553, Dus4 of *S. pombe*; P32695, DusA of *E. coli*; P0ABT5, DusB of *E. coli*; P33371, DusC of *E. coli*; Q57608, archaeal Dus of *Methanocaldococcus jannaschii*), OrthoInspector [34] were used to extract sequences of Dus homologs. OrthoInspector maintains a benchmark set of genomes, which it uses to consistently determine the absence or presence of orthologs, including

both model and non-model organisms across archaea, bacteria, and eukaryota. This allows, with each query, additional lists of organisms in which homologs of the query sequence are "not present in", as well. All lists resulting from these queries were concatenated, and redundancies were removed. The sum of organisms from both list types, "present" and "not present in" was used to derive the final list of organisms to use in these analyses (Supplemental Table S1). To confirm subfamily membership of each sequence, an arbitrary number of sequences per batch (50–100 sequences) were checked by performing sequence alignments and generating sequence trees containing positive controls for each subfamily (i.e., the controls used were equivalent to the Dus sequences used in retrieval queries) (for an example of this approach, see the Supplemental example in Figure S1). The latter was completed using ClustalO (EMBL-EBI; https://www.ebi.ac.uk/Tools/msa/clustalo/ accessed on 21 April 2022) for each checked subset [35]. The assignment of subfamily membership for each sequence was determined according to their phylogenetic proximity to the aforementioned control (i.e., query) sequences. After subgroup membership was assigned, these were then used to determine the absence or presence (counts) per subgroup and per organism. Taxonomic identifiers were mapped using the UniProt sequence entries.

2.2. Dus2 Fusion Analysis

Using the Dus2 sequence of *S. pombe* (O74731), the BLAST tool of OrthoMCL (release 6.10, accessed on 21 April 2022) [36] was used to gather an initial batch of Dus2 family sequences of various architectures (orthologous group: OG6_102617; 527 sequences, total). Sequences were then mapped to UniProt Accession Identifiers and taxonomic IDs using the UniProt mapping tool [37], with the total number of sequences equaling 383. The CDD batch search tool [38] was used to map recognizable domains across all sequences, assigning either the highest fidelity specific hit fusion domain or CDD clan cd02801 (DUS_like_FMN). Dus2 homolog sequences of closely related organisms were used to BLAST the genomes suspected of gene losses/oversight (i.e., Dus2 homolog is missing; check via NCBI BLAST). Dus2 family members and fusions exported from InterPro [39], distinct from the sequences already curated, were then merged with the master list (total of 390 sequences) to give Supplemental Table S2 after being verified for Dus2 subfamily membership, again using the Dus homologs 1–4 of *S. pombe* as positive controls in an alignment and then a sequence tree (example, Supplemental Figure S2). Lengths of Dus family domains were determined from UniProt domain annotation and, if UniProt annotations were lacking, CDD Search. The phylogenetic tree was generated using PhyloT (database 2022.1; https://phylot.biobyte.de, accessed on 26 August 2022) and iToL [40]. Data (i.e., sequence lengths, domain fusions) were mapped using the iToL tree editor (accessed 26 August 2022).

2.3. Dus2 Sequence Logos

JalView was used to perform the multiple sequence alignments (ClustalO program within JalView) for the Dus2-specific set of homologs (see Dus2 Fusion Analysis methods subsection) [41]. Sequence logos were made using WebLogo (https://weblogo.threeplusone.com, accessed on 24 August 2022) [42].

2.4. AlphaFold Models

All AlphaFold models were generated using AlphaFold2, which is hosted through ColabFold (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb accessed on 1 June 2022) [43]. Atomic coordinates for these files with a pLDDT score less than 70 were not analyzed.

2.5. Cloning of dsRBD of Dus2 from Amphimedon Queenslandica

We obtained a commercially supplied synthetic plasmid of the dsRBD of Dus2 from *Amphimedon queenslandica* (residues 266–371) (pEX- dsRBD_Aq ampicillin-resistant) from Eurofins. We used this plasmid to amplify by PCR and clone into pET22b the gene of *A. queenslandica* Dus2 dsRBD (dsRBDAq), which contains a sequence encoding for

a 6-histidine tag placed at the C-terminal region of the protein. The PCR fragment was amplified with these primers (Forward aagaaggagatatacatatgAAATCGAAAATG-GATCCAGAAG; Rerverse gcggtcggcagcaggtattttcagtggtggtggtggtggtgTGAATTGCTG-GCAGTTGAC) and purified with QIAquick PCR purification kit (Qiagen, Germantown, MD USA), before cloning with SLIC cloning method into pET22b previously linearized with PCR using these primers (Forward AAATACCTGCTGCCGACC; Reverse CATATG-TATATCTCCTTCTTAAAGTTAAAC) and gel purified with QIAquick gel purification kit (Qiagen). Chemically competent DH5α cells were transformed with the plasmid, and gene integrity was verified by DNA sequencing (Eurofins).

2.6. Overexpression and Purification of dsRBDAq

Overexpression of dsRBDAq was achieved using a chemically competent BL21 (DE3) *E. coli* strain (Novagen, Madison, IA USA) transformed with pET22b-dsRBDAq. 100 mL (LB medium) of the overnight cultures were used to inoculate a larger scale cell culture (6 L) at 37 °C until the optical density at 600 nm reached 0.5. Protein synthesis was induced by the addition of Isopropyl 1-thio-β-Dgalactopyranoside to a final concentration of 1 mM. Cells were grown for an additional 16 h at 16 °C, collected by centrifugation (9000× *g* at 4 °C for 10 min), and stored at −80 °C until use. Cells were resuspended in 50 mM Tris-HCl pH 8, 500 mM NaCl, 10% glycerol, 5 mM imidazole, and 1% triton X-100 and discontinuously sonicated for 15 min in a water-ice batch. Cellular extracts were centrifuged for 45 min at 193,000× *g*, which yielded a soluble fraction of dsRBDAq.

The resulting supernatant was loaded onto a Hitrap Excel column (5 mL, GE Health-care) equilibrated with buffer, 50 mM Tris-HCl pH 8, 300 mM NaCl, 10% glycerol, and 5 mM imidazole. Bound proteins were eluted with a gradient of imidazole (0–500 mM) in buffer: 50 mM Tris-HCl pH 8, 300 mM NaCl, 10% glycerol, and 500 mM imidazole; dsRBDAq was eluted at 150 mM imidazole. SDS- PAGE allowed the identification of dsRBDAq, and the purest protein was pooled and concentrated with Amicon Ultra 10K cut-off concentrators (Millipore) until a volume of 5 mL was reached. 45 mg of protein was loaded onto a HiLoad 16/600 Superdex 75 pg equilibrated with 50 mM Tris-HCl, pH 8, 300 mM NaCl, and 5 mM 2-Mercaptoethanol. Protein concentrations were determined by a Bradford assay (Bio-Rad), with BSA used as a standard. The elution volume of dsRBDAq was 78 mL. The pure and monomeric protein fractions were pooled, concentrated to 11 mg/mL, frozen in liquid N2, and then stored at −80 °C.

2.7. Crystallization, Data Collection, and Structure Determination

Crystals of dsRBDAq were grown by the hanging drop vapor diffusion method at 292 K by mixing the purified dsRBDAq (at 15 mg/mL in 50 mM Tris, pH 8, 300 mM NaCl, and 5 mM β mercaptoethanol) with an equal amount (1 µL) of reservoir solution (0.1 M Hepes sodium salt, pH 7.5, and 1.4 M tri-sodium citrate as precipitant) and seeds previously prepared. After a few days, crystals appeared and were swept through a reservoir solution complemented with 30% glycerol and immediately flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the synchrotron SOLEIL on the beamline Proxima 2 and were indexed, integrated, and scaled using the autoPROC pipeline [44] (ref). The structure was solved by molecular replacement with phenix.phaser [45] using a search model processed with sculptor [46] and based on 4wft. Phases and models were improved with phenix.autobuild [47]. The final model and phases were obtained by alternating manual building in Coot [48] and refinement in BUSTER (www.globalphasing.com/buster/ version v2.10.4 (8-JUN-2022)). Data collection and refinement statistics are shown in Table 2.

Table 2. Data collection and refinement statistics of the dsRBD of *A. queenslandica* Dus2.

	dsRBDaq
PDB code	8B02
Data collection	
Wavelength (Å)	0.9801
Resolution range (Å)	42.37–1.68 (1.70–1.68)
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	29.077, 56.895, 63.587
α, β, γ (°)	90.00, 93.05, 90.00
Multiplicity	2.9 (2.0)
Completeness (%)	97.9 (79.7)
Mean I/sigma(I)	7.8 (0.8)
Wilson B-factor ($Å^2$)	24.07
R-meas	0.090 (1.245)
R-pim	0.051 (0.762)
CC1/2	0.997 (0.351)
Refinement	
Reflections used in refinement	23356 (1170)
R-work / R-free (%)	20.35/22.78 (32.80/36.10)
Number of non-hydrogen atoms	
macromolecules	1548
ligands	28
solvent	139
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.02
Ramachandran plot (%)	
favored	98.96
allowed	1.04
outliers	0.00
Average B-factor ($Å^2$)	
Overall	28.45
macromolecules	26.91
ligands	42.77
solvent	42.75

3. Results and Discussion

3.1. Phylogenetic Distribution of Dus2 and Domain Analyses

To understand the distribution and more readily interpret the plausible evolutionary history of Dus2, a benchmark subset of 324 genomes were investigated for the absence or presence of all Dus family subgroups, both derived from OrthoInspector (see methods and Supplemental Data Table S1). In batches of 50–100 sequences, homologs were checked for subgroup membership through multiple sequence alignments and sequence trees, including described control sequences (see methods for sequences; example of approach, Figure S1). Approximately 41% of organisms considered in this study were found to possess at least one Dus2 family member, making it the third most common of all Dus subfamilies. Similarly, 40% of taxonomic groups (see partitions and groups in Supplemental Data Table S1) were observed to have at least one organism possessing Dus2. As anticipated, organisms found to be without a single Dus2 homolog included all archaeal taxonomic groups (i.e., Asgard Group, DPANN, Euryarchaeota, TACK Group) and all bacteria (i.e., FCB Group, Proteobacteria, PVC Group, Terrabacteria Group, Other Bacteria) (Supplemental Data Table S1). Of organisms with Dus2 and another of a different Dus family subgroup, they were more likely to also have a Dus3 homolog than any other Dus subfamily member (Supplemental Data Table S1).

To better understand the within-family diversity of Dus2, domain architectures and fusions were collected from various functional annotation databases. Using a Dus2 control

sequence of *S. pombe*, a BLAST query of OrthoMCL was used to determine a precise Dus2 orthologous group (OG6_102617; 527 sequences). Subsequently, the InterPro and Pfam databases were used to acquire any additional unique architectural variants of Dus2, all of which were confirmed, individually, as proper subfamily members by using the four *S. pombe* control sequences (Dus1–4) in alignments and sequence trees (Supplemental Figure S2). Ultimately, these sequences were concatenated into a Dus2 master list of 390 sequences (Supplemental Data Table S2). Using CDD batch search, all sequences were assigned a name referring to the specific hit domains present. These fusion "names" were then used in the binary determination of absence or presence for unique fusions per organism (Figure 3).

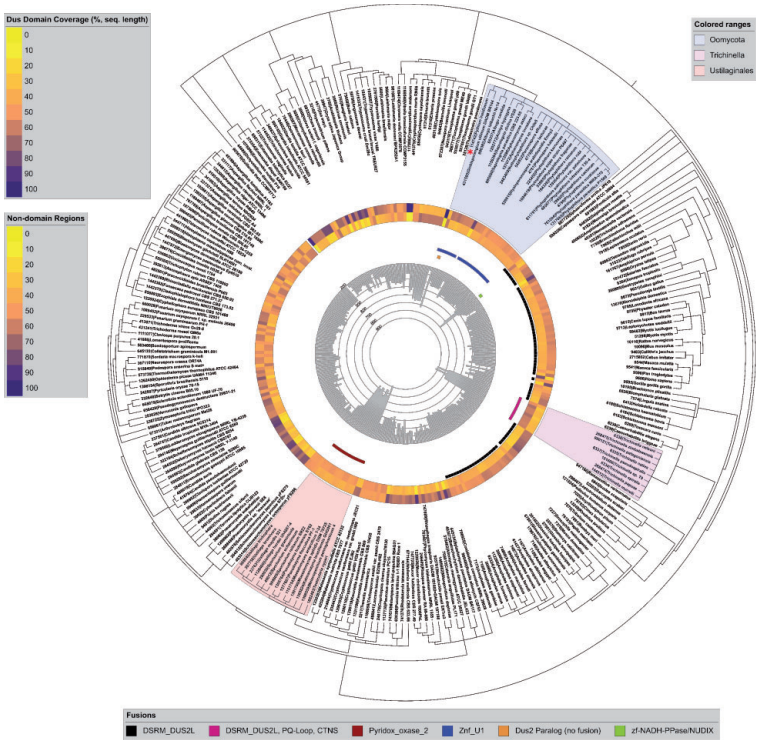


Figure 3. Phylogenetic tree of 280 Dus2 family members illustrating domain architecture diversity. Moving from the inside of the figure to the outside: (1) Sequence lengths are denoted by the central histogram; (2) Colored bars (one square per organism) indicate the presence of distinct fusion proteins (note: given multiple squares for one organism, they indicate the instance of separately encoded homologs); (3) Percent coverage of sequences for both the remaining lengths not identified as belonging to any recognizable domain (inside) and the identified Dus domain (outside) are shown. A red asterisk among the labels of the tree's leaves denotes the zinc finger domain-containing homolog of *Pythium insidiosum* that, in addition to ZnD_U1, also contains a separate ICL_KPHMT fusion domain.

The sequence lengths, domain lengths, and occurrence of N- and C-termini (binary) were acquired (Supplemental Data Table S2). Across all Dus2 sequences, the average sequence size was found to be 379 amino acids in length, with a maximum length of 793 aa and a minimum of 154 aa (Supplemental Data Table S2; Figure 3). In examining the diversity within the Dus2 subfamily, six unique domains were found to be fused with Dus2 domains (Figure 3): ZnF_U1 (smart00451, PSSMID 197732), ZnF_U1 + ICL_KPHMT (cd00377, PSSMID 119340), DSRM_DUS2L (cd19871, PSSMID 380700), zf-NADH/PPase/NUDIX (PF09297, PSSMID 401294), DSRM_DUS2L + [PQ-loop(x2)] + CTNS (PF04193, PSSMID 398045; smart00679, PSSMID 128923), and Pyridox_oxase_2 (PF12766; PSSMID 403846). The first two are combined in labeling within Figure 3, as the second fusion of the two was only found in a single sequence. Fusion domains, if present, were always found to occur within the C-terminal region of each sequence. With such notable levels of diversity, it was necessary to perform further structural analyses of Dus2 hybrid proteins.

3.2. Structural Analysis of the Dus2 Canonical Domains

Of the 30 structural models of Dus2 calculated along the evolutionary tree from bikonts to animals, all proteins analyzed retained the canonical folding, i.e., the TIM-barrel on the N-terminal side followed by the HD domain (Figures 4 and S3). The per-residue confidence score (pLDDT) values produced by AlphaFold2 for these two domains are very high, generally above 90, indicating strong confidence in the structural patterns of these regions. In addition to these domains, there are N-terminal and C-terminal extensions of proteins that we will discuss in the following section (see below).

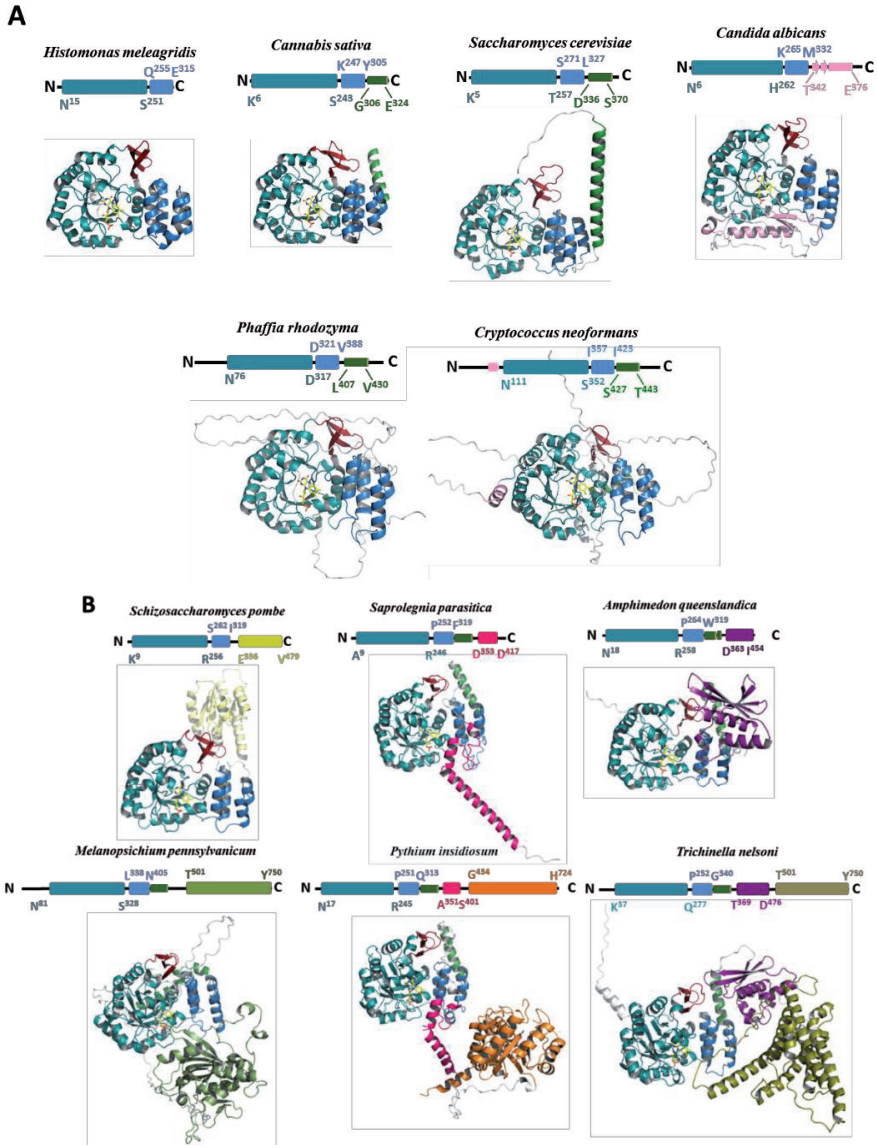


Figure 4. 3D structural models of various Dus2. (**A**) Models of Dus2 showing minimal modularity. The TIM-barrel domain (TBD) appears in teal, while the inserted beta-sheet is in red, the helical domain (HD) is in blue, the connective alpha-helix (c-αH) is in green, and the other structural extensions are in pink. The FMN coenzyme is denoted in yellow. (**B**) Models of Dus2 showing complex modularity with the addition of an extra domain. The same color codes for the canonical domains are followed (TBD + inserted beta sheet + HD). Rossman, zinc finger, dsRBD, PyrOX_2, ICL_KPHMT, and CTNS domains are in yellow, pink, purple, light green, orange, and olive, respectively. Above each of model is represented the schematic modular organization of Dus2 and is indicated the boundary of each domain. We have chosen not to show the delineation of the inserted beta-sheet in order to avoid figure overload. However, this structural element is colored in red in each of the 3D models presented.

3.2.1. Catalytic Domain (TIM-Barrel)

When the AlphaFold models are structurally aligned with the 4XP7 or 4WFS crystallographic structures, the RMSD values are all less than one. For example, the RMSD values obtained for *Candida albicans* Dus2 versus 4XP7, *Cryptococcus neoformans*, *S. pombe*, *Aspergillus awamori*, *Fusarium oxysporum*, and *S. cerevisiae* are 0.886 (over 202 atoms), 0.855 (over 223 atoms), 0.801 (240 atoms), 0.783 (231 atoms), 0.713 (227 atoms), and 0.702 (202 atoms), respectively (Figure S4). The N-terminal region of Dus2 consists of a conserved α11/β11 TIM-barrel fold, in which a central barrel composed of eight parallel beta strands is surrounded by 11 alpha helices (Figures 2B and 4). This structural arrangement is partly reminiscent of that found in flavoproteins such as dihydroorotate dehydrogenase and domain IV of dihydropyrimidine dehydrogenase, enzymes catalyzing comparable reactions [49]. As expected for α/β-barrel flavoenzymes, the central barrel offers a cavity for the binding of the redox coenzyme, FMN. The TIM-barrel core diverges from the classical α8/β8 fold because three additional β-strands insert at the N-terminal region of the TIM-barrel forming a new antiparallel β-sheet (β3-β5) (Figures 2B and 4), which seems strictly conserved among the Dus2 family. For example, this β-sheet inserts between C46 and Y76 in *S. cerevisiae*, between S46 and H77 in *C. albicans*, and between T49 and H76 in *S. pombe*. This structural element was first reported in hDus2 (Figure 2B) [25]. However, it was thought to be a peculiarity of hDus2 because it was not present in the crystallographic structures of DusA, DusB, and DusC (Figure 2A). Thus, we can clearly identify this β-sheet as a distinguishing feature of the canonical Dus2 fold. Note that this distinctive additional β sheet has not been predicted by the previous homology model of Dus2 [20].

3.2.2. Active Site

Since AlphaFold was recognized to predict all-atom accuracy of 1.5 Å rmsd95 compared to experimentally determined structures meaning that it can also produce highly accurate side chains [50], we decided to examine Dus2 active sites. The models produced were carried out on the apoprotein, although Dus2 is a flavoenzyme that uses FMN as a prosthetic group. In order to examine the active site in more detail, we superimposed the calculated models on the 4XP7 or 4WFS crystallographic structure. It is interesting to note that, first of all, none of the side chains of the models clash with the FMN present in the crystallographic structure of hDus2. Moreover, most of the conserved residues adopt a conformation similar to those observed in 4XP7. Thus, the analysis of the active sites does not seem to be biased by the fact that one is examining an apoprotein model. The resulting structural superpositions for three Dus2 are shown in Figure S5. As expected, the FMN cofactor lies inside a deep, positively charged crevice, stabilizing the negative charge of the phosphate moiety of the cofactor. All the constituting parts of the FMN, i.e., the isoalloxazine ring and the ribityl phosphate chain, make extensive interactions with surrounding amino acid residues, ensuring a tight binding and a proper orientation of the entire coenzyme. The strictly conserved Met (M19 in hDus2, M13 in *S. cerevisiae* and *C. albicans*, and M15 in *S. pombe*) stacks against the re-face of the isoalloxazine ring and is placed at ∼4 Å from the FMN C8 carbon. The side chains of three extremely conserved residues, a Q, N, and K (Q87/N113/K155 in hDus2, Q88/N114/K160 in *S. cerevisiae*, and Q87/N113/K159 in *C. albicans* and *S. pombe*, Supplemental Figure S6), interact with the pyrimidine moiety of FMN. The specifically positively charged lysine residue in this triad is likely to stabilize the negative density of $FMNH^-$ (Figure S7) obtained upon flavin reduction by NADPH. A 13-amino-acid loop spanning residues 116 to 128 and acting as a lid on the active site and inserted in the TIM-barrel between β6 and the small α5 helix is disordered in the crystal structures of hDus2. This loop contains the highly conserved C116 for hDus2, which has been proposed to function as a proton exchange site [2,18,22]. The corresponding loop gets ordered in the presence of tRNA [22]. Interestingly, in all the models, this active site loop appeared structured, forming a short eight-residue alpha-helix. While this cysteine is oriented away from the active site in the crystallographic structures of hDus2, in the models, this key residue (C116 in *S. cerevisiae*, *S. pombe*, and C117 in *C. albicans*) faces the flavin by being positioned above the isoalloxazine.

3.2.3. Helical Domain

The TIM-barrel is connected to the HD by a short linker whose length varies within the Dus2 family. For example, this linker is composed of thirteen residues in *S. cerevisiae* (T257 to D268), *H. meleagridis* (S251 to K256), *S. pombe* (R256 to S262), six in *H. sapiens* (L253 to R258) based on X-ray crystallography, and in *S. parasitica* (R246 to L251), whereas in *C. sativa*, it is composed of five amino acids (S243-K247). The helical subdomain is formed by a bundle of four helices, a feature that also appears to be conserved in bacterial Dus [19,22,26,27]. The TIM-barrel and HD domains of all Dus2 share a predominantly hydrophobic interaction surface but stabilized by additional electrostatic contacts, including hydrogen bonds and π-cationic interactions as observed in the crystallographic structure of the canonical domains of Dus2. The aromatic residues present in the interface may participate in the orientation of the HD relative to the catalytic domain. The relative orientation of these two domains is globally conserved among Dus2, as shown by the structural alignments between the 4XP7 and Dus2 models (Figure S4). However, this orientation differs from that of DusA, which is the bacterial homolog that, like Dus2, catalyzes D20 formation (Figure 2A). The presence of the additional β-sheet (β3-β5) that is inserted into the TIM-Barrel of Dus2, increases the surface area of interaction between the TIM-Barrel and the HD compared to that of bacterial Dus. On the basis of the crystallographic structures of hDus2, the increase in this interface was estimated to be 15% of the surface area [25]. Although globally the HD is well conserved in Dus2, we noted that the loop connecting helix 2 to helix 3 may exhibit size variability. While this loop consists of 5 residues in humans (R290-E294), it extends to 11 residues in *S. parastica,* for example, making it 16 residues long (S284-D299).

3.3. Structural Extensions of Dus2

Beyond the canonical Dus architecture, many Dus2s carry N- and C-terminal extensions or only an N-terminal or C-terminal extension (Figure 4). Some Dus2s such as *H. meleagridis*, have no extensions, suggesting that there are Dus2s that can obviously function only with the TIM-barrel and HD domains like in the bacterial Dus. In Dus2, extensions can be classified into two types: (i) unstructured sequence additions; (ii) minimal order structural extensions such as the addition of a helix, such as the connecting α-helix (c-αH) placed just after the HD (see below) or other structural elements that do not constitute a protein domain.

3.3.1. Unstructured Extensions

Structureless extensions can be observed at both ends of the protein (Figure 4A,B). The size of these extensions can vary from a few amino acids to much longer lengths, as is the case, for example, with Dus2 from *C. neoformans,* which has an unstructured extension of 80 amino acids at the N-terminus (M1 to S81) but also another at the C-terminus of about 50 amino acids (R445 to S494) (Figure 4A). All the analyzed unstructured extensions showed pLDDT values <50. The five models proposed by AlphaFold (ranked by score) for each of the Dus2, present extensions that adopt several positions in space. The lack of structure could be in agreement with a sampling of the conformations, as seen with the intrinsically disordered proteins (IDP), or it could be that AlphaFold is not able to predict a structure for these regions because it did not find interchain contacts due to its intrinsic limitation. However, the first explanation appears to be the most convincing. This is supported by structural evidence obtained from the C-terminal region of mouse Dus2. Indeed, a PDB of the solution structure of the isolated dsRBD domain of mouse Dus2 under the code 1WHN and annotated "Solution structure of the dsRBD from hypothetical protein BAB26260" is available [2]. In this NMR models, the long C-terminal extension is disordered likely due to the lack of restraints, a consequence of the intrinsic flexibility of this region. Hence, low-confidence residues may be explained by some form of disorder, although one should be cautious about it. Indeed, IDPs are common in the proteomes of eukaryotes, and a study estimated that the percentage of disordered residues in the human proteome is between 37% and 50%. These disorder predictions could also encompass both regions that are intrinsically disordered and regions that are structured only in complexes with cellular partners [51].

3.3.2. Structured Extensions and the Connecting α-Helix

Some structured extensions show good pLDDT scores. The HD is very often followed by an additional long alpha helix, as in the case of 27 out of 30 generated models, including S. *cerevisiae* for example, suggesting the important conservation of this specific structural entity of Dus2 (Figure 4). Among the thirty Dus2 models generated, only *H. meleagridis*, *C. albicans*, and *S. pombe* Dus2 lack this helix. This helix can be fragmented into two helices as in Dus2 of *F. oxysporum* and *A. awamori,* for example. We propose here to name this helix the connecting helix (c-αH), as it connects the additional domains (see below) to the canonical domains. We and Antson's group had observed the presence of this helix in hDus2, which is absent in bacterial Dus, but it was not clear whether this was specific to the human enzyme [25,28]. It can now be stated that c-αH is a characteristic feature of Dus2 proteins. The overall good results of the *S. cerevisiae* model, including the c-αH, allow us to analyze this extension in more detail. This helix begins with residue D336 and ends with S370 (Figure 4A). Indeed, the prediction of the c-αH of about 30 residues is robust, as shown by the scores (>90 between P337-N353, between 90 and 70 for residues from A354 to Q364) (Figure S8). C-αH is attached to the HD by a linker of ten amino acids. In a sense, this alpha helix extends the HD on the opposite side of the TIM-barrel. The orientation of c-αH is maintained by ionic and hydrophobic interactions engaging its N-terminal region with certain residues of helices-3 and -4 of the HD. C-αH is followed by fourteen unstructured amino acids (S370-I384), with I384 being the last residue of the protein.

Beyond the c-αH, other structural extensions may exist. For instance, the extension observed in *C. albicans* Dus2 is very unique because it includes several novel structural elements and does not carry the c-αH (Figure 4A). First, the HD helix-4 is elongated by thirteen residues (K320-Q331) projecting down the domain, in the opposite direction to the c-αH observed in *S. cerevisiae* Dus2. This helix is attached to a beta-sheet formed by 2 strands by a linker. This beta-sheet is itself linked to a long alpha helix of 23 amino acids. These novel structural features extend the TIM-barrel on the proximal side and are stabilized by numerous interactions between the TIM-barrel and the HD domains. As with *S. cerevisiae* Dus2, the structural extension ends with an unstructured region of approximately fifty residues. Similarly, hDus2 ends with an unstructured region, which appears to be often the case in Dus2 enzymes.

3.4. Modularity of Dus2

Beyond simple structural extensions, some Dus2 have an entire domain appended to the canonical domain. Our phylogenetic analysis identified five domains present in Dus2 (Figure 3). However, we were able to identify an additional domain by modeling Dus2 with sizes larger than those of a protein containing only the canonical domains (between 300 and 320 amino acids), a domain that had not been picked up by phylogenetic analysis alone (Figure 4B, case of *S. pombe*). All these domains are always added after the HD, and more particularly after the c-αH when this helix is present. Although for the majority of Dus only the addition of a single domain is observed, there are very few cases where we find the grafting of two additional domains to the two existing canonical domains, as is the case, for example, for *P. insidiosum* or *Trichinella nelsoni* (Figure 4B). These added domains are generally attached to the rest of the protein by long, flexible linkers that provide these modules with a large degree of freedom (Figure 5). This is perfectly illustrated on the different models generated because, for a given Dus2, the positioning of these domains is generally not conserved and thus orients differently from one model to the next due to the absence of inter-domain contacts found by AlphaFold2 (Figure 5). Aside from the zinc finger domain (ZnFD) and the double stranded RNA binding domain (dsRBD), which are well known for their ability to bind dsRNAs, the involvement of the other domains in RNA and nucleic acid metabolism in general has not been documented. It should be noted that we were not able to obtain a reliable model of the nudix domain due to the very low scores.

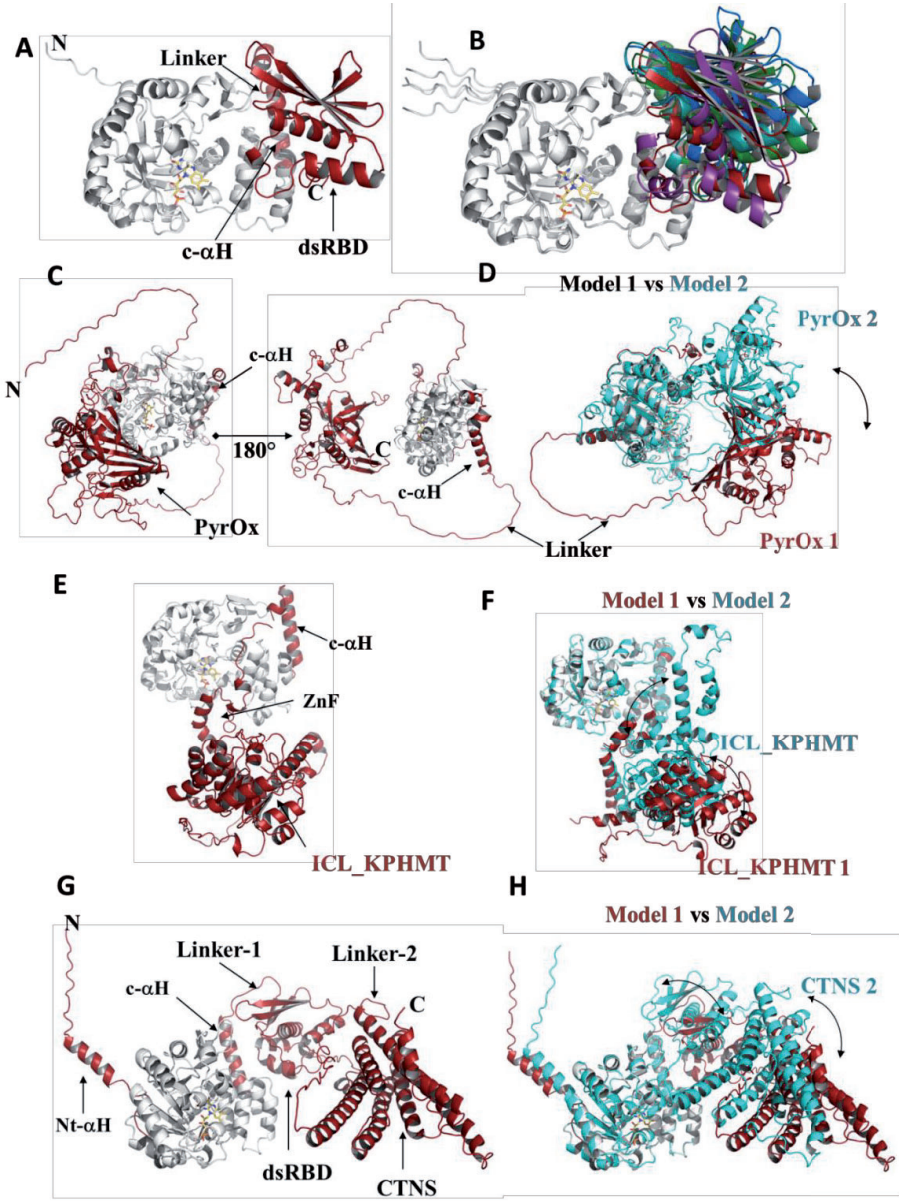


Figure 5. Modularity and domain orientation in Dus2. (**A**) Structural 3D model of *A. queenslandica* Dus2. The canonical domains are in gray while the c-αH and the dsRBD are in red. The linker attaching the c-αH to the dsRBD is indicated. (**B**) Structural superposition of the 5 *A. queenslandica* Dus2 model generated by AlphaFold2. The TIM-barrel and the HD of each model are in gray while the dsRBD are colored in a different color for each model. (**C**) Structural 3D model of *M. pennsylcanicum* Dus2 showing to different view that differs by a rotation of 180° around the z-axis. The canonical domains are in gray while the c-αH and PyrOx domains are in red. (**D**) Structural superposition of models 1 and 2 of *M. pennsylcanicum* Dus2 generated by AlphaFold2. Model 1 is colored as indicated in (**C**) while model 2 is in cyan. The double-headed arrow indicates the different orientation of the PyrOx domain in the two models. (**E**) Structural 3D model of *P. insidiosum* Dus2. The canonical domains are in gray while the c-αH, the ZnFD, and the ICL-KPHMT domains are in red. (**F**) Structural superposition of models 1 and 2 of *P. insidiosum* Dus2 generated by AlphaFold2. Model 1 is colored as indicated in (**E**) while model 2 is in cyan. The double-headed arrows indicate the difference in orientation of the ZnFD and ICL-KPHMT domains in the two models. (**G**) Structural 3D model of *T. nelsoni* Dus2. The canonical domains are in gray while the c-αH, the dsRBD, and the CTNS domains are in red. (**H**) Structural superposition of models 1 and 2 of *T. nelsoni* Dus2 generated by AlphaFold2. Model 1 is colored as indicated in (**G**) while model 2 is in cyan. The double-headed arrows indicate the respective difference in orientation of the dsRBD and CTNS domains in the two models.

3.4.1. Zinc-Finger Domain

The ZnFD follows the c-αH, the latter acting as a connector between the HD and the ZnFD. The size of this domain is globally conserved; e.g., *S. parasitica* has a ZnFD of about 64 amino acids in length. In all cases, this domain serves as the C-terminal domain, except in the case of *P. insidiosum*, where the ZnFD is followed by an additional ICL-KHMPT-like domain (Figure 4B). The ZnFD of Dus2 consists of a 3–10 helix followed by a 2-stranded beta-sheet that takes the form of a finger, and then comes a first helix (H1) with two turns, a kink, and a second larger helix (H2) (for example, nine turns for the case of *S. parasitica*). The positioning of this ZnFD relative to the rest of the protein can vary between Dus2 models. To illustrate this, we present the Dus2 of *A. candida* and *S. parasitica*. As shown in Figure 6, the *A. candida* ZnFD is facing the c-αH extension (Figure 6A), whereas in the case of *S. parasitica*, the ZnFD is positioned in front of the TIM barrel and the HD (Figure 6B). It seems therefore difficult to determine the exact location of the domain, however it is quite possible to imagine that this ZnFD may have a degree of freedom allowing it to move and adapt its position in particular in the presence of the RNA substrate thanks to a long linker of about 15 amino acids that attaches this domain to the HD. The ability to move is a known property of ZnFDs. The ZnF motif of Dus2 is of the CX2CX12HX5H type, part of the large family of C2H2 class zinc fingers, the most commonly used in transcription factors with the ability to bind preferentially to double-stranded RNA (dsRNA) [52,53].

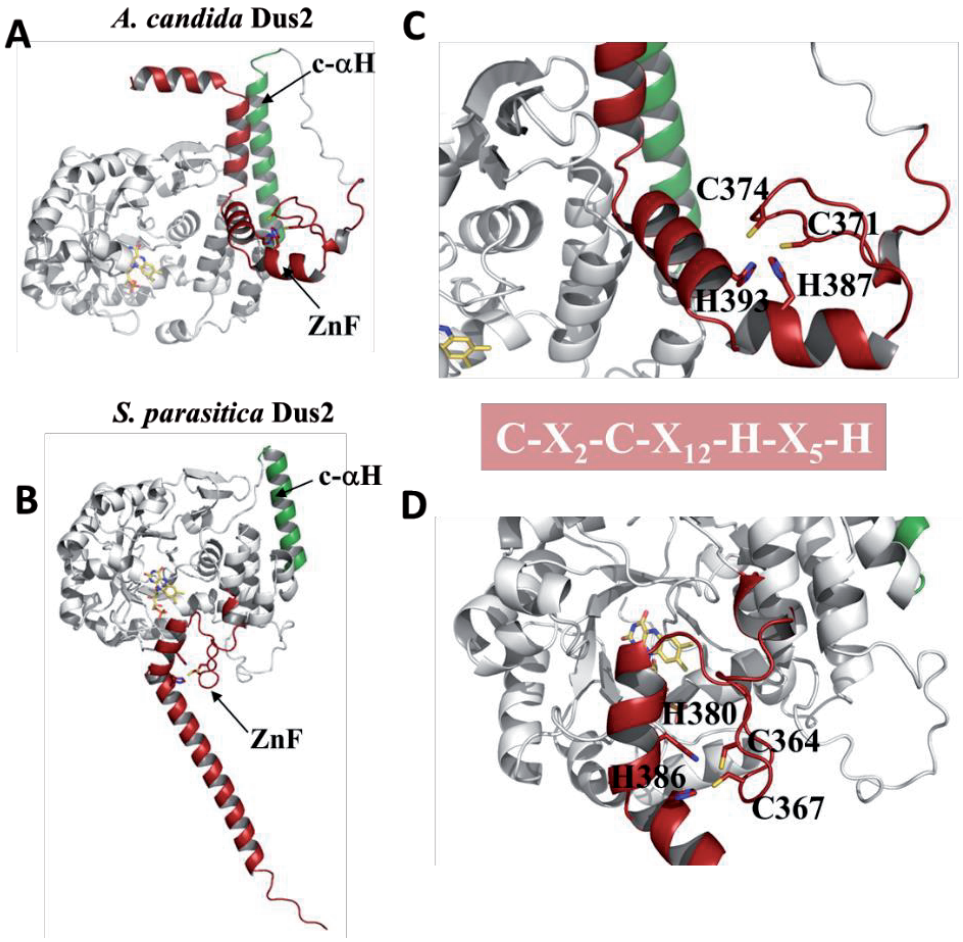


Figure 6. *Cont.*

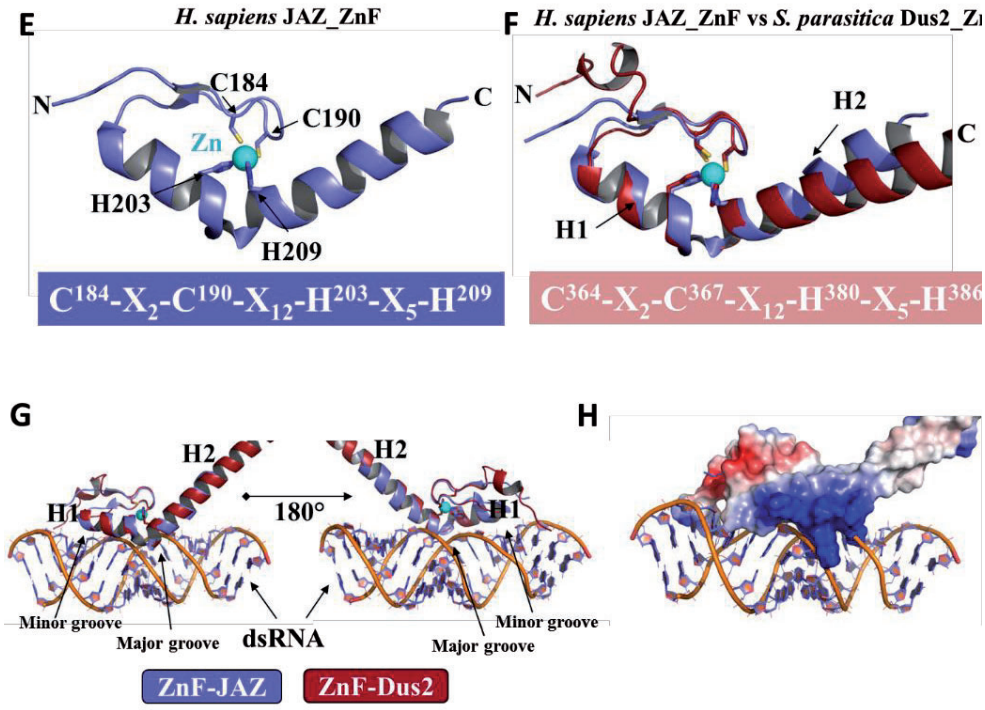


Figure 6. The Zinc finger domain in Dus2. (**A**) Structural 3D model of *A. candida* Dus2. The canonical domains are in gray while the c-αH and the ZnFD are in green and red, respectively. (**B**) Structural 3D model of *S. parasitica* Dus2. (**C**,**D**) Zoom on the zinc finger motifs of *A. candida* and *S. parasitica* Dus2, respectively. The signature of the motif is indicated in the red box. (**E**) Solution NMR structure of a ZnFD of human JAZ protein (PDB: 2MKN). The zinc atom is represented as a ball colored in cyan. The ZnF motif for this domain is indicated below the structure. (**F**) Structural superposition of 2MKD (colored in purple) with the ZnFD of *S. parasitica* Dus2 (in red). The ZnF motif of *S. parasitica* Dus2 is shown below the figure of the structural alignment. (**G**) Structural superposition of 2MKN (PDB code for the structure of JAZ ZnFD in complex with a dsRNA with the ZnFD of *S. parasitica* Dus2 (in red). Two different views that differ by a rotation of 180° around the z-axis are shown. (**H**) Model of the *S. parastica* Dus2/dsRNA complex. The ZnFD is represented in the electrostatic surface mode.

The two cysteines of the motif are present in the central beta-sheet while the two histidines are found in each of the two helices, H1 and H2 (Figure 6C,D). Interestingly, the ZnFD of Dus2 shares a similar structure to the zinc fingers of the human Just Another Zinc finger (JAZ) protein [54] (Figure 6E). This four zinc finger protein is known to bind endogenous and exogenous dsRNAs such as adenoviral VAI RNA [54]. Indeed, superimposing the ZnFD of Dus2 *S. parasitica* on the NMR structure of JAZ ZnFD in the 2MKD PDB gives a very low RMSD (~0.696 Å over 25 Cα) consistent with the structural conservation of this domain in the 2 proteins (Figure 6F). Even more, the side chains of the ligands of the zinc atom, namely C2H2, adopt perfectly identical orientations to those of JAZ, suggesting that these residues are oriented in their functional form in the models generated by AlphaFold (Figure 6F). Structural alignment of the ZnFD of Dus2 *S. parasitica* with the NMR structure of the JAZ:dsRNA complex (PDB: 2MKN [54]) shows that the two proteins share the same dsRNA recognition mode (Figure 6G). Indeed, dsRNA binding occurs mainly via H1, kink, and H2. H1 and H2 recognize the minor grooves, while the kink and the N-terminus of H2 recognize the major grooves. Interactions between dsRNA and the ZnFD domain of Dus2 are largely driven by interactions of an electrostatic nature, as shown in Figure 5H. Specifically, H1 and H2 have charged residues that will generate a highly positively charged surface to accommodate phosphates on the RNA backbone. In JAZ, the dipole moment of H2 also contributes to this interaction [54].

The portion of the tRNA that will be recognized by the Dus2 ZnFD remains undetermined at this stage. However, it should be noted that the tRNA contains many double-

stranded, structured regions. In addition, there are RNA modification enzymes that also use ZnFDs as additional tRNA recognition modules. One example is Mod5, an isopentenyl transferase that catalyzes the formation of i6A37 in *S. cerevisae* tRNAs, and that in addition to its catalytic domain and insertion domain has a C2H2-type C-terminal ZnFD that recognizes the tRNA anticodon stem as shown in the crystal structure of Mod5:tRNACys [55] (Figure 7). Similar to JAZ, the ZnFD of Dus2 from *S. parasitica* overlaps with the ZnFD of Mod5 except for the central beta-sheet. However, despite this discrepancy, the C2H2 residues of Dus2 and Mod5 overlap perfectly and adopt an identical orientation with respect to the zinc atom. Furthermore, no clash was detected between the ZnFD of Dus2 and tRNA.

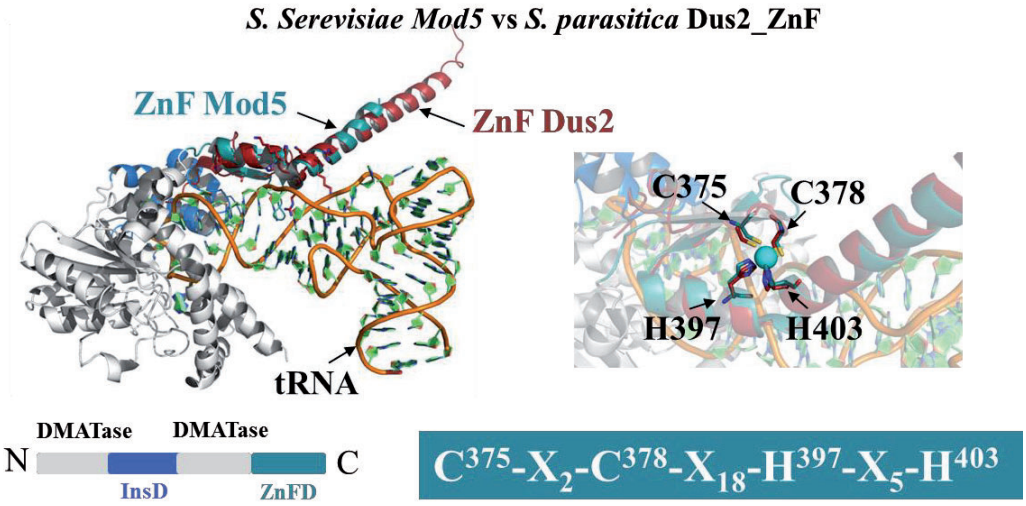


Figure 7. Structural comparison of the ZnFD of Dus2 and that of *S. cerevisiae* Mod5. ZnFD of *S. parasitica* Dus2 (in red) is superimposed on the ZnFD of Mod5 (in deep teal color) in the crystal structure of yeast Mod5/tRNACys (PDB: 3EPH). The below the figure is the schematic representation of the domain modularity of Mod5. The catalytic, inserted, and ZnFD of Mod5 are in gray, blue, and deep teal, respectively. A zoom of the superposition on the ZnF motif region is shown on the right. The ZnF motif of Mod5 is indicated below the zoom in the deep teal colored box.

3.4.2. The Double-Stranded Binding Domain

The other domain acquired by more evolved Dus2 is the dsRBD (Figures 4B and 5A). Initially observed in human Dus2, a phylogenetic analysis showed that the dsRBD is indeed more widely distributed in nature than expected since it is present in Dus2 of animals [20]. However, where exactly this additional domain appeared in the phylogenetic tree of Dus2 remains enigmatic. By reanalyzing the sequences of eukaryotic Dus2, we found out that the dsRBD does not seem to be restricted to animals but is also present in Choanoflagellates and in Filasterea, all of which are part of the Filozoa clade. A more rigorous search of the sequences allowed us to trace the phylogenetic tree and show that an Ichthyosporea, *S. arctica*, and a Cristidiscoidea, *F. alba* also possess a Dus2 with a dsRBD. We produced seven models of Dus2 carrying dsRBDs, two in animals (hDus2 and Dus2 from *A. queenslandica*), two Choanoflagellates, one Filastera, one Ichthyosporea and one Fonticuli. Interestingly, the genome of *C. owczarzaki* was found to have two paralog fusions: one containing the dsRBD domain and the other containing the aforementioned zf-NADH/PPase/NUDIX (PF09297, PSSMID 401294) domain. The former, typically around 68 amino acids, is well-known for its functional versatility by means of a particular α1-β1β2β3-α2 canonical structure that allows the recognition of a variety of simple RNA structures ranging from A-form RNA helices to hairpins or tetraloops in shape-dependent manners [56–58], even though a sequence-specific mode of recognition has been invoked for a few of them [59,60]. We showed that the dsRBD of hDus2 has an additional extension at the N-terminal, named NTE, which is also involved

in the binding of dsRNA and specifically the acceptor-TΨC stem region [29], which is the longest dsRNA region of the tRNA.

The models obtained for the various full lengths of Dus2 show that the dsRBD is attached to the HD indirectly via a c-αH like that identified in *S. cerevisiae* Dus2. However, a notable difference lies in the fact that in dsRBDs-containing Dus2, c-αH protrudes outside the plane formed by the HD domain so that it points towards the back of the protein. This difference in orientation could be concomitant with the acquisition of the dRBD in order to properly position the dsRBD. The orientation of the dsRBD relative to the rest of the protein is not conserved within the five models proposed by AlphaFold2 (Figure 5B). We had previously shown that in hDus2, the dsRBD is connected to c-αH by a flexible linker whose role could be the adjustment of the dsRBD functional position in the presence of the tRNA substrate [25].

A focus on specifically the dsRBD revealed that the overall structure is conserved among all the models. To determine if the dsRBDs of animals at the bottom of the evolutionary tree retain the same double-stranded RNA recognition mechanism compared to the dsRBD of hDus2, we crystallized the dsRBD of Dus2 from *A. queenslandica*, *Monosiga brevicolis*, and *F. alba*. Unfortunately, only *A. quenslandica* dsRBD led to diffracting crystals that allowed structural resolution. Interestingly, the validity of these models is supported by the superposition of the crystallographic structure of *A. quenslandica* dsRBD, which we solved at 1.68 angstroms in this study, with the dsRBD coming from the model (RMSD = 0.426 over 78 atoms) (Figure 8A). It is equally amazing to see that even the orientations of the side chains observed in the crystal structure are globally preserved in the model (Figure 8A). To investigate if these newly identified dsRBDs present functional similarities with those of hDus2, we structurally aligned them with the crystal structure 5OC6, which is the crystal structure of hDus2′s dsRBD (construct T339-K451) in complex with an eleven palindromic oligo-ribonucleotide that we have recently published (Figure 8B) [29]. Again, no major structural clashes can be detected between the dsRNA and the bound dsRBD. We showed that recognition of dsRNA is essentially achieved via three major canonical regions, namely helix-α1, helix-α2, and the C-terminal part of the β1–β2 loop of the dsRBD. These regions are all present in all dsRBDs analyzed here. In the case of hDus2, we showed that three residues of helix-α1 (T369, E376, and R379) interact with ribose 2′-OH groups in dsRNA's minor groove, while K371 of helix-α1 together with K419, K420, and Q424 located in the N-terminal extremity of helix α2 recognize exclusively the phosphodiester backbone of the major groove. The C-terminal part of the β1–β2 loop binds to the minor groove via R397, which makes hydrogen bonds with both the ribose and a nucleobase. These interactions seem to be conserved in all dsRBDs. We noted that instead of M371 in the human dsRBD at dsRNA recognition region 1, in some dsRBDs like those of *S. rosetta* (E426) and *F. alba* (Q410), the hydrophobic residue is replaced by a hydrophilic residue capable of interacting with the dsRNA, providing an additional anchoring point to the dsRNA. Although the majority of dsRNA recognition is ensured by interactions involving the dsRBD canonical fold, we showed that in hDus2, two positively charged arginines (R360/R361) act synergistically to recognize the tRNA. These two positively charged residues are also present in *S. arctica* (R426/K427) and *S. rosetta*. In contrast, only one of these positively charged residues is observed in *F. alba* (R400) and *A. queenslandica* (K381). Collectively, all dsRBDs of Dus2 analyzed so far likely carry a dsRNA binding capacity via the cooperative action of both their canonical structure and NTE.

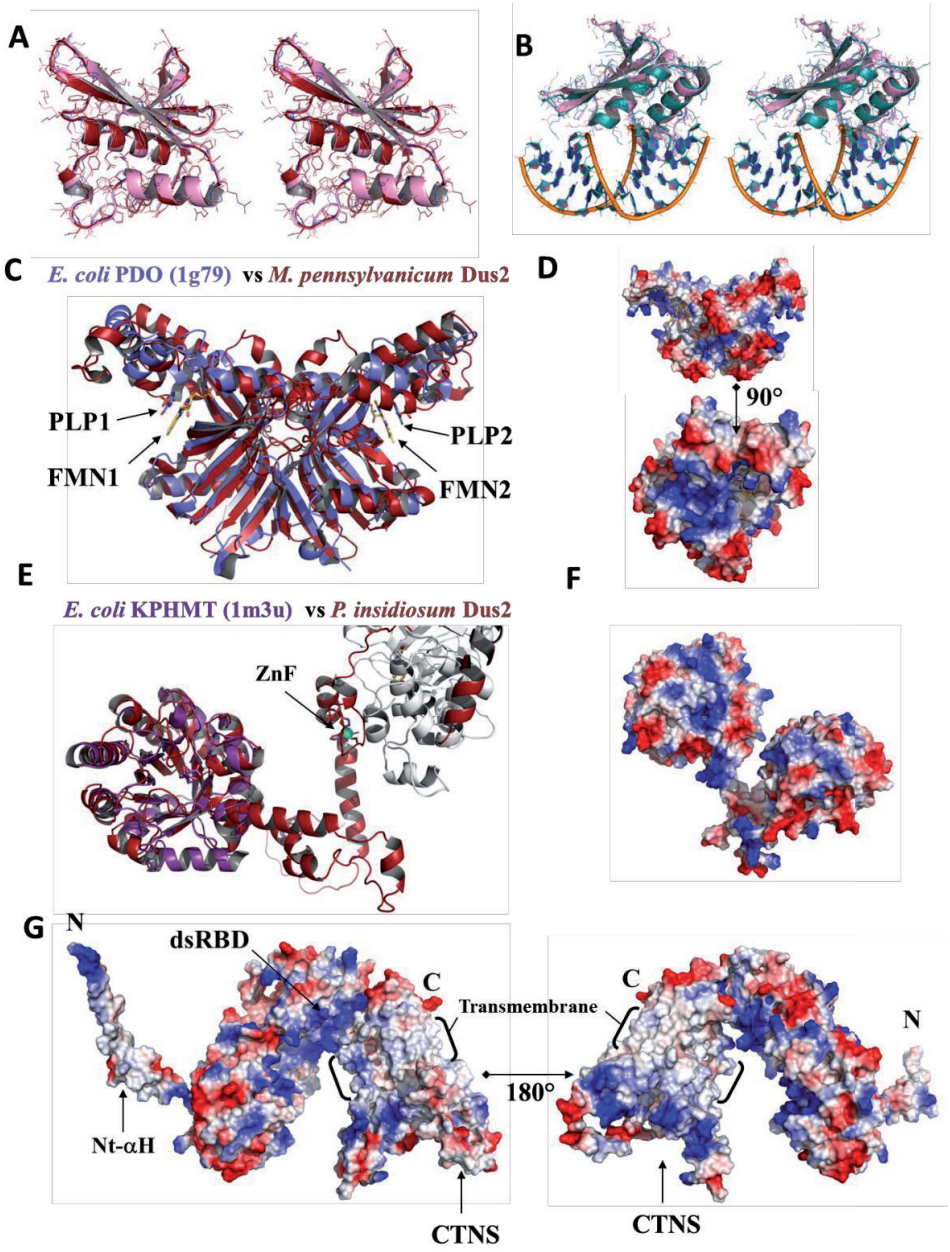


Figure 8. Structural analysis of various Dus2 domains. (**A**) Stereoview of the structural superimposition of the crystal structure of the dsRBD of *A. queenslandica* Dus2 (colored in pink) with the dsRBD obtained from the 3D model of *A. queenslandica* Dus2 (colored in red) generated by AlphaFold2. The side chains are shown as lines. (**B**) Stereoview of the structural superimposition of the crystal structure of the dsRBD of *A. queenslandica* Dus2 (colored in pink) with the crystal structure of the hDus2 dsRBD (colored in deep teal) in complex with a dsRNA (PDB: 5OC6). The backbone of the dsRNA is orange, while the nucleosides are deep teal. (**C**) Structural superimposition of the PyrOx domain of *M. pennsylvanicum* from the Dus2 model with the X-ray structure of the dimer of *E. coli* pyridoxine 5′-phosphate oxidase complexed with pyridoxal 5′-phosphate (PLP) and flavin mononucleotide (FMN) (PDB: 1G79). The dimer of PyrOx generates two equivalent active sites, each containing a PLP and an FMN. PyrOx of Dus2 and 1G79 are in red and purple, respectively. The FMN and PLP represented as sticks are in yellow and purple, respectively. (**D**) Electrostatic surface of PyrOx domain of *M. pennsylvanicum* Dus2. (**E**) Structural superimposition of the KPHMT domain of *P. insidiosum* from the Dus2 model (in red) with the crystal structure of ketopantoate hydroxymethyltransferase complexed the product ketopantoate (PDB: 1M3U, colored in violet). (**F**) Electrostatic surface of the KPHMT domain of *P. insidiosum* from the Dus2 model. (**G**) Electrostatic surface of Dus2 model from *T. nelsoni*.

3.4.3. Pyridoxamine 5′-Phosphate Oxidase Domain

This domain is typically that of an FMN flavoprotein that catalyzes the oxidation of pyridoxamine-5-P (PMP) and pyridoxine-5-P (PNP) to pyridoxal-5-P (PLP). This type of protein is involved in the last step of PLP cofactor anabolism. The structural alignment of the PyrOx domain of Dus2 from *P. pensylvanicum* with 1G79, the *E. coli* PyrOx shows a good conservation of the global folding (RMSD: 3.87 Å on 120 Cα), except for some loops whose length is more important in Dus2; however, the core, i.e., the central beta-sheet as well as the helices, are perfectly aligned (Figure 8C). The cleft that acts as an active site in PyrOx containing FMN and PNP is also present in the PyrOx domain of Dus2, however a difference distinguishes them. Although this cleft is large enough to accommodate both FMN and PNP in the PyrOx of Dus2, its hydrophobic nature does not allow the binding of FMN and PNP in Dus2 (Figure 8D). Indeed, the active site of the PyrOx is generally positively charged, which allows the stabilization of the phosphate groups present on the FMN coenzyme and PNP substrate. This polarity inversion raises serious questions about the exact role of this domain in Dus2. A more detailed analysis of this domain in Dus2 shows the existence of another large pocket above the hydrophobic crevice that is positively charged (Figure 8D). It remains to be seen whether this could serve as an RNA binding site, which would imply that the function of this domain has evolved to allow for RNA recognition, which does not seem incongruous given that the phosphates play a key role in enzyme/substrate recognition in RNA modifying enzymes. Although this remains speculative, it is an interesting hypothesis to test experimentally.

3.4.4. A Newly Identified Rossman Fold Domain

We have been able to identify for the first time that all the Dus2 of Schizosaccharomyces species, which are four in number (*S. pombe*, *S. cryophilus*, *S. octoporus,* and *S. japonicus*), carry after their HD an additional domain of about 145 amino acids (Figures 3 and 4). This domain is organized around a central β-sheet made up of four parallel β-strands surrounded by six alpha-helices. A quick analysis by Dali suggests that this domain adopts a Rossman fold type domain. Its functional role in RNA binding is unlikely since the analysis of the electrostatic surface does not delineate positive patches that are expected to be observed for RNA binding sites. At this stage, it is not possible to discard this possibility or a potential other role, such as its implication in a regulatory process, but this would require biochemical validation.

3.4.5. ICL_KPHMT and CTNS Domains

Members of the ICL/PEPM_KPHMT enzyme superfamily catalyze the formation and cleavage of either P-C or C-C bonds. Typical members are phosphoenolpyruvate mutase (PEPM), phosphonopyruvate hydrolase (PPH), carboxyPEP mutase (CPEP mutase), oxaloacetate hydrolase (OAH), isocitrate lyase (ICL), 2-methylisocitrate lyase (MICL), and ketopantoate hydroxymethyltransferase (KPHMT). In Dus2, this domain adopts a TIM-barrel, with 8 alpha helices surrounding the central barrel (Figure 8E). This domain is connected to the *P. insidiosum* ZnFD by a long linker of 34 amino acids (S401-I435). The terminal beta-sheet of the barrel gives way to a helix of about 20 amino acids that is followed by three other small helices, thus completing the sequence of the protein. On the basis of the electrostatic surface, it is difficult to predict any role for this domain (Figure 8F).

In *Trichinella nelsoni*, the dsRBD of Dus2 is followed by a CTNS domain, a cystine/H+ symporter known as a mediator in the export of cystine, the oxidized dimer of cysteine, from lysosomes. Importantly, no structure for such a domain has been reported in the PDB database. In the case of *T. nelsoni*, the CTNS is connected to the dsRBD by a 27 amino acid linker (N475-I501) (Figure 5G). Interestingly, the CTNS domain of Dus2 has a highly hydrophobic central ring with a width of more than 25 Å that serves as an anchoring zone across the membrane (Figure 8G). This means that Dus2 in these organisms is localized to the cell membrane or to the membrane of a cell organelle. This is, to date, the first case of an RNA-modifying enzyme that has a transmembrane domain. The canonical domains

of these Dus2 as well as the dsRBD are probably cytosolic to allow modification of RNAs. However, the precise localization of these Dus2 remains to be clarified. In humans, hDus2 has been localized in the endoplasmic reticulum [13].

3.5. The functional Role of Dus2 Extensions and the Evolution of tRNA Binding

To determine whether these extensions have functional relevance in D20 biosynthesis, we conducted a comparative structural study of Dus/tRNA complexes from three different organisms (Figures 9 and S9). The first one is *T. thermophilus* DusA, for which an X-ray crystallographic structure of a DusA/tRNA^{Phe} complex is available [22]. In this structure, DusA captures the tRNA in a positively charged crevice, where the HD contributes by providing most of the positive residues while others come from the catalytic domain (Figures 9A and S9A,B).

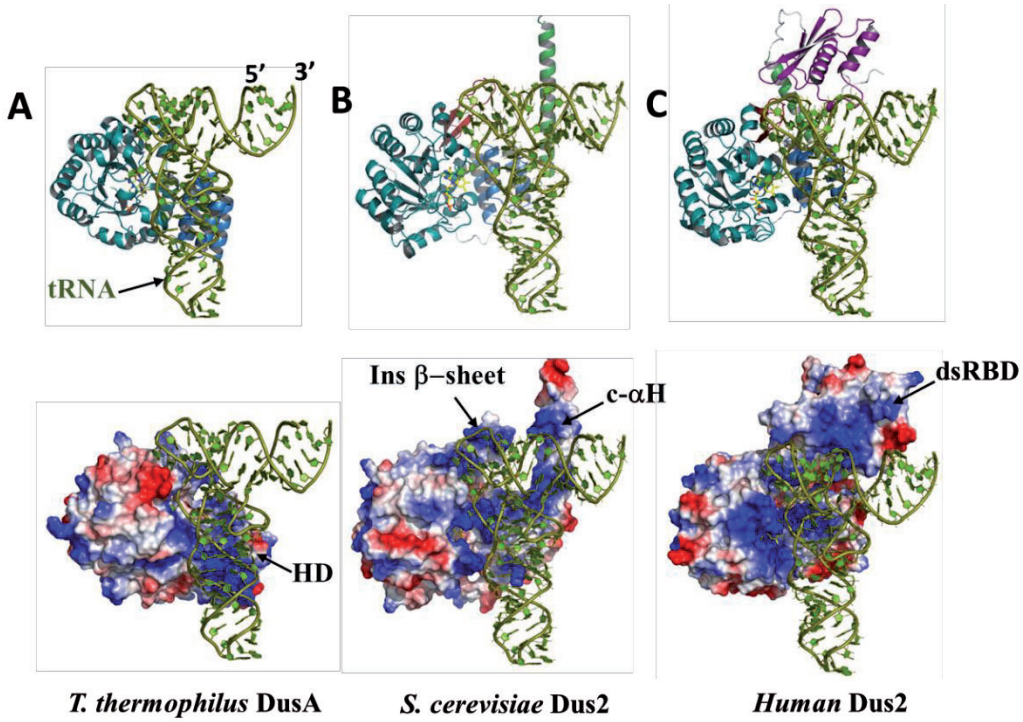


Figure 9. Evolution of tRNA binding mode in Dus enzymes catalyzing D20 biosynthesis. (**A**) Crystal structure of *T. thermophilus* DusA in complex with tRNA (PDB: 3B0V). (**B**,**C**) Structural models of *S. cerevisiae* Dus2/tRNA and hDus2/tRNA complexes, respectively. The TIM-barrel domain (TBD) appears in teal, while the helical domain (HD) is in blue, the inserted beta-sheet in red, the connecting c-αH in green, and the dsRBD in purple. The FMN coenzyme is denoted in yellow. The electrostatic surface of each Dus protein is shown below the protein or tRNA.

Interestingly, no major conformational changes are observed between the free and bound states of either the protein or the tRNA substrate. Nonetheless, slight distortions of U16, U17, and U20 are observed, as is a flipping of the latter base, which found itself buried in the active site pocket stacked on the si-face of the isoalloxazine. DusA recognizes almost the complete tRNA elbow region formed by the junction of the D- and T-loops, stabilized by critical tertiary interactions, and the enzyme flips the target base without unwinding this tRNA structure. The absence of a drastic conformational change during catalysis in this family of enzymes is also supported by two DusC/tRNA structures [27]. These observations indicate that flipping of the target uridine to enter the active site can occur without altering the tertiary structure of tRNA, and this could be the case for the different types of tRNA substrates. Instead, these enzymes, as is the case for many tRNA-modifying enzymes, make use of a complementary surface to charge and shape the tRNA target region. In light of this information, we set out to model a *S. severvisiae* Dus2/tRNA

complex. *T. thermophilus* $tRNA^{Phe}$ from the DusA/tRNA structure was chosen because it is already in a productive conformation and ready for modification.

In order to generate this model, we aligned the DusA/$tRNA^{Phe}$ structure with the *S. cerevisiae* Dus2 model. Remarkably, this alignment did not generate any major clashes; however, some slight manual readjustments were necessary. At first glance, *S. cerevisiae* Dus2 does not need to resort to some drastic conformational changes in its polypeptide to accommodate the tRNA (Figures 9B and S9C,D). The TIM-barrel in cooperation with the HD, which carries several positive patches in a large crevice formed by the junction of the two domains, provides the tRNA binding site. Again, Dus2 embraces much of the elbow, the D- and T-loops. Surprisingly, the c-αH, which carries positive charges for more than half of its length, appears to be perfectly positioned to recognize an additional tRNA region (Figures 9B and S9C,D), namely the back of the acceptor stem, whereas in the DusA/$tRNA^{Phe}$ complex, this region is entirely free and solvent accessible. Thus, in cooperation with the canonical domains, this extension acts as a new tRNA anchor, allowing a broader recognition surface than that engaged by DusA. Finally, we examined Dus2 with dsRBDs, taking the human enzyme as our preferred choice because we have previous experimental data that allowed us to delineate the exact areas of interactions on both the protein and tRNA [29]. By reproducing the same approach used for *S. cerevisiae* Dus2, we were able to generate a structural model of full-length hDus2 in complex with tRNA (Figures 9C and S9E,F). As for *S. cerevisiae*, a slight repositioning of the tRNA was required to remove the few collisions observed between the protein and tRNA. Interestingly, the resulting model is in close agreement with the previously proposed model using isolated domains of hDus2, which was inferred from mutagenesis, crystallography, NMR, and SAXS experiments [29]. The dsRBD recognizes a larger surface area of the tRNA acceptor region. Through the acquisition of this additional domain, hDus2 recognizes almost the entire tRNA molecule, except the anticodon region. Interestingly, c-αH no longer carries a positive surface charge since it is not involved in tRNA binding (Figure S9F). In a way, it is the dsRBD that compensates for this loss of charge compared to the yeast Dus2. Finally, it is worth mentioning that we have not been able to produce reliable models of the Dus2/tRNA complex with ZnFD-containing enzymes because this domain is not positioned in an orientation that prevents it from colliding with tRNA. However, it can be assumed that, as in all Dus, the canonical domains of these Dus2 should recognize the same tRNA regions as their counterparts. Under these conditions, ZnFD could, as with c-αH and dsRBD, recognize the acceptor stem and the T arm, both of which form the longest dsRNA region of the tRNA.

4. Conclusions

Artificial intelligence via AlphaFold has revolutionized structural biology due to the accuracy of the structural models generated by this methodology [50,51]. Combining this approach with phylogenetic analysis has proven to be an interesting strategy to study the evolutionary and functional features of enzyme systems. The application of this methodology to Dus2 from different organisms along the evolutionary tree and the analysis of the resulting structural models allowed us to unexpectedly discover that this enzyme presents a great deal of structural diversity through the presence of various extensions appended to the canonical fold that have an obvious functional relevance, at least for some of them. Analyses of enzyme/tRNA models perfectly illustrate the impact of the structural evolution of Dus on their tRNA recognition mode. What stands out from these models is that over the course of evolution, recognition of a larger surface area of tRNA by the Dus2 enzyme appears to have been deemed necessary for D20 biosynthesis. Although the analyses of protein surface electrostatics and modeling of Dus2/tRNA complexes suggest that some of these extensions are likely involved in RNA recognition, one can wonder the reason for such extensions from the standpoint of evolution. Recent publications have shown that eukaryotic Dus from *S. cerevisiae*, *S. pombe*, and humans can dihydrouridylate other substrates than tRNA, such as mRNAs, and long non-coding RNAs [4–6]. In bacteria,

there seems to be no D in mRNAs and therefore the bacterial Dus may probably have only tRNA as a substrate. Thus, the need for D in other types of RNAs and consequently the additional substrates recognized by Dus2 could partly explain the increased complexity of the protein's modularity compared to its bacterial counterpart. Another possibility could be an evolutionary shift towards a gain in stability of the Dus/tRNA complex during evolution, but this remains to be demonstrated experimentally.

In general, it is rare for tRNA-modifying enzymes catalyzing modifications targeting areas other than the anticodon to use this recognition mode. In contrast, enzymes that modify the anticodon use a tRNA recognition mechanism that involves large interaction surfaces [61]. This mode of recognition is also shared by the amino-acyl tRNA synthetases, which recognize almost the entire tRNA molecule. Moreover, this class of enzyme has seen its modularity increase in complexity during evolution by the decoration of additional domains, some for regulatory purposes [62]. More generally, there are many other examples of enzymes that gain modularity during evolution, with human proteins remaining the pinnacle of modularity complexity. However, in the case of tRNA-modifying enzymes catalyzing modifications in the tRNA body, the evolution of the structural diversity of Dus2 remains a unique example to our knowledge.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom12121760/s1, Table S1: List of Dus Family Subgoups; Table S2: Dus2 Fusions and Sequence Disorders; Figure S1: Example of Dus subfamily confirmation for Table S1 sequences; Figure S2: Example of Dus2 subfamily confirmation for Table S2 sequences; Figure S3: 3D-model obtained by AlphaFold2 of the Dus2 enzyme from different eukaryotic organisms; Figure S4: Structural alignment of AlphaFold models of selected Dus2 with the crystal structure of hDus2; Figure S5: Active sites alignment; Figure S6: Motifs conserved across Dus2 family members; Figure S7: Role of conserved residues interacting with the FMN; Figure S8: Predicted IDDT score per position for *S. cerevisiae* Dus2 models; Figure S9: Evolution of tRNA binding mode in Dus enzymes catalyzing D20 biosynthesis.

Author Contributions: M.L., V.d.C.-L. and D.H. conceptualized, supervised, and managed the project; M.L. and D.H. performed the 3D models; C.J.R. and V.d.C.-L. performed the phylogeny; B.F. cloned, expressed, purified, and crystalized the dsRBDaq; L.P. collected the synchrotron diffraction data and resolved the X-ray crystal structure of dsRBDaq. M.L., V.d.C.-L., L.P, B.F., S.T., C.S., D.B. and D.H. analyzed the data. All the authors wrote, edited, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Centre National de la Recherche Scientifique, the university Pierre et Marie Curie, the French State Program 'Investissements d'Avenir' (Grants ANR-DFG, 20-CE92-0030, to D.H and D.B) as well as the by National Institutes of Health (Grant R01 GM132254 to V.d.C.-L.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw data collection and refinement for the X-ray structure of dsRBD (PDB: 3B02) are available in the PDB.

Acknowledgments: We also acknowledge SOLEIL for the provision of synchrotron radiation facilities (proposal ID 20150780), and we would like to thank Proxima 2 for assistance in using the beamline.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Finet, O.; Yague-Sanz, C.; Marchand, F.; Hermand, D. The Dihydrouridine landscape from tRNA to mRNA: A perspective on synthesis, structural impact and function. *RNA Biol.* **2022**, *19*, 735–750. [CrossRef] [PubMed]
2. Bregeon, D.; Pecqueur, L.; Toubdji, S.; Sudol, C.; Lombard, M.; Fontecave, M.; de Crecy-Lagard, V.; Motorin, Y.; Helm, M.; Hamdane, D. Dihydrouridine in the Transcriptome: New Life for This Ancient RNA Chemical Modification. *ACS Chem. Biol.* **2022**, *17*, 1638–1657. [CrossRef] [PubMed]

3. Boccaletto, P.; Stefaniak, F.; Ray, A.; Cappannini, A.; Mukherjee, S.; Purta, E.; Kurkowska, M.; Shirvanizadeh, N.; Destefanis, E.; Groza, P.; et al. MODOMICS: A database of RNA modification pathways. 2021 update. *Nucleic Acids Res.* **2022**, *50*, D231–D235. [CrossRef] [PubMed]
4. Draycott, A.S.; Schaening-Burgos, C.; Rojas-Duran, M.F.; Wilson, L.; Scharfen, L.; Neugebauer, K.M.; Nachtergaele, S.; Gilbert, W.V. Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA. *PLoS Biol.* **2022**, *20*, e3001622. [CrossRef]
5. Dai, W.; Li, A.; Yu, N.J.; Nguyen, T.; Leach, R.W.; Wuhr, M.; Kleiner, R.E. Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat. Chem. Biol.* **2021**, *17*, 1178–1187. [CrossRef]
6. Finet, O.; Yague-Sanz, C.; Kruger, L.K.; Tran, P.; Migeot, V.; Louski, M.; Nevers, A.; Rougemaille, M.; Sun, J.; Ernst, F.G.M.; et al. Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Mol. Cell* **2022**, *82*, 404–419.e9. [CrossRef]
7. Kim, S.H.; Suddath, F.L.; Quigley, G.J.; McPherson, A.; Sussman, J.L.; Wang, A.H.; Seeman, N.C.; Rich, A. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **1974**, *185*, 435–440. [CrossRef]
8. Suddath, F.L.; Quigley, G.J.; McPherson, A.; Sneden, D.; Kim, J.J.; Kim, S.H.; Rich, A. Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0angstroms resolution. *Nature* **1974**, *248*, 20–24. [CrossRef]
9. Dalluge, J.J.; Hashizume, T.; Sopchik, A.E.; McCloskey, J.A.; Davis, D.R. Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Res.* **1996**, *24*, 1073–1079. [CrossRef]
10. Dyubankova, N.; Sochacka, E.; Kraszewska, K.; Nawrot, B.; Herdewijn, P.; Lescrinier, E. Contribution of dihydrouridine in folding of the D-arm in tRNA. *Org. Biomol. Chem.* **2015**, *13*, 4960–4966. [CrossRef]
11. Alexandrov, A.; Chernyakov, I.; Gu, W.; Hiley, S.L.; Hughes, T.R.; Grayhack, E.J.; Phizicky, E.M. Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* **2006**, *21*, 87–96. [CrossRef] [PubMed]
12. Kuchino, Y.; Borek, E. Tumour-specific phenylalanine tRNA contains two supernumerary methylated bases. *Nature* **1978**, *271*, 126–129. [CrossRef] [PubMed]
13. Kato, T.; Daigo, Y.; Hayama, S.; Ishikawa, N.; Yamabuki, T.; Ito, T.; Miyamoto, M.; Kondo, S.; Nakamura, Y. A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer Res.* **2005**, *65*, 5638–5646. [CrossRef] [PubMed]
14. Dalluge, J.J.; Hamamoto, T.; Horikoshi, K.; Morita, R.Y.; Stetter, K.O.; McCloskey, J.A. Posttranscriptional modification of tRNA in psychrophilic bacteria. *J. Bacteriol.* **1997**, *179*, 1918–1923. [CrossRef] [PubMed]
15. Bishop, A.C.; Xu, J.; Johnson, R.C.; Schimmel, P.; de Crecy-Lagard, V. Identification of the tRNA-dihydrouridine synthase family. *J. Biol. Chem.* **2002**, *277*, 25090–25095. [CrossRef]
16. Xing, F.; Hiley, S.L.; Hughes, T.R.; Phizicky, E.M. The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.* **2004**, *279*, 17850–17860. [CrossRef]
17. Lombard, M.; Hamdane, D. Flavin-dependent epitranscriptomic world. *Arch. Biochem. Biophys.* **2017**, *632*, 28–40. [CrossRef]
18. Rider, L.W.; Ottosen, M.B.; Gattis, S.G.; Palfey, B.A. Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *J. Biol. Chem.* **2009**, *284*, 10324–10333. [CrossRef]
19. Bou-Nader, C.; Montemont, H.; Guerineau, V.; Jean-Jean, O.; Bregeon, D.; Hamdane, D. Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: Perspectives on the evolution scenario. *Nucleic Acids Res.* **2018**, *46*, 1386–1394. [CrossRef]
20. Kasprzak, J.M.; Czerwoniec, A.; Bujnicki, J.M. Molecular evolution of dihydrouridine synthases. *BMC Bioinform.* **2012**, *13*, 153. [CrossRef]
21. Faivre, B.; Lombard, M.; Fakroun, S.; Vo, C.D.; Goyenvalle, C.; Guerineau, V.; Pecqueur, L.; Fontecave, M.; De Crecy-Lagard, V.; Bregeon, D.; et al. Dihydrouridine synthesis in tRNAs is under reductive evolution in Mollicutes. *RNA Biol.* **2021**, *18*, 2278–2289. [CrossRef] [PubMed]
22. Yu, F.; Tanaka, Y.; Yamashita, K.; Suzuki, T.; Nakamura, A.; Hirano, N.; Yao, M.; Tanaka, I. Molecular basis of dihydrouridine formation on tRNA. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 19593–19598. [CrossRef] [PubMed]
23. Kusuba, H.; Yoshida, T.; Iwasaki, E.; Awai, T.; Kazayama, A.; Hirata, A.; Tomikawa, C.; Yamagami, R.; Hori, H. In vitro dihydrouridine formation by tRNA dihydrouridine synthase from Thermus thermophilus, an extreme-thermophilic eubacterium. *J. Biochem.* **2015**, *158*, 513–521. [CrossRef] [PubMed]
24. Xing, F.; Martzen, M.R.; Phizicky, E.M. A conserved family of Saccharomyces cerevisiae synthases effects dihydrouridine modification of tRNA. *RNA* **2002**, *8*, 370–381. [CrossRef]
25. Bou-Nader, C.; Pecqueur, L.; Bregeon, D.; Kamah, A.; Guerineau, V.; Golinelli-Pimpaneau, B.; Guimaraes, B.G.; Fontecave, M.; Hamdane, D. An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic Acids Res.* **2015**, *43*, 9446–9456. [CrossRef]
26. Chen, M.; Yu, J.; Tanaka, Y.; Tanaka, M.; Tanaka, I.; Yao, M. Structure of dihydrouridine synthase C (DusC) from Escherichia coli. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2013**, *69*, 834–838. [CrossRef]
27. Byrne, R.T.; Jenkins, H.T.; Peters, D.T.; Whelan, F.; Stowell, J.; Aziz, N.; Kasatsky, P.; Rodnina, M.V.; Koonin, E.V.; Konevega, A.L.; et al. Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6033–6037. [CrossRef]
28. Whelan, F.; Jenkins, H.T.; Griffiths, S.C.; Byrne, R.T.; Dodson, E.J.; Antson, A.A. From bacterial to human dihydrouridine synthase: Automated structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71*, 1564–1571. [CrossRef]

29. Bou-Nader, C.; Barraud, P.; Pecqueur, L.; Perez, J.; Velours, C.; Shepard, W.; Fontecave, M.; Tisne, C.; Hamdane, D. Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Res.* **2019**, *47*, 3117–3126. [CrossRef]
30. Mittelstadt, M.; Frump, A.; Khuu, T.; Fowlkes, V.; Handy, I.; Patel, C.V.; Patel, R.C. Interaction of human tRNA-dihydrouridine synthase-2 with interferon-induced protein kinase PKR. *Nucleic Acids Res.* **2008**, *36*, 998–1008. [CrossRef]
31. Chen, X.; Ji, B.; Hao, X.; Li, X.; Eisele, F.; Nystrom, T.; Petranovic, D. FMN reduces Amyloid-beta toxicity in yeast by regulating redox status and cellular metabolism. *Nat. Commun.* **2020**, *11*, 867. [CrossRef] [PubMed]
32. Bou-Nader, C.; Bregeon, D.; Pecqueur, L.; Fontecave, M.; Hamdane, D. Electrostatic Potential in the tRNA Binding Evolution of Dihydrouridine Synthases. *Biochemistry* **2018**, *57*, 5407–5414. [CrossRef] [PubMed]
33. Bou-Nader, C.; Pecqueur, L.; Barraud, P.; Fontecave, M.; Tisne, C.; Sacquin-Mora, S.; Hamdane, D. Conformational Stability Adaptation of a Double-Stranded RNA-Binding Domain to Transfer RNA Ligand. *Biochemistry* **2019**, *58*, 2463–2473. [CrossRef] [PubMed]
34. Nevers, Y.; Kress, A.; Defosset, A.; Ripp, R.; Linard, B.; Thompson, J.D.; Poch, O.; Lecompte, O. OrthoInspector 3.0: Open portal for comparative genomics. *Nucleic Acids Res.* **2019**, *47*, D411–D418. [CrossRef] [PubMed]
35. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef]
36. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef]
37. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
38. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef]
39. Blum, M.; Chang, H.Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [CrossRef]
40. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
41. Procter, J.B.; Carstairs, G.M.; Soares, B.; Mourao, K.; Ofoegbu, T.C.; Barton, D.; Lui, L.; Menard, A.; Sherstnev, N.; Roldan-Martinez, D.; et al. Alignment of Biological Sequences with Jalview. In *Multiple Sequence Alignment*; Springer: New York, NY, USA, 2021; Volume 2231, pp. 203–224. [CrossRef]
42. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]
43. Mirdita, M.; Schutze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. [CrossRef] [PubMed]
44. Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **2011**, *67*, 293–302. [CrossRef] [PubMed]
45. Mccoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Cryst.* **2007**, *40*, 658–674. [CrossRef] [PubMed]
46. Bunkoczi, G.; Read, R.J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D* **2011**, *67*, 303–312. [CrossRef]
47. Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkoczi, G.; Chen, V.B.; Croll, T.I.; Hintze, B.; Hung, L.W.; Jain, S.; McCoy, A.J.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D* **2019**, *75*, 861–877. [CrossRef]
48. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501. [CrossRef]
49. Dobritzsch, D.; Schneider, G.; Schnackerz, K.D.; Lindqvist, Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *EMBO J.* **2001**, *20*, 650–660. [CrossRef]
50. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
51. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [CrossRef]
52. Lu, D.; Searles, M.A.; Klug, A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **2003**, *426*, 96–100. [CrossRef] [PubMed]
53. Hall, T.M. Multiple modes of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 367–373. [CrossRef]
54. Burge, R.G.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Structural Characterization of Interactions between the Double-Stranded RNA-Binding Zinc Finger Protein JAZ and Nucleic Acids. *Biochemistry* **2014**, *53*, 1495–1510. [CrossRef]
55. Zhou, C.; Huang, R.H. Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: Insight into tRNA recognition and reaction mechanism. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16142–16147. [CrossRef] [PubMed]
56. Tian, B.; Bevilacqua, P.C.; Diegelman-Parente, A.; Mathews, M.B. The double-stranded-RNA-binding motif: Interference and much more. Nature reviews. *Mol. Cell Biol.* **2004**, *5*, 1013–1023. [CrossRef]

57. Chang, K.Y.; Ramos, A. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J.* **2005**, *272*, 2109–2117. [CrossRef]
58. Masliah, G.; Barraud, P.; Allain, F.H. RNA recognition by double-stranded RNA binding domains: A matter of shape and sequence. *Cell. Mol. Life Sci. CMLS* **2013**, *70*, 1875–1895. [CrossRef] [PubMed]
59. Stefl, R.; Oberstrass, F.C.; Hood, J.L.; Jourdan, M.; Zimmermann, M.; Skrisovska, L.; Maris, C.; Peng, L.; Hofr, C.; Emeson, R.B.; et al. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **2010**, *143*, 225–237. [CrossRef] [PubMed]
60. Jayachandran, U.; Grey, H.; Cook, A.G. Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. *Nucleic Acids Res.* **2016**, *44*, 1924–1936. [CrossRef]
61. Goto-Ito, S.; Ito, T.; Kuratani, M.; Bessho, Y.; Yokoyama, S. Tertiary structure checkpoint at anticodon loop modification in tRNA functional maturation. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1109–1115. [CrossRef]
62. Guo, M.; Yang, X.L.; Schimmel, P. New functions of aminoacyl-tRNA synthetases beyond translation. Nature reviews. *Mol. Cell Biol.* **2010**, *11*, 668–674. [CrossRef]

See Article 2: Functional redundancy in tRNA dihydrouridylation.

Nucleic Acids Research, 2024, **52**, 5880–5894
https://doi.org/10.1093/nar/gkae325
Advance access publication date: 29 April 2024
Nucleic Acid Enzymes

Functional redundancy in tRNA dihydrouridylation

Claudia Sudol[1,2], Lea-Marie Kilz[3], Virginie Marchand[4,5], Quentin Thullier[4,5], Vincent Guérineau[6], Catherine Goyenvalle[1], Bruno Faivre[2], Sabrine Toubdji[1,2], Murielle Lombard[2], Olivier Jean-Jean[1], Valérie de Crécy-Lagard[7,8], Mark Helm[3], Yuri Motorin[4,5], Damien Brégeon[1,*] and Djemel Hamdane[2,*]

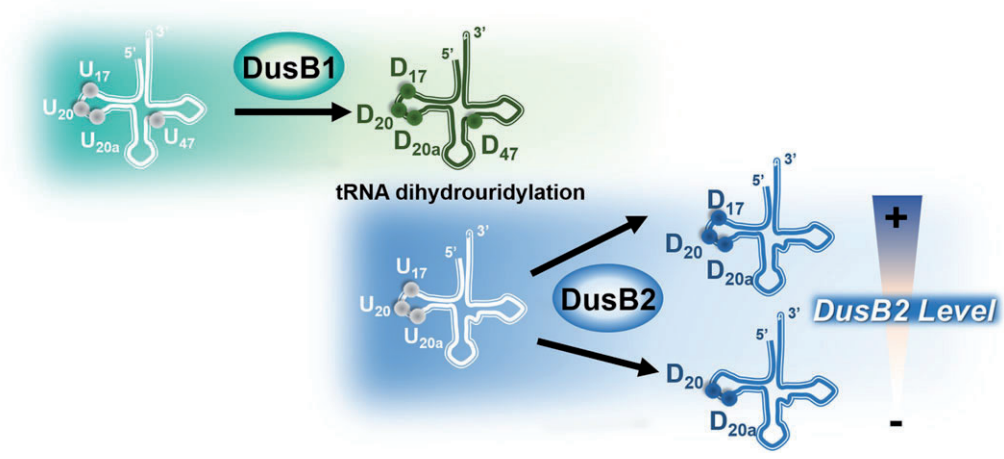
[1]Sorbonne Université, CNRS, Institut de Biologie Paris Seine, Biology of Aging and Adaptation, Paris 75252, France
[2]Collège De France, Sorbonne Université, CNRS, Laboratoire de Chimie des Processus Biologiques, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France
[3]Institut für pharmazeutische und biomedizinische Wissenschaften (IPBW), Johannes Gutenberg-Universität, Mainz 55128, Germany
[4]Université de Lorraine, CNRS, INSERM, UMS2008/US40 IBSLor, EpiRNA-Seq Core Facility, Nancy F-54000, France
[5]Université de Lorraine, CNRS, UMR7365 IMoPA, Nancy F-54000, France
[6]Université Paris-Saclay, CNRS, Institut de Chimie des Substances Naturelles, UPR 2301, 91198, Gif-sur-Yvette, France
[7]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA
[8]University of Florida, Genetics Institute, Gainesville, FL 32610, USA

*To whom correspondence should be addressed. Tel: +33 1 44271645; Email: djemel.hamdane@college-de-france.fr; djemel.hamdane@sorbonne-universite.fr
Correspondence may also be addressed to Damien Brégeon. Email: damien.bregeon@sorbonne-universite.fr

Abstract

Dihydrouridine (D) is a common modified base found predominantly in transfer RNA (tRNA). Despite its prevalence, the mechanisms underlying dihydrouridine biosynthesis, particularly in prokaryotes, have remained elusive. Here, we conducted a comprehensive investigation into D biosynthesis in *Bacillus subtilis* through a combination of genetic, biochemical, and epitranscriptomic approaches. Our findings reveal that *B. subtilis* relies on two FMN-dependent Dus-like flavoprotein homologs, namely DusB1 and DusB2, to introduce all D residues into its tRNAs. Notably, DusB1 exhibits multisite enzyme activity, enabling D formation at positions 17, 20, 20a and 47, while DusB2 specifically catalyzes D biosynthesis at positions 20 and 20a, showcasing a functional redundancy among modification enzymes. Extensive tRNA-wide D-mapping demonstrates that this functional redundancy impacts the majority of tRNAs, with DusB2 displaying a higher dihydrouridylation efficiency compared to DusB1. Interestingly, we found that *Bs*DusB2 can function like a *Bs*DusB1 when overexpressed *in vivo* and under increasing enzyme concentration *in vitro*. Furthermore, we establish the importance of the D modification for *B. subtilis* growth at suboptimal temperatures. Our study expands the understanding of D modifications in prokaryotes, highlighting the significance of functional redundancy in this process and its impact on bacterial growth and adaptation.

Graphical abstract



Introduction

All RNA transcripts undergo a series of post-transcriptional processes tailored to optimize their functionality (1–3). These processes include the addition of various chemical groups appended to the base and/or ribose moieties at conserved positions within the RNA polymer, and catalyzed *de novo* by

Received: August 31, 2023. **Revised:** March 26, 2024. **Editorial Decision:** April 5, 2024. **Accepted:** April 24, 2024

specific enzymes (4). Over 170 chemical modifications have been documented thus far, with ongoing advancements in transcriptome analysis, particularly through high-throughput sequencing technologies combined with chemical labeling and mass spectrometry, continually unveiling novel modifications (5). At the forefront of the most extensively modified RNA species lie tRNAs, small non-coding RNA molecules involved in decoding genetic information during translation (6,7). tRNAs undergo complex modifications predominantly clustered at positions 34 and 37 within the anticodon loop. These modifications are not only acknowledged for their indispensable role in ensuring the accuracy and efficiency of translation processes (8), but are also emerging as vital regulatory elements (9,10). Equally significant, the chemical modifications located outside the anticodon and scattered throughout the polymer stabilize the peculiar and essential L-shaped tRNA structure formed by the kissing dihydrouridine (D) and ribothymidine (rT = m^5U at position 54) loops (11–13), both of which represent conserved modified bases.

Unlike all other modified bases, D is a non-aromatic base that cannot participate in stacking interactions or engage in base pairing via hydrogen bonding. Nevertheless, dihydrouridine fulfills a distinctive role by promoting the flexible C2'-endo conformation of the ribose (14). The exact function of D in RNA remains somewhat elusive, although several assumptions have been proposed. It is widely accepted that because the D base is not aromatic and thus disinclined to stacking interactions, it confers a certain degree of flexibility around its position, thereby allowing favorable tertiary interactions in the tRNA elbow region (14,15). This notion of flexibility finds support in studies showing that psychrophilic organisms, thriving in low-temperature environments, generally exhibit higher D content than thermophilic counterparts (16). In addition, higher D content may confer a growth advantage to cancer cells over healthy cells, perhaps by enhancing translational efficiency (17). However, the exact mechanisms underlying such effect remain unclear and require further exploration.

D is commonly present at multiple canonical sites in tRNAs (D16–D17–D20–D20a–D20b–D47), for both bacteria and eukaryotes (Figure 1), with its abundance depending on the organism and tRNA type (4,30). Its biosynthesis is achieved through the reduction of the C5=C6 uridine double bond, catalyzed by the dihydrouridine synthases (Dus), which belong to the COG0042 (Cluster of Orthologous Group) family of flavoenzymes (18–23). All hitherto investigated Dus enzymes use NADPH as a hydride source to reduce flavin mononucleotide (FMN) to $FMNH^-$, which then donates its hydride to the electrophilic C6 atom of the uridine substrate (22,24). Initially regarded as a modified base commonly seen in tRNA, recent studies have reported D residues in mRNA and certain long non-coding RNAs in yeast and human cells (23,25–27). Dus enzymes responsible for introducing D into tRNA also participate in mRNA dihydrouridylation, highlighting their substrate promiscuity, a property shared with other RNA-modifying enzymes such as pseudouridine synthases, m^1A and m^5C methyltransferases (28,29). The site-specificities of the Dus enzymes have been established in various organisms including yeast (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*), humans, *Escherichia coli*, *Thermus thermophilus* and recently in *Mycoplasma capricolum* (18,26,30–33) (Figure 1). These enzymes have been categorized into eight subfamilies, including three bacterial Dus (DusA, DusB, DusC), four eukaryotic Dus (Dus1, Dus2, Dus3, Dus4), and an archaeal Dus (34). The distribution of Dus enzymes is less uniform in prokaryotes and varies between organisms. For instance, Gammaproteobacteria such as *E. coli* possess the three bacterial Dus enzymes, namely DusA, B and C, involved in D20-D20a, D17 and D16, modifications (Figure 1), respectively (31). In contrast, *T. thermophilus* has only one Dus enzyme of the DusA type, which synthesizes the D20–D20a modifications (32) (Figure 1).

DusB emerged as the first Dus from the ancestral Dus, giving rise subsequently to DusA and DusC through duplication events (34). Our recent phylogenetic analysis revealed that Gram-positive bacteria exclusively carry DusB homologs, categorized into three subgroups: DusB1, DusB2 and DusB3 (33). While most of the examined genomes carry either a *dusB1* or *dusB2* gene, approximately 40% of these organisms contain both *dusB1* and *dusB2* genes, with *dusB3* being restricted to a subset of Clostridia. *Bacillus* species generally retained both *dusB1* and *dusB2* (BSU00810 and BSU08030 annotated as *dusB* and *dusC*, respectively), while Mollicutes conserved only *dusB1* and *Staphylococcus* species kept *dusB2*. Both DusB subgroups likely originated from an ancestral DusB duplication event, which probably occurred in the common ancestor of the Firmicutes. In addition, the limited distribution of DusB3 suggests more recent origin. Biochemical characterization of DusB1 from *M. capricolum* (MCAP_0837) revealed its multisite specificity, catalyzing dihydrouridylation at U17, U20 and U20a positions, (Figure 1), consistent with sequenced tRNAs from this Mollicute species (33). The multi-site specificity feature of Gram-positive Dus, likely shared by both DusB1 and DusB2, is also supported by the tRNA modification profiles of three other bacteria: *Lactococcus lactis*, *Streptomyces griseus* and *S. aureus*, all displaying D17, D20 and D20a modifications.

All cases studied show that a given D residue is specifically synthesized by a single Dus. However, while several Dus have been reported to synthesize D at different positions, the redundancy of synthesis in terms of overlapping specificities has not yet been documented. In this investigation, we explore the contribution of Dus homologs, specifically DusB1 and DusB2, in D biosynthesis. Using *B. subtilis* as our model organism, we reveal a significant level of functional redundancy in D biosynthetic pathways catalyzed by both DusB homologs.

Materials and methods

Deletion of *dusB1* and *dusB2* of *B. subtilis* and complementation

The *B. subtilis* strains used in this study were derived from strain W168, obtained from Chastanet's lab (INRAE, Jouy en Josas, France), and listed in Supplementary Table S1. All the primers used for mutant strain and plasmid constructions in this study are listed in Supplementary Table S2. Mutant strains were obtained from Bacillus Genetic Stock Center. Double mutant Δ*dusB1::kan*, Δ*dusB2::erm* strain was generated by transforming single Δ*dusB1::kan* with the PCR product amplified from the single Δ*dusB2::erm* genome using BSU08030-5pL/BSU08030-3pR (35). *B. subtilis* strains expressing *SadusB2* (SACOL0067) under the control of *BsdusB1* promoter (Δ*dusB1::SadusB2-kan*, Δ*dusB2::erm*) was obtained by transforming Δ*dusB2::erm* strain with a PCR

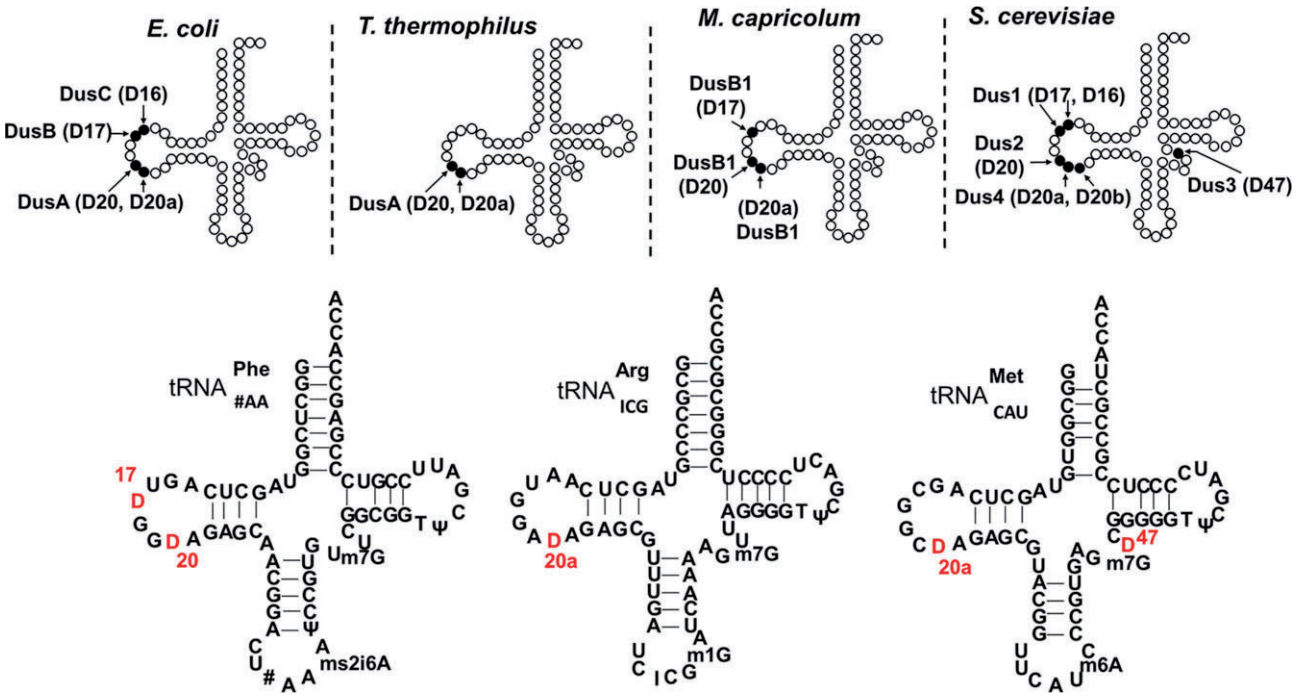


Figure 1. Location of D-sites in tRNA and the corresponding enzyme involved in site dihydrouridylation determined experimentally. Schematic representation of the secondary structure of tRNA, showing the location of D residues and the corresponding Dus enzyme responsible for their synthesis in *E. coli*, *T. thermophilus* and *M. capricolum* for eubacteria and *S. cerevisiae* for eukaryotes. In the lower panel is shown the sequence of *B. subtilis* tRNAs used to analyze the D-sites in the MALDI-MS experiments.

fragment containing (i) the 5′ *BsdusB1* genomic sequence, (ii) *SadusB2* CDS, (iii) a kanamycin resistance cassette and (iv) the 3′ *BsdusB1* genomic sequence. The same strategy was used to express *McdusB1* (MCAP_0837). All Bacillus transformations were performed following the protocol described by Koo *et al.* (35). Strain selections were done on LB-agar containing kanamycin (40 μg ml^{-1}) and/or erythromycin (5 μg ml^{-1}). All strains were verified by PCR and sequencing. *E. coli* strains and growth conditions are detailed in previous studies (31,33).

Cloning *dus*B1 and *dus*B2 from *B. subtilis*, *dus*B1 from *M. capricolum* and *dus*B2 from *S. aureus*

Plasmids containing *dusB1* and *dusB2* genes of *B. subtilis* (pEX-*BsdusB*1 and pEX-*BsdusB2*) and *dusB2* of *S. aureus* (pEX- *SadusB2*) were obtained from Eurofins. We used these plasmids to amplify by PCR *dusB* gene sequences using the primer pairs listed in Supplementary Table S2. The *dusB1* and *dusB2* genes of *B. subtilis* were cloned as follow into pET15b with a sequence encoding for a 6-histidine tag and a thrombin protease site placed at the 5′ end of the genes. After amplification, PCR fragments purified with QIAquick PCR purification kit (Qiagen) were cloned into PCR-linearized pET15b plasmid using the SLIC cloning method (36). Similarly, *dusB1* and *dusB2* genes of *B. subtilis* were cloned in pDG148 for overexpression in *B. subtilis* strains (37). In the case of *dusB2* from *S. aureus* (*SadusB2*), the gene was cloned into the pET28a plasmid containing a sequence encoding for a 6-histidine tag placed at 5′ end of *SadusB2* gene using the same strategy as described above. After cloning, *dusB* gene integrity was verified by DNA sequencing (Eurofins).

RNA extraction and tRNA purification

Bulk tRNA was extracted from *B. subtilis* W168, and its derivative Δ*dusB1::kan* and Δ*dusB2::erm* or double mutant Δ*dusB1::kan*, Δ*dusB2::erm*. Purification of specific tRNA has been previously described (31). Here, $tRNA^{Arg}_{ICG}$, $tRNA^{Phe}_{GAA}$ and $tRNA^{Met}_{CAU}$, from *B. subtilis* strains was performed with 5′ biotinylated complementary oligonucleotide (5′-biot-TGGCGCGCCCGAGGGGAGTCGAACCCCTAA-3′, 5′-biot-TGGTGGCTCGGGACGGAATCGAACCGCCGA-3′ and 5′-biot-TGGTAGCGGCGGAGGGGATCGAACCCCCG-3′ respectively) while $tRNA^{Arg2}_{ICG}$, $tRNA^{Ile1}_{GAU}$ and $tRNA^{Leu1}_{CAG}$, from *E. coli* were purified as described previously (31). For AlkAniline-Seq and LC-MS experiments, total RNA was isolated using hot phenol or Trizol according to manufacturer's instructions.

Activity assay and dihydrouridine quantification

In vitro activity was assayed for 1 h at 37°C in 50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT, 10 mM $MgCl_2$, 100 μM FMN and 15% glycerol under air. Bulk tRNAs (25 μM) issued from the Δ*dusB1::kan*, Δ*dusB2::erm* strain were incubated with various concentration of protein ranging from 0.05 to 50 μM in a total volume of 100 μl and reaction was started upon addition of NADPH at a final concentration of 2 mM. Quenching was performed by adding 100 μl of acidic phenol (Sigma-Aldrich) followed by centrifugation at 16 000×*g* for 10 min. tRNAs in the aqueous phase were ethanol precipitated and further purified using a MicroSpin G-25 column (GE-healthcare). Dihydrouridine quantification was carried out by LC–MS spectrometry analysis.

Liquid chromatography–tandem mass spectrometry (LC–MS)

1 μg of tRNA per sample was digested to nucleoside level using 0.6 units (U) nuclease P1 from *P. citrinum* (Sigma-Aldrich), 0.2 U snake venom phosphodiesterase from *C. adamanteus* (Worthington), 2 U FastAP (Thermo Scientific), 10 U benzonase (Sigma-Aldrich) and 200 ng Pentostatin (Sigma-Aldrich) in 25 mM ammonium acetate buffer at pH 7.5 (Sigma-Aldrich) overnight at 37°C. LC–MS/MS analysis was performed using an Agilent 1260 series LC with a Synergi Fusion RP18 column (4 μM particle size, 80 Å pore size, 250 × 2.0 mm; Phenomenex) and an Agilent 6460A Triple Quadrupole mass spectrometer equipped with an electrospray ion source (ESI). 5 mM ammonium acetate buffer at pH 5.3 was used as solvent A and LC–MS grade acetonitrile (Honeywell) served as solvent B. The elution started with 100% solvent A with a flow rate of 0.35 mL/min, followed by a linear gradient to 8% solvent B at 10 min, raising to 40% solvent B after 20 min and subsequent three-minute restoration of the initial conditions. 100% solvent A was held for further 10 min before starting the next elution. During elution a diode array detector (DAD) recorded the UV signal at 254 nm to monitor the main nucleosides and the ESI parameters were set as follows: gas temperature 350°C, gas flow 8 l min^{-1}, nebulizer pressure 50 psi, sheath gas temperature 350°C, sheath gas flow 12 l min^{-1} and capillary voltage 3000 V. The mass spectrometer was run in the dynamic multiple reaction monitoring (dMRM) mode using Agilent MassHunter software. The quantitative analysis was performed as described in Kellner *et al.* (38) using internal calibration. For internal calibration 300 ng of digested sample were spiked with 50 ng of 13C stable isotope-labelled nucleosides from *E. coli* and subjected to analysis.

MALDI-TOF spectrometry analysis

For mass spectrometry analysis, about 50 μg of tRNAs were digested with either 10 μg of RNAse A (Euromedex) or RNAseT1 (Sigma-Aldrich), which generates 3′-phosphate nucleosides, in a final volume of 10 μl at 37°C for 4 h. One microliter of digest was mixed with 9 μl HPA (40 mg/ml in water: acetonitrile 50:50) and 1 μl of the mixture was spotted on the MALDI plate and air-dried ('dried droplet' method) as previously described (31). MALDI-TOF MS analyses were performed directly on the digestion products using an UltrafleXtreme spectrometer (Bruker Daltonique, France). Acquisitions were performed in positive ion mode. An identical strategy was applied for RNase T1 digests (cleavage after G generating 3′-phosphate nucleosides).

Bioinformatic analyses

The FASTA sequences of 203 proteins annotated in BV-BRC (39) as tRNA-dihydrouridine (20/20a) synthase (EC 1.3.1.91) (or DusA), tRNA-dihydrouridine (16) synthase (or DusC), tRNA-dihydrouridine synthase DusB (or DusB/DusB1) and tRNA-dihydrouridine synthase 2 (or DusB2) where extracted from 120 reference bacterial genomes using the BV-BRC filtering tools. A phylogenetic tree was generated using the MAFFT version 7 (40) web-based pipeline (https://mafft.cbrc.jp/alignment/server/) using the default parameters with boostrap sampling size of 100. The final newick tree is given as Supplemental data 2. The newick tree was then visualized and annotated in iTol (41).

Results

Contribution of *Bs*DusB1 and *Bs*DusB2 to tRNA dihydrouridylation in *B. subtilis*

B. subtilis, complete modification profiles have been established for 24 tRNA sequences over a total of 35 different isoacceptors, allowing us to compile a more or less accurate distribution of the D sites present in this organism (4). The predominant positions where D is found include the canonical positions 17, 20 and 20a, along with position 47 for a single tRNA, $tRNA^{Met}_{CAU}$. A quick survey shows that residues D20 and D20a are the most frequent D residues, followed by D17 and D47 (Supplementary Table S3). Notably, D20 stands out as the most prevalent across all tRNA sequences from all organisms (4). While the abundance of each specific tRNA still needs to be determined, it is reasonable to assume that D20 and D20a account for most of the D content in *B. subtilis* tRNAs.

The D content was determined using liquid chromatography–mass spectrometry (LC–MS) in tRNAs extracted from wild type *B. subtilis* W168 strain or from the isogenic single mutants (Δ*dusB1::kan* and Δ*dusB2::erm*) or double mutant (Figure 2A). The double deletion led to a complete depletion of D content in bulk *B. subtilis* tRNA, indicating that one or both DusB enzymes cover all D biosynthesis in tRNA. However, intriguingly, in the Δ*dusB1::kan* strain, D content decreased by 34%, while in the Δ*dusB2::erm* strain, it only decreased by 18%. In other words, in the Δ*dusB1::kan* strain, *Bs*DusB2 was responsible for 66% of the D content, whereas *Bs*DusB1 synthesized 82% of the D content in the Δ*dusB2::erm* strain (Figure 2B). These seemingly contradictory results may in fact be explained by an overlapping specificity shared by the two DusB enzymes. Complementation assays in the *B. subtilis* Δ*dusB1*::kan,Δ*dusB2*::erm strain showed that the expression of *McdusB1* and *SadusB2* from the *dusB1* promoter restored 76% and 22% of the D content of wild type tRNAs, respectively (Figure 2B). This indicates that both genes encode for Dus enzymes, and that the *M. capricolum* enzyme is more active than the *S. aureus* enzyme in the *B. subtilis* heterologous system.

Functional redundancy of the DusB enzymes in *B. subtilis* determined by MALDI-MS

The *in vivo* specificity of *Bs*DusB1 and *Bs*DusB2 dihydrouridylation sites was determined by comparing the D content in the tRNAs of four *B. subtilis* strains, including the W168 (wild type) and the single or double deletion strains. The approach involved a three-step workflow: (i) purification of specific tRNA types from various *B. subtilis* cells, (ii) fragmentation of the tRNA using RNAseA or RNAseT1 and (iii) analysis of the resulting fragments by MALDI-TOF. Deletion of *dusB* genes was expected to generate fragments containing U residues at the positions targeted by the corresponding enzymes, resulting in a −2Da shift relative to fragments in tRNAs extracted from wild type cells. We selected three tRNAs to cover all D sites, namely, $tRNA^{Phe}_{GAA}$ for D17 and D20, $tRNA^{Arg}_{ICG}$ for D20a and $tRNA^{Met}_{CAU}$ for D20a and D47 (Figure 1). The mass profiles of these tRNAs are depicted in Figure 3A–D and Supplementary Figure S1. Analysis of tRNAs from the W168 strain confirmed the presence of all distinct D-containing fragments at the expected positions, validating the

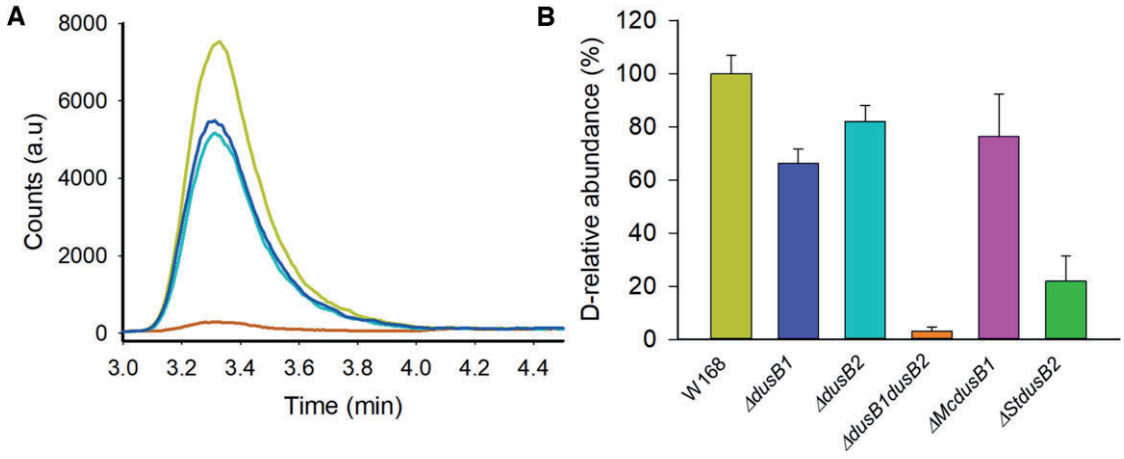


Figure 2. Quantification of D-level in tRNA from *B. subtilis*. (**A**) Extracted ion chromatograms of dihydrouridine in tRNAs isolated from *B. subtilis* WT strain (W168 in light green), Δ*dusB1::kan*, Δ*dusB2::erm* double deletion strain (orange) and Δ*dusB1::kan* (blue) and Δ*dusB2::erm* single mutant strains (cyan). The signals were normalized to the respective UV signal of Adenosine. (**B**) D levels determined in bulk tRNAs of *B. subtilis* WT strains (W168 in light green), Δ*dusB1* (blue) and Δ*dusB2* (cyan) single mutants, Δ*dusB1*Δ*dusB2* (orange) or double deletion complemented with either DusB1 of *M. capricolum* (magenta) or DusB2 from *S. aureus* (green). The strains were grown in LB media at 37°C. Results are shown as average of three biological replicates in relation to the *wild type* strain W168.

approach. Analysis of the D17 modification was made possible by monitoring the *m/z* 978 fragment corresponding to the $UD_{17}G$ trinucleotide generated by digestion of $tRNA^{Phe}_{GAA}$ by RNAseT1 (Figure 3A). This fragment (its corresponding intensity showing background level) was absent in the double mutant strain while the intensity of the *m/z* 976 peak increased. Similar results were observed for $tRNA^{Phe}_{GAA}$ from the Δ*dusB1::kan* strain. In contrast, the $UD_{17}G$ fragment was detected in the Δ*dusB2::erm* strain with intensity comparable to that of the W168 control. Therefore, these results suggested that *Bs*DusB1 was responsible for D17 biosynthesis. D20 was probed with two distinct fragments of D-containing $tRNA^{Phe}_{GAA}$ from two different digestions. The first digestion, performed with RNAseA, yielded the GGD_{20} trinucleotide (*m/z* 1017) (Figure 3B). The second digestion, performed with RNAseT1, generated the trinucleotide $D_{20}AG$ (*m/z* 1001). In both scenarios, these two fragments did not disappear in $tRNA^{Phe}_{GAA}$ from the two *dusB* single deletion strains, although a more consequent decrease in intensity was observed in the case of Δ*dusB2::erm*. In contrast, in the case of the double mutant, the peak was no longer detectable. We concluded that D20 was inserted into $tRNA^{Phe}_{GAA}$ using both *Bs*DusB1 and *Bs*DusB2, with a dihydrouridylation efficiency that appeared to be higher for *Bs*DusB2. D20a was detected in two different tRNAs: $tRNA^{Arg}_{ICG}$ via the $GGAD_{20a}$ (*m/z* 1346) fragments obtained by RNAseA treatment and $AD_{20a}AG$ generated by RNAseT1 (*m/z* 1346) (Supplementary Figure S1), and $tRNA^{Met}_{CAU}$ via the $CD_{20a}AG$ fragment (*m/z* 1306) obtained by RNAseT1 (Figure 3C). In the case of D20a in $tRNA^{Met}_{CAU}$, both DusBs participated in its synthesis as neither mutant caused a substantial decrease in the intensity of the *m/z* 1306 peak, and their profiles were quite similar to that of the wild type. However, for D20a in $tRNA^{Arg}_{ICG}$, only the deletion of *Bsdusb2* or the double mutant led to a significantly decreased peak at *m/z* 1346, accompanied by an increase in the peak at *m/z* 1344 corresponding to the non-dihydrouridylated fragment (Supplementary Figure S1). These results suggest that the involvement of the two *Bs*DusB paralogs in D20a biosynthesis may depend on the tRNA substrate. Lastly, D47 was assayed by following the $D_{47}CG$ fragment (*m/z* 977), derived from treatment of $tRNA^{Met}_{CAU}$ with RNAseT1 (Figure 3D). This analysis was carried out following the same analytical grid as before. The $tRNA^{Met}_{CAU}$ from wild type and Δ*dusB2::erm B. subtilis* strains retained the prominent peak at *m/z* 977. In contrast, in the case of the Δ*dusB1::kan* or the double mutant strains, the intensity of this peak drastically decreased concomitantly with the increase in the *m/z* 975 peak, suggesting that *Bs*DusB1 was also responsible for D47 biosynthesis.

Dihydrouridylation redundancy targets several tRNAs as investigated by deep-sequencing based AlkAnilineSeq method

An analysis of *B. subtilis Bs*DusB *in vivo* specificities was performed using the AlkAnilineSeq method (see supplementary methods for details) (42). This method exploits the D-ring's instability under alkaline conditions (20), leading to its cleavage and the formation of β-ureidopropionic acid. This instability results in aniline-driven RNA cleavage, generating a 5′-phosphate group (5′-P) on the neighboring N + 1 residue, which serves as an input for highly selective ligation of sequencing adapters. Alongside D-residue detection, AlkAnilineSeq also allows parallel detection of 7-methylguanosine (m^7G), 3-methylcytidine (m^3C) and 5-hydroxycytidine (ho^5C), which share some degree of fragility in their base rings and/or N-glycosidic bonds, present in these modified residues. Mapping was achieved for all D containing tRNAs from the four *B. subtilis* strains, including the W168 strain, as well as the single and double *dus* deletion strains. It is important to emphasize that none of the D residues detected by this method was present at stoichiometric levels, suggesting partial dihydrouridylation of the target uridines. Importantly, the results obtained by AlkAnilineSeq were consistent with the MALDI-MS mapping experiments. For example, the disappearance of D17 in $tRNA^{Ala}_{GGC}$ and $tRNA^{Ala}_{UGC}$ of *B. subtilis* W168 was observed only in Δ*dusB1::kan* and double deletion strains, suggesting that *Bs*DusB1 was involved in the reduction of U17 in these tRNAs. In $tRNA^{Arg}_{ACG}$, the loss of both D17

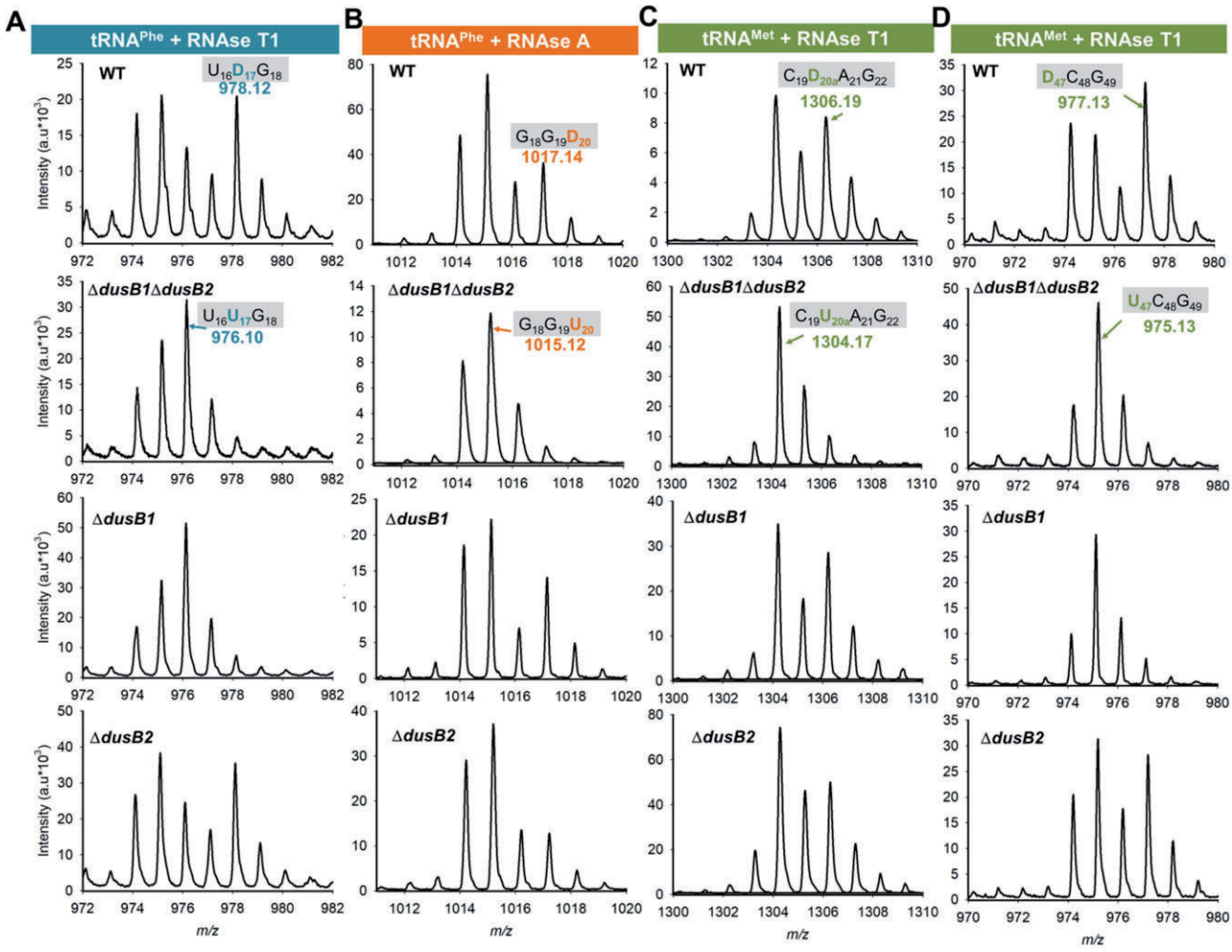


Figure 3. MALDI-TOF analysis of position 17, 20, 20a and 47 in tRNAs from *B. subtilis WT* and Dus deletion mutants. (**A**) D17-containing MS relative isotope patterns of derived oligonucleotides after RNAse T1 treatment of $tRNA^{Phe}_{\#AA}$ isolated from wild type, Δ*dusB1*Δ*dusB2*, Δ*dusB1 and* Δ*dusB2*, respectively. (**B**) D20-containing MS relative isotope patterns of derived oligonucleotides after RNAse A treatment of $tRNA^{Phe}_{\#AA}$ isolated from wild type, Δ*dusB1*, Δ*dusB2 and* Δ*dusB1*Δ*dusB2*, respectively. (**C**) D20a-containing MS relative isotope patterns of derived oligonucleotides after RNAse T1 treatment of $tRNA^{Met}_{CAU}$ isolated from wild type, Δ*dusB1*, Δ*dusB2 and* Δ*dusB1*Δ*dusB2*, respectively. (**D**) D47-containing MS relative isotope patterns of derived oligonucleotides after RNAse T1 treatment of $tRNA^{Met}_{CAU}$ isolated from wild type, Δ*dusB1*, Δ*dusB2 and* Δ*dusB1*Δ*dusB2*, respectively. Further details of the tRNA-derived oligonucleotide fragments and their sizes (*m*/*z*) used for the identification of DusB specificities are shown in supplementary figures.

and D20a was seen in the double deletion strain, whereas in the Δ*dusB1::kan* strain, only the loss of D17 was observed (Supplementary Figure S2). In contrast, in the Δ*dusB2::erm* strain, the signal attributed to D20a declined when compared to the signal observed in Δ*dusB1::kan*, while D17 remained unchanged (Supplementary Figure S2). This is consistent with the fact that *Bs*DusB1 was responsible for the formation of both D17 and D20a, whereas *Bs*DusB2 formed only D20a in this tRNA (Supplementary Figure S2). Moreover, *Bs*DusB1 was implicated in the biosynthesis of all three D17/D20/D20a residues in $tRNA^{Asp}_{GUC}$, whereas *Bs*DusB2 participated only in the latter two positions. In the case of $tRNA^{Glu}_{UUC}$, *Bs*DusB1 was only capable of forming D20, while *Bs*DusB2 could form both D20 and D20a. These findings suggested that the two enzymatic dihydrouridylation activities did overlap.

To gain a comprehensive view of both *Bs*Dus enzymes' activity, we generated an activity profile heatmap, as presented in Figure 4. The heatmap clearly demonstrates that only the double mutant lacked all D residues in tRNAs, consistent with both LC–MS and MALDI-TOF data. This supports the earlier observation that both Dus enzymes are essential for dihydrouridylation across the full range of tRNA substrates. Moreover, it is evident from the heatmap that only the *Bs*DusB1 enzyme was involved in the formation of D17, whereas both enzymes contributed to the formation of D20 and D20a. Further analysis of the D signal intensities revealed that while most of the D20 and D20a residues were synthetized by both *Bs*DusB1 and *Bs*DusB2, a few dihydrouridylation events preferentially used *Bs*DusB2 (such as for D20 in $tRNA^{Gly}_{UCC}$ and $tRNA^{Tyr}_{GUA}$ and for D20a in $tRNA^{Ile}_{CAU}$, $tRNA^{Ser}_{UGA}$, $tRNA^{Arg}_{CCG}$). We did not detect D-signal for three tRNA, namely $tRNA^{Ile}_{GAU}$, $tRNA^{Pro}_{UGG}$ and $tRNA^{Val}_{UAC}$. Also the AlkAniline-Seq method did not detect the presence of D47, unlike the experiments performed by MALDI-MS on $tRNA^{Met}_{CAU}$. This discrepancy could be explained by interference

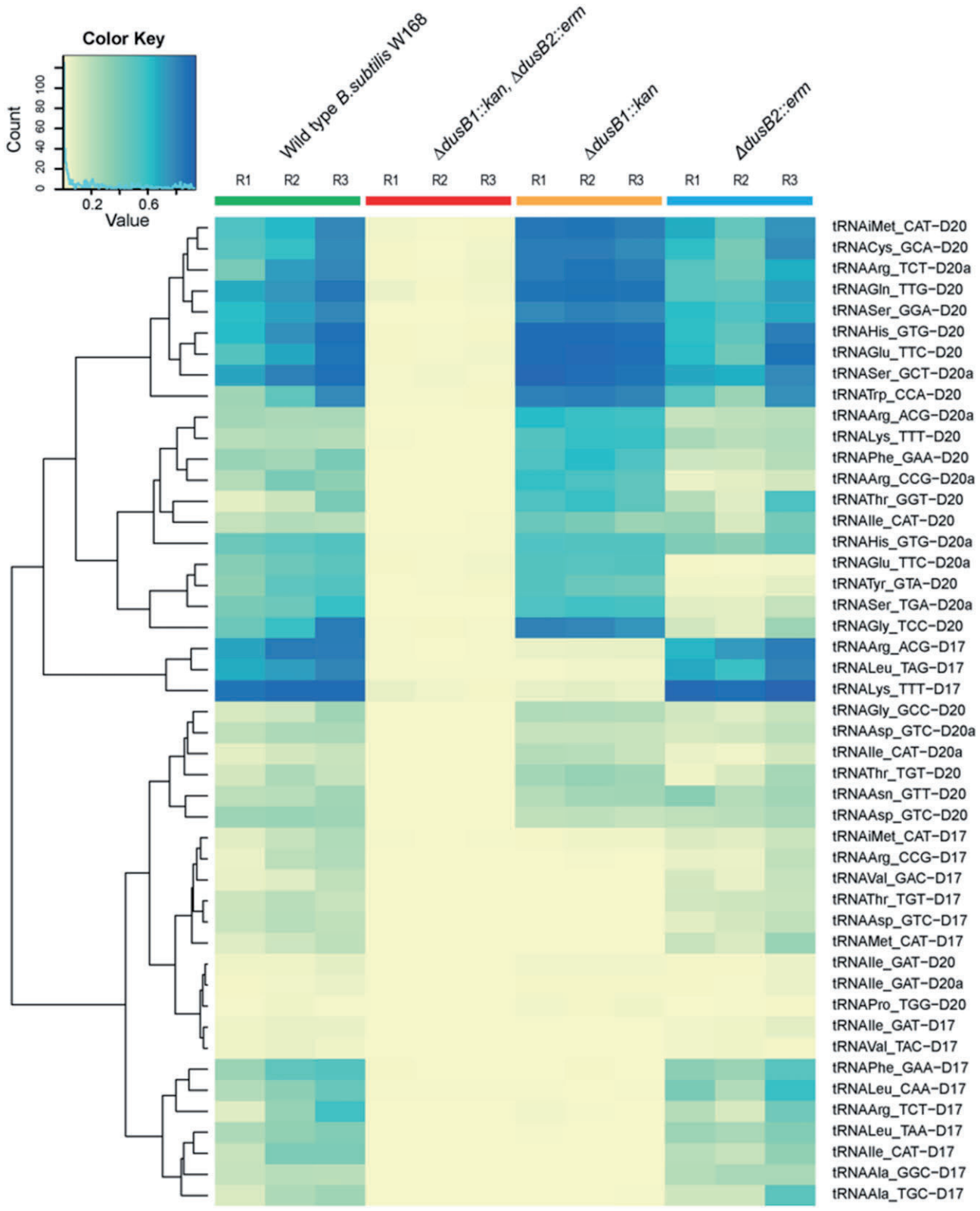


Figure 4. Heatmaps for the assessment of dihydrouridylation changes in individual modified sites in tRNAs from *B. subtilis* and its DusB mutants. The heatmap displays one specific D-modification's stoichiometry across the different samples (in X-axis) and the different D-sites retained for analysis (in Y-axis). The stoichiometry is blue-coded and relies on through stop ratio of the AlkAnilineSeq detection method, which detects m^7G, m^3C and D. R1, R2 and R3 represent the results for the three different replicas.

caused by m^7G46, which produces a strong AlkAnilineSeq signal.

*Bs*DusB1 and *Bs*DusB2 are flavoproteins characterized by a distinct polarity of their active site

The *Bs*DusB1 and *Bs*DusB2 proteins share a relatively low sequence identity of 26% (Supplementary Figure S3). To characterize these two proteins *in vitro*, the genes encoding *Bs*DusB1 (BSU00810, Uniprot Id P37567) and *Bs*DusB2 (BSU08030, Uniprot Id O31546) were cloned into expression vectors, expressed in *E. coli*, and subsequently purified to homogeneity (Supplementary Figure S4). To determine their oligomeric state in solution, gel filtration on a Superdex increase 75 10/300 column was performed, revealing that both proteins exist as monomers with an estimated molecular weight (Mw) of approximately 40 kDa for *Bs*DusB1 (elution volume ∼ 11.2 ml) and 39 kDa for *Bs*DusB2 (elution volume ∼ 11.7 ml). *Bs*DusB proteins were found to be copurified with their flavin coenzyme, evident from the yellowish color of the protein samples and characteristic absorbance spectra (Figure 5A). The latter featured two absorption bands typical for flavin: the S0-S2 bands exhibited a maximum at 372 nm, while the S0–S1 band in *Bs*DusB1 and *Bs*DusB2 showed a maximum at 450 and 458 nm, respectively. The difference in the wavelength maximum of the S0–S1 transition between the two proteins suggests dissimilarity in the polarity of their active sites. Upon the addition of sodium dodecyl sulfate (SDS), the proteins denatured, releasing flavin into the solution. The resulting flavin in solution displayed an absorption spectrum similar to that of free FMN, confirming that both *Bs*DusB enzymes are flavoproteins with the FMN non-covalently bound to the apoprotein. FMN fluorescence in both holoproteins was also monitored and showed a slight red shift in the maximum fluorescence emission band of *Bs*DusB2, at 530 nm, compared to that of *Bs*DusB1 observed at 527 nm, supporting the existence of distinct environments for the two FMN coenzymes (Supplementary Figure S5). This polarity contrast is substantiated by our analysis of the active sites in the holoprotein forms of *Bs*DusB1 and *Bs*DusB2 Alphafold models (see supplementary results and Figure 5B).

An unusual behavior of Dus pyrimidine discrimination and dihydrouridylation activity of tRNA

Dus enzymes share a highly conserved catalytic mechanism that involves two redox reactions (22,24). NADPH reduces FMN to yield $FMNH^-$, which is then oxidized to upon reduction of uridine to dihydrouridine. We measured the NAD(P)H oxidase activity of the two *Bs*DusB enzymes independently by monitoring the consumption of NADH or NADPH under aerobic conditions using absorbance spectrophotometry at 340 nm and steady-state conditions. The data were analyzed using the Michaelis–Menten formalism and the related kinetics parameters are presented in Table 1. The results revealed that *Bs*DusB1 oxidized NADPH and NADH with identical catalytic constants (k_{cat} ∼ 0.013 s^{-1}) and comparable K_M values, indicating that the enzyme did not discriminate between NADH or NADPH and could use both equally. This result was unexpected, because all previously studied Dus enzymes, both prokaryotic and eukaryotic, showed a preference for NADPH over NADH (30,33,43). In contrast, for *Bs*DusB2, NADPH was a better substrate than NADH due to a lower K_M for NADPH (2 μM) than for NADH (22 μM) and ∼ a 3-fold higher catalytic constant for NADPH than for NADH. Overall, NADPH exhibited a 5-fold higher catalytic efficiency than NADH.

To examine the *Bs*DusB activity of *B. subtilis*, *in vitro* dihydrouridylation assays were performed with bulk tRNAs from the double deletion strain, and the reaction products were traced using LC/MS. In the presence of 1 μM protein, *Bs*DusB2 was able to restore a 40% higher D level compared to *Bs*DusB1 after 1 hour, indicating that *Bs*DusB2 is the more active enzyme (Figure 5C).

Structural characterization of DusB enzymes and RNA binding

The structural models of *Bs*DusB1, *Bs*DusB2, *Ec*DusB and *Mc*DusB1 were examined using models generated through Alphafold Colab2 (Figure 5D). The derived models exhibited per-residue confidence scores exceeding 90% across most of their respective regions, as illustrated in Supplementary Figure S6A. As anticipated, these enzymes display a conserved canonical folding of the Dus family, i.e. (i) a catalytic domain adopting a TIM-Barrel type structure (TBD) where the flavin coenzyme binding site lies at the entrance of the barrel, (ii) a helical domain (HD) composed of a 4-helix bundle, and (iii) a short linker of about 10 amino acids connecting the two domains. Conducting a structural alignment and comparing the models revealed low RMSD values within the *Bs*DusB1 subfamily (Supplementary Figure S6). This supports the notion that the models for *Bs*DusB1, *Mc*DusB1 and *Ec*DusB1 exhibit highly similar structures. A broad distribution of positive surface charges accessible to the solvent, most likely engaged in interactions with the tRNA substrates, can be distinguished (Figure 5E). This distribution is arranged on both sides of a line of demarcation (LOD) that can be drawn from the left extremity of the TBD throughout the active site cavity, ending at the lower tip of the HD at the C-terminus. Several interesting points can be observed based on this spatial arrangement. *Bs*DusB1 has a continuous, positive electrostatic surface stretched on both sides of the LOD, whereas *Bs*DusB2 is distinguished by a positive surface forming an elongated stripe parallel to the LOD and spanning almost on all its length, but primarily found on the proximal side of this line. In *Ec*DusB, a significant portion of the positive area forms an off-center globular area on the distal edge of the LOD, involving predominantly the apical region of the TBD. *Mc*DusB1 shows a certain similarity to *Bs*DusB1 but with a distinctive feature, namely the presence of several rather isolated positive charge patches. Thus, each of the studied DusB seems to have its own tRNA binding pattern, likely adapted to its site specificity. Likewise, each Dus will probably orientate the tRNA in a distinct way to allow the active site of the enzyme to gain access to the correct uridine substrate to be modified (21,44). To evaluate whether this difference in positive surface area affects the stability of the enzyme/tRNA complex, we examined the ability of *Bs*DusB1 and *Bs*Dus2 to bind to tRNA by specifically monitoring the impact of tRNA titration on flavin fluorescence. Addition of tRNA resulted in an increase in FMN fluorescence of both *Bs*DusB describing a cooperative process

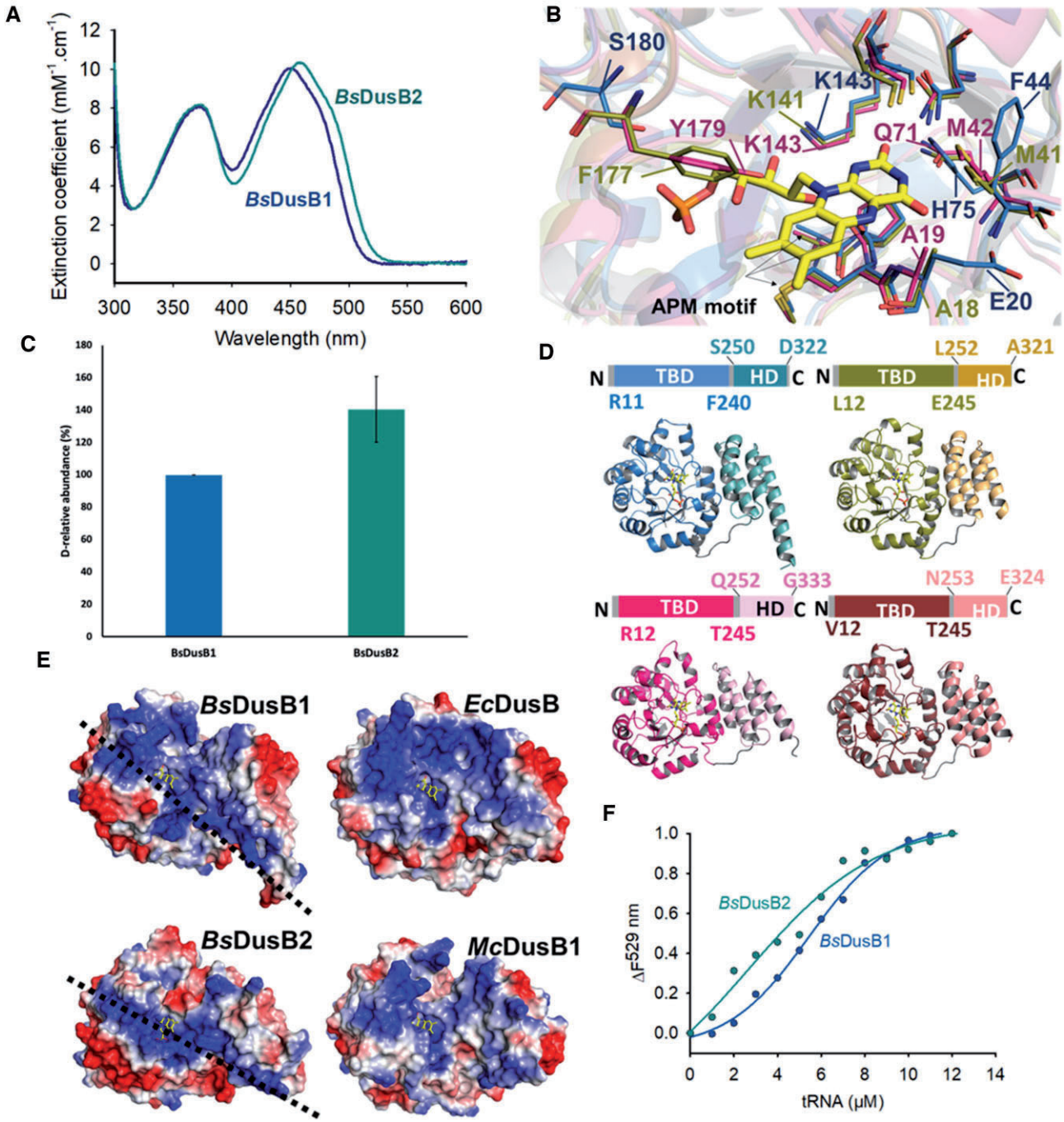


Figure 5. Structural and functional characterization of *B. subtilis* DusB. (**A**) UV-visible absorption spectra of *Bs*DusB1 (blue) and *Bs*DusB2 (teal) holoproteins. (**B**) Comparative structural models of the active sites of *Bs*DusB1, *Bs*DusB2 and DusB of *E. coli* (*Ec*DusB). The active site view is centered on the overlay of a section encompassing the FMN isoalloxazine (yellow) of the respective active site of the three DusB (*Bs*DusB1 in pink, *Bs*DusB2 in blue and *Ec*DusB in deep olive). Residues around the FMN are shown in stick in the respective color codes of the Dus. (**C**) *In vitro* dihydrouridylation activity test of recombinant *Bs*DusB at 1 μM of enzyme after 1 hour incubation at 37°C. Dihydrouridine levels were determined by LC-MS/MS and normalized to the UV signal of adenosine. To compare the activity of *Bs*DusB, the activity of *Bs*DusB1 was set to 100%. Results are shown as average of biological duplicates. (**D**) Structural models of the DusB holoenzymes from *B. subtilis*, *E. coli* and *M. capricolum*. Except for *Ec*DusB, which is a crystallographic structure (PDB, 6EI9), the other three models are from Alphafold. TBD = TIM Barrel Domain, HD = Helical domain. The FMN is shown in yellow stick. (**E**) Electrostatic surface of the Dus model. The dashed line represents the line of demarcation (LOD) mentioned in the text. (**F**) Isotherm of tRNA binding to *Bs*DusB. ΔF529nm is the change in FMN fluorescence at 529 nm resulting from tRNA titration to *Bs*DusB1 (blue) and *Bs*DusB2 (teal).

Table 1. Kinetic parameters for NAD(P)H oxidase activity of *B. subtilis* Dus

	NADH			*NADPH*		
	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$)
*Bs*DusB1	0.013 ± 0.0014	21 ± 4	6×10^{-4}	0.013 ± 0.002	18 ± 3	7×10^{-4}
*Bs*DusB2	0.23 ±	22 ± 4	6×10^{-2}	0.7 ± 0.1	2.1 ± 0.2	0.3

(Figure 5F). Half-transition was observed at about 3 and 5 μM for *Bs*DusB1 and *Bs*DusB2, respectively, showing no significant differences.

RNA dihydrouridylation broadening specificity depends on enzyme concentration

Our complementation results in the *E. coli* triple *dus* mutant strain (Δ*dusA*::kan, Δ*dusB*::Ø, Δ*dusC*::Ø) with *BsdusB1* or *BsdusB2* demonstrated that both enzymes could dihydrouridylate positions U17, 20 and 20b, acting as both *Ec*DusB and *Ec*DusA (see supplementary results & Supplementary Figure S7). While the outcomes for *Bs*DusB1 were anticipated, the unexpected capability of *Bs*DusB2 to catalyze D17 formation in *E. coli* was intriguing. This finding suggested several possibilities in a heterologous context: (i) *Bs*DusB2 lost its substrate specificity due to the differences in tRNA nature (sequence and modification profile) between both organisms; (ii) a protein partner, RNA, or other compounds in *B. subtilis* controlled the site specificity; or (iii) the intracellular concentrations of Dus proteins differed between *E. coli* and *B. subtilis*. Indeed, complementation assays in *E. coli* were performed with *BsdusB1* or *BsdusB2* under the control of an arabinose-inducible promoter with concentration of inducer adjusted to allow for the detection of dihydrouridylation. In contrast, in *B. subtilis* both genes are expressed from the chromosome by their own promoter.

To further explore these possibilities, we assessed the effect of increasing enzyme concentrations on *Bs*DusB's dihydrouridylation activity *in vitro* using tRNA from the *B. subtilis* double deletion strain as a substrate. Additionally, we performed the experiments in the presence of *B. subtilis* Δ*dusB1::kan*, Δ*dusB2::erm* cell extract to examine the existence of a potential partner for *Bs*DusB2 that might be essential for its site specificity. AlkAnilineSeq quantifications showed that the level of D17 inserted by *Bs*DusB1 and *Bs*DusB2 increased with enzyme concentration (Figure 6A), confirming that *Bs*DusB2 can synthetize D17 *in vitro* on tRNA from *B. subtilis*. AlkAnilineSeq also provided insights into the dihydrouridylation efficiency for all D-sites (Figure 6B). Dihydrouridylation efficiency seemed to depend on the nature of the tRNA and the modified position. As expected, *Bs*DusB1 formed D17/D20/D20a. Except for $tRNA^{Lys}_{UUU}$ and $tRNA^{Ala}_{GGC}$, the dihydrouridylation efficiency was higher at positions 20 and 20a than at position 17. Experiments conducted with crude *B. subtilis* extracts revealed that *Bs*DusB2 retained its ability to synthesize D17 even at higher enzyme concentrations (data not shown), suggesting the absence of a cellular partner that regulates the specificity of this Dus enzyme. To validate these findings *in vivo*, both wild type and mutant strains were transformed with plasmids overexpressing either *Bs*DusB1 or *Bs*DusB2. AlkAnilineSeq profiles from these strains clearly demonstrated that overexpression of *Bs*DusB2 in dusB1-deficient strains or *Bs*DusB1 in dusB2-deficient strains was able to restore the dihydrouridylation profile for a significant subset of tRNAs (Supplementary Figure S8). Moreover, *Bs*DusB2 exhibited the capability to introduce D17 residues into several tRNAs, indicating its functional equivalence to *Bs*DusB1 upon overexpression (Supplementary Figure S8). Taken together, these results demonstrate that specificity likely depends on both the nature of the tRNAs and the enzyme concentration.

Effect of *Bs*DusB deletions on cell growth

The optimal growth temperature of *B. subtilis* ranges from 35 to 37°C. The influence of the lack of *Bs*DusB and by extrapolation of D on the growth of *B. subtilis* was investigated in LB medium at 23, 30 and 37°C (Table 2). At the standard growth temperature of 37°C, *B. subtilis* W168 exhibited a generation time of 21 minutes. However, in the case of the three strains with deletions in either one or both *dus* genes, there was a slight increase in generation time. The effect was slightly more visible when cells were grown at 30°C, with the generation time rising from 31 min for the wild type to 39 and 40 min for *ΔdusB1* and *ΔdusB2*, respectively. The effect was even more pronounced in the double mutant strain, where this doubling-time increased to 43 min. A more significant difference in growth was observed when the temperature was lowered to 23°C. Here, the generation time increased from 49 minutes for W168 to 87 min for the three mutant strains. Thus, the absence of D does not seem to have too great of an impact on *B. subtilis* at physiological growth temperatures, but becomes significant at low temperature such as 23°C. This observation aligns with the role of this modified base in promoting structural flexibility at the tRNA level, a feature that is more crucial at lower temperatures than at higher ones.

Generation time is just one among several growth parameters for bacteria, serving as an indicator of potential fitness loss. Therefore, we conducted competition experiments between mutants and the wild-type strain to evaluate the impact of tRNA dihydrouridylation loss on mutant fitness. Surprisingly, all mutant strains exhibited decreased fitness compared to the wild type, even at 37°C, with the *ΔdusB1* strain showing the lowest competitive index (Supplementary Figure S9). However, observed differences in fitness among mutants were not statistically significant (*t*-test, $P > 0.03$), suggesting a potential role of the redundancy in specificity of *Bs*DusB enzymes.

Discussion

We investigated the role of the two homologs, DusB1 and DusB2, in D base biosynthesis in *B. subtilis* tRNAs. Both *Bs*DusB enzymes are FMN-dependent flavoenzymes with a conserved canonical structure of bacterial Dus, retaining key catalytic residues (Figure 5A,B). However, they differ in the

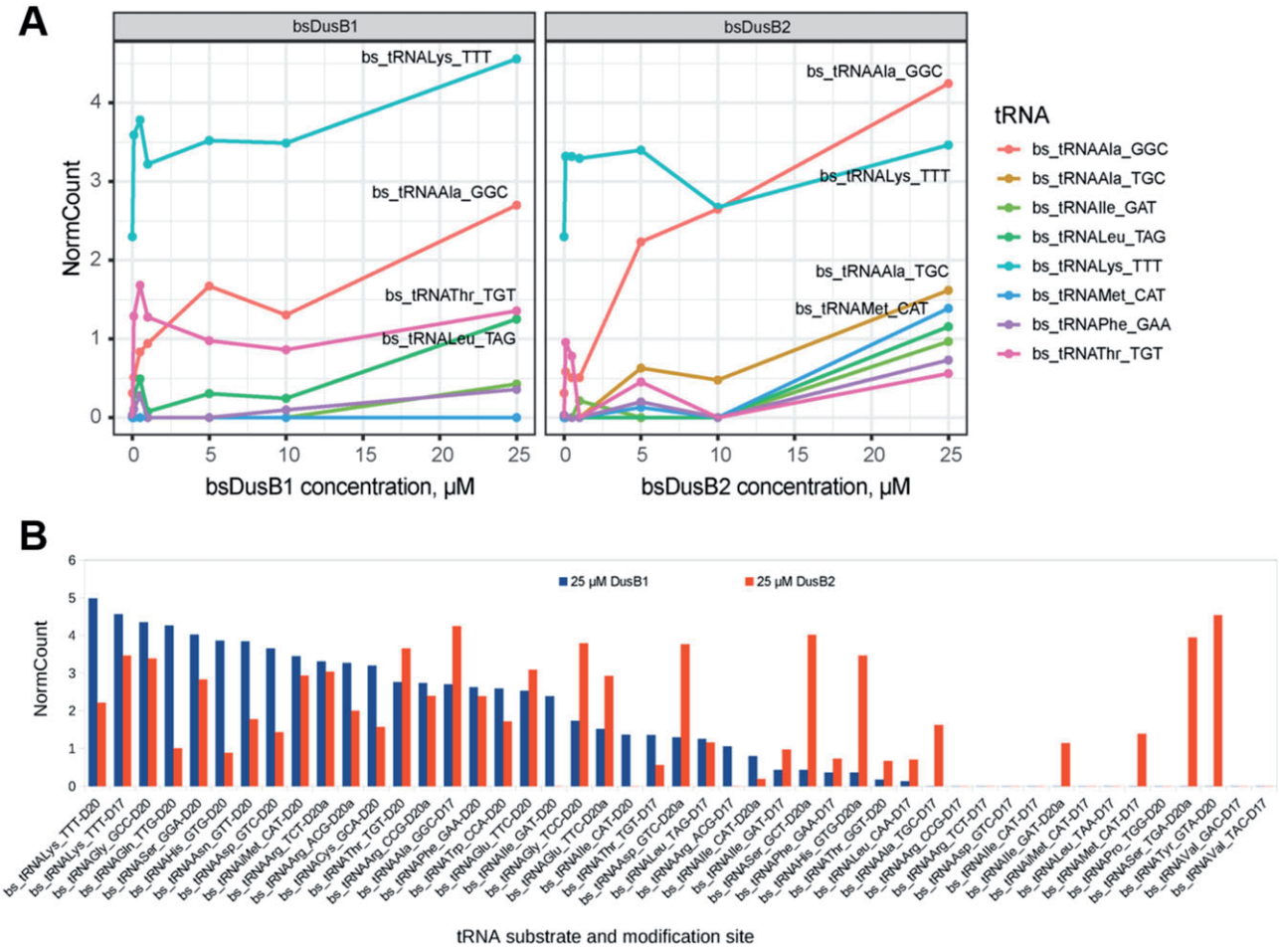


Figure 6. *In vitro* biosynthesis of D in *B. subtilis* tRNAs catalyzed by the recombinant *Bs*DusB1 and *Bs*DusB2 proteins. (**A**) Recombinant enzymes expressed in *E. coli* and purified were incubated with D-unmodified *B. subtilis* total RNA fraction extracted from Δ*dusB1::kan*,Δ*dusB2::erm* strain. Quantification of D17 level was done using NormCount score of AlkAnilineSeq (the signal normalized to median of background cleavages in the surrounding 10 nucleotides). NormCount score (as well as other AlkAnilineSeq Scores) does not show linear dependence from D content, but provides good compromise between sensitivity and specificity of detection for low D levels in tRNA. Only 8 best modified tRNA sites are shown (out of 18 altogether). Concentration of the recombinant *Bs*DusB1 and *Bs*DusB2 is expressed in µM. Identity of tRNA substrates analyzed is shown at the right. (**B**) Modification efficiency of the D-sites measured at 25 µM of enzymes. Quantification of D level was done using NormCount score of AlkAnilineSeq.

polarity of their active sites and preference for the reducing agent, NAD(P)H (see supplementary discussion and Table 1). Most modification enzymes are highly site-specific and modify only one position. However, a small number of enzymes exhibits promiscuous site specificity, targeting either adjacent bases, or multiple positions scattered along the nucleotide sequence of their RNA substrate, or even have both capabilities (45–52) (see also Supplementary Table S4). The Dus enzymes also display the two cases of targeting juxtaposed uridines as observed with bacterial DusA (31) and *Bs*DusB2 for U20–U20a, and with eukaryotic Dus1 (U16–U17) and Dus4 (U20a–U20b)(20). Gram$^+$ Dus enzymes show a wider multisite specificity as seen with the *Mc*DusB1 that modifies the U17–U20–U20a triplet (33) and reinforced here with the discovery that *Bs*DusB1 modifies not only the same bases as *Mc*DusB1 but also the U47 (Figure 3). D47 is located in the variable loop which, in eukaryotes, is catalyzed by Dus3, an enzyme that differs from all Dus by its size and complex modularity (22).

Remarkably, we uncovered an unprecedented property in modification enzymes namely, functional redundancy. This property remains very enigmatic since *Bs*DusB1 can introduce almost the entire D content while *Bs*DusB2 provides a backup activity for positions 20–20a with an efficiency largely in favor of this enzyme. It is worth mentioning that this overlap in activity concerns most tRNAs (Figure 4). Nevertheless, *Bs*DusB2 also have its proper tRNA substrates not shared by *Bs*DusB1 suggesting that this D20–D20a redundancy in dihydrouridylation activity targets a specific set of tRNAs. Surprisingly, *Bs*DusB2 has also the ability to modify U17 only at a certain enzyme concentration, which could probably be consistent with a lower dihydrouridylation efficiency for this site (Figure 6). Of note, Dus enzymes involved exclusively in D20 (or D20–D20a) modification seem to be always more ac-

Table 2. Effect of *dus* deletion on the generation time of *B. subtilis*

	37°C	30°C	23°C
W168	21 ± 0.2	31 ± 0.8	49 ± 3
ΔdusB1ΔdusB2	25 ± 0.3	39 ± 2.2	87 ± 0.7
ΔdusB1	26 ± 0.7	40 ± 4	87 ± 2
ΔdusB2	26 ± 0.6	43 ± 1.5	87 ± 1.5

*The generation times are expressed in minutes.

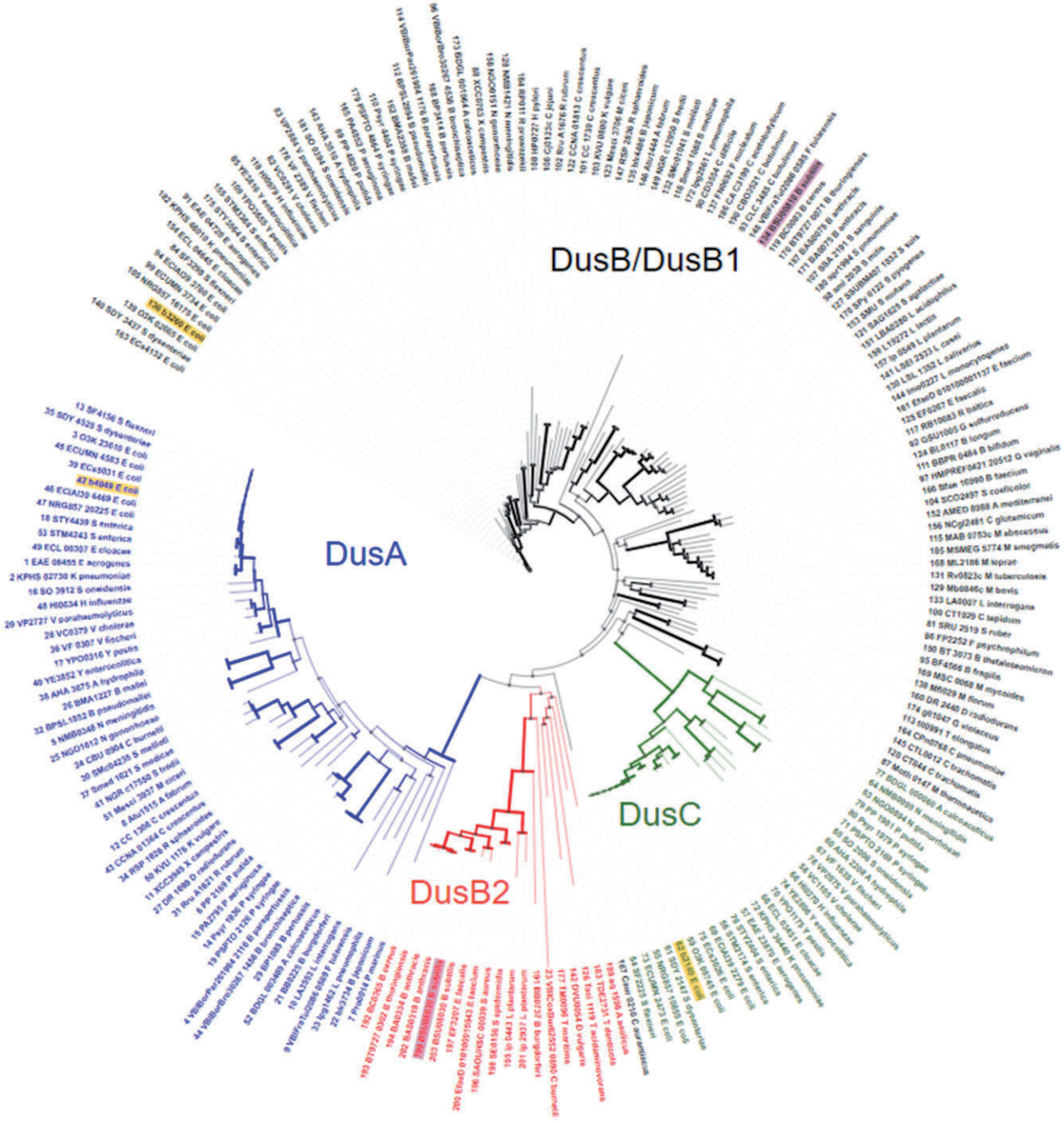


Figure 7. Phylogenetic analysis of DusB1, DusB2, DusC and DusA proteins in 120 reference and complete Bacteria. DusA proteins are in blue. DusC proteins are in green. DusB/DusB1 proteins are in black. DusB2 proteins are in red. The BV-BRC annotations seem to correctly group the proteins with one exception, the Caur_0210 protein annotated as DusB but clustering with the DusB2 proteins. This section of the tree has however very low bootstrap values as the thickness of the tree branches are reflective of the bootstrap percentage values. *E. coli* proteins are highlighted in yellow and *B. subtilis* proteins in purple.

tive than those catalyzing other D (22) and this has indeed remained verified again with *B. subtilis* enzymes. A recent analysis of dihydrouridylation in the *S. pombe* transcriptome found instances where the dependence of several tRNA sites on Dus enzymes couldn't be statistically determined (23). This suggests either the necessity of a D-site at a position for modifying another site or rare cases where multiple Dus enzymes target the same site. This could imply broader dihydrouridylation redundancy among Dus enzymes, requiring further clarification.

The physiological significance of this redundancy in *B. subtilis* raises intriguing questions. In general, homolog-based functional redundancy can provide functional resilience or flexibility to cope with varying conditions or stresses (53–55). This could indeed apply to *Bs*DusB taking the advantage of having one enzyme more efficient than the other, especially when dealing with redox reactivity issues. It is tempting to propose that this backup functionality could be a more efficient way to dihydrouridylate tRNAs under conditions or

events leading to significant tRNA damage requiring rapid maturation of newly transcribed pools of tRNAs to afford the cell to cope with abrupt environmental changes notably under limiting NADPH concentration for example. In such a scenario, up-regulation of *Bs*DusB2 could also be an additional mean by which the cell boosts tRNA-dihydrouridylation activity but also extends its site specificity to compensate for the low *Bs*DusB1 activity. Interestingly, such type of regulation has precedent as exemplified by the downregulation of the gene coding for the mesophilic *Clostridium botulinum* DusB homolog during a heat shock stress at 45°C (56). In that specific case, D has probably no utility at high temperatures, and thus this bacterium would naturally require less D and would therefore decrease the expression of its cognate enzyme. DusC is also differentially regulated in response to the growth temperature in the thermophilic *B. manusensis* (57). In *B. subtilis*, our studies revealed a visible impact of the absence of *Bs*DusB1 or *Bs*DusB2 on the growth phenotype of this organism (Table 2), suggesting that loss of D can have significant effects on cell physiology.

Another speculative yet intriguing possibility for this functional redundancy is related to the evolutionary process of these enzymes. Both *Bs*DusB1 and *Bs*DusB2 originated from a duplication event of an ancestral Dus enzyme likely multisite specific. *Bs*DusB1 has retained the functional features of this ancestral enzyme, while *Bs*DusB2 might be undergoing a process of functional speciation. This could explain why *Bs*DusB2's dihydrouridylation activity at position 17 is detectable only under high enzyme concentration (Figure 6). Comparative analysis of the presumed tRNA binding interfaces on *Bs*DusB models suggests that the *Bs*DusB2 interface is clearly different from the others with nonetheless some positive charges that remain common to these enzyme systems (Figure 5D). This agrees with the fact that *Bs*DusB2 may preferentially bind its tRNAs according to its own recognition mode. Previous phylogenetic analyses proposed that DusB/DusB1 was the common ancestor to all bacterial Dus proteins (34), a finding that we reproduce in a small-scale analysis with reference bacterial genomes (Figure 7). While the exact timing of the branching of the DusB2 subgroup from the DusB group remains uncertain, it is clearly distinct from both the DusA and DusC subgroups. Further comprehensive phylogenetic analyses would be needed to understand this evolutionary relationship. However, this data suggests that *Bs*DusB2 might be converging towards DusA-type activities, specifically modifying the 20/20a position, while potentially losing its capacity to modify the U17 positions like its DusB homologues. This suggests a possible transitional state in the evolution of *Bs*DusB2's enzymatic specificity.

Data availability

All NGS data associated to this manuscript are deposited and made publicly available in ENA (The European Bioinformatics Institute EMBL-EBI) under the accession number PRJEB74134.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors thank all past and present members of their laboratories who participated in the dihydrouridine project. We would also like to thank Chastanet's lab (INRAE, Jouy en Josas, France) for the generous gift of strain W168 and Ciaran Condon (IBPC, Paris, France) for the pDG148 plasmid.

Funding

Deutsche Forschungsgemeinschaft [439669440-TRR-319-TP C01 and HE 3397/21-1]; Agence Nationale de la Recherche [20-CE92-0030]; National Institute of Health [GM132254]. Funding for open access charge: ANR [20-CE92-0030].

Conflict of interest statement

None declared.

References

1. Shepherd,J. and Ibba,M. (2015) Bacterial transfer RNAs. *FEMS Microbiol. Rev.*, **39**, 280–300.
2. Berg,M.D. and Brandl,C.J. (2021) Transfer RNAs: diversity in form and function. *RNA Biol.*, **18**, 316–339.
3. Sekulovski,S. and Trowitzsch,S. (2022) Transfer RNA processing - from a structural and disease perspective. *Biol. Chem.*, **403**, 749–763.
4. Boccaletto,P., Stefaniak,F., Ray,A., Cappannini,A., Mukherjee,S., Purta,E., Kurkowska,M., Shirvanizadeh,N., Destefanis,E., Groza,P., *et al.* (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.*, **50**, D231–D235.
5. Helm,M. and Motorin,Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.
6. Agris,P.F., Eruysal,E.R., Narendran,A., Vare,V.Y.P., Vangaveti,S. and Ranganathan,S.V. (2018) Celebrating wobble decoding: half a century and still much is new. *RNA Biol.*, **15**, 537–553.
7. Barraud,P. and Tisné,C. (2019) To be or not to be modified: miscellaneous aspects influencing nucleotide modifications in tRNAs. *IUBMB Life*, **71**, 1126–1140.
8. 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.
9. Agris,P.F. (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.*, **9**, 629–635.
10. de Crécy-Lagard,V. and Jaroch,M. (2021) Functions of bacterial tRNA modifications: from ubiquity to diversity. *Trends Microbiol.*, **29**, 41–53.
11. Motorin,Y. and Helm,M. (2010) tRNA stabilization by modified nucleotides. *Biochemistry*, **49**, 4934–4944.
12. Lorenz,C., Lunse,C.E. and Morl,M. (2017) tRNA modifications: impact on structure and thermal adaptation. *Biomolecules*, **7**, 35.
13. Hori,H. (2019) Regulatory factors for tRNA modifications in extreme- thermophilic bacterium *Thermus thermophilus*. *Front. Genet.*, **10**, 204.
14. Dalluge,J.J., Hashizume,T., Sopchik,A.E., McCloskey,J.A. and Davis,D.R. (1996) Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic Acids Res.*, **24**, 1073–1079.
15. Dyubankova,N., Sochacka,E., Kraszewska,K., Nawrot,B., Herdewijn,P. and Lescrinier,E. (2015) Contribution of dihydrouridine in folding of the D-arm in tRNA. *Org. Biomol. Chem.*, **13**, 4960–4966.
16. Dalluge,J.J., Hamamoto,T., Horikoshi,K., Morita,R.Y., Stetter,K.O. and McCloskey,J.A. (1997) Posttranscriptional modification of tRNA in psychrophilic bacteria. *J. Bacteriol.*, **179**, 1918–1923.

17. Kato,T., Daigo,Y., Hayama,S., Ishikawa,N., Yamabuki,T., Ito,T., Miyamoto,M., Kondo,S. and Nakamura,Y. (2005) A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer Res.*, **65**, 5638–5646.
18. Bishop,A.C., Xu,J., Johnson,R.C., Schimmel,P. and de Crécy-Lagard,V. (2002) Identification of the tRNA-dihydrouridine synthase family. *J. Biol. Chem.*, **277**, 25090–25095.
19. Xing,F., Martzen,M.R. and Phizicky,E.M. (2002) A conserved family of *Saccharomyces cerevisiae* synthases effects dihydrouridine modification of tRNA. *RNA*, **8**, 370–381.
20. Xing,F., Hiley,S.L., Hughes,T.R. and Phizicky,E.M. (2004) The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.*, **279**, 17850–17860.
21. Yu,F., Tanaka,Y., Yamashita,K., Suzuki,T., Nakamura,A., Hirano,N., Yao,M. and Tanaka,I. (2011) Molecular basis of dihydrouridine formation on tRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 19593–19598.
22. Brégeon,D., Pecqueur,L., Toubdji,S., Sudol,C., Lombard,M., Fontecave,M., de Crecy-Lagard,V., Motorin,Y., Helm,M. and Hamdane,D. (2022) Dihydrouridine in the transcriptome: new life for this ancient RNA chemical modification. *ACS Chem. Biol.*, **17**, 1638–1657.
23. Finet,O., Yague-Sanz,C., Marchand,F. and Hermand,D. (2022) The dihydrouridine landscape from tRNA to mRNA: a perspective on synthesis, structural impact and function. *RNA Biol.*, **19**, 735–750.
24. Rider,L.W., Ottosen,M.B., Gattis,S.G. and Palfey,B.A. (2009) Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *J. Biol. Chem.*, **284**, 10324–10333.
25. Dai,W., Li,A., Yu,N.J., Nguyen,T., Leach,R.W., Wuhr,M. and Kleiner,R.E. (2021) Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat. Chem. Biol.*, **17**, 1178–1187.
26. Finet,O., Yague-Sanz,C., Kruger,L.K., Tran,P., Migeot,V., Louski,M., Nevers,A., Rougemaille,M., Sun,J., Ernst,F.G.M., *et al.* (2022) Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Mol. Cell*, **82**, 404–419.
27. Draycott,A.S., Schaening-Burgos,C., Rojas-Duran,M.F., Wilson,L., Scharfen,L., Neugebauer,K.M., Nachtergaele,S. and Gilbert,W.V. (2022) Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA. *PLoS Biol.*, **20**, e3001622.
28. McCown,P.J., Ruszkowska,A., Kunkler,C.N., Breger,K., Hulewicz,J.P., Wang,M.C., Springer,N.A. and Brown,J.A. (2020) Naturally occurring modified ribonucleosides. *Wiley Interdiscip. Rev. RNA*, **11**, e1595.
29. Sun,H., Li,K., Liu,C. and Yi,C. (2023) Regulation and functions of non-m(6)A mRNA modifications. *Nat. Rev. Mol. Cell. Biol.*, **24**, 714–731.
30. Bou-Nader,C., Pecqueur,L., Bregéon,D., Kamah,A., Guérineau,V., Golinelli-Pimpaneau,B., Guimaraes,B.G., Fontecave,M. and Hamdane,D. (2015) An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic Acids Res.*, **43**, 9446–9456.
31. Bou-Nader,C., Montemont,H., Guérineau,V., Jean-Jean,O., Brégeon,D. and Hamdane,D. (2018) Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: perspectives on the evolution scenario. *Nucleic Acids Res.*, **46**, 1386–1394.
32. Kusuba,H., Yoshida,T., Iwasaki,E., Awai,T., Kazayama,A., Hirata,A., Tomikawa,C., Yamagami,R. and Hori,H. (2015) In vitro dihydrouridine formation by tRNA dihydrouridine synthase from *Thermus thermophilus*, an extreme-thermophilic eubacterium. *J. Biochem.*, **158**, 513–521.
33. Faivre,B., Lombard,M., Fakroun,S., Vo,C.D., Goyenvalle,C., Guérineau,V., Pecqueur,L., Fontecave,M., De Crécy-Lagard,V., Brégeon,D., *et al.* (2021) Dihydrouridine synthesis in tRNAs is under reductive evolution in Mollicutes. *RNA Biol.*, **18**, 2278–2289.
34. Kasprzak,J.M., Czerwoniec,A. and Bujnicki,J.M. (2012) Molecular evolution of dihydrouridine synthases. *BMC Bioinformatics*, **13**, 153.
35. Koo,B.M., Kritikos,G., Farelli,J.D., Todor,H., Tong,K., Kimsey,H., Wapinski,I., Galardini,M., Cabal,A., Peters,J.M., *et al.* (2017) Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.*, **4**, 291–305.
36. Holland,R.A. (2023) A sequence- and ligation-independent cloning (SLIC) procedure for the insertion of genes into a plasmid vector. *Methods Mol. Biol.*, **2633**, 25–32.
37. Joseph,P., Fantino,J.R., Herbaud,M.L. and Denizot,F. (2001) Rapid orientated cloning in a shuttle vector allowing modulated gene expression in *Bacillus subtilis*. *FEMS Microbiol. Lett.*, **205**, 91–97.
38. Thuring,K., Schmid,K., Keller,P. and Helm,M. (2016) Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods*, **107**, 48–56.
39. Olson,R.D., Assaf,R., Brettin,T., Conrad,N., Cucinell,C., Davis,J.J., Dempsey,D.M., Dickerman,A., Dietrich,E.M., Kenyon,R.W., *et al.* (2023) Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.*, **51**, D678–D689.
40. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
41. Letunic,I. and Bork,P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
42. Marchand,V., Ayadi,L., Ernst,F.G.M., Hertler,J., Bourguignon-Igel,V., Galvanin,A., Kotter,A., Helm,M., Lafontaine,D.L.J. and Motorin,Y. (2018) AlkAniline-Seq: profiling of m(7) G and m(3) C RNA modifications at single nucleotide resolution. *Angew. Chem. Int. Ed. Engl.*, **57**, 16785–16790.
43. Bou-Nader,C., Brégeon,D., Pecqueur,L., Fontecave,M. and Hamdane,D. (2018) Electrostatic potential in the tRNA binding evolution of dihydrouridine synthases. *Biochemistry*, **57**, 5407–5414.
44. Byrne,R.T., Jenkins,H.T., Peters,D.T., Whelan,F., Stowell,J., Aziz,N., Kasatsky,P., Rodnina,M.V., Koonin,E.V., Konevega,A.L., *et al.* (2015) Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6033–6037.
45. Motorin,Y., Keith,G., Simon,C., Foiret,D., Simos,G., Hurt,E. and Grosjean,H. (1998) The yeast tRNA: pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA*, **4**, 856–869.
46. Motorin,Y. and Grosjean,H. (1999) Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast *saccharomyces cerevisiae*: identification of the gene and substrate specificity of the enzyme. *RNA*, **5**, 1105–1118.
47. Pintard,L., Lecointe,F., Bujnicki,J.M., Bonnerot,C., Grosjean,H. and Lapeyre,B. (2002) Trm7p catalyses the formation of two 2 '-O-methylriboses in yeast tRNA anticodon loop. *EMBO J.*, **21**, 1811–1820.
48. Hur,S. and Stroud,R.M. (2007) How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Mol. Cell*, **26**, 189–203.
49. Awai,T., Kimura,S., Tomikawa,C., Ochi,A., Ihsanawati Bessho,Y., Yokoyama,S., Ohno,S., Nishikawa,K., Yokogawa,T., *et al.* (2009) *Aquifex aeolicus* tRNA (N-2, N-2-Guanine)-dimethyltransferase (Trm1) catalyzes transfer of methyl groups not only to Guanine 26 but also to guanine 27 in tRNA. *J. Biol. Chem.*, **284**, 20467–20478.
50. Roovers,M., Wouters,J., Bujnicki,J.M., Tricot,C., Stalon,V., Grosjean,H. and Droogmans,L. (2004) A primordial RNA modification enzyme: the case of tRNA (m(1)A) methyltransferase. *Nucleic Acids Res.*, **32**, 465–476.
51. Hamdane,D., Guelorget,A., Guérineau,V. and Golinelli-Pimpaneau,B. (2014) Dynamics of RNA modification by

a multi-site-specific tRNA methyltransferase. *Nucleic Acids Res.*, **42**, 11697–11706.
52. Kawamura,T., Hirata,A., Ohno,S., Nomura,Y., Nagano,T., Nameki,N., Yokogawa,T. and Hori,H. (2016) Multisite-specific archaeosine tRNA-guanine transglycosylase (ArcTGT) from *thermoplasma acidophilum*, a thermo-acidophilic archaeon. *Nucleic Acids Res.*, **44**, 1894–1908.
53. Ghosh,S. and O'Connor,T.J. (2017) Beyond paralogs: the multiple layers of redundancy in bacterial pathogenesis. *Front. CelL. Infect. Microbiol.*, **7**, 467.
54. Louca,S., Polz,M.F., Mazel,F., Albright,M.B.N., Huber,J.A., O'Connor,M.I., Ackermann,M., Hahn,A.S., Srivastava,D.S., Crowe,S.A., *et al.* (2018) Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.*, **2**, 936–943.
55. Fenech,E.J., Ben-Dor,S. and Schuldiner,M. (2020) Double the fun, Double the trouble: paralogs and homologs functioning in the endoplasmic reticulum. *Annu. Rev. Biochem.*, **89**, 637–666.
56. Liang,W.D., Bi,Y.T., Wang,H.Y., Dong,S., Li,K.S. and Li,J.S. (2013) Gene expression profiling of *Clostridium botulinum* under heat shock stress. *Biomed. Res. Int.*, **2013**, 760904.
57. Sun,Q.L., Sun,Y.Y., Zhang,J., Luan,Z.D., Lian,C., Liu,S.Q. and Yu,C. (2019) High temperature-induced proteomic and metabolomic profiles of a thermophilic *Bacillus manusensis* isolated from the deep-sea hydrothermal field of Manus Basin. *J. Proteomics*, **203**, 103380.

Received: August 31, 2023. **Revised:** March 26, 2024. **Editorial Decision:** April 5, 2024. **Accepted:** April 24, 2024

PhD project and goals

During my PhD, I focused on ribosomal dihydrouridylation in *E. coli*. My research was organized into four main areas:

1-Developing a specific detection technique for D.

2-Conducting genetic studies using *E. coli* deletion strains to identify the gene responsible for ribosomal dihydrouridine synthase by several D-detection technique, MALDI-Mass spectrometry with the help of Professor Stephen Douthwaite and Simon Rose. Alkalinine-Seq with Professor Yuri Motorin's team of the University of Lorraine.

3-Biochemical characterization of the newly discovered E. coli dihydrouridine synthase. (Toubdji et al., 2024).

At the end of this part of the project, two questions remained unsolved, which led to the next chapters of my research.

- What is the enzymatic mechanism of the dihydrouridylation reaction of the novel ribosomal dihydrouridine synthase RdsA?

- Are D-2449 and/or the dihydrouridine synthase RdsA implicated in the translation process of the ribosome?

4-Investigating the biological significance of ribosomal dihydrouridylation and dihydrouridine synthase.

These efforts aimed to deepen our understanding of how dihydrouridine modification influences ribosomal function and cellular processes in *E. coli.* To this end, I used polysome profiling analysis and ribosome Cryo-Em with the help of Nicolas Leulliot and Bruno Bonettaz.

Results

Chapter II: “Exploring a Novel Class of Flavoenzymes: Identification and Biochemical Characterization of Ribosomal RNA Dihydrouridine Synthase”.

1. Preamble

The identification of a gene encoding an RNA modification enzyme begins with the deletion of this gene from the genome and the confirmation that the corresponding modification disappears in the RNA isolated from this strain. While this step is crucial, it is not sufficient to unequivocally prove that the gene in question is directly responsible for the biosynthesis of the modification. To establish the function of this gene, it is necessary to express the recombinant protein and demonstrate in vitro that this enzyme is active in the presence of the appropriate RNA substrate (usually an in vitro transcript) as well as any necessary cofactors or additional substrates. The recombinant enzyme must show in vitro activity and act directly on the target nucleotide. This method is generally applicable to tRNA modification enzymes but becomes more complex when dealing with rRNA modification enzymes. The additional challenges include the size of the substrate and the multitude of factors to consider, such as ribosomal proteins or a particular intermediate ribosomal assembly.

In the exploration of ribosomal RNA modifications, dihydrouridine (D) stands out due to its evolutionary conservation and prevalence within the transcriptome, especially in tRNAs and, to a lesser extent, mRNAs. Recent studies have identified D at position 2449 in the 23S rRNA of *E. coli*, near the critical peptidyl transferase site of the ribosome. Despite knowing the presence of dihydrouridine in various RNA species, the enzyme responsible for its incorporation into rRNA has eluded researchers until now.

Following my arrival to the lab of Dr Jean-Jean as a Master student in 2021, and soon after as a PhD student, I started tackling this gap by introducing a rapid and specific method for detecting dihydrouridine in rRNA, named Rho-RT-PCR. This method involves the selective blocking of reverse transcriptase at rhodamine-labeled D2449, followed by PCR amplification. The D residues of the purified rRNA are subjected to reductive cleavage using sodium borohydride (NaBH4) in an alkaline medium. This step opens the non-aromatic dihydrouracil

ring, producing a 3-ureidopropanol moiety attached to the ribose at C-1' position (Finet, Yague-Sanz, Krüger, et al., 2022) (Betteridge et al., 2007).

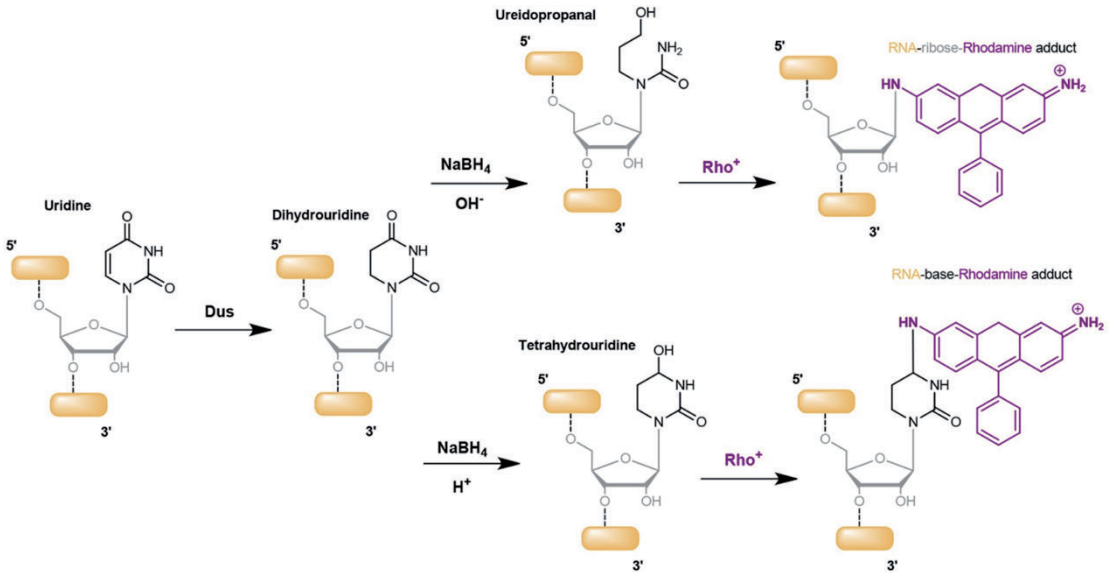


Figure 50: Chemical Labeling Reaction of Dihydrouridine with Rhodamine-110, adapted from (Brégeon et al., 2022)**.**

This 3-ureidopropanol moiety serves as a reactive intermediate that can be readily replaced by a fluorophore bearing a primary amine group. Common fluorophores used include rhodamine 110 and proflavin, which bind to the opened dihydrouracil ring, enabling the fluorescent labeling of the rRNA. This final adduct causes steric hindrance in the active site of the reverse transcriptase enzyme which can be revealed by PCR, using specific primers that amplifies targeted region.

Utilizing this innovative approach, we conducted a thorough investigation of rRNA from various *E. coli* strains, including those with chromosomal deletions and single-gene knockouts.

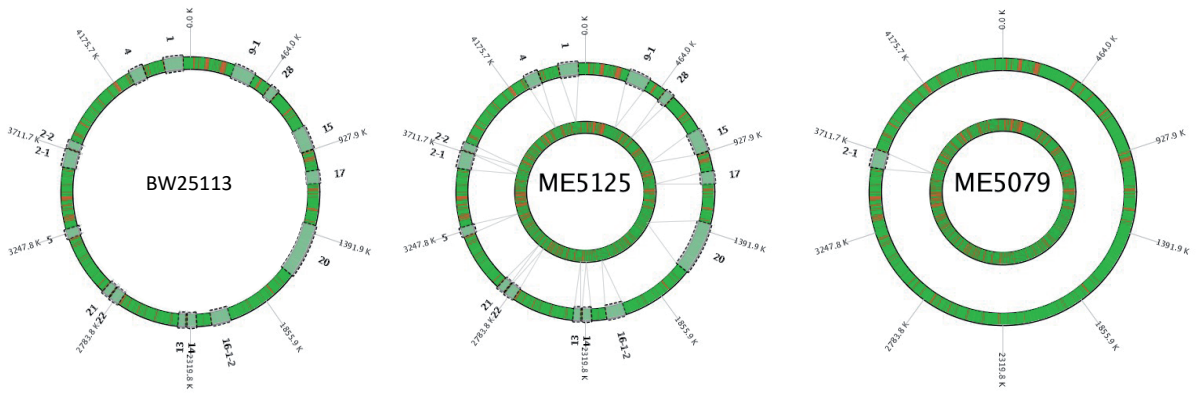


Figure 51: Chromosomal DNA of *E. coli* strain. wild-type BW 25113 with the complete genome, ME5125 harboring 15 major deletion and ME5079 harboring one genomic deletion.

Our findings pinpoint the *yhiN* gene as the ribosomal dihydrouridine synthase, which we have designated as RdsA for **R**ibosomal **d**ihydrourdine **s**ynthase **A.**

Biochemical characterizations reveal that RdsA represents a novel class of flavoenzymes, utilizing FAD and NADH for its redox-dependent catalytic activity. This enzyme efficiently catalyzes the dihydrouridylation of a short rRNA transcript that mimics the peptidyl transferase center's local structure, suggesting that this modification occurs early, possibly prior to ribosome assembly. Phylogenetic analysis further demonstrates the widespread distribution of the *yhiN* gene across the bacterial kingdom, highlighting the evolutionary conservation of rRNA dihydrouridylation.

The implications of these findings are significant. First, they provide a deeper understanding of the biosynthetic pathways involved in rRNA modification, specifically the elusive dihydrouridylation process. Second, they underscore the critical role of dihydrouridine in ribosomal function, potentially influencing ribosome biogenesis and protein synthesis. Finally, the discovery of RdsA opens new avenues for exploring the biological significance of rRNA modifications and their potential as targets for novel antibiotic therapies.

2. Outlook

Following article 4 it would be verry interesting to test RdsA mutant to further elucidate its mechanism. Since the ribosomal dihydrouridine do not have a cysteine as is the case for classical Dus of tRNA, we identified E338 residue which could be involved in the hydride transferer and act as a cysteine. So, we designed two mutants E338-A and E338-Cys.

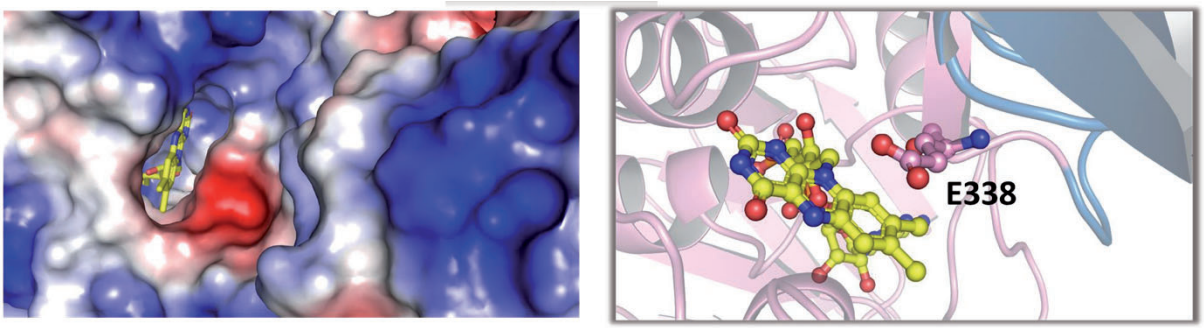


Figure 52: structural representation of E. coli RdsA holoprotein generated from the AlphaFold model. showing the binding of the FAD coenzyme (ball sticks, yellow) at the Rossman fold domain.

A major question that remains unanswered is how RdsA binds rRNA? In the future, efforts should be focused on resolving this issue. Mutant were designed targeting the three positively charged patches likely involved in rRNA binding.

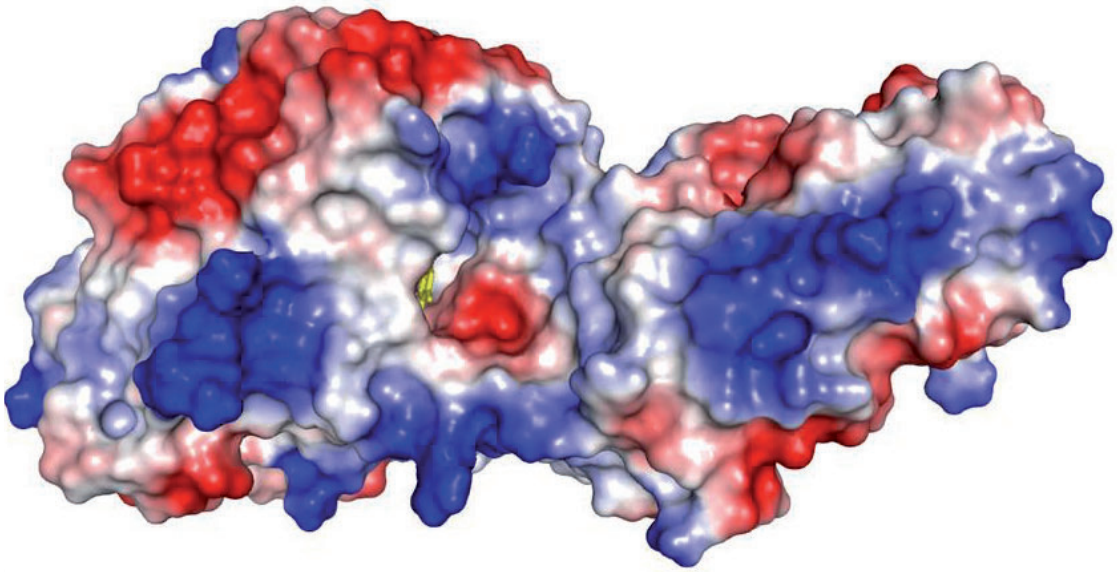


Figure 53: Electrostatic surface representation of *E. coli* RdsA generated from the AlphaFold model.

Furthermore, it would be interesting to test the activities of all this RdsA mutants on a plethora of substrates. One strategy would be to reconstitute the entire domain V of *E. coli* 23S rRNA in vitro, or/and to use small molecules. We are currently investigating this issue.

PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Exploring a unique class of flavoenzymes: Identification and biochemical characterization of ribosomal RNA dihydrouridine synthase

Sabrine Toubdji[a,b], Quentin Thullier[c,d], Lea-Marie Kilz[e], Virginie Marchand[c,d], Yifeng Yuan[f], Claudia Sudol[a,b], Catherine Goyenvalle[a], Olivier Jean-Jean[a], Simon Rose[g], Stephen Douthwaite[g], Léo Hardy[h], Zeynep Baharoglu[h], Valérie de Crécy-Lagard[f,i], Mark Helm[e], Yuri Motorin[c,d], Djemel Hamdane[b,1], and Damien Brégeon[a,1]

Affiliations are included on p. 8.

Edited by Paul Schimmel, Scripps Research Florida, Jupiter, FL; received January 29, 2024; accepted June 20, 2024

Dihydrouridine (D), a prevalent and evolutionarily conserved base in the transcriptome, primarily resides in tRNAs and, to a lesser extent, in mRNAs. Notably, this modification is found at position 2449 in the *Escherichia coli* 23S rRNA, strategically positioned near the ribosome's peptidyl transferase site. Despite the prior identification, in *E. coli* genome, of three dihydrouridine synthases (DUS), a set of NADPH and FMN-dependent enzymes known for introducing D in tRNAs and mRNAs, characterization of the enzyme responsible for D2449 deposition has remained elusive. This study introduces a rapid method for detecting D in rRNA, involving reverse transcriptase-blockage at the rhodamine-labeled D2449 site, followed by PCR amplification (RhoRT-PCR). Through analysis of rRNA from diverse *E. coli* strains, harboring chromosomal or single-gene deletions, we pinpoint the *yhiN* gene as the ribosomal dihydrouridine synthase, now designated as RdsA. Biochemical characterizations uncovered RdsA as a unique class of flavoenzymes, dependent on FAD and NADH, with a complex structural topology. In vitro assays demonstrated that RdsA dihydrouridylates a short rRNA transcript mimicking the local structure of the peptidyl transferase site. This suggests an early introduction of this modification before ribosome assembly. Phylogenetic studies unveiled the widespread distribution of the *yhiN* gene in the bacterial kingdom, emphasizing the conservation of rRNA dihydrouridylation. In a broader context, these findings underscore nature's preference for utilizing reduced flavin in the reduction of uridines and their derivatives.

dihydrouridine | rRNA | dihydrouridine synthase | RNA modification | flavoenzyme

Significance

Posttranscriptional modifications of rRNA play a crucial role in ribosome biogenesis and fine-tuning of protein synthesis. These are predominantly concentrated at decoding and peptidyl transferase center (PTC) sites. To date, 35 of the 36 identified ribosomal modifications have been characterized, although the biosynthetic pathway for U2449 dihydrouridylation within the PTC has remained elusive. To fill this gap in knowledge, we developed a technique for monitoring dihydrourylation reaction of 23S rRNA. This method involves chemical labeling of dihydrouridine and RT-PCR blockage. By integrating this approach with epitranscriptomic, genetic, and biochemical methodologies, we successfully identified a unique class of flavin-dependent enzymes involved in rRNA dihydrouridylation. This finding opens up different avenues for exploring the biological significance of PTC dihydrouridylation in ribosomal function.

Posttranscriptional maturation processes shape RNA molecules into functional biomolecules. Among these processes, the incorporation of chemical groups at the nucleoside level, primarily within the bases, is catalyzed by modification enzymes. Over 170 modifications have been identified to date, with tRNAs and rRNAs being the most extensively modified RNA species (1). The chemical diversity of these modifications can significantly alter the physicochemical properties of nucleosides, exerting a profound impact on local and even global RNA structures. For the majority of modifications, the effect is stabilizing in nature (2, 3). For instance, the introduction of methyl groups tends to strengthen base stacking. One of the most abundant modifications across the transcriptome is dihydrouridine (D) (4, 5) (Fig. 1*A*). This modified base is evolutionarily conserved and is predominantly found in bacterial and eukaryotic tRNAs. Recently, D has also been detected in certain eukaryotic mRNAs (6–8) and long noncoding RNAs (6, 7, 9). In the case of rRNAs, sporadic observations of D have been reported, notably at position 2449 (D2449) in domain V of the 23S ribosomal RNA of *Escherichia coli*, in the peptidyl transferase center (PTC) (10, 11) (Fig. 1 *B* and *C*). D2449 has also been identified at positions 2449 and 2500 in *Clostridium sporogenes* (with D2449 being partially methylated to m^5D) (12) and at position 1211 or 1212 in the 16S rRNA of *Clostridium acetobutylicum* (13). Comprehensive studies to investigate the presence of dihydrouridine in the rRNAs of bacteria and eukaryotes remain to be made.

The study of D synthases involved in tRNA dihydrouridylation has revealed that the conversion of U to D is a redox-dependent reaction relying on NADPH as a reducing agent (4, 14–21) (Fig. 1*A*). Dus enzymes are flavoenzymes, utilizing FMN as a coenzyme for redox reactions. In the initial step, the flavin is reduced to an anionic hydroquinone ($FMNH^-$), which then transfers a hydride to the C5–C6 double bond of uridine, forming an enolate intermediate (4, 22). This intermediate accepts a proton from the strictly conserved catalytic cysteine, leading to the breakdown of D's aromaticity,

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: djemel.hamdane@college-de-france.fr or damien.bregeon@sorbonne-universite.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2401981121/-/DCSupplemental.

Published July 30, 2024.

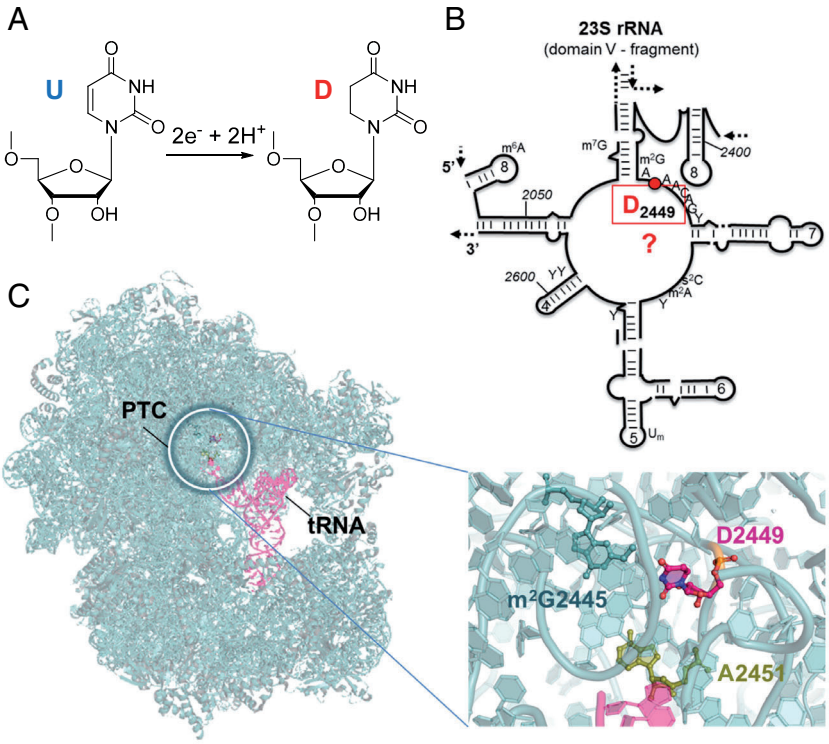


Fig. 1. Dihydrouridylation of the PTC of the *E. coli* ribosome: (*A*) Biosynthetic reaction of dihydrouridine involves the reduction of the C5=C6 double bond of uridine. This redox reaction requires 2 electrons + 2 protons. (*B*) Secondary structure of the V domain of *E. coli* 23S rRNA. D2449 is highlighted in red, and the question mark indicates that the associated modification enzyme was unknown before our current study. (*C*) High-resolution structure of the *E. coli* ribosome in complex with tRNA (magenta). The PDB associated with this structure is 8B0X (11). The *Inset* on the *Right* provides a close-up of the PTC, showing the positioning of D2449 (pink) in proximity to A2451 and m^2G2445.

resulting in a nonplanar base and a more flexible ribose conformation. The significance of D's base flexibility is highlighted by the fact that psychrophilic organisms generally have more D in their tRNAs than thermophiles (23). Structurally, D destabilizes small stem loops, promoting the formation of larger loops (24). The presence of D2449 in the central large loop of domain V in *E. coli*'s 23S rRNA could potentially serve this purpose, although addressing this issue is challenging due to the unidentified gene encoding ribosomal Dus. Beyond its presumed structural role, the biological significance of D remains obscure. Recent discoveries of D residues in yeast and human mRNAs, also introduced by tRNA dihydrouridine synthases, suggest potential roles in meiosis and translational efficiency through yet-to-be-determined molecular mechanisms (6, 8). These studies indicate that Dus enzymes are promiscuous, although, at least in the case of *E. coli*, they do not act on rRNA.

The *E. coli* rRNA encompasses 36 modifications, and to date, all modification enzymes for *E. coli* rRNA have been identified, except for the rRNA D synthase (1) (*SI Appendix*, Table S1). In this study, we identified this elusive enzyme, hereafter referred to as RdsA (Ribosomal Dihydrouridine Synthase A), responsible for the biosynthesis of D2449 in 23S rRNA. Biochemical characterizations of RdsA unveiled its identity as a flavin-dependent D synthase with redox activity reliant on FAD and NADH. The proposed chemical mechanism of dihydrouridylation, derived from our findings, outlines a likely hydride transfer mechanism, underscoring nature's preference for utilizing flavin to reduce uridine or its derivatives. Moreover, our phylogenetic data demonstrate that the RdsA gene is widely distributed among bacteria, including many pathogens, while absent in eukaryotes. This observation suggests that ribosomal D may be pervasive in microbial rRNAs, rendering it a potential target for selective antibiotic intervention.

Results

Development of a Rapid D-ribosomal Detection Test. To facilitate the swift detection of D2449, we have developed a methodology termed RhoRT-PCR. This innovative approach involves the blocking of reverse transcriptase (RT) through dihydrouridine labeling with rhodamine 110, following a well-established mechanism (*SI Appendix*, Fig. S1). Subsequently, this step was followed by PCR amplification of RT-products employing a primer pair to target a specific region (Fig. 2*A*). The primer pair 2F/3R was meticulously designed to amplify the region downstream of position 2449 in the 23S rRNA, yielding a 248 bp fragment. It functions as a positive control for RT activity. Conversely, the region encompassing position 2449 was amplified using primers 1F/3R, generating a 429 bp product. Therefore, the absence of rhodamine labeling on D2449 or the lack of dihydrouridylation at position 2449 (indicative of U2449) should result in the presence of both 248 and 429 bp fragments, discernible on an agarose gel. In contrast, the presence of D2449 would result solely in the 248 bp band. The validity of this approach was tested using purified rRNA from wild-type strains of *E. coli* (known to contain D2449) and *Bacillus subtilis* (known to harbor a uridine at the corresponding rRNA position). As depicted in Fig. 2*B*, in the case of *E. coli*, the absence of Rho labeling showed the anticipated two fragments. However, prelabeling D2449 with Rho resulted in the disappearance of the 429 bp band, strongly indicating the presence of D2449. Conversely, in the case of *B. subtilis*, whether labeled or not, both bands persisted, suggesting the absence of D in this region of rRNA. Recently, it has been demonstrated that the three Dus enzymes in *E. coli* (DusA, B, and C) are not responsible for the synthesis of D2449 (7). To confirm this result, we applied our RhoRT-PCR method to purified rRNA from the *E. coli* triple

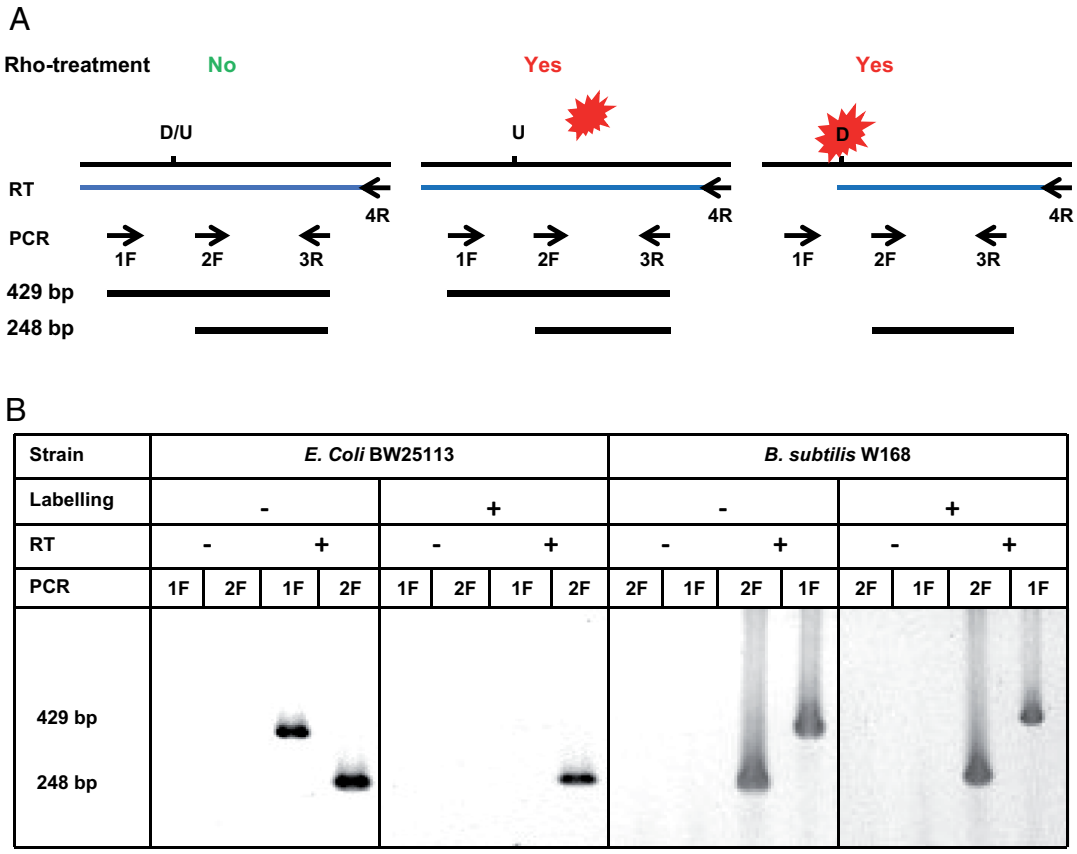


Fig. 2. Principle of RhoRT-PCR. (*A*) Diagram showing the main steps of RhoRT-PCR and the expected results depending on the presence of D in the analyzed region. The red scrambled oval depicts specific D-labeling with rhodamine-110. (*B*) Representative results obtained with rRNA containing D at position 2449 (*E. coli* BW25113, 14 repeats) or U at the same position (*B. subtilis* W168, 2 repeats). RT was performed with species-specific 4R oligonucleotides. PCR amplification was performed with species-specific 1F/3R (1F) and 2F/3R (2F) oligonucleotides.

mutant strain *ΔdusABC*. As shown in *SI Appendix*, Fig. S2*A*, the result revealed a single band at 248 bp, clearly demonstrating the presence of D2449 in this mutant. This indicates that the classical Dus enzymes are not involved in the dihydrouridylation of *E. coli* 23S rRNA. Thus, the RhoRT-PCR technique emerges as an effective method for the precise detection of D2449.

Identification of the Gene Encoding rRNA Dihydrouridine Synthase. In the quest to pinpoint the gene responsible for dihydrouridine synthesis in rRNA, we applied the RhoRT-PCR technique to rRNA extracted from the *E. coli* strain ME5125 (25, 26). This strain is distinguished by 15 major deletions in its chromosomal DNA, making it an ideal candidate due to its ability to maintain the smallest genome among *E. coli* strains while retaining sufficient viability for rRNA isolation. As illustrated in Fig. 3*A*, the deletions exhibited both fragments, whether treated with rhodamine or not, strongly suggesting the probable absence of D. To substantiate these findings, we subjected BW25113 and ME5125 rRNA to MALDI mass spectrometry after RNAse T1 treatment. The analysis revealed the ^{2448}ADAACAGp^{2454} fragment with an *m/z* of 2293.3 in the case of BW25113, while in ME5125, the corresponding fragment displayed an *m/z* of 2291.3, indicative of the non-dihydrouridylated fragment (Fig. 3*B*). Moreover, these results were validated by AlkAniline-Seq, underscoring the presence of D2449 in WT rRNA and its conspicuous absence in the ME5125 strain (Fig. 3*C*). Individual testing of the 15 deletion strains using RhoRT-PCR and AlkAniline-Seq showed that only the ME5079 strain, harboring a genomic deletion encompassing 77 genes (*SI Appendix*, Fig. S3), lacked D2449 (Fig. 3 *A* and *C*). A similar approach was used for the analysis of 10 strains bearing single-gene knockouts that were annotated with unknown functions but potentially linked to redox processes or ribosome biogenesis (*SI Appendix*, Table S2 and Fig. S2). Comprehensive testing across all these strains unequivocally demonstrated that only the rRNA from the ΔyhiN strain lacked D2449 (Fig. 3*C* and *SI Appendix*, Fig. S2*A*). Notably, *yhiN* is annotated as a putative oxidoreductase with an FAD/NAD(P)H domain and operates outside an operon (*SI Appendix*, Table S2). However, it is situated in proximity to two rRNA methyltransferase genes, namely *rsmJ* and *rlmJ*, and a putative ribosome-associated ATPase gene (*rbbA*) (*SI Appendix*, Table S2 and Fig. S3).

The Gene *yhiN* Encodes for an FAD Binding Protein. The *E. coli yhiN* gene encodes a 400-amino acid protein with a theoretical mass of 43.7 kDa (*SI Appendix*, Fig. S4). To elucidate the structural organization of YhiN, an AlphaFold model was generated (Fig. 4*A*). Remarkably, YhiN showcases a distinct fold compared to classical Dus proteins (4, 15–17, 19, 20). A PDB search revealed homology between YhiN and various flavoproteins, with crystallographic structures available for holoprotein forms (2GQF, 3V76, 4CNJ, 2I0Z), including *Streptococcus oligofermentes* L-aminoacetone oxidase. Structural alignments of the YhiN model with homologous structures revealed RMSD values below 2.3 Å (*SI Appendix*, Fig. S5), signifying significant structural conservation within this flavoprotein class. YhiN comprises three distinct domains with an intricately complex topology (Fig. 4*A*). The Rossmann fold domain (FBD), housing the flavin and putatively the catalytic site, accommodates the insertion of the other two domains between G190 and A337, resulting in the Nt and Ct ends being borne by this domain. The second domain (ins1D),

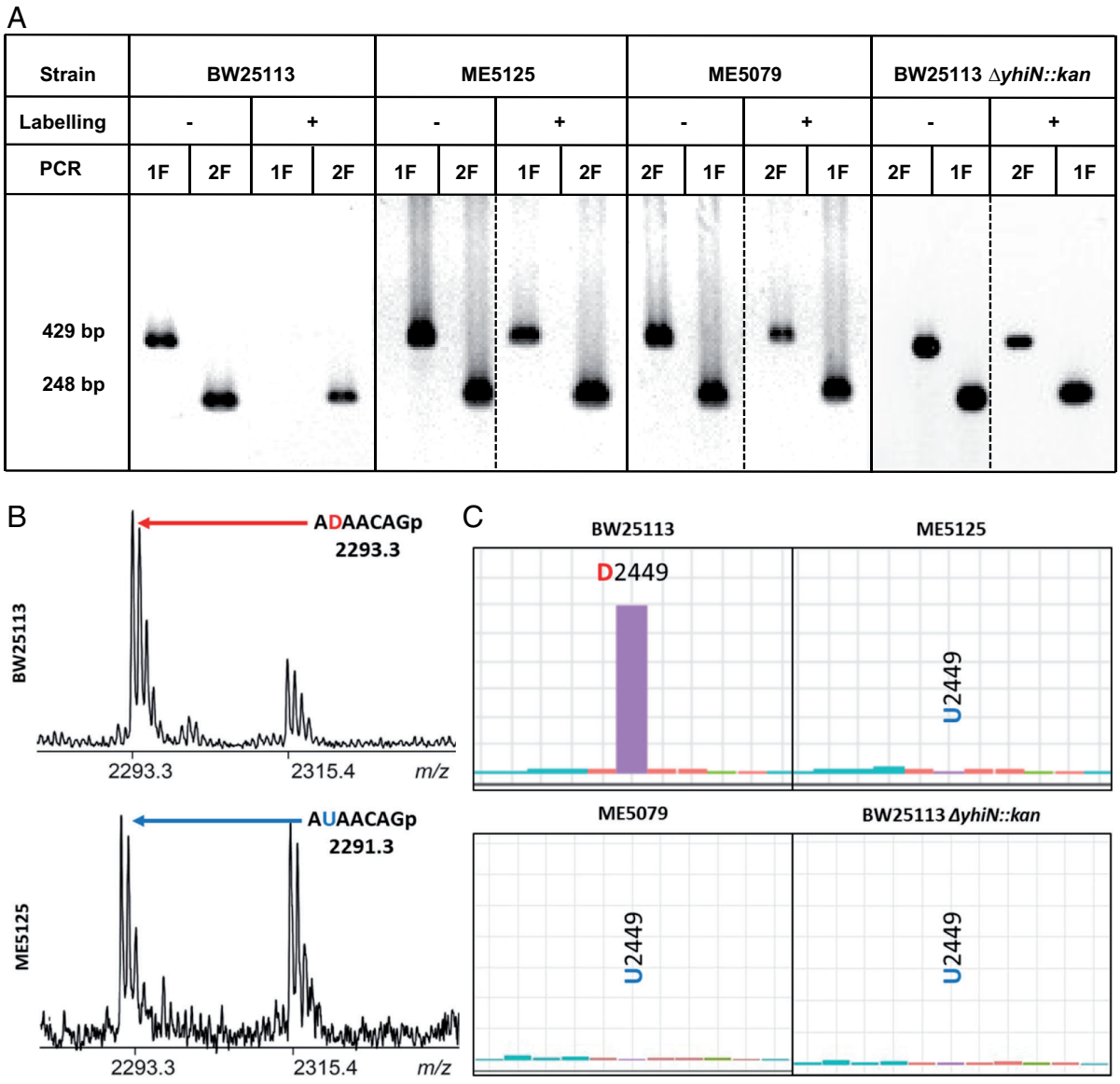


Fig. 3. Identification of D2449 deficient *E. coli* strains. (*A*) Representative RhoRT-PCR results with the indicated strains (BW25113 and ME5125, 14 repeats; ME5079 and BW25113 *ΔyhiN::kan*, 3 repeats). Primers used are the same as in Fig. 2*A*. The absence of 1F/3R amplification with BW25113 rRNA as a substrate indicates the presence of D2449, which was confirmed by MALDI-TOF analysis of RNAse digestion fragments (2 repeats) (*B*) and AlkAniline-Seq (*C*) (for the latest, BW25113, ME5125, and ME5079, 3 repeats; BW25113 *ΔyhiN::kan*, 2 repeats).

homologous to the Ct domain of *E. coli* EF-Tu, is interrupted at residue L257 by the insertion of the third domain (ins2D), exhibiting structural homology with the Helix-2-turn-helix domain of the topoisomerase VI-B subunit of *Sulfolobus shibatae*. Ins2D spans L258 to L322, succeeded by a segment of ins1D and FBD. Structural alignment of YhiN with flavoprotein homologs facilitated the identification of the conserved FAD binding site (Fig. 4*B* and *SI Appendix*, Fig. S5). No clashes were observed between residues forming the FAD pocket and the coenzyme. The FAD assumes an elongated conformation upheld by a series of interactions, with the isoalloxazine, the redox center, situated within a solvent-accessible crevice. To assess the flavoprotein nature of *E. coli* YhiN (RdsA), recombinant protein production was undertaken. As depicted in Fig. 4 *C* and *D*, the UV/visible spectrum of freshly purified YhiN indicated the presence of oxidized flavin in 61% of holoprotein. The addition of SDS induced coenzyme release, yielding a UV/visible spectrum akin to free FAD. Additionally, mutating key residues in this identified FAD binding pocket led to a lower ratio of holoproteins ranging from 37%, when affecting residues interacting with the adenine moiety, to almost 0% when affecting residues in the vicinity of the ribityl (Fig. 4*D*). These findings conclusively establish YhiN as a noncovalently bound FAD flavoprotein characterized by a sophisticated structural topology.

The Flavoprotein YhiN: Unraveling Its Role as the rRNA Dihydrouridine Synthase. The conversion of uridine to dihydrouridine is a redox reaction requiring 2 electrons + $2H^+$ (or hydride + H^+) (Fig. 1*A*) (4). In the case of classical Dus proteins, their RNA substrates are reduced by $FMNH^-$, produced through the prior reduction of FMN by NADPH (4, 22). Initially, we investigated whether NADPH could serve as a reducer of flavin by monitoring the spectrophotometric NADPH oxidase activity at 340 nm of recombinant YhiN. However, no NADPH consumption was observed. Conversely, we observed efficient NADH oxidation with a k_{cat} of 0.027 ± 0.003 s^{-1} and a K_M of 6.5 ± 1.5 μM (Fig. 4*E*). With a conclusive reducing source identified, our focus shifted to

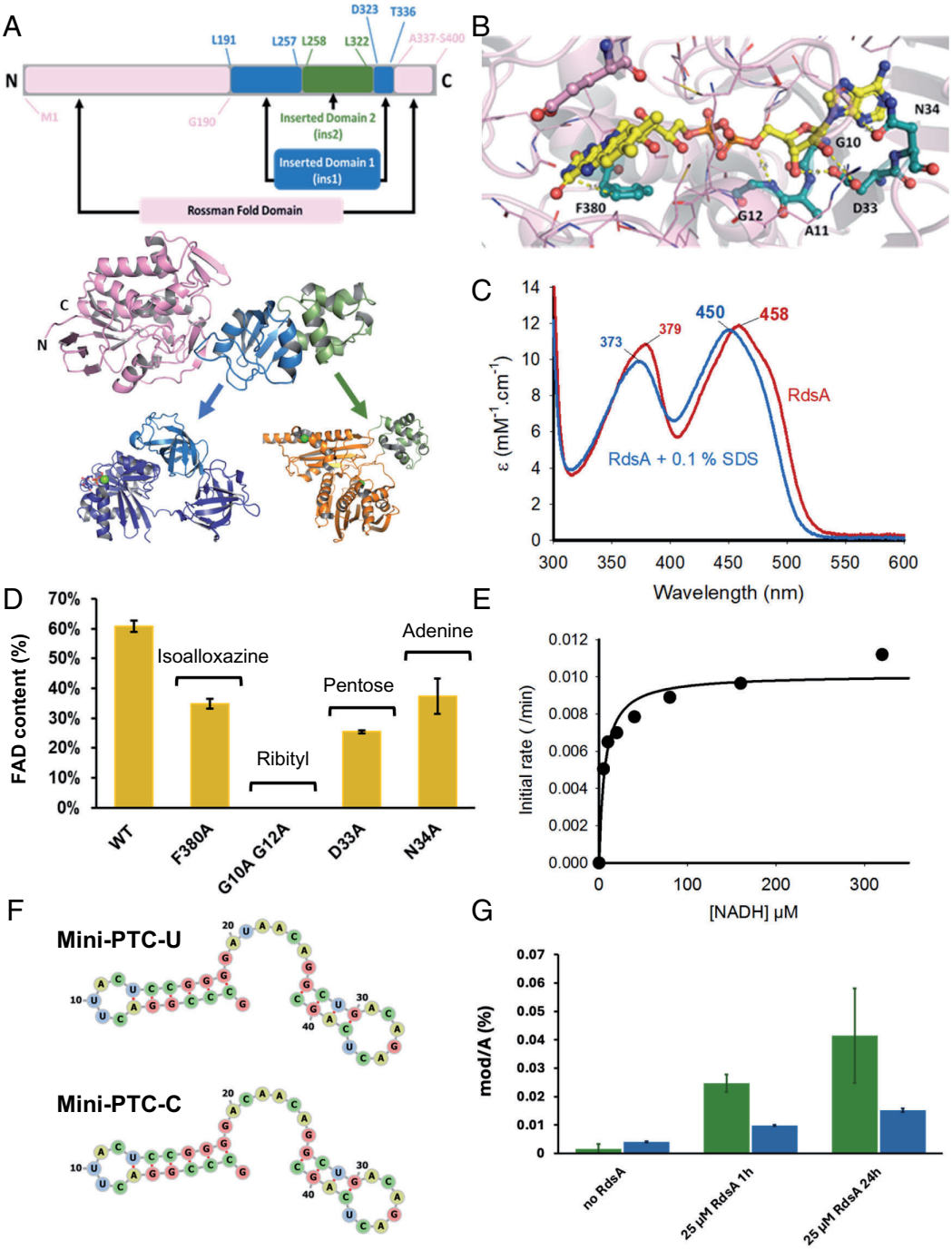


Fig. 4. Structural and biochemical characterization of *E. coli* rRNA dihydrouridine synthase. (*A*) Structural organization of *E. coli* RdsA. At the *Top*, a schematic diagram depicting the modularity of RdsA and delineating various domains—Rosman fold domain in pink, Ins1 domain in blue, and Ins2 domain in green. At the *Bottom*, an AlphaFold model of *E. coli* RdsA in a cartoon representation with the same color code as above. The Ins1 domain is homologous to the C-terminal domain of *E. coli* EF-Tu (PDB: 1EFC, colored in blue), while the Ins2 domain is homologous to the helix-2-turn-helix domain of *S. shibatae* DNA topoisomerase VI-B subunit (PDB: 1MU5, in green). (*B*) Holoprotein model of RdsA showing the binding of the FAD coenzyme (ball sticks, yellow) at the Rossman fold domain and key residues potentially interacting with FAD. (*C*) UV-Visible spectrum of native (red) and 0.1% SDS-denatured (blue) recombinant RdsA protein from *E. coli*. (*D*) FAD/protein molar ratio of recombinant RdsA harboring no change (WT) or mutations changing key residues interacting with FAD shown in B. This ratio ± SEM of 3 repeats was determined by spectrophotometric quantification of released FAD from a given amount of SDS-treated protein. For each mutant, the targeted interaction with FAD is indicated. (*E*) Michaelis–Menten curve for the NADH oxidation by RdsA. (*F*) Sequence and potential structure of the in vitro transcript Mini-PTC-U (*Top*) and Mini-PTC-C (*Bottom*) established with MXFold2 server. Nucleotide at position 21 is the anticipated target of dihydrouridylation. (*G*) Dihydrouridylation activity of RdsA on a Mini-PTC-U (blue) and Mini-PTC-C (green). D was quantified by LC–MS after incubation in reaction buffer of the designated transcript without or with 25 µM of RdsA for the indicated time. Results are expressed as the percentage of quantified D per quantified A ± SEM of 3 to 4 repeats (%mod/A).

detecting potential dihydrouridylation activity of rRNA by YhiN. In our initial endeavors, we aimed to reconstitute in vitro activity using rRNA isolated from an *E. coli* strain with the *yhiN* deletion. This involved coincubating the rRNA with recombinant YhiN and an excess of NADH. Subsequently, the generated rRNA product was enzymatically digested to its nucleoside level, and the presence of D was analyzed using LC/MS. Despite our concerted efforts, definitive detection of activity proved challenging, primarily due to notable contamination of D most probably originating from tRNA (*SI Appendix*, Fig. S6). To overcome this challenge, we

designed and produced an in vitro transcribed 40-mer RNA (Mini-PTC-U), encompassing nucleotides 2438 to 2459 and mimicking stems and loops of *E. coli*'s 23S rRNA (Fig. 4*F*), serving as a refined substrate in subsequent activity assays. Our results unequivocally established that YhiN dihydrouridylates this mini-RNA, strategically mimicking a segment of the ribosomal peptidyl-transferase center (Fig. 4*G*). The D level increased with prolonged incubation time (Fig. 4*G*) and, after 24 h of incubation, 0.04% of D per quantified adenosine was formed (%mod/A). As the synthetic RNA contains 10 adenosines, the number of D per molecule is then 0.4%. However, this artificial substrate might adopt alternative structures prone to dihydrouridylation of other uridines found in this Mini-PTC as a much lower, but detectable, activity was measured when changing the target uridine by a cytidine (Mini-PTC-C in Fig. 4 *F* and *G*). Nonetheless, these findings conclusively position YhiN as the dihydrouridine synthase for *E. coli*'s ribosome, catalyzing the reduction of U2449 to D2449, leveraging FAD as a redox coenzyme and NADH as a source of reducing equivalents.

Predicted RdsA Are widespread in Bacteria and Members of a Nonisofunctional Superfamily. The *E. coli* YhiN protein is a member of the COG2081 and PF03486 (HI0933-like protein) families. An initial analysis of the family in the EggNog database (http://eggnog6.embl.de/search/ogs/COG2081/) suggested that it was not isofunctional. Indeed, the 9,168 proteins listed in this family are from 6,723 species, hence 2,245 are paralogs (Dataset S1). The only two experimentally characterized protein in this family are 3-dehydro-bile acid delta(4,6)-reductase (BaiN) from *Clostridium scindens* involved in bile degradation (27) and aminoacetone oxidase (AAO) from *Streptococcus cristatus*, involved in resistance to endogenous metabolites-generated ROS and mutagens (28, 29). Hence, the RdsA activity of YhiN must be shared by only a subset of the COG2081/PF03486 members. To identify the RdsA subgroup more precisely, we constructed a sequence similarity networks (SSN) of PF03486 with different alignment score thresholds (AST, see *SI Appendix*, *Supplementary Materials and Methods*). An AST of 80 allowed the separation of the cluster containing $YhiN_{Ec}$ from the cluster containing bonafide BaiN (indicated by a yellow line in Fig. 5*A*) and the cluster containing AAO_{So} (indicated by a green line in Fig. 5*A*). Here, we analyzed the rRNA of *Vibrio cholerae* species by AlkalinineSeq and showed that its 23S rRNA indeed harbors D2432 while carrying a RdsA (*SI Appendix*, Fig. S7). These bacteria do encode proteins highly similar to YhiNEc (cyan nodes in Fig. 5*A*), supporting their function as RdsA enzymes. In contrast, *B. subtilis* 168 rRNA does not show any D residues (1), but its genome does encode a PF03486 family protein (YtfP or BSU30060). This protein is not located in the same SSN cluster as YhiNEc (Fig. 5*A*) and the gene is in an operon with the *opuD* gene encoding the glycine betaine/arsenobetaine transporter (Fig. 5*A* and *SI Appendix*, Fig. S8*A*). Most *Acinetobacter* species encode two members of the PF03486 family (Dataset S2). These separate nicely into two groups in the SSN (Fig. 5*A*). Group 1 proteins (YhiN1) are part of the YhiNEc cluster and hence would be predicted to catalyze the formation of D in rRNA. Group 2 proteins (YhiN2) are part of separate cluster, but the corresponding genes are located downstream *trmB* genes encoding tRNA (guanine(46)-N(7))-methyltransferase) (Fig. 5*B*) still indicating a possible link to RNA modifications for this YhiN2 group. Other physical association with RNA modifications were observed: y*hiN* homolog are located next to *dusA* genes in organisms such as *Synechococcus elongatus* (yellow nodes in Fig. 5 and *SI Appendix*, Fig. S9), or to genes encoding ribosomal pseudouridine synthase RsuA-like proteins (orange nodes in Fig. 5 and *SI Appendix*, Fig. S8*B*). In summary, the COG2081/PF03486 family is clearly not isofunctional and, because gene neighborhood information does not give definitive functional clues, further experimental characterization of the proteins from the different subgroups is required to establish whether they harbor RdsA activity.

To survey the taxonomic distribution of the PF03486 proteins conservatively predicted to be isofunctional with YhiNEc, we further analyzed the proteins within the same cluster (boxed in Fig. 5*A* and Dataset S3) in an SSN with a more stringent alignment score threshold (*SI Appendix*, Fig. S9) and constructed a phylogenetic tree of the same protein set (*SI Appendix*, Fig. S10). These analyses show that members of the YhiNEc subgroup are mostly found in bacteria with only two eukaryotic homologs from dinoflagellates (Uniprot IDs: A0A812V8P8 and A0A9P1GL58) and a small group found in Archaea (mainly Methanobacteriales and Methanomicrobiales, Dataset S3). The protein tree diverges from the species tree in several places (for example, the inclusion of Thermodesulfobacteriota in the middle of the gammaproteobacteria proteins) suggesting horizontal gene transfer events. The gene neighborhood information does not link YhiN to translation or RNA modifications as genes encoding universal stress protein, transporters, and RNA helicase-like proteins are mostly found surrounding these sets of YhiN encoding genes (*SI Appendix*, Fig. S10).

Discussion

In the course of this investigation, we uncovered a unique dihydrouridine synthase, referred to as RdsA, responsible for catalyzing the NADH and FAD-dependent reduction of U2449 in the 23S rRNA of *E. coli*. Traditional Dus enzymes, encompassing bacterial DusA, B, C, and eukaryotic Dus 1 to 4, also operate as flavoenzymes but rely on FMN and NADPH (4). This parallels their counterparts, dihydropyrimidine deshydrogenase (DHPD) (30), and dihydroorotate deshydrogenases (DODH) (31), which facilitate the reduction of uracil to dihydrouracil and orotate to dihydroorotate (a uracil derivative), respectively. This implies that nature utilizes reduced flavin for uridine reduction. However, in the tertiary structure generated by AlphaFold, RdsA distinguishes itself from these FMN-dependent flavoenzyme systems due to its predicted distinct structure, displaying a unique topology. Unlike Dus, DHPD, and DODH, where the catalytic site binding FMN is carried by a TIM BARREL, RdsA's flavin is held by a Rossman fold domain (FBD) housing two additional inserted domains (ins1D and ins2D). Surface charge representation on the AlphaFold model unveils several patches of positive charges concentrated on the three domains, arranging around the active site pocket and likely defining the RNA binding site (*SI Appendix*, Fig. S11). This model, if proved to be correct, would suggest that FBD, ins1D, and ins2D probably play a role in substrate recognition. RNA modifications follow a nonrandom, stepwise progression at distinct stages during ribosome biogenesis. Despite the complexity of this process, our observation of dihydrouridylation activity on a mini-ARN containing a small PTC-like region by RdsA should imply a potential role for this enzyme in acting on a ribosome precursor.

Phylogenetic analysis reveals the widespread distribution of YhiN among bacteria, implying the presence of D in the rRNA of numerous bacterial species. Indeed, our analyses of *V. cholerae* rRNA demonstrate the presence of D2432 in this organism (corresponding to the D2449 in *E. coli*), and furthermore, this species harbors a gene encoding RdsA. Nevertheless, our study brings attention to a perplexing observation: the absence of YhiN in the genome of *C. sporogenes*, despite the detection of m5D2449 and D2500 in its 23S rRNA (12). This suggests the potential existence of another unidentified type of Dus. In contrast, *B. subtilis* harbors

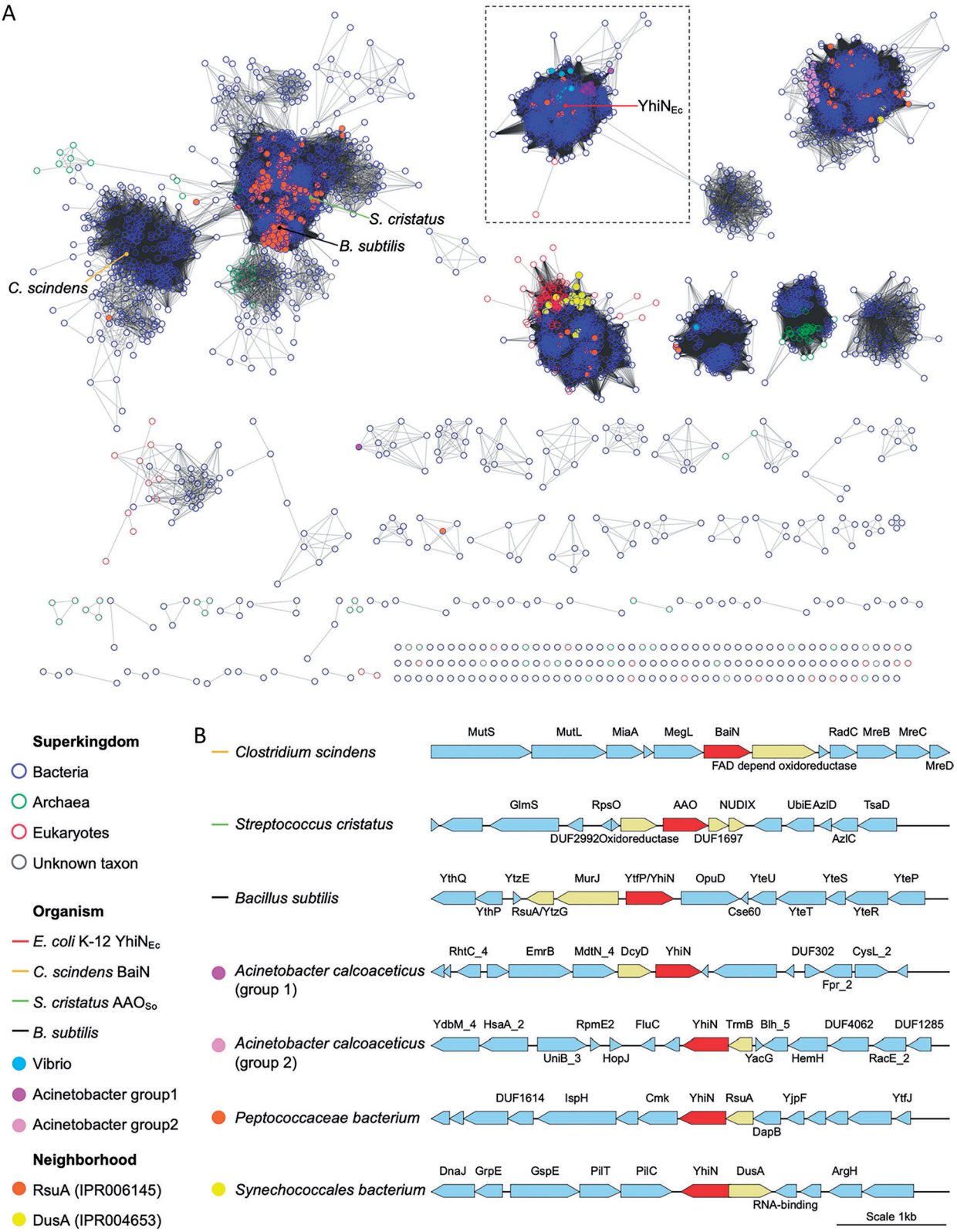


Fig. 5. Protein SSN of PF03486 family proteins. (*A*) SSN of 7,930 PF03486 family proteins with the cutoff of alignment score of 80. Each node in the network represents one or multiple PF03486 family proteins that share 90% or more identity. An edge (represented as a line) is drawn between two nodes with a BLAST *E*-value cutoff of better than 10^{-80}. The nodes are colored based on taxonomy or the gene neighborhood context of the corresponding genomes. The cluster containing YihN$_{Ec}$ is boxed. (*B*) Examples of gene neighborhoods organization. Uniprot IDs of PF03486 proteins: *C. scindens* BaiN, B0NAQ4; *S. cristatus* AAO, E8JVH0; *B. subtilis* YhiN, Q795R8; *Acinetobacter calcoaceticus* YhiN1, A0A0A8XKZ3; *A. calcoaceticus* YhiN2, A0A446ZNN4; *Peptococcaceae bacterium* YhiN, A0A942G053; *Synechococcales bacterium* YhiN, A0A354AWX7.

a *yhiN* gene, yet its rRNA modification profile does not show any D residues (1), hinting at potential alternative biological functions for this flavoprotein. This proposition gains support from our findings, which reveal that RdsA belongs to a nonisofunctional superfamily. This is further underscored by the identification of a homologous gene to *E. coli yhiN*, operating as an aminoacetone

oxidase in *Streptococcus oligofermentas* (28), albeit with measured activities that are notably weak.

Our genetic studies show that a functional RdsA enzyme is not essential for *E. coli* survival, and this is in keeping with the majority of rRNA modification enzymes. However, the biological role of ribosomal D remains elusive. Multiple studies have demonstrated that the lack of rRNA modifications can result in changes of ribosomal active sites (32), leading to a reduction in translation speed, an increase in mRNA reading errors by the ribosome (33), reduced virulence in pathogenic bacterial species, and a disrupted response to metabolites and antibiotics (34, 35). Limited to microbes, including certain pathogens, our research lays the groundwork for exploring the impact of dihydrouridylation on bacterial ribosome maturation and the potential targeting of this process by inhibitors.

Materials and Methods

Strains Media and Oligonucleotides Used in This Study. All strains are listed in *SI Appendix*, Table S3. BW25113 were obtained from the Keio collection, and genomic deletion strains were obtained from NBRP (Japan). Cells were usually grown in LB medium at 37 °C. Oligonucleotides used for RhoRT-PCR and MALDI-TOF analysis are listed in *SI Appendix*, Table S4.

Rhodamine Labeling, Reverse Transcription, and PCR. The method for rhodamine-110 labeling was derived from the one described in ref. 7. From the RNA stock, 50 μg was extracted, to which 16 μL of 1 M Tris HCl pH 7.5 and 40 μL of 100 mg/mL NaBH4 diluted in 10 mM KOH were added, reaching a final volume of 400 μL. After 1 h of incubation at room temperature (open tubes in the dark), the reaction was neutralized by adding 80 μL of 6 M acetic acid. The treated RNAs were then precipitated with 0.3 M NaCl and two volumes of 100% ethanol. The rRNAs were subsequently resuspended in 550 μL of 0.1 M formate buffer pH 3.15, containing 0.5 mM Rhodamine 110, and incubated at 37 °C for 1.5 h for fluorochrome conjugation. The reaction was stopped by adding 0.2 M Tris-HCl, pH 8.5, and the excess fluorochrome was removed by acidic phenol extraction (pH 4.3). Finally, the RNA was precipitated and resuspended in RNase-free H_2O. Reverse transcription and PCR conditions are detailed in *SI Appendix*.

Analysis of Ribosomal Dihydrouridylation at Position 2449. Dihydrouridine quantification was done by tandem LC–MS as detailed in supplementary information. Ribosomal modification profiles were analyzed by the AlkAniline-Seq method (36). This method exploits the instability of the D-ring under alkaline conditions leading to its cleavage and formation of β-ureidopropionic acid (21). This instability leads to aniline-driven RNA cleavage generating a 5′-phosphate group (5′-P) on the neighboring N + 1 residue, which serves as an input for highly selective ligation of sequencing adapters.

Biochemical Characterization of RdsA. *E. coli yhiN* gene was cloned in pET15b for production and purification in BL21-DE3 *E. coli* cells. The NADH oxidase activity was determined by following decrease of A_{340} under aerobic conditions and dihydrouridylation assays were performed by incubating the recombinant RdsA with the target RNA in a buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT, 10 mM MgCl2, 100 μM FAD, and 15% glycerol. The detailed procedure for LC–MS analysis of the reaction is detailed in *SI Appendix*.

General Bioinformatic resources. For literature and sequence retrievals, the resources at NCBI (https://www.ncbi.nlm.nih.gov/) (37), UniProt (https://www.uniprot.org) (38), EggNog (http://eggnog6.embl.de/) (39), and BV-BRC (https://www.bv-brc.org) (40) were routinely used. PaperBlast was used to find published papers on members of the COG2018 family (papers.genomics.lbl.gov/) (38). The SubtiWiki/CoreWiki (http://corewiki.uni-goettingen.de/) (41), the String database (https://string-db.org/) (42) and the EFI Gene Neighborhood webtool (43) were used to explore physical clustering. Protein sequences were aligned using MUSCLE v5.1 (44) and visualized using Weblogo3 (45). SSN and phylogenetic analysis are detailed in supplementary information.

Data, Materials, and Software Availability. All study data are included in the article and/or supporting information. All NGS data associated to this manuscript are deposited and made publicly available in ENA (The European Bioinformatics Institute EMBL-EBI) under the accession number PRJEB77571 (46).

ACKNOWLEDGMENTS. This research is funded by ANR/DFG Grant DERASE (#20-CE92-0030), by Grant GM132254 to V.d.C.-L. M.H. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–Project-ID 439669440–TRR-319–TP C01 and, HE 3397/21-1. S.D. was funded by a grant from the Danish Research Agency (FNU-rammebevilling 10-084554). S.T. was funded by ED515-Complexité du vivant (Sorbonne Université).

Author affiliations: [a]Sorbonne Université, CNRS, Institut de Biologie Paris Seine, Biology of Aging and Adaptation, Institut de Biologie Paris-Seine, F-75252 Paris Cedex 05, France; [b]Collège De France, Sorbonne Université, CNRS, Laboratoire de Chimie des Processus Biologiques, F-75231, Paris Cedex 05, France; [c]Université de Lorraine, CNRS, Institut National de la Santé et de la Recherche Médicale, Ingénierie-Biologie-Santé en Lorraine, Epitranscriptomique et Séquençage Core Facility, F-54000 Nancy, France; [d]Université de Lorraine, CNRS, Ingénierie Moléculaire, Cellulaire et Physiopathologie, F-54000 Nancy, France; [e]Institut für Pharmazeutische und Biomedizinische Wissenschaften, Johannes Gutenberg-Universität, Mainz D-55128, Germany; [f]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611; [g]Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark; [h]Institut Pasteur, Université Paris Cité, CNRS UMR3525, Unité Plasticité du Génome Bactérien, F-75015 Paris, France; and [i]Genetics Institute, University of Florida, Gainesville, FL 32610

Author contributions: O.J.-J., S.D., Z.B., V.d.C.-L., M.H., Y.M., D.H., and D.B. designed research; S.T., Q.T., L.-M.K., V.M., Y.Y., C.S., C.G., O.J.-J., S.R., S.D., L.H., Z.B., V.d.C.-L., M.H., Y.M., and D.H. performed research; S.T., Q.T., L.-M.K., V.M., Y.Y., C.S., C.G., O.J.-J., S.R., S.D., L.H., Z.B., V.d.C.-L., M.H., Y.M., D.H., and D.B. analyzed data; and S.T., L.-M.K., V.M., Y.Y., C.G., O.J.-J., S.D., Z.B., V.d.C.-L., M.H., Y.M., D.H., and D.B. wrote the paper.

1. A. Cappannini *et al.*, MOMICS: A database of RNA modifications and related information. 2023 update. *Nucleic Acids Res.* **52**, D239–D244 (2024).
2. C. Lorenz, C. E. Lünse, M. Mörl, tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules* **7**, 35 (2017).
3. Y. Motorin, M. Helm, tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934–4944 (2010).
4. D. Brégeon *et al.*, Dihydrouridine in the transcriptome: New life for this ancient RNA chemical modification. *ACS Chem. Biol.* **17**, 1638–1657 (2022).
5. O. Finet, C. Yague-Sanz, F. Marchand, D. Hermand, The dihydrouridine landscape from tRNA to mRNA: A perspective on synthesis, structural impact and function. *RNA Biol.* **19**, 735–750 (2022).
6. W. Dai *et al.*, Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat. Chem. Biol.* **17**, 1178–1187 (2021).
7. O. Finet *et al.*, Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Mol. Cell.* **82**, 404–419.e9 (2022).
8. A. S. Draycott *et al.*, Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA. *PLoS Biol.* **20**, e3001622 (2022).
9. X. Chen, S. Sim, E. J. Wurtmann, A. Feke, S. L. Wolin, Bacterial noncoding Y RNAs are widespread and mimic tRNAs. *RNA.* **20**, 1715–1724 (2014).
10. J. A. Kowalak, E. Bruenger, J. A. McCloskey, Posttranscriptional modification of the central loop of domain V in *Escherichia coli* 23 S ribosomal RNA. *J. Biol. Chem.* **270**, 17758–17764 (1995).
11. S. A. Fromm *et al.*, The translating bacterial ribosome at 1.55 Å resolution generated by cryo-EM imaging services. *Nat. Commun.* **14**, 1095 (2023).
12. F. Kirpekar *et al.*, Mapping of ribosomal 23S ribosomal RNA modifications in *Clostridium sporogenes*. *RNA Biol.* **15**, 1060–1070 (2018).
13. G. Emmerechts, S. Barbé, P. Herdewijn, J. Anné, J. Rozenski, Post-transcriptional modification mapping in the *Clostridium acetobutylicum* 16S rRNA by mass spectrometry and reverse transcriptase assays. *Nucleic Acids Res.* **35**, 3494–3503 (2007).
14. F. Yu *et al.*, Molecular basis of dihydrouridine formation on tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19593–19598 (2011).
15. H. Kusuba *et al.*, In vitro dihydrouridine formation by tRNA dihydrouridine synthase from *Thermus thermophilus*, an extreme-thermophilic eubacterium. *J. Biochem.* **158**, 513–521 (2015).
16. C. Bou-Nader *et al.*, An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic Acids Res.* **43**, 9446–9456 (2015).
17. R. T. Byrne *et al.*, Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6033–6037 (2015).
18. A. C. Bishop, J. Xu, R. C. Johnson, P. Schimmel, V. de Crécy-Lagard, Identification of the tRNA-dihydrouridine synthase family. *J. Biol. Chem.* **277**, 25090–25095 (2002).
19. C. Bou-Nader, D. Brégeon, L. Pecqueur, M. Fontecave, D. Hamdane, Electrostatic potential in the tRNA binding evolution of dihydrouridine synthases. *Biochemistry.* **57**, 5407–5414 (2018).
20. C. Bou-Nader *et al.*, Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: Perspectives on the evolution scenario. *Nucleic Acids Res.* **46**, 1386–1394 (2018).
21. F. Xing, S. L. Hiley, T. R. Hughes, E. M. Phizicky, The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.* **279**, 17850–17860 (2004).

22. L. W. Rider, M. B. Ottosen, S. G. Gattis, B. A. Palfey, Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *J. Biol. Chem.* **284**, 10324–10333 (2009).
23. J. J. Dalluge, T. Hashizume, A. E. Sopchik, J. A. McCloskey, D. R. Davis, Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Res.* **24**, 1073–1079 (1996).
24. N. Dyubankova *et al.*, Contribution of dihydrouridine in folding of the D-arm in tRNA. *Org. Biomol. Chem.* **13**, 4960–4966 (2015).
25. M. Hashimoto *et al.*, Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* **55**, 137–149 (2005).
26. J. Kato, M. Hashimoto, Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* **3**, 132 (2007).
27. S. C. Harris *et al.*, Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium *Clostridium scindens* ATCC 35704. *Biochim Biophys. Acta, Mol. Cell. Biol. Lipids* **1863**, 276–283 (2018).
28. G. Molla *et al.*, Aminoacetone oxidase from *Streptococcus oligofermentans* belongs to a new three-domain family of bacterial flavoproteins. *Biochem. J.* **464**, 387–399 (2014).
29. P. Zhou, L. Liu, H. Tong, X. Dong, Role of operon aaoSo-mutT in antioxidant defense in *Streptococcus oligofermentans*. *PLoS ONE* **7**, e38133 (2012).
30. D. Dobritzsch, G. Schneider, K. D. Schnackerz, Y. Lindqvist, Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *EMBO J.* **20**, 650–660 (2001).
31. R. A. G. Reis, F. A. Calil, P. R. Feliciano, M. P. Pinheiro, M. C. Nonato, The dihydroorotate dehydrogenases: Past and present. *Arch. Biochem. Biophys.* **632**, 175–191 (2017).
32. H. Demirci *et al.*, Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function. *RNA* **16**, 2319–2324 (2010).
33. A. Baudin-Baillieu *et al.*, Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res.* **37**, 7665–7677 (2009).
34. C. P. van Buul, W. Visser, P. H. van Knippenberg, Increased translational fidelity caused by the antibiotic kasugamycin and ribosomal ambiguity in mutants harbouring the ksgA gene. *FEBS Lett.* **177**, 119–124 (1984).
35. S. Okamoto *et al.*, Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol. Microbiol.* **63**, 1096–1106 (2007).
36. V. Marchand *et al.*, AlkAniline-Seq: Profiling of m7 G and m3 C RNA modifications at single nucleotide resolution. *Angew. Chem. Int. Ed. Engl.* **57**, 16785–16790 (2018).
37. E. W. Sayers *et al.*, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **52**, D33–D43 (2024).
38. UniProt Consortium, UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
39. A. Hernández-Plaza *et al.*, eggNOG 6.0: Enabling comparative genomics across 12535 organisms. *Nucleic Acids Res.* **51**, D389–D394 (2023).
40. R. D. Olson *et al.*, Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): A resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* **51**, D678–D689 (2023).
41. L. A. Flórez, S. F. Roppel, A. G. Schmeisky, C. R. Lammers, J. Stülke, A community-curated consensual annotation that is continuously updated: The *Bacillus subtilis* centred wiki SubtiWiki. *Database (Oxford)* **2009**, bap012 (2009).
42. D. Szklarczyk *et al.*, The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
43. R. Zallot, N. Oberg, J. A. Gerlt, The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182 (2019).
44. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
45. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
46. Q. Thullier, V. Marchand, Y. Motorin, Data from "New Ribosomal RNA Dihydrouridine Synthase in E.coli." European Nucleotide Archive. https://www.ebi.ac.uk/ena/browser/view/PRJEB77571. Deposited 13 July 2024.

Suplementary information for:

Exploring a Novel Class of Flavoenzymes: Identification and Biochemical Characterization of Ribosomal RNA Dihydrouridine Synthase

Sabrine Toubdji[1,2], Quentin Thullier[3,4], Lea-Marie Kilz[5], Virginie Marchand[3,4], Yifeng Yuan[6], Claudia Sudol[1,2], Catherine Goyenvalle[1], Olivier Jean-Jean[1], Simon Rose[7], Stephen Douthwaite[7], Léo Hardy[8], Zeynep Baharoglu[8], Valérie de Crécy-Lagard[6,9], Mark Helm[5], Yuri Motorin[3,4], Djemel Hamdane*[2] and Damien Brégeon*[1]

[1] Sorbonne Université, CNRS, Institut de Biologie Paris Seine, Biology of Aging and Adaptation, IBPS-B2A, F-75252 Paris cedex 05, France

[2] Collège De France, Sorbonne Université, CNRS, Laboratoire de Chimie des Processus Biologiques, LCPB, F-75231 Paris Cedex 05, France.

[3] Université de Lorraine, CNRS, INSERM, IBSLor, EpiRNA-Seq Core Facility, F-54000 Nancy, France.

[4] Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France.

[5] Institut für Pharmazeutische und Biomedizinische Wissenschaften (IPBW), Johannes Gutenberg-Universität, D-55128 Mainz, Germany

[6] University of Florida, Department of Microbiology and Cell Science, FL-32611Gainesville, Florida, USA.

[7] University of Southern Denmark, Department of Biochemistry and Molecular Biology, DK-5230 Odense M, Denmark.

[8] Institut Pasteur, Université Paris Cité, CNRS UMR3525, Unité Plasticité du Génome Bactérien, Paris, France

[9] University of Florida, Genetics Institute, Gainesville, FL-32610Gainesville, Florida, USA.

*To whom correspondence should be addressed:

Damien Brégeon, Sorbonne Université, CNRS UMR-8256, Institut de Biologie Paris Seine, Biology of Aging and Adaptation, F-75252 Paris cedex 05, France.
Email : damien.bregeon@sorbonne-universite.fr

Djemel Hamdane, Laboratoire de Chimie des Processus Biologiques, CNRS-UMR 8229, Collège de France, 11 place Marcelin Berthelot, F-75231 Paris Cedex 05, France.
Email : djemel.hamdane@college-de-france.fr

Supplementary materials and methods

Preparation of ribosomal rRNA

150 ml of LB medium was inoculated with 0.3 mL of the preculture and incubated at 37°C with agitation at 300 rpm until reaching OD_{450} = 0.3-0.4. The culture was then centrifuged (6000xg at 4°C) for 10 minutes. Subsequent steps are performed on ice. The pellet was resuspended in 50 mL of cold TMNr (50 mM Tris-HCl, pH 7.5-7.8, 10 mM $MgCl_2$, 100 mM NH_4Cl), then centrifuged under the same conditions. The cell pellet was again resuspended in 50mL of cold TMNr and centrifuged (6000xg at 4°C) for 10 minutes. Finally, the cells were resuspended in 10 ml of cold TMNr supplemented with 1 mg of lysozyme, followed by a 20-minute incubation on ice.

Sonication was performed with a customized program for each bacterial species used (*E. coli*: 3 pulses, 20% amplitude, 10 sec. each separated by a 5 sec. pause, repeated 4 times with a 30 sec. interval between each series, and *B. subtilis*: 3 pulses, 20% amplitude, 30 sec. each separated by a 10 sec. pause, repeated 3 times with a 30 sec. interval between each series) to lyse any remaining cells. Cellular debris were removed by two successive centrifugations of 10 minutes at 13,000xg at 4°C. Ribosomes were then sedimented by ultracentrifugation at 25,000 g (4°C) for 16 to 18 hours. The ribosome pellet was resuspended in 1 mL of cold TMNr, and proteins were removed from the preparation by phenol extraction (pH 4.3), followed by phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1) extractions. rRNA was then precipitated by the addition of sodium acetate pH 5.3 (final concentration 0.3 M) and 2 volumes of 100% ethanol. After washing with 80% ethanol, rRNA pellets were resuspended in 200 µl of water and quantified by measuring absorbance at 260 nm.

For RNA extraction of *Vibrio cholerae*, overnight cultures were diluted 1:1000 in MH (Mueller-Hinton) medium and grown with agitation at 37°C until an OD_{600} of 0.25 (exponential phase). 2 mL of these cultures were centrifuged and the supernatant removed. Pellets were homogenized by resuspension with 1.5 mL of room temperature TRIzol Reagent. 300 µL chloroform was then added to the samples, which were mixed by vortexing. Samples were then centrifuged at 4°C for 10 minutes at 13000 rpm. Upper (aqueous) phase was transferred to a new 2 mL tube and mixed with 1 volume of 70% ethanol. From this point, the homogenate was loaded into a RNeasy Mini kit (Qiagen) column and RNA purification proceeded according to the manufacturer's instructions. Samples were then subjected to DNase treatment using TURBO DNA-free Kit (Ambion) according to the manufacturer's instructions.

Enzymatic Digestion with HaeIII and DNase I

To eliminate any DNA contamination, our samples underwent: i) enzymatic digestion (3 hours at 37°C) with the restriction enzyme HaeIII (20 units), which cleaves DNA in the region targeted for PCR, and ii) enzymatic digestion with DNase I (1 unit per µg of RNA, 2 hours at 25°C), followed by inactivation at 65°C for 10 minutes. This second digestion was performed either on the rRNA stock or after the labeling step described below. At the end of each of the two enzymatic digestions, samples were quantified to determine the concentration of our RNAs, which, upon completion of this step, were ready for RT-PCR.

D-Labeling procedure

From the RNA stock, 50 µg was extracted, to which 16 µL of 1 M Tris HCl pH 7.5 and 40 µL of 100 mg/mL $NaBH_4$ diluted in 10 mM KOH are added, reaching a final volume of 400 µl. After one hour of incubation at room temperature (open tubes in the dark), the reaction was neutralized by adding 80 µL of 6 M acetic acid. The treated RNAs were then precipitated with

0.3 M NaCl and two volumes of 100% ethanol. The rRNAs were subsequently resuspended in 550 µL of 0.1 M formate buffer pH 3.15, containing 0.5 mM Rhodamine-110, and incubated at 37°C for 1.5 hours for fluorochrome conjugation. The reaction was stopped by adding 0.2 M Tris-HCl, pH 8.5, and the excess fluorochrome was removed by acidic phenol extraction (pH 4.3). Finally, RNA was precipitated and resuspended in RNase free H_2O.

Reverse transcription

The reverse transcription is performed with 50 units of M-MLV reverse transcriptase (Moloney-Murine Leukemia Virus), 1 µg of RNA was incubated for 5 minutes at 65°C in the presence of 1 pmole of oligonucleotide 4R (**Table S7**) and 0.5 mM dNTP, then placed on ice. The reaction buffer and the enzyme at varying concentrations were then added. The samples were incubated for 50 minutes at 37 °C and 15 minutes at 70°C.

PCR

PCR reactions were carried out with GoTaq2 (Promega) following the supplier's instructions and using 50ng of RNA per sample, along with the oligonucleotides listed in the table S7. The reaction was incubated for 3 minutes at 95°C for complete denaturation, followed by 23 cycles of amplification (20 seconds at 95°C, 30 seconds at 55°C, and 30 seconds at 72°C), followed by a final elongation step of 5 minutes at 72°C.

MALDI TOF mass spectrometry

The sequence from A2423 to A2473 within domain V of the *E. coli* 23S rRNA was isolated by hybridization to a complementary 50-mer deoxyoligonucleotide (**Table S7**). Each time, 100 pmol of total rRNA was hybridized to 500 pmol of deoxyoligonucleotide; unprotected regions of the rRNAs were digested away with mung bean nuclease and RNase A; and the rRNA sequences paired to the deoxyoligonucleotide were separated by gel electrophoresis and extract (1, 2). The rRNA sequences of around 50 nt were digested with either RNase A or RNase T1 in aqueous solution containing 3-hydroxypicolinic acid and analysed by MALDI-MS (Voyager Elite, Perseptive Biosystems) recording in reflector and positive ion mode (3). Spectra were analysed using the program m/z (Proteometrics Inc).

Cloning of yhiN the putative RdsA protein

The *yhiN* gene was cloned into the pET15b plasmid using primers specified in Table S4, which contained sequence homologies with the plasmid and introduced two restriction sites (NdeI, BamHI) to amplify the coding sequence of *yhiN* from genomic DNA extracted from an *E. coli* BW25113 strain. The PCR fragment was purified using the Wizard SV Gel and PCR Clean-Up System. Subsequently, both the plasmid and the PCR fragment were digested with NdeI and BamHI. After an over-night ligation using T4 DNA ligase (Neb Biolabs), chemically competent DH5α cells were then transformed with the plasmid, and the integrity of the gene was confirmed through DNA sequencing conducted by Eurofins.

Expression and purification of RdsA protein

Chemically competent *E. coli* BL21DE3 star cells, transformed with the pET15b-yhiN plasmid, were cultivated in LB medium supplemented with ampicillin (100 µg/mL) at 37°C until the optical density at 600 nm reached 0.6. Protein synthesis was induced by adding isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 220 µM, followed by overnight growth at 16°C. The harvested cells were centrifuged (6000 rpm, 4°C, 15 min) and stored at -80°C.

For protein extraction, cells were resuspended in a buffer containing 50 mM sodium phosphate pH 8, 300 mM NaCl, 5 mM DTT, 25 mM imidazole, 10% glycerol (v/v), 100 µM free FMN and

FAD, and one antiprotease tablet (Roche). Discontinuous sonication for 15 minutes in a water ice batch was performed, and cellular extracts were centrifuged for 1 hour at 13000 rpm, resulting in a soluble fraction of RdsA proteins.

The soluble fraction underwent Ni-NTA affinity chromatography using a column equilibrated with Buffer A (50 mM sodium phosphate pH 8, 300 mM NaCl, 25 mM imidazole, 10% glycerol). After extensive washing with Buffer A, elution was carried out with Buffer A supplemented with 250 mM imidazole. Fractions containing RdsA proteins were pooled and concentrated by ultrafiltration.

Additional purification steps involved the use of PD-10 Desalting Columns containing Sephadex G-25 resin equilibrated in 50 mM HEPES pH 7.5, 150 mM NaCl, and 15% glycerol (v/v). The purity of the proteins was assessed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Finally, the purified proteins were concentrated to 22.8 mg/ml, flash-frozen in liquid nitrogen, and stored at -80°C until use. Protein concentrations were determined using the Bradford assay (Biorad), with BSA serving as a standard.

NAD(P)H Oxidase Activity

The ability of RdsA to oxidize NADH/NADPH under steady state conditions was determined under air in 50 mM HEPES pH 7.5, 150 mM NaCl and 10% glycerol. Assays were performed using 2 µM of RdsA protein in the presence of various concentrations of NAD(P)H from 10 to 320 µM. The amount of NAD(P)H oxidized was monitored by following the decrease of absorbance at 343 nm (ε343 = 6.21 mM-1.cm-1). The initial rate versus NAD(P)H concentration was analyzed according to Michaelis-Menten formalism.

In vitro transcription of a mini-PTC

The in vitro transcription was carried out using the in-house T7 RNA polymerase (Tfu) in a total volume of 10 mL, comprising 1X transcription buffer, 10% PEG 8000, 22 mM MgCl2, 4 mM NTPs (GTP, ATP, CTP, UTP), Oligo1, which includes the promoter sequence for Polymerase recruitment, at a concentration of 0.6 µM, Oligo 2 mini-PTC-U or Mini-PTC-C at a concentration of 0.6 µM with two 2'O-methylation modifications at the 5' end to halt transcription, and 0.4 mg of Tfu RNA polymerase. The incubation was carried out for 4 hours in a water bath at 37°C. After 2 hours of incubation, the reaction was assessed by RNA gel. For this, 10 µL of the reaction mixture was taken, to which 10 µL of stop mix (2X) was added. The resulting mixture was loaded onto a 12% polyacrylamide gel to control the reaction. The remaining reaction mixture was filtered through a DEAE resin column to purify the transcribed RNA. The RNA was eluted with 3M NaAc, pH 5.2, then precipitated by adding 3 volumes of 100% ethanol. The RNA was recovered by centrifugation for 1 hour at 12,000 g at 4°C. Finally, the RNA underwent a desalting step using a PD10 column.

Activity assay and dihydrouridine quantification

In vitro activity was assayed for 1 and 24 hours at room temperature (25°C) in 50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT, 10 mM $MgCl_2$, 100 µM FAD and 15 % glycerol in aerobic condition. 50 µM of 40-mer In vitro transcribed RNA were incubated with 25 µM of protein RdsA in a total volume of 100 µL and reaction was started upon addition of 5 mM of NADH. Quenching was performed by adding one volume of acidic phenol followed by centrifugation. RNA in the aqueous phase were ethanol precipitated. Dihydrouridine quantification was carried out by LC-MS spectrometry analysis.

Liquid chromatography-tandem mass spectrometry (LC-MS)

From 2300 to 8000 ng of RNA per sample was digested to nucleoside level using 0.6 U nuclease P1 from P. citrinum (Sigma-Aldrich), 0.2 U snake venom phosphodiesterase from C. adamanteus (Worthington), 0.2 U bovine intestine phosphatase (Sigma-Aldrich), 10 U benzonase (Sigma-Aldrich) and 200 ng Pentostatin (Sigma-Aldrich) in 5 mM Tris (pH 8) and 1 mM magnesium chloride at 37 °C for two hours. LC-MS/MS analysis was performed using an Agilent 1260 series LC with a Synergi Fusion RP18 column (4 µM particle size, 80 Å pore size, 250 × 2.0 mm; Phenomenex) and an Agilent 6470B Triple Quadrupole mass spectrometer equipped with an electrospray ion source (ESI). 5 mM ammonium acetate buffer at pH 5.3 was used as solvent A and LC-MS grade acetonitrile (Honeywell) served as solvent B. The elution started with 100% solvent A with a flow rate of 0.35 mL/min, followed by a linear gradient to 8% solvent B at 10 min, raising to 40% solvent B after 20 min and subsequent three-minute restoration of the initial conditions. 100% solvent A was held for further 10 minutes before starting the next elution. During elution a diode array detector (DAD) recorded the UV signal at 254 nm to monitor the main nucleosides and the ESI parameters were set as follows: gas temperature 300 °C, gas flow 7 L/min, nebulizer pressure 60 psi, sheath gas temperature 400 °C, sheath gas flow 12 L/min and capillary voltage 3000 V. The mass spectrometer was run in the dynamic multiple reaction monitoring (dMRM) mode using Agilent MassHunter software searching for the mass transition of dihydrouridine (m/z 247 → 115) in positive ion mode. The quantitative analysis was performed as described in (4) using a combined external and internal calibration. For internal calibration 2300 ng of digested sample were spiked with 50 ng of 13C stable isotope-labelled nucleosides from *E. coli* and subjected to analysis.

Sequence Similarity Networks (SSNs)

Sequence Similarity Networks (SSNs) were generated with the Enzyme Function Initiative (EFI) suite of webtools (https://efi.igb.illinois.edu/) (5) and visualized using Cytoscape 3.10.1 (6). The PF02486 identifier was given as input to the SSN building tool of the EFI platform using a minimum protein length cut of 300 aa and the family reduction option that selects every 5th sequence in the family reducing the number of sequences from ~42,000 to 7,930. The initial SSN was generated with an alignment score cutoff set such that each connection (edge) represents a sequence identity of above approximately 40%. Then more stringent SSNs were created by increasing the Alignment Score Threshold (AST) in small increments (usually by 5). This process was continued until the Clostridia BaiN sequences (7) were no longer in the same group as YhiN$_{Ec}$. The cluster containing YihN$_{Ec}$ was further analyzed with a more stringent AST of 150. PF03486 family members in the SSN are listed in **Table S4**. Proteins in the subgroup that contains YhiN$_{Ec}$ are listed in **Table S5**. Genomic region data of representative *yhiN* neighborhoods was extracted using EFI-Genome Neighborhood Diagrams tools and subjected to visualization using Gene Graphics (8). The genomic region data are listed in Table S8.

Phylogenetic analyses

The sequences of 2,311 YhiN homologs located in the YhiNEc cluster (boxed in **Figure 5A**) were retrieved from Uniprot using the Enzyme Function Initiative (EFI) suite of web tools (5). After adding three sequences of BaiN-like proteins (Uniprot IDs B0NAQ4, A0A850HAK8, and R9NAC7) as outgroups, the sequences were aligned using MUSCLE v5.1 (9) and trimmed using BMGE v1.12 using default settings (10). A maximum likelihood tree was built using FastTree (11) using the following settings: -lg -cat 20 -gamma with a hundred bootstraps. Trees were visualized using the iTOL platform (https://itol.embl.de) (version 6.8.1) (12).

Supplementary references

1. T. E. Andersen, B. T. Porse, F. Kirpekar, A novel partial modification at C2501 in *Escherichia coli* 23S ribosomal RNA. *RNA.* **10**, 907–913 (2004).

2. S. Douthwaite, F. Kirpekar, Identifying modifications in RNA by MALDI mass spectrometry. *Methods Enzymol.* **425**, 3–20 (2007).

3. F. Kirpekar, S. Douthwaite, P. Roepstorff, Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *RNA.* **6**, 296–306 (2000).

4. S. Kellner, *et al.*, Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).

5. R. Zallot, N. Oberg, J. A. Gerlt, The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry.* **58**, 4169–4182 (2019).

6. P. Shannon, *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

7. S. C. Harris, *et al.*, Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium *Clostridium scindens* ATCC 35704. *Biochim Biophys. Acta. Mol. Cell. Biol. Lipids.* **1863**, 276–283 (2018).

8. K. J. Harrison, V. de Crécy-Lagard, R. Zallot, Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics.* **34**, 1406–1408 (2018).

9. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

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

11. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One.* **5**, e9490 (2010).

12. I. Letunic, P. Bork, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475-478 (2011).

13. J. Wrzesinski, A. Bakin, K. Nurse, B. G. Lane, J. Ofengand, Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. *Biochemistry.* **34**, 8904–8913 (1995).

14. M. Liu, G. W. Novotny, S. Douthwaite, Methylation of 23S rRNA nucleotide G745 is a secondary function of the RlmAI methyltransferase. *RNA.* **10**, 1713–1720 (2004).

15. J. M. Lövgren, P. M. Wikström, The *rlmB* gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in *Escherichia coli*. *J. Bacteriol.* **183**, 6957–6960 (2001).

16. S. Okamoto, *et al.*, Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol. Microbiol.* **63**, 1096–1106 (2007).

17. S. Raychaudhuri, L. Niu, J. Conrad, B. G. Lane, J. Ofengand, Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *J. Biol. Chem.* **274**, 18880–18886 (1999).

18. K.-T. Wang, *et al.*, Structure of the bifunctional methyltransferase YcbY (RlmKL) that adds the m7G2069 and m2G2445 modifications in *Escherichia coli* 23S rRNA. *Nucleic Acids Res.* **40**, 5138–5148 (2012).

19. D. V. Lesnyak, *et al.*, Methyltransferase that modifies guanine 966 of the 16 S rRNA: functional identification and tertiary structure. *J. Biol. Chem.* **282**, 5880–5887 (2007).

20. C. T. Madsen, J. Mengel-Jørgensen, F. Kirpekar, S. Douthwaite, Identifying the methyltransferases for m(5)U747 and m(5)U1939 in 23S rRNA using MALDI mass spectrometry. *Nucleic Acids Res.* **31**, 4738–4746 (2003).

21. X. R. Gu, C. Gustafsson, J. Ku, M. Yu, D. V. Santi, Identification of the 16S rRNA m5C967 methyltransferase from *Escherichia coli*. *Biochemistry.* **38**, 4053–4057 (1999).

22. J. Conrad, D. Sun, N. Englund, J. Ofengand, The *rluC* gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. *J. Biol. Chem.* **273**, 18562–18566 (1998).

23. M. Del Campo, Y. Kaya, J. Ofengand, Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA.* **7**, 1603–1615 (2001).

24. J. S. Tscherne, K. Nurse, P. Popienick, J. Ofengand, Purification, cloning, and characterization of the 16 S RNA m2G1207 methyltransferase from *Escherichia coli*. *J. Biol. Chem.* **274**, 924–929 (1999).

25. P. V. Sergiev, M. V. Serebryakova, A. A. Bogdanov, O. A. Dontsova, The *ybiN* gene of *Escherichia coli* encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. *J. Mol. Biol.* **375**, 291–300 (2008).

26. E. Purta, M. O'Connor, J. M. Bujnicki, S. Douthwaite, YgdE is the 2'-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA. *Mol. Microbiol.* **72**, 1147–1158 (2009).

27. S. Kimura, T. Suzuki, Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. *Nucleic Acids Res.* **38**, 1341–1352 (2010).

28. P. V. Sergiev, D. V. Lesnyak, A. A. Bogdanov, O. A. Dontsova, Identification of *Escherichia coli* m2G methyltransferases: II. The *ygjO* gene encodes a methyltransferase specific for G1835 of the 23 S rRNA. *J. Mol. Biol.* **364**, 26–31 (2006).

29. S. Kimura, Y. Sakai, K. Ishiguro, T. Suzuki, Biogenesis and iron-dependency of ribosomal RNA hydroxylation. *Nucleic Acids Res.* **45**, 12974–12986 (2017).

30. J. Wrzesinski, A. Bakin, J. Ofengand, B. G. Lane, Isolation and properties of *Escherichia coli* 23S-RNA pseudouridine 1911, 1915, 1917 synthase (RluD). *IUBMB Life.* **50**, 33–37 (2000).

31. S.-M. Toh, L. Xiong, T. Bae, A. S. Mankin, The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA. *RNA.* **14**, 98–106 (2008).

32. N. M. Andersen, S. Douthwaite, YebU is a m5C methyltransferase specific for 16 S rRNA nucleotide 1407. *J. Mol. Biol.* **359**, 777–786 (2006).

33. E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, S. Douthwaite, YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA.* **14**, 2234–2244 (2008).

34. G. N. Basturea, K. E. Rudd, M. P. Deutscher, Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *RNA.* **12**, 426–434 (2006).

35. T. Caldas, *et al.*, The FtsJ/RrmJ heat shock protein of *Escherichia coli* is a 23 S ribosomal RNA methyltransferase. *J. Biol. Chem.* **275**, 16414–16419 (2000).

36. G. N. Basturea, D. R. Dague, M. P. Deutscher, K. E. Rudd, YhiQ is RsmJ, the methyltransferase responsible for methylation of G1516 in 16S rRNA of *E. coli*. *J. Mol. Biol.* **415**, 16–21 (2012).

37. B. Poldermans, L. Roza, P. H. Van Knippenberg, Studies on the function of two adjacent N6,N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of *Escherichia coli*. III. Purification and properties of the methylating enzyme and methylase-30 S interactions. *J. Biol. Chem.* **254**, 9094–9100 (1979).

38. E. Purta, M. O'Connor, J. M. Bujnicki, S. Douthwaite, YccW is the m5C methyltransferase specific for 23S rRNA nucleotide 1962. *J. Mol. Biol.* **383**, 641–651 (2008).

39. A. Y. Golovina, *et al.*, The last rRNA methyltransferase of *E. coli* revealed: the *yhiR* gene encodes adenine-N6 methyltransferase specific for modification of A2030 of 23S ribosomal RNA. *RNA.* **18**, 1725–1734 (2012).

Supplementary tables

Modification	Enzyme	Ref.	Modification	Enzyme	Ref.	Modification	Enzyme	Ref.
Ψ -516	RsuA	(13)	m^1G -745	RlmA	(14)	Gm -2251	RlmB	(15)
m^7 -G527	RsmG	(16)	Ψ -746	RluA	(17)	m2G -2445	RlmKL	(18)
m^2 -G966	RsmD	(19)	m^5 -U747	RlmC	(20)	D -2449	RdsA	This paper
m^5 -C967	RsmB	(21)	Ψ -955	RluC	(22)	Ψ -2457	RluE	(23)
m^2 -G1207	RsmC	(24)	m^6A -1618	RlmF	(25)	Cm -2498	RlmM	(26)
m^4 -C1402	RsmH	(27)	m^2G -1835	RlmG	(28)	ho^5C -2501	RlhA	(29)
Cm -1402	RsmI	(27)	Ψ -1911	RluD	(30)	m^2A -2503	RlmN	(31)
m^5 -C 1407	RsmF	(32)	$m^3Ψ$ -1915	RlmH	(33)	Ψ -2504	RluC	(22)
m^3 -U1498	RsmE	(34)	Ψ -1917	RluD	(30)	Um -2552	RlmE	(35)
m^2 -G1516	RsmJ	(36)	m^5U -1939	RlmD	(20)	Ψ -2580	RluC	(22)
m_6^2 -A1518	RsmA	(37)	m^5C -1962	RlmI	(38)	Ψ -2604	RluF	(23)
m_6^2 -A1519	RsmA	(37)	m^6A -2030	RlmJ	(39)	Ψ -2605	RluB	(23)
			m^7G -2069	RlmKL	(18)			

Table S1: List of *E.coli* ribosomal 23S RNA post-transcriptional modifications and their modifying enzymes. In light blue 16S rRNA modifications, green 23S rRNA modifications.

Gene name	Function
rhsB	rhsB element core protein RshB
yhhH	hypothetical protein
yrhC	
yhhI	putative transposase
yhhJ	putative permease subunit, ABC transporter superfamily, inner membrane protein
rbbA	fused ribosome-associated ATPase: ATP-bindingprotein/ATP-binding protein/predicted membrane protein
yhiI	putative membrane fusion protein (MFP) of efflux pump
yhiJ	hypothetical protein
yhiL	
yhiM	acid resistance protein, inner membrane
yhiN	putative oxidoreductase with FAD/NAD(P)-binding domain
pitA	phosphate transporter, low-affinity; tellurite importer
uspB	universal stress (ethanol tolerance) protein B
uspA	universal stress global response regulator
dtpB	dipeptide and tripeptide permease B
rsmJ	16S rRNA m(2)G1516 methyltransferase, SAM-dependent
prlC	oligopeptidase A
rlmJ	23S rRNA m(6)A2030 methyltransferase, SAM-dependent
gor	glutathione oxidoreductase
dinQ	damage inducible protein
arsR	DNA-binding transcriptional repressor
arsB	arsenite/antimonite transporter
arsC	arsenate reductase
yhiS	
insH1	IS5 transposase and trans-activator
slp	outer membrane lipoprotein
dctR	putative DNA-binding transcriptional regulator
yhiD	putative Mg(2+) transport ATPase, inner membrane protein
hdeB	acid-resistance protein
hdeA	stress response protein acid-resistance protein
hdeD	acid-resistance membrane protein
gadE	DNA-binding transcriptional activator
mdtE	anaerobic multidrug efflux transporter, ArcA-regulated
mdtF	anaerobic multidrug efflux transporter, ArcA-regulated
gadW	transcriptional activator of gadA and gadBC; repressor of gadX
gadX	DNA-binding transcriptional dual regulator
gadA	glutamate decarboxylase A, PLP-dependent
yhjA	putative cytochrome C peroxidase
treF	cytoplasmic trehalase
yhjB	putative two-component system response regulator, LuxR-like HTH domain, function unknown
yhjC	putative DNA-binding transcriptional regulator
yhjD	putative alternate lipid exporter, suppressor of msbA and KDO essentiality, inner membrane protein
yhjE	inner membrane protein, predicted transporter
yhjG	Inner membrane protein, AsmA family
yhjH	cyclic-di-GMP phosphodiesterase, FlhDC-regulated
kdgK	ketodeoxygluconokinase
yhjJ	putative zinc-dependent peptidase
dctA	C4-dicarboxylic acid, orotate and citrate transporter
yhjK	cyclic-di-GMP phosphodiesterase
bcsC	cellulose synthase subunit
bcsZ	endo-1,4-D-glucanase
bcsB	regulator of cellulose synthase, cyclic di-GMP binding
bcsA	cellulose synthase, catalytic subunit
bcsQ	
yhjR	hypothetical protein

Gene name	Function
bcsE	cellulose production protein
bcsF	hypothetical protein
bcsG	inner membrane protein, predicted endoglucanase, DUF3260 family
ldrD	toxic polypeptide, small
yhjV	putative transporter
dppF	dipeptide transporter
dppD	dipeptide/heme transporter
dppC	dipeptide/heme transporter
dppB	dipeptide/heme transporter
dppA	dipeptide transporter
eptB	KDO phosphoethanolamine transferase, Ca(2+)-inducible
yhjX	inner membrane protein, predicted oxalate-formate antiporter
yhjY	hypothetical protein
tag	3-methyl-adenine DNA glycosylase I, constitutive
yiaC	putative acyltransferase with acyl-CoA N-acyltransferase domain
bisC	biotin sulfoxide reductase
yiaD	multicopy suppressor of bamB; outer membrane lipoprotein
ghrB	glyoxylate/hydroxypyruvate reductase B

Table S2: List of genes comprised in genomic the deletion of the ME5079 strain and their known and/or predicted function(s).

Table S3, S4 and S5 are downloadable Excel files.

Strain	Genotype
E. coli	
BW25113	∆*(araD-araB567* ∆ *(rhaD-rhaB) 568* ∆*lacZ4787 (::rrnB-3) hsdR514 rph-1*
∆yhiN	BW25113 ∆*yhiN::kan*
∆yhhA	BW25113 ∆*yhhA::kan*
∆yhhB	BW25113 ∆*yhhB::kan*
∆yhhC	BW25113 ∆*yhhC::kan*
∆yhhI	BW25113 ∆*yhhI::kan*
∆yhiD	BW25113 ∆*yhiD::kan*
∆yhiE	BW25113 ∆*yhiE::kan*
∆yhiF	BW25113 ∆*yhiF::kan*
∆yhiG	BW25113 ∆*yhiG::kan*
∆rbbA	BW25113 ∆*rbbA::kan*
∆dusABC	BW25113 ∆*dusA::kan,* ∆*dusB::0,* ∆*dusC::0*
ME5125	*rpsL* ∆ *(yjgR-yjjA, yqjL-b3122, yabC-arcC, yjcD-cadC, yiaH-yibI, rhsB-yiaE, yeeE-mrp, yccC-yceE, ygaD-b2740, intA-emrB, yejO-alkB, ompC-yfaA, yfaE-b2275, ycjD-b1505, hipA-b1579, b0245-tauD, galM-ybjZ)::tet, bla, cat*
ME5079	*rpsL* ∆*(rhsB-yiaE)*
ME5081	*rpsL* ∆*(yiaH-yibI)*
ME5082	∆*(recB ptr recC recD)::Plac-(bet exo kan) rpsL hsdR::bla* ∆*(yjcD-cadC)::CRS*
ME5085	*rpsL* ∆*(yqjL-b3122)*
ME5086	*rpsL* ∆*(b0245-tauD)::bla*
ME5088	∆*(recB ptr recC recD)::Plac-(bet exo kan) rpsL hsdR::bla* ∆*(yfaE-b2275)*
ME5091	*rpsL* ∆*(yejO-alkB, ompC-yfaA)*
ME5092	*rpsL* ∆*(galM-ybjZ)::cat*
ME5094	∆*(recB ptr recC recD)::Plac-(bet exo kan) rpsL hsdR::bla* ∆*(yeeE-mrp)*
ME5099	*rpsL* ∆*(yccC-yceE)*
ME5102	*rpsL* ∆*(ycjD-b1505, hipA-b1579)::tet*
ME5106	*rpsL* ∆*(intA-emrB)*
ME5108	*rpsL* ∆*(yabC-arcC)*
ME5110	*rpsL* ∆*(yjgR-yjjA)*
V. cholerae	
N169161	*hapR*+
B.subtilis	
W168	*trpC2*

Tables S6: List of strains and their genotype used in this study

Name	Sequence
rRNA *B.subtilis* 1F	GCCCAAAGGTTCCCTCAGA
rRNA *B.subtilis* 2F	ATC TCC CCC AAG AGT CCA CAT CG
rRNA *B.subtilis* 3R	TTG GCA GAA CAA CTG GTA CA
rRNA *B.subtilis* 4R	AGT ATC TGT CAG CTC CAT GTG T
rRNA *E.coli* 1F	AGG AGC ACG AAG GTT GGC TAA TC
rRNA *E.coli* 2F	ATA CCG CCC AAG AGT TCA TAT CG
rRNA *E.coli* 3R	TTG GCA TGA CAA CCC GAA CA
rRNA *E.coli* 4R	AGT ACC GGT TAG CTC AAC GCA T
yhiN In pET15b UP	CCGCGCGGCAGCCATATGGAAAGGTTTGATGCCATTATTATAGGCGCT
yhiN In pET15b DN	GCAGCCGGATCCTCAGGACGACTTTGCTGCAATCAAATCCT
Oligo 1 for in vitro transcription	TAATACGACTCACTATA
Oligo 2 for in vitro transcription Mini-PTC-U	mGmCTGAGTCTGTCAGCCTGTTATCCCCGGAGTAAGTCCGGGCTATAGTGAGTCGTATTA (m is 2' O-methylation)
Oligo 2 for in vitro transcription Mini-PTC-C	mGmCTGAGTCTGTCAGCCTGTTATCCCCGGAGTAAGTCCGGGCTATAGTGAGTCGTATTA (m is 2' O-methylation)
rRNA protection (*E. coli*) 2423-2473	ACTCTTGGGCGGTATCAGCCTGTTATCCCCGGAGTACCTTTTATCCGTTGA

Table S7: List of oligonucleotides used in this study

Table S8 is a downloadable Excel file.

Supplementary figures

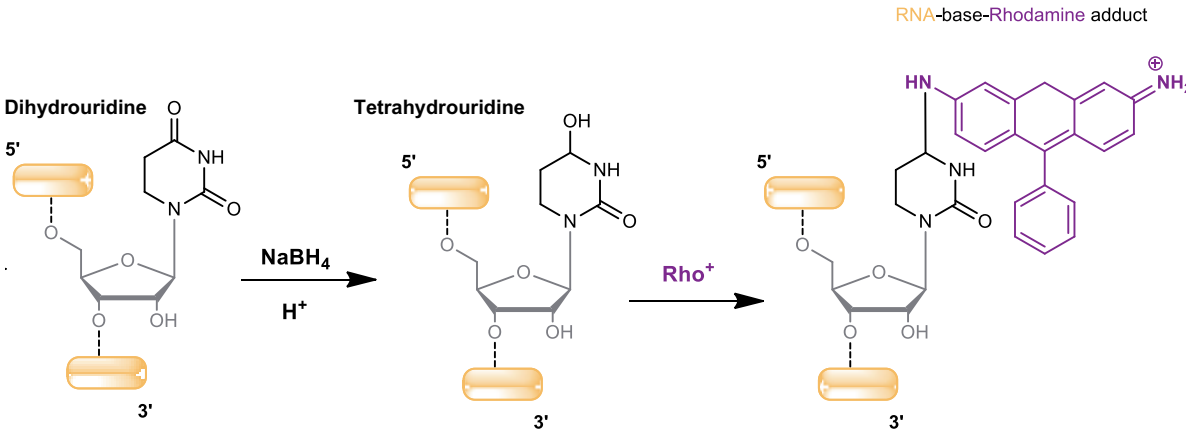


Figure S1: Chemical labeling reaction of dihydrouridine with Rhodamine-110

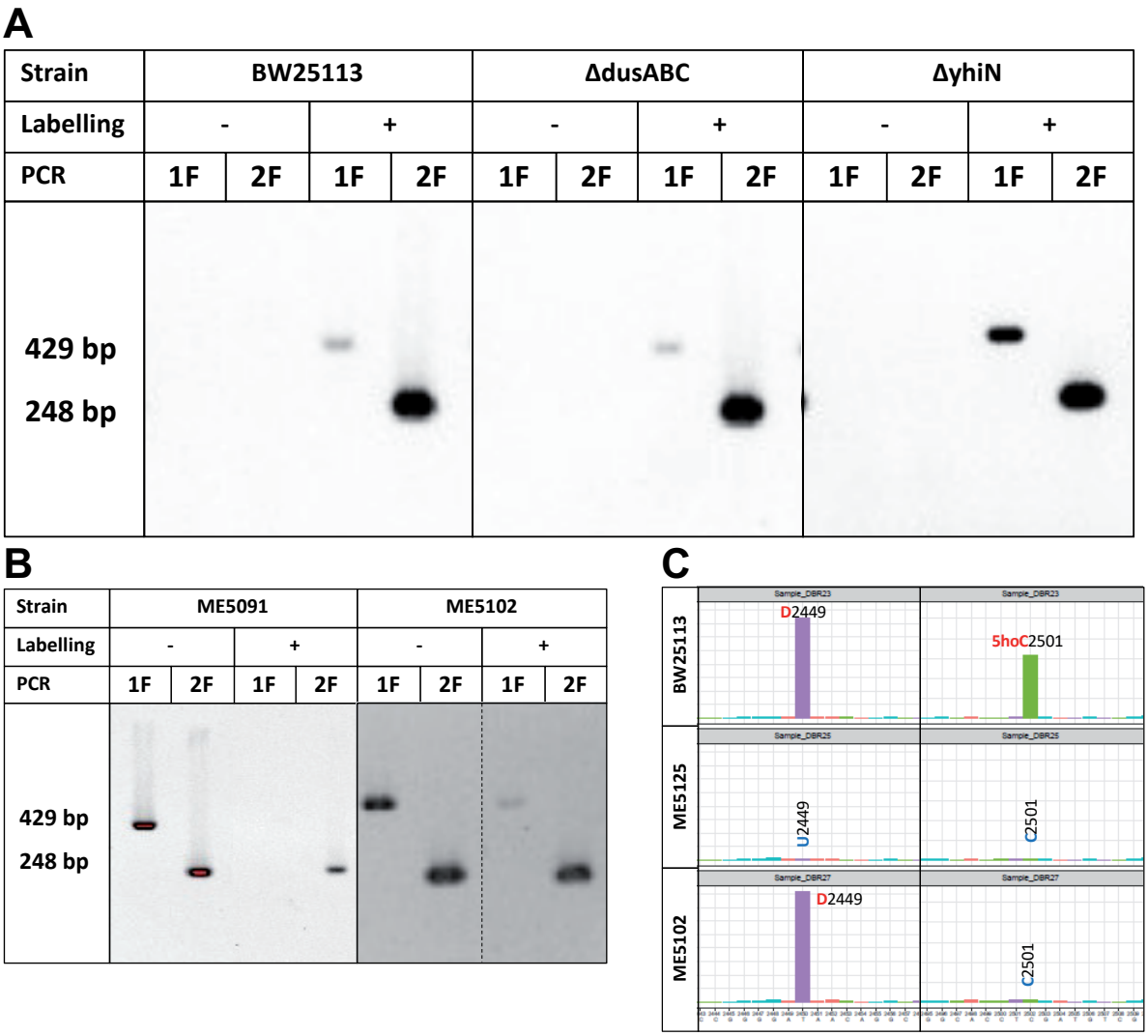


Figure S2: Detection of D2449 in other strains

RhoRT-PCR was applied to detect the presence of D2449 in ΔdusABC strain (2 repeats) (A) that shows here the same profile as the BW25113 with strongly reduced amplification of the 429 bp band compared to the ΔyhiN strain. RhoRT-PCR was also applied to detect the presence of D2449 in ME5082, ME5085, ME5086, ME5088, ME5081, ME5091, ME5092, ME5094, ME5099, ME5106, ME5108, ME5110, BW25113 Δ*yhhA::kan*, BW25113 Δ*yhhB::kan*, BW25113 Δ*yhhC::kan*, BW25113 Δ*yhhI::kan*, BW25113 Δ*yhiD::kan*, BW25113 Δ*yhiE::kan*, BW25113 Δ*yhiF::kan*, BW25113 Δ*yhiG::kan* and BW25113 Δ*rbbA::kan*. The representative result for all those strains (ME5091) is shown here (A). RhoRT-PCR shows a slight amplification of the 429 bp band which is attributed to the absence of 5hoC2501 in ME5102 (2 repeats) where the genomic deletion encompasses the *ydcP* gene coding the RlhA (5hoC2501 synthase) protein (B).

In (A and B), RT was performed with species specific 4R oligonucleotides. PCR amplification was performed with species specific 1F/3R (1F) and 2F/3R (2F) oligonucleotides. Dashed line indicates lanes reordering for the sake of clarity.

1

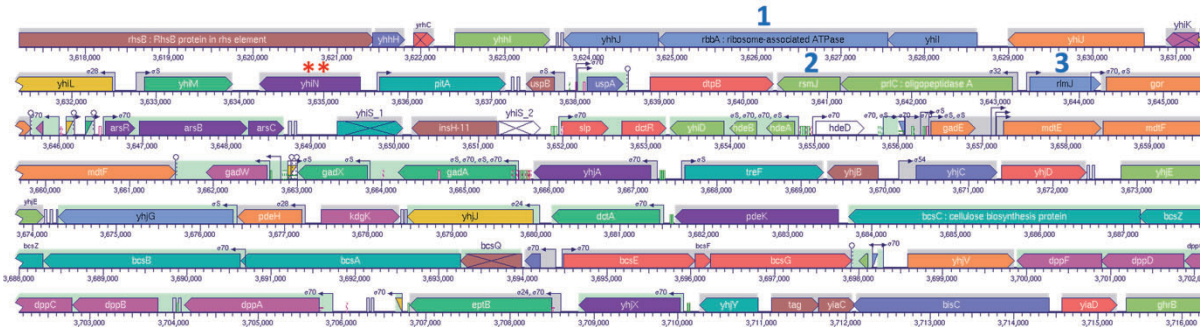


Figure S3: Genomic structure of deletion L2-1 deletion of the ME5079 strain
Picture was obtained from (http://vm-trypanocyc.toulouse.inra.fr/ECOLI/NEW-IMAGE?type=MAPTICK&object=3617215/3716307&chromosome=COLI-K12&redirect=T&bp-range=3617215/3716307). The yhiN gene is indicated a double red-asterisc. Other genes involved, or potentially involved, in ribosome biogenesis or function are indicated with blue numbers 1) rbbA ribosome-associated ATPase, 2) rsmJ 16S rRNA m(2)G1516 methyltransferase and 3) rlmJ 23S rRNA m(6)A2030 methyltransferase.

1
2
3

```
        10         20         30         40         50         60         70         80         90
MERFDAIIIG AGAAGMFCSA LAGQAGRRVL LIDNGKKPGR KILMSGGGRC NFTNLYVEPG AYLSQNPHFC KSALARFTQW DFIDLVNKHG

       100        110        120        130        140        150        160        170        180
IAWHEKTLGQ LFCDDSAQQI VDMLVDECEK GNVTFRLRSE VLSVAKDETG FTLDLNGMTV GCEKLVIATG GLSMPGLGAS PFGYKIAEQF

       190        200        210        220        230        240        250        260        270
GLNVLPTRAG LVPFTLHKPL LEELQVLAGV AVPSVITAEN GTVFRENLLF THRGLSGPAV LQISSYWQPG EFVSINLLPD VDLETFLNEQ

       280        290        300        310        320        330        340        350        360
RNAHPNQSLK NTLAVHLPKR LVERLQQLGQ IPDVSLKQLN VRDQQALIST LTDWRVQPNG TEGYRTAEVT LGGVDTNELS SRTMEARKVP

       370        380        390        400
GLYFIGEVMD VTGWLGGYNF QWAWSSAWAC AQDLIAAKSS
```

Figure S4: Amino acid sequence of RdsA

4
5

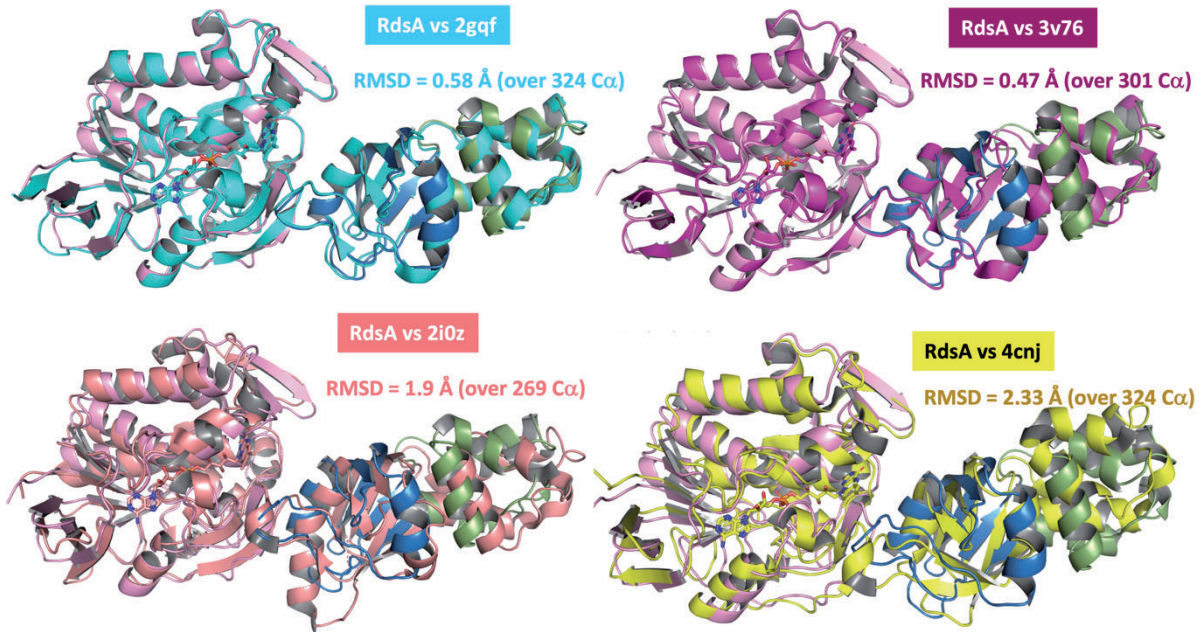


Figure S5: Structural alignment of the AlphaFold model of E. coli RdsA with the crystallographic structure of homologous counterparts. The FDB, Ins1D, and Ins2D domains of Ec RdsA are colored in pink, blue, and green, respectively.

1

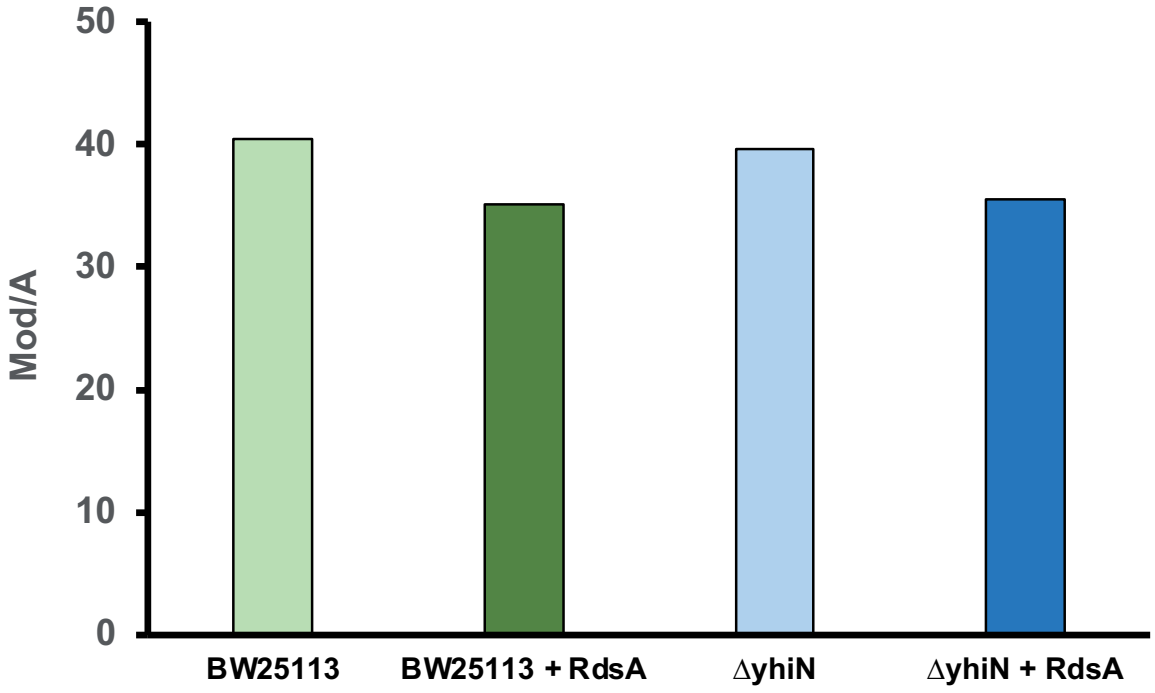


Figure S6: D quantification in ribosomes incubated with RdsA
Ribosomes extracted (10 µM) from BW25113 (WT) or BW25113 Δ*yhiN::kan* strains were incubated in reaction buffer with purified RdsA (25 µM). D was quantified by LC-MS. Results are expressed as the relative amount of quantified D per quantified A. As mentioned in the main text, quantifications show here a notable contamination of D most probably originating from tRNA and were therefore attempted only once.

2

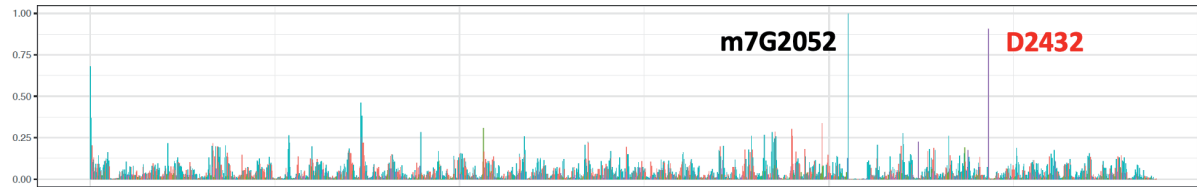


Figure S7: AlkAlinine-Seq profile of *V. cholerae* 23S rRNA

1

2

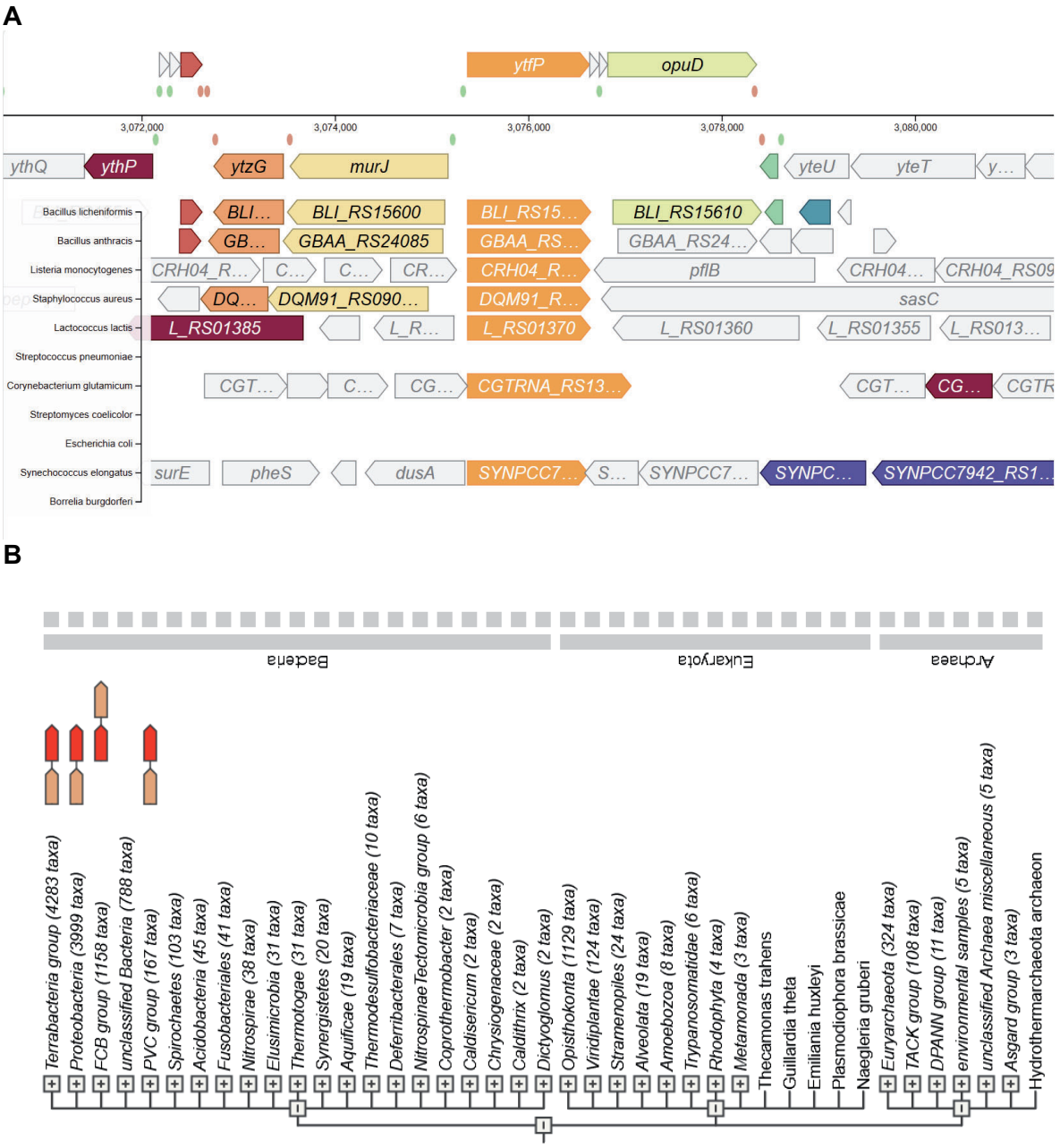


Figure S8: Examples of Sporadic gene neighborhood associations between of COG2081 and RNA modifications enzymes.
(A) Physical clustering between *dusA* and the COG2081-encoded homolog in *Synecoccus elongatus* extracted from (http://corewiki.uni-goettingen.de/gene/3184); (B) Physical clustering associations in the String Database using the family search tool between COG2081 (Predicted flavoprotein YhiN) in red and COG1187 (Pseudouridylate synthase RsuA) in orange. https://string-db.org/cgi/network?taskId=bU9CkNV4A16g&sessionId=bplhPKknoWLb&allnodes=1

1

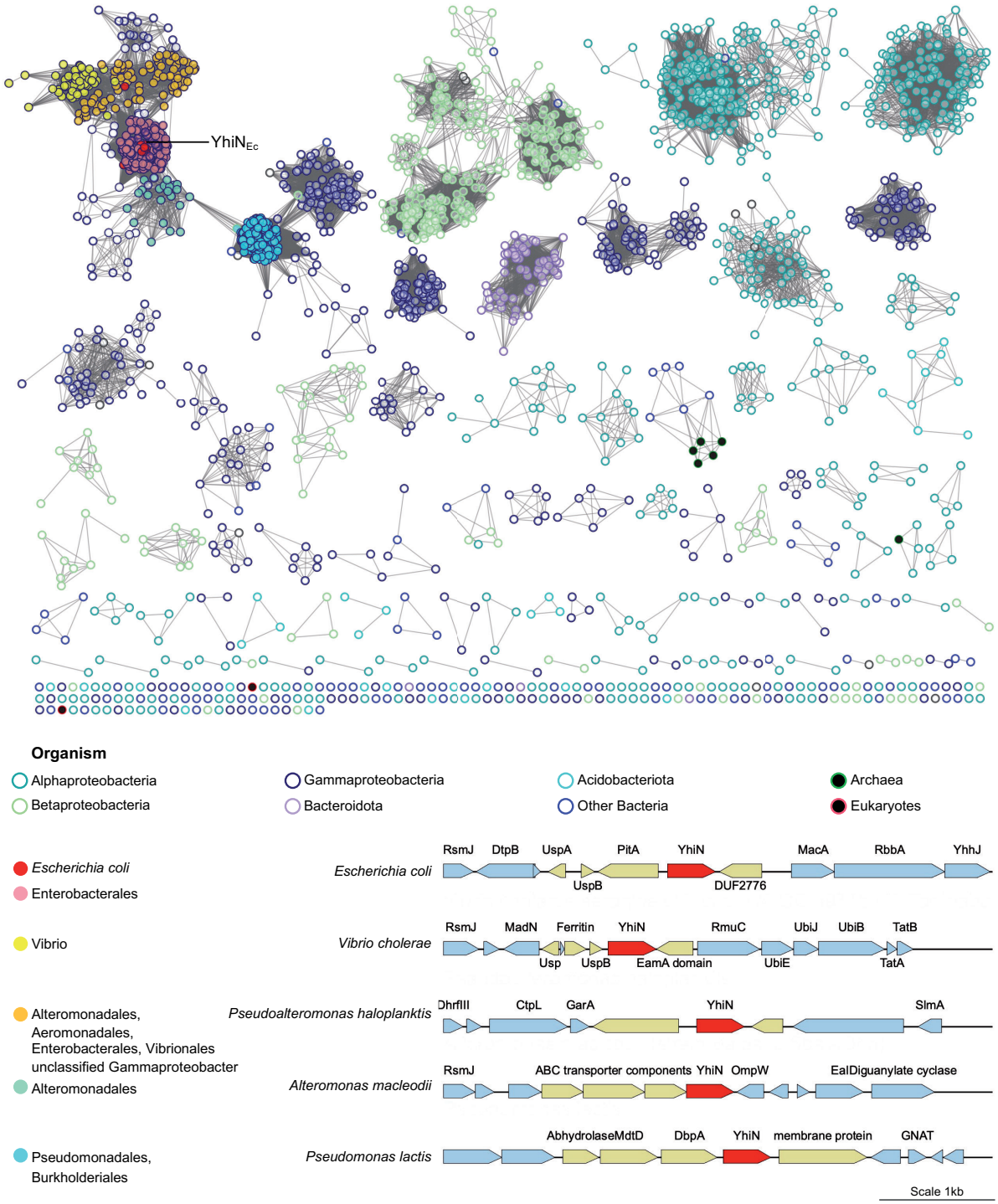


Figure S9: SSN of PF03486 family proteins in the cluster containing YihNEc.
SSN with the alignment score threshold of 150 of 2,403 PF03486 family proteins in the YihNEC cluster (boxed in Figure 5). Each node in the network represents one or multiple PF03486 family proteins that share 100% identity. An edge (represented as a line) is drawn between two nodes with a BLAST E-value cutoff of better than 10–150. The nodes are colored based on taxonomy of the corresponding genomes. Nodes in the proximity of YhiNEc are filled and the corresponding representative gene neighborhood organization are shown. Uniprot IDs for YhiN proteins, *E. coli* A0A066SWL7, *Vibrio cholerae* Q9KVQ9, *Pseudoalteromonas haloplanktis* A0A9W4QTT1, *Alteromonas macleodii* K0EB93, and *Pseudomonas lactis* I4K9X8.

1

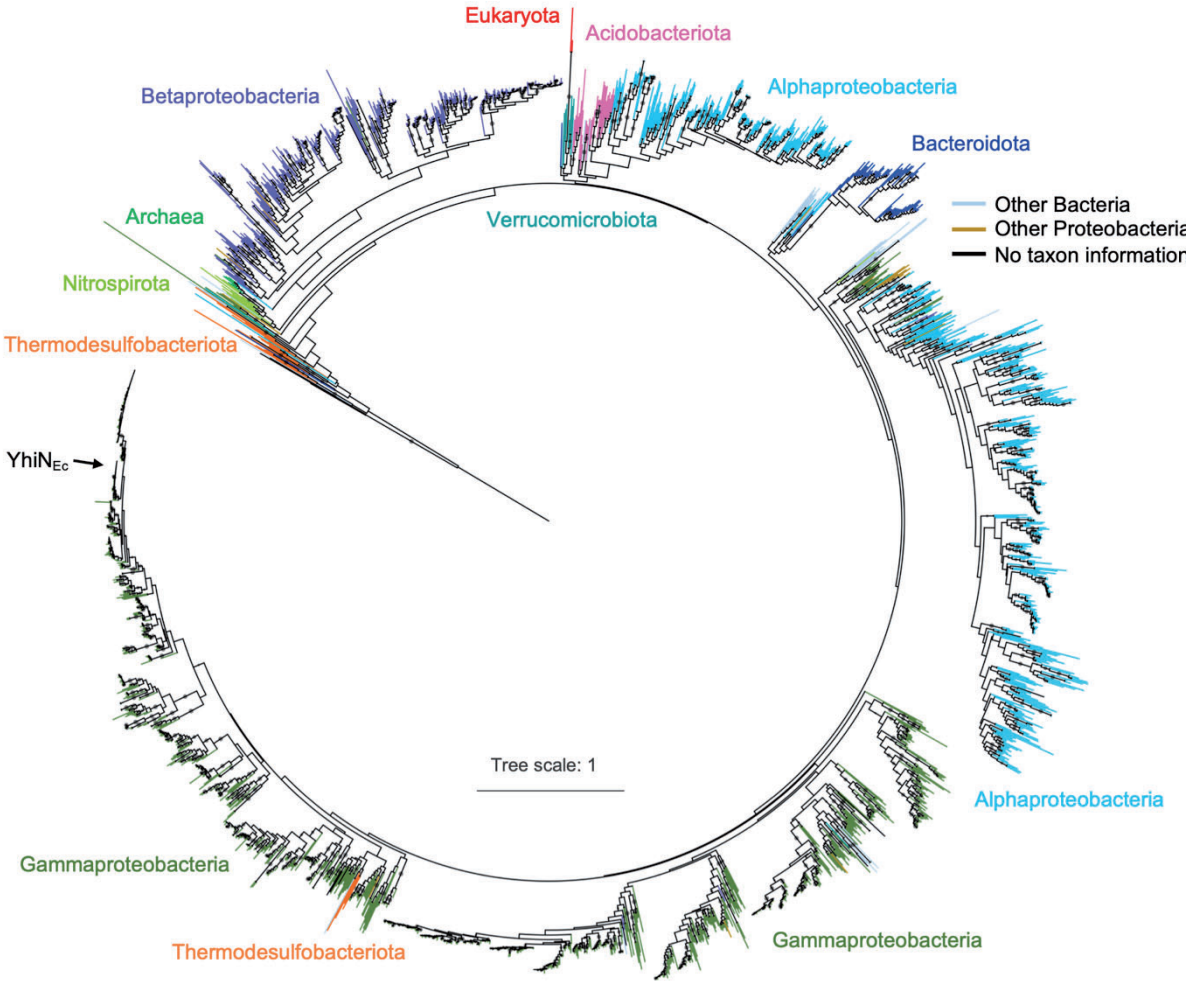


Figure S10: The phylogeny of YhiN proteins.
The maximum likelihood tree of 2,310 YhiN proteins in the YihN$_{EC}$ cluster (boxed in Figure 5). For better visualization, the branches are grouped and colored by kingdom, phyla and class. The branches with bootstrap support value more than 50 were labeled by dots.

1
2

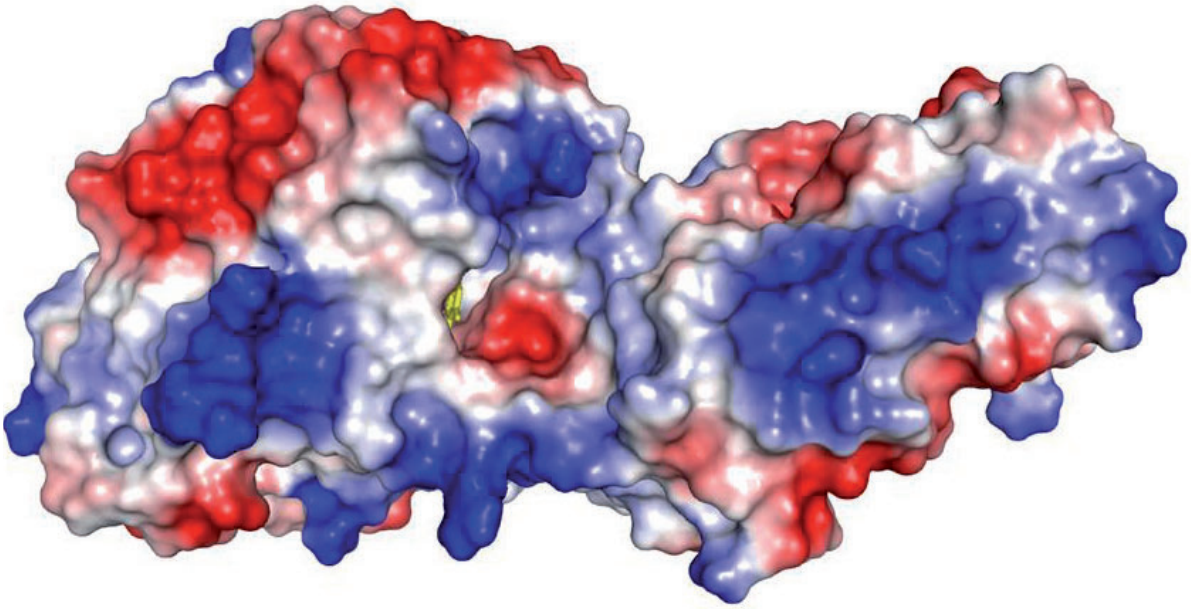


Figure S11: Electrostatic surface representation of *E. coli* RdsA generated from the AlphaFold model.

Chapter III: The biological significance of ribosomal RNA dihydrouridylation and dihydrouridine synthase.

1. Preamble

Post-transcriptional RNA modifications have gained increasing attention in recent years as a cellular mechanism to regulate RNA structure, function, and stability. The physiological importance of RNA modifications is highlighted by the severe consequences that can arise from their misregulation. Ribosomal RNAs are no exception, as they are decorated with numerous modifications. In *Escherichia coli*, there are 36 modification sites on rRNA: 11 on the small subunit 16S rRNA and 25 on the large subunit 23S rRNA (Popova & Williamson, 2014). The most common modifications include the conversion of uridine to pseudouridine and various ribose and nucleobase methylations. These modifications tend to cluster around functionally important sites on the ribosome, such as the mRNA decoding center, tRNA binding sites, the nascent peptide exit tunnel, and the peptidyl transferase center (PTC). While the specific functions of individual modifications remain largely unknown, they are thought to contribute to proper ribosome biogenesis (Ero et al., 2024) , stability, interactions with ribosomal ligands like tRNAs, and antibiotics resistance against drugs targeting the ribosome. However, no single modification has been shown to be essential (Fasnacht et al., 2022) and deletion of these modifications generally results in mild phenotypes, suggesting that post-transcriptional rRNA modifications play a fine-tuning role in translational activity.

Unlike in eukaryotes and archaea, where rRNA modifications are predominantly facilitated by small nucleolar RNA–protein complexes (snoRNPs), *E. coli* relies on protein-only enzymes that are site- or region-specific (Decatur & Fournier, 2002) (see introduction). This means that almost every rRNA modification site has its own dedicated enzyme. All but one of the rRNA modification enzymes in *E. coli* have been identified, with the enzyme responsible for dihydrouridine formation at residue 2449 of the 23S rRNA remaining elusive. However, thanks to our work, RdsA was identified as the ultimate modifying enzyme for *E. coli* rRNA modifications (Toubdji et al PNAS 2024). It catalyzes the reduction of the C5-C6 double bond of the uridine residue U-2449. Nucleotide U2449 is located in the PTC and adjacent to nucleotide A2451, a crucial component of the PTC.

The importance of dihydrouridine (D) in RNA conformational dynamics was initially demonstrated through structural studies. Its biological significance was further highlighted by MacCloskey's analysis of post-transcriptional modifications in tRNAs from psychrophilic organisms, which thrive in extremely low temperatures (−5 to 12 °C) (J. J. Dalluge et al., 1997) (Noon et al., 2003). These organisms employ evolutionary strategies to maintain molecular flexibility in cold environments, incorporating biochemical components that enhance molecular mobility. Notably, while psychrophilic organisms have fewer post-transcriptional modifications in their tRNAs, they maintain normal levels of pseudouridine, m5U, and m7G, and show significantly higher levels of D (40-70% more) compared to mesophilic organisms like *E. coli*. These findings support the role of D in enhancing nucleic acid flexibility.

Depletion of D in *E. coli* by deleting *dus* genes does not significantly affect growth (Bishop et al., 2002). Similarly, the absence of the enzyme responsible for rRNA dihydrouridine synthesis prompted (O'Connor, 2001) to mutate U2449 to C2449 to assess its impact. Their study revealed that D2449 is not essential for *E. coli* physiology or ribosome function. This aligns with observations that many "nonessential" RNA modifications, often located within the RNA body, have minor phenotypic effects when removed. These modifications are part of an interconnected network, where compensation or functional redundancy can mitigate the loss of a single modification. The biological significance of these modifications becomes apparent under stress or when the network is disrupted beyond the loss of an individual modification (Fasnacht et al., 2022)(see introduction)

We therefor set out to first analyze growth curve and competitions between wild type *E. coli* and mutant *ΔyhiN* at different temperatures. Our data showed a slight delay in generation time for the mutant at low temperature 23°C and a competitive disadvantage at high temperature 42°C. We showed, however, that the presence of the RdsA-catalyzed modification at U2449 is beneficial for *E. coli* grown under oxidative stress. Finally, we revealed a fine-tuning role of D-2449 for translation utilizing a β-gal/Luciferase reporter system allowing the measurement of translation errors. Taken together our results showed that reduction of the C5-C6 double bound at 23S rRNA position 2449 fine-tunes translation and provides a growth benefit to *E. coli* during oxidative stress.

2. Results

2.1 RdsA-catalyzed reduction of the U2449 provides cells with a growth advantage

To compare the cellular fitness of strains expressing or lacking the *yhiN* gene, growth curves were performed in rich medium under optimal growth conditions at different temperature from 23°C to 42°C. A slight growth defect of the knockout strain of 17 minutes compared to the wildtype strain could be observed at low temperature 23°C **Figure 54**.

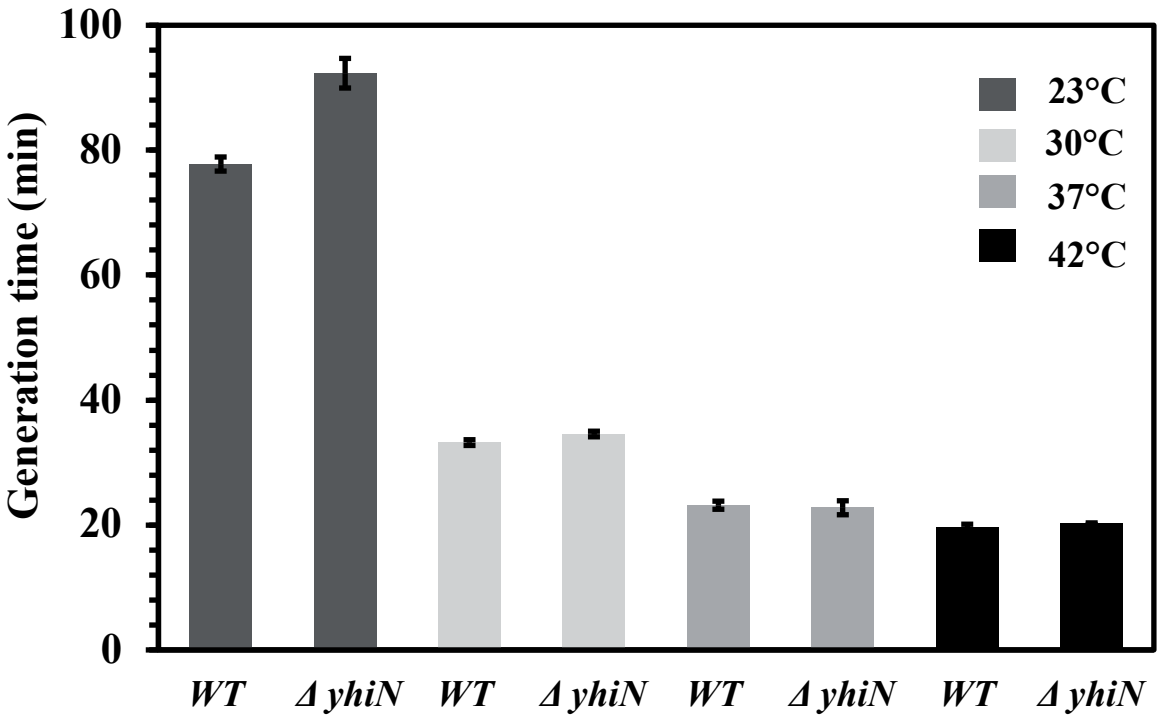


Figure 54: Growth phenotype of *E. coli WT* and mutant strain *ΔyhiN*. Average growth curves of the knockout strain ΔyhiN and the wild type *E. coli*. All strains were incubated in LB medium at different temperatures 23°C, 30°C, 37°C, 42°C.

The minor growth deficiency of the *ΔyhiN* strain was enhanced in growth competition experiments. In such assays, equal cell numbers (based on OD600 measurements) of the *ΔyhiN* and WT strains were grown for 24 h. (**Figure 55**). The ratio of wildtype to knockout strain was monitored by plating in parallel on LB-Agar plates (which allows growth of both WT and *ΔyhiN::Kan* strains) and kanamycin selection plates (where only the *ΔyhiN::Kan* cells can form colonies). A clear and steady decrease of the *ΔyhiN* population became apparent in the

redilution cultures at high temperature 42°C, demonstrating successful growth competition of cells expressing RdsA over the knockout strain.

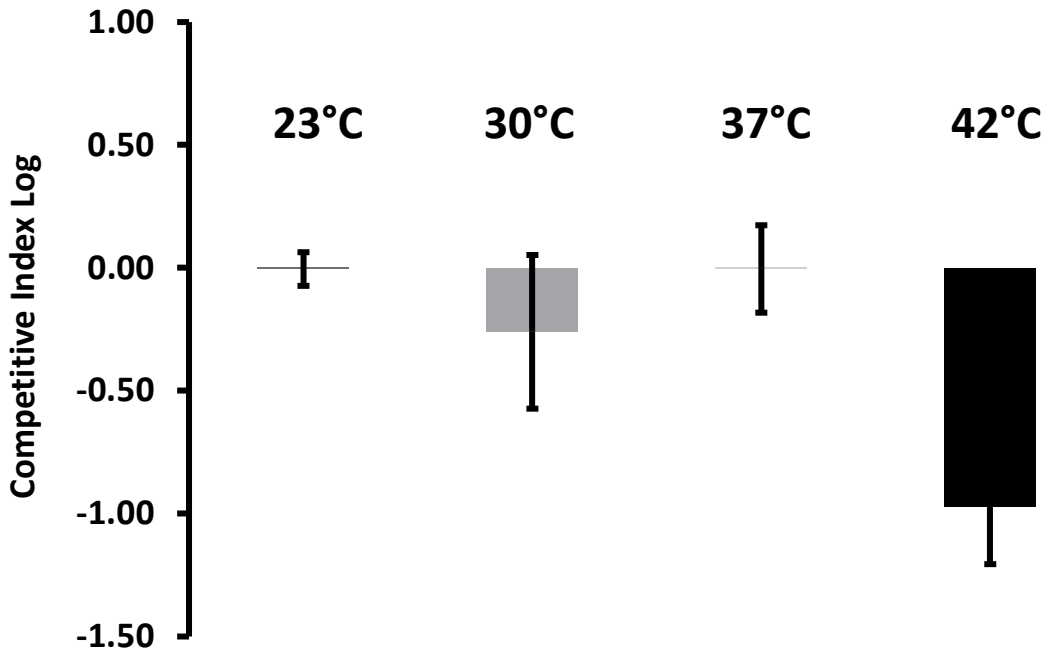


Figure 55: Growth competition between *E. coli* BW25113 wild type strain and *ΔyhiN* Mutant at 42°C, 37°C, 30°C and 23°C. Initial bacterial inoculum determined by preculture titration. After 24 hours of incubation, total bacterial count assessed on LB-agar plates and the number of mutant cells by plating on LB-agar plates containing Kanamycin *(*Δ*yhiN::Kan*). Competitive index calculated by dividing mutant cells by wild-type cells post-growth, normalized to initial ratio. Y-axis: Log competitive index.

2.2 D2449 benefits cells under oxidative stress

It was shown that the presence of ho5C2501, nucleobase C5 hydroxylation of the 23S rRNA residue C2501 catalyzed by RlhA, provides a strong growth advantage during oxidative stress. Deletion of *rlhA* causes mild growth-defects under normal and temperature-stress conditions but confers a strong growth disadvantage during oxidative stress. While ho5C2501 slightly inhibits in vitro protein synthesis, it enhances in vivo translation under oxidative stress, preparing cells for stress by boosting protective proteins like Dps, helping cells to conserve energy. This modification, therefore, fine-tunes translation rates to match environmental challenges. In summary *E. coli* regulates ho5C2501 modification to balance the need for rapid growth and survival under oxidative stress, demonstrating a sophisticated adaptation mechanism through ribosome modification (Fasnacht et al., 2022)

In order to evaluate the effect D2449 under oxidative stress, I measured growth curve for both *E. coli* strain wild type and mutants *ΔyhiN, ΔrlhA* and under oxidative stress H_2O_2.

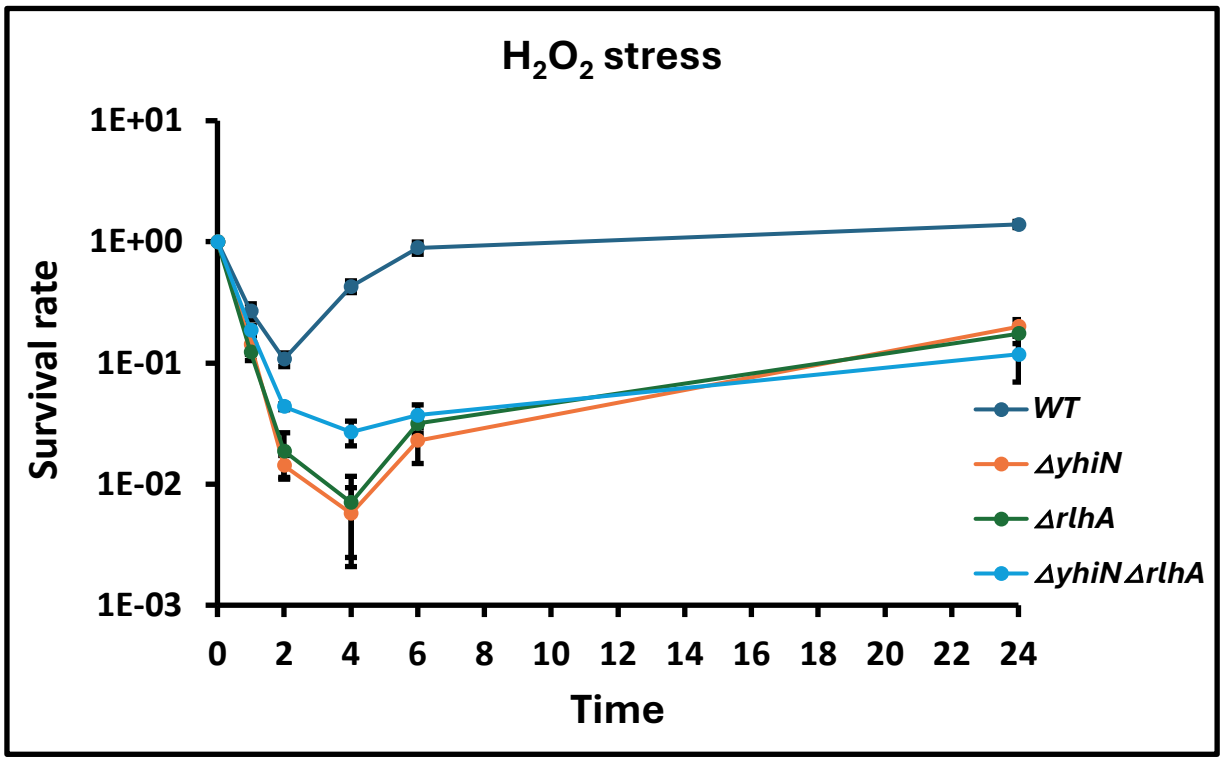


Figure 56: Survival to H_2O_2 treatment *E. coli* WT and deletion mutant *ΔyhiN* and *ΔrlhA*. cultures were grown without H_2O_2 up to early exponential phase. Total number of bacteria was determined by plating on LB-agar plates before addition of 2mM of H_2O_2, at time T0. After 1 h, 2h, 4h, 6h and 24h of incubation with H_20_2, the number of surviving bacteria was determined by plating on LB-Agar plates. Survival was calculated for each strain by dividing the number of surviving bacteria at T(x) by the initial number of bacteria. The y-axis represents the log10 survival ratio.

The analysis of cell survival as shown in **Figure 56**, it appears that the deletion of *yhiN* has a strong effect on survival under oxidative stress. Surprisingly, the effect is similar when *rlhA* is deleted but no addition effects are observed (Fasnacht et al., 2022)

To investigate whether D2449 influenced the expression of specific proteins, whole proteome analysis was performed on the *ΔyhiN* and the wild type *E. coli* strains in collaboration with Mariette Matondo Head of the Proteomics Core Facility Mass Spectrometry for Biology (MSBio) the proteomic analysis were performed by Rayen Elj (**Figure 57**).

Protein samples were taken from unstressed cells. Comparing both strains under unstressed conditions, only few proteins were found to be differentially expressed. The most significantly upregulated protein in the *ΔyhiN* strain compared to the wt strain, was the sodA sodC

(superoxide dismutase) strongly indicating that the *ΔyhiN* cells are already combating oxidative stress.

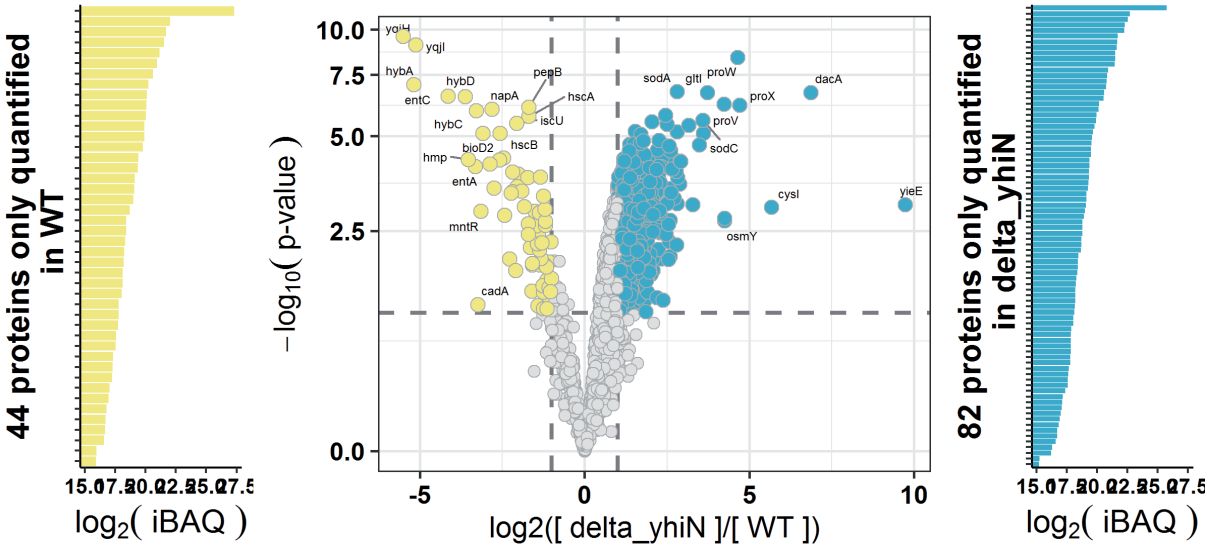


Figure 57: Volcano plot of the comparative proteome analysis of unstressed wt cells versus unstressed *ΔyhiN* cells. Differential Analysis with log2(Fold-Change) = 1 and FDR = 1 %.

2.3 Investigation of D2449 potential functions in translation

D2449 loss does not affect polysomes in a global manner

Polysome fractionation analysis (polysome profiling) allows to study the association of mRNAs with ribosomes and provide a snapshot of the translation activity in the cell. It also allows the analysis of the ribosomal components of the mutant strain, and particularly, the accumulation or underproduction of 30S and 50S ribosomal subunits (Charollais et al., 2003) (Charollais, 2004) (Proux et al., 2011). Thus, we compared the polysome profiles of *E. coli ΔyhiN* and WT cells to detect variations in ribosomal subunit abundance, ribosome assembly and translation efficiency (Figure 58). After growth in LB medium at 37°C, cells were lysed, and ribosomal particles were separated by sucrose gradient sedimentation under conditions favoring ribosome association (see Materials and Methods section).

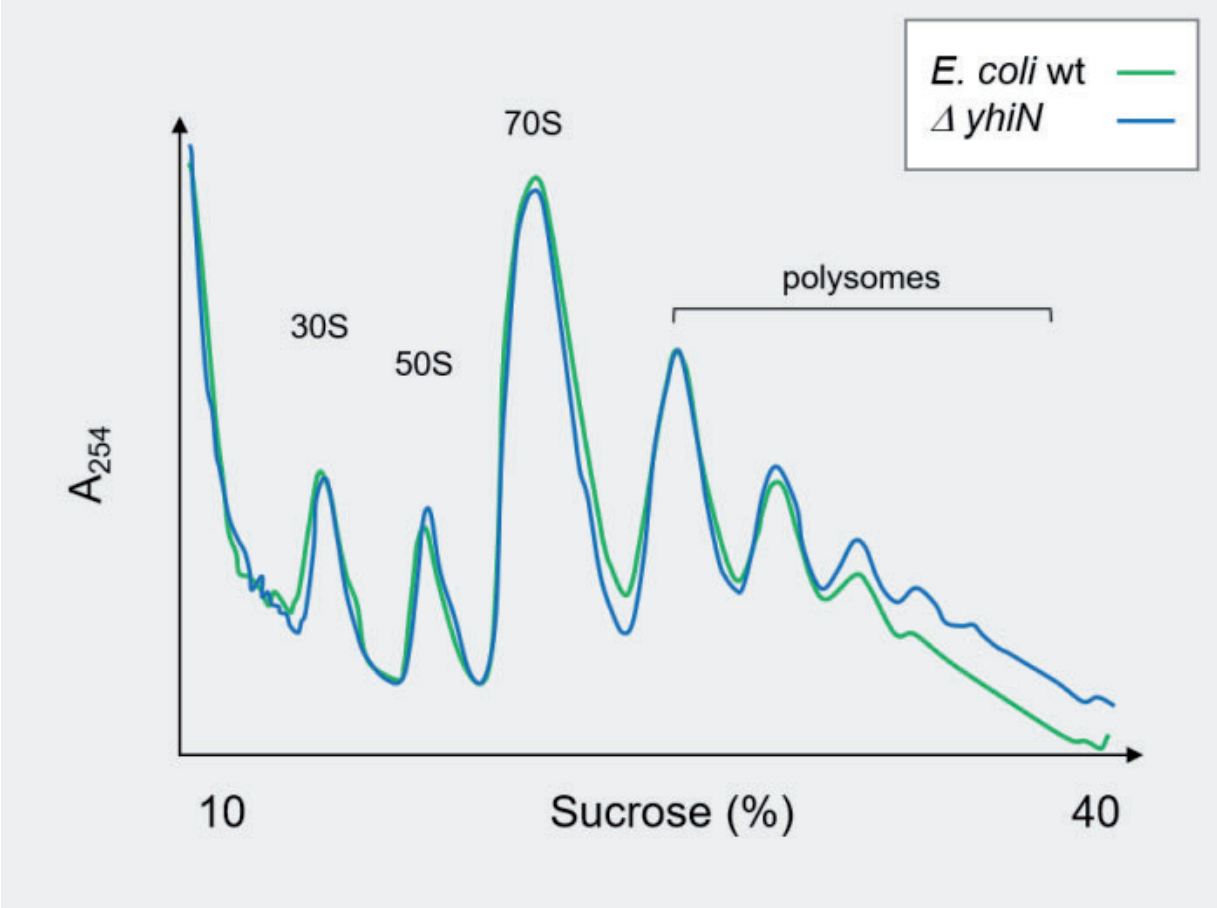


Figure 58: Polysome profile of *ΔyhiN* and wild type *E. coli* strains. After growth in LB medium at 37°C, cells were lysed, and cell extracts were layered onto a 10–40% (w/v) linear sucrose gradient. Ribosomal particles were separated by centrifugation at 39,000 rpm in a SW41 Beckman rotor for 2.75 h at 4°C. Gradients were analyzed by continuous monitoring of optical density at 254 nm. Peaks corresponding to free 30S and 50S subunits, 70S ribosomes (free 70S and monosomes), and polysomes are indicated above the absorbance profile of the gradients.

The polysome profile observed with the *ΔyhiN* strain strongly resembles the profile obtained with the wild-type strain. The peaks of the free ribosomal 30S and 50S subunits, 70S ribosomes (mixture of monosomes and free ribosomes), and light polysomes of the mutant strain are strongly similar to that of the wild-type strain. The small differences in heavy polysome peaks between *ΔyhiN* and wt strains was not reproducibly observed and was likely not due to differences in translation efficiency. These results suggest that ribosomal particle assembly and translation efficiency are not strongly affected by *ΔyhiN* gene deletion.

D2449 involved in translation fine tuning?

The above results argue against a general translation failure upon D2449 loss. In order to investigate potential translation defects, I performed translation errors measurement using the reporter system of the pBAD33LacLuc vector which contain the lacZ-luc gene fusion containing a unique MscI site at the junction of both CDS. The gene is expressed under arabinose inducible promoter and terminator.

In addition to the wild-type *E. coli* strain and the *ΔyhiN* mutant, we decided to test two other mutant strains: *ΔrlhA,* as described above, and *ΔrlmL*, a deletion strain for the RlmL methyltransferase enzyme, which is responsible for modifying the 23S rRNA at position G2447. All three mutant strains lack genes encoding enzymes that modify the rRNA involved in the central loop of domain five of the 23S rRNA, which constitutes the peptidyl-transferase center of the ribosome.

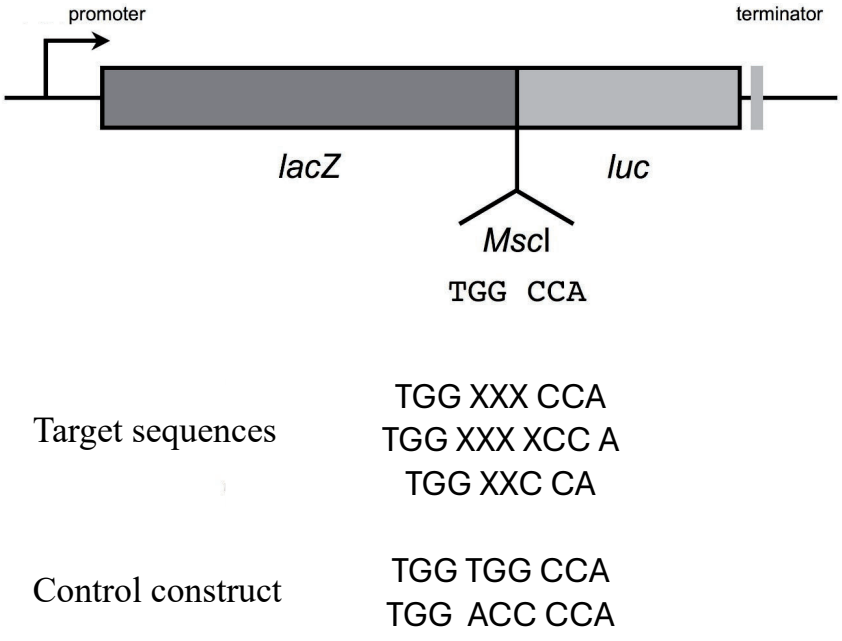


Figure 59: Schematic representation of the reporter system pBAD33LacLuc.

To biologically assess the propensity of chosen sequences and DnaX to induce translation errors, complementary synthetic oligonucleotides were annealed and cloned into the unique MscI site of a plasmid located between the ß-galactosidase and luciferase open reading frames using the SLIC technique (Jeong et al., 2012). For each sequence tested (**Table 7**), a control sequence was also constructed.

Table 7: Translation errors sequence.

Error sequence name	Sequence
TGG Control	AGGGGATCCCGCTAGCTGG**TGG**CCAAACATAAAGAAAG
TGA STOP codon	AGGGGATCCCGCTAGCTGG**TGA**CCAAACATAAAGAAAG
TAG STOP codon	AGGGGATCCCGCTAGCTGG**TAG**CCAAACATAAAGAAAG
TAA STOP codon	AGGGGATCCCGCTAGCTGG**TAA**CCAAACATAAAGAAAG
Frameshift +1 TGG	AGGGGATCCCGCTAGCTGG**TGGC**CCAAACATAAAGAAAG
Frameshift -1 TGG	AGGGGATCCCGCTAGCTGG**TG**CCAAACATAAAGAAAG
ACC Control	AGGGGATCCCGCTAGCTGG**ACC**CCAAACATAAAGAAAG
Frame shift -1 ACC	AGGGGATCCCGCTAGCTGG**AC**CCAAACATAAAGAAAG
DnaX +1	TggCAGGGAGCAACCAAAGCAAAAAAGGAGCGAACCGGCAGCCGCTACCCGCGCGCGGCCGGTcca
DnaX	TggCAGGGAGCAACCAAAGCAAAAAAGAGCGAACCGGCAGCCGCTACCCGCGCGCGGCCGGTcca

After strains transformation with this vector, each translation event will give rise to a protein with ß-galactosidase activity, but the luciferase activity of this protein will only be present if error event occurs during the translation of the target sequence.

Therefore, for each sequence to be tested, enzyme activities expressed from the control construct were also quantified. So that no translational recoding event is required to obtain a ß-galactosidase-luciferase fusion protein. For each target sequence, the ratio of luciferase activity versus ß-galactosidase activity expressed from the control constructs is set as the 100% reference point. Thus, comparison of the ratio of luciferase activity versus ß-galactosidase activity obtained from target sequence-containing and control constructs gives an accurate estimate of the error rate.

The translation errors measurement was performed in normal condition and under oxidative stress 2mM H_2O_2 at least four times and the results were reproducible. We obtained a significant increase in -1 frame shift level for the *ΔyhiN* 28.58% and *ΔrlmL* 25.20% at normal condition and under oxidative stress.

Table 8: Translation errors rates of *in silico*-selected sequences and Dnax.

		Normal condition			Under oxydative stress		
Strains	**Construction**	**Erreur rate**	**Ratio △/WT**	**Errors**	**Erreur rate**	**Ratio△/WT**	**Errors**
WT	TGA stop codon	27.95%	**1.00**	±1.00	31.76%	**1.00**	±1.00
ΔyhiN	TGA stop codon	23.59%	**0.84**	±0.65	34.87%	**1.10**	±0.95
ΔrlhA	TGA stop codon	11.96%	**0.43**	±0.48	15.73%	**0.50**	±0.65
ΔrlmL	TGA stop codon	11.54%	**0.41**	±0.29	26.80%	**0.84**	±1.28
WT	TAG stop codon	0.33%	**1.00**	±1.00	0.57%	**1.00**	±1.00
ΔyhiN	TAG stop codon	0.59%	**1.77**	±0.48	0.85%	**1.50**	±1.30
ΔrlhA	TAG stop codon	0.21%	**0.62**	±0.69	0.23%	**0.41**	±0.53
ΔrlmL	TAG stop codon	0.37%	**1.11**	±0.78	3.01%	**5.31**	±8.04
WT	TAA stop codon	0.27%	**1.00**	±1.00	0.33%	**1.00**	±1.00
ΔyhiN	TAA stop codon	0.36%	**1.31**	±1.01	0.40%	**1.22**	±1.06
ΔrlhA	TAA stop codon	0.17%	**0.63**	±0.70	0.11%	**0.34**	±0.44
ΔrlmL	TAA stop codon	0.11%	**0.41**	±0.29	0.21%	**0.65**	±0.98
WT	Frameshift+1TGG	0.95%	**1.00**	±1.00	1.38%	**1.00**	±1.000
ΔyhiN	Frameshift+1TGG	0.77%	**0.81**	±0.62	1.29%	**0.94**	±0.810
ΔrlhA	Frameshift+1TGG	0.29%	**0.30**	±0.34	0.54%	**0.39**	±0.512
ΔrlmL	Frameshift+1TGG	0.54%	**0.57**	±0.40	1.06%	**0.77**	±1.166
WT	Frameshift-1TGG	0.02%	**1.00**	±1.00	0.03%	**1.00**	±1.00
ΔyhiN	Frameshift-1TGG	28.58%	**1272.76**	±979.59	23.89%	**695.74**	±602.31
ΔrlhA	Frameshift-1TGG	0.01%	**0.52**	0.58	0.02%	**0.64**	±0.84
ΔrlmL	Frameshift-1TGG	25.20%	**1122.14**	±788.52	36.66%	**1067.45**	±1615.60
WT	Frameshift -1ACC	2.24%	**1.00**	±1.00	1.85%	**1.00**	±1.00
ΔyhiN	Frameshift -1ACC	15.64%	**7.00**	±11.76	13.56%	**7.35**	±10.70
ΔrlhA	Frameshift -1ACC	1.39%	**0.62**	±1.04	0.93%	**0.50**	±0.08
ΔrlmL	Frameshift -1ACC	1.15%	**0.52**	±1.01	2.83%	**1.53**	±0.01
WT	Dnax	35.32%	**1.00**	±1.00	38.55%	**1.00**	±1.00
ΔyhiN	Dnax	41.50%	**1.17**	±0.79	83.39%	**2.16**	±1.94
ΔrlhA	Dnax	41.89%	**1.19**	±1.17	68.48%	**1.78**	±2.04
ΔrlmL	Dnax	41.78%	**1.18**	±0.75	39.35%	**1.02**	±0.38

3. Discussion

Our study set out to analyze the functional impact of the RdsA-catalyzed dihydrouridine modification at uridine 2449 (U2449) within the 23S rRNA of *E. coli*. We began by examining the growth rates of Δ*yhiN* mutant at different temperature. Even though a slight increase of the generation time is observed at 23°C, 77±1.1 min for WT and 92±2.3 min for the mutant, no obvious impact of the absence of D2449 on exponential growth is observed. We then decided to grow mutant cells in competition with wild type cells. This kind of experiments is widely used to evaluate the impact of a mutation on the fitness of a population of cells and is not restricted only to exponential growth parameters. Mutant cells deprived of D at position 2449 on the 23S rRNA clearly show a growth disadvantage at high temperature (42°C) with about 10 times fewer mutant cells than wild type one after 24h of growth in competition. No difference is observed at other temperatures meaning that mutant cells have similar growth parameters in those conditions.

Furthermore, we investigated the effect of the U2449 modification on *E. coli* grown under oxidative stress conditions. Our data revealed that the presence of the RdsA-catalyzed modification at U2449 is beneficial for *E. coli* under oxidative stress, as evidenced by improved growth rates compared to the mutant strain lacking this modification. This highlights the role of the U2449 modification in conferring stress resistance to the bacterial cells and is consistent with those of (Fasnacht et al., 2022). The results from the proteomic analysis we obtained shows that several genes of unknown functions are differentially expressed in mutant cells compared to wild type cells. Interestingly, this analysis also shows that two detoxifying gene, *sodA* and *sodC*, are overexpressed in Δ*yhiN* cells. These observations of thermosensitivity and poor growth under oxidative stress point out to the fact that Δ*yhiN* mutant cells tend to maintain their fitness under regular growth condition despite a perturbed homeostasis. However, when a cumulative stress is applied, a threshold is passed and lead a kind of error catastrophe that is reflected by the observed phenotype under stress (Orgel, 1963).

In the case of D2449 absence, the affected cell process most probably translation. In this regard, we attempted to evaluate different parameters linked to translation. First, polysome profiling of the *ΔyhiN* strain was conducted to examine ribosomal assembly and translation efficiency. The profiles obtained for the mutant strain closely resembled those of the wild-type strain, with similar peaks for the free ribosomal 30S and 50S subunits, 70S ribosomes (a mixture of monosomes and free ribosomes), and light polysomes. Small differences observed in the heavy

polysome peaks were not reproducible and likely not due to translation efficiency differences. These results indicate that ribosomal particle assembly and overall translation efficiency are not significantly impaired by the deletion of the *yhiN* gene. This is not particularly surprising, it aligns with observations that many "nonessential" RNA modifications, often located within the RNA body, have minor phenotypic effects when removed. These modifications are part of an interconnected network, where the loss of a single modification can be mitigated through compensation or functional redundancy. Only the individual deletion of modification enzymes exhibits severely compromised growth and ribosome assembly such as RlmE, also RlmB, RlmKL, RlmN and RluC previously known to facilitate large ribosome subunit assembly (Liljeruhm et al., 2022). Recent study shows that RlmB and RlmKL have functions in ribosome assembly independent of their modification activities. While the assembly stage specificity of rRNA modification enzymes is well established, in this study the authors demonstrates that there is a mutual interdependence between the rRNA modification process and large ribosome subunit assembly (Ero et al., 2024). However, it is important to repeat these experiments under stress conditions to really conclude on that part.

To further investigate potential translation defects, we measured translation errors using the pBAD33LacLuc reporter system, which contains a lacZ-luc gene fusion. Translation error measurements were performed under normal conditions and oxidative stress (2mM H_2O_2). The results were consistent and reproducible, showing a significant increase in -1 frameshift levels for the Δ*yhiN* (28.58%) and Δ*rlmL* (25.20%) mutants under both conditions. These translational frameshifts cannot be explained by a poor codon-anticodon accommodation of cognate or near-cognate tRNA in the ribosomal A-site (Hill & Brierley, 2023). However, these high -1 frameshift levels observed in Δ*yhiN* mutants might be a consequence of subtle changes in PTC conformation that might affect kinetic parameters of the ribosome (Caliskan et al., 2014). Indeed, crystallographic studies revealed that the N3 of eukaryotic A2486 (A2451 in *E. coli*) is the nearest titratable group to the peptide bond being synthesized and likely functions as a general base to facilitate the nucleophilic attack by the α-amino group of the A-site substrate, the mechanism is well described (see introduction section). For A2451 to act as a general base, its pKa needs to be roughly 5 units higher than normal. Typically, the pKa of the N1 of adenosine monophosphate is about 3.5, with the N3 being 2 pH units lower. For A2451 to function effectively, the pKa of the N3 would need to be around 7 or higher. At a crystal pH of 5.8, this implies a pKa of >6 for the N3 of A2451, which is unusually high (Nissen et al., 2000). This unusual pKa is likely due to interactions between A2451 and G2447, which hydrogen

bonds with the buried phosphate of A2450. The modified D2449, described in our study, is located immediately before A2450 and then might slightly affect A2450 position and thus impact kinetic parameters of translation. Altogether, this significant increase in translation errors underscores the importance of the U2449 modification in maintaining translational fidelity and kinetic parameters, particularly under stress conditions. Structural determination of mutant ribosome by cryo-EM is currently under way.

In conclusion, this study provides valuable insights into the functional significance of the RdsA-catalyzed dihydrouridine modification at U2449 in the 23S rRNA of *E. coli*. While this modification is not essential for basic ribosomal function and translation efficiency, it plays a critical role in fine-tuning translation and providing a growth advantage under oxidative stress conditions. The absence of this modification leads to increased translational errors, highlighting its importance in maintaining the accuracy of protein synthesis. These findings contribute to a deeper understanding of the role of rRNA modifications in bacterial physiology and stress response. Future research should focus on elucidating the detailed molecular mechanisms by which these modifications influence ribosomal function and translational accuracy, as well as exploring their potential impact on antibiotic resistance and cellular adaptation to various environmental stresses.

4. Materials and Methods

Growth Curve and Competition Experiments

To determine the generation time of both the wild type and mutant strains at various temperatures, we monitored OD600 nm during cell growth in LB medium. For the competition experiments, overnight cultures from single colonies of the wild-type strain (BW25113) and the mutant strain were serially diluted in water for titration, allowing us to assess the initial ratio for the competition inoculation. The competition cultures were then inoculated with approximately 100 cells from each pre-culture and incubated with agitation at 42°C, 37°C, 30°C, and 23°C for 24 hours. To monitor the final ratio of mutant to wild type cells, we plated appropriate dilutions on LB-agar and LB-agar containing kanamycin (40 µg/mL). The plates were incubated overnight at 37°C, and the number of colony-forming units (CFUs) was determined. The competitive index was calculated by establishing the final ratio of wild type to mutant strain and normalizing this to the initial ratio for each competition experiment**.**

H_20_2 test treatment on *Escherichia coli*

To determine the effect of hydrogen peroxide (H_2O_2) on the growth of E. coli strains, an additional test was performed alongside the growth curve analysis. The strains were grown overnight at 37°C with shaking at 300 rpm in Luria-Bertani (LB) medium, either with or without kanamycin (40 µg/mL). Growth was assessed by measuring optical density at 600 nm (OD600 nm). The cells were then exposed to 2 mM H_2O_2 for periods ranging from 1 hour to 24 hours. Cell survival during hydrogen peroxide treatment was measured by plating aliquots of serial dilutions in water on LB agar plates, which were then incubated overnight at 37°C.

Proteomic analysis, protein extraction from bacterial cells

Five isolated colonies were picked and inoculated into 100 ml of LB medium. The culture was incubated at 37°C with agitation until the optical density at 600 nm (OD600) reached 0.4-0.5. The culture was then centrifuged at 5000 rpm for 10 minutes at 4°C. The supernatant was discarded, and the cell pellet was resuspended in 15 ml of freshly prepared lysis buffer composed of 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% NP-40, 0.1% SDS, protease inhibitor, and lysozyme. The resuspended cells were incubated on ice for 1 hour. The cell suspension was sonicated on ice, ensuring appropriate settings to avoid overheating and achieve efficient lysis.

The lysed cell suspension was centrifuged at 5000 rpm for 2 hours at 4°C, and the supernatant containing the soluble proteins was carefully collected. 20 µL of the supernatant was taken for SDS-PAGE analysis on a 15% acrylamide gel, and 10 µL of the supernatant (diluted 1/10) was used for the Bradford protein assay to determine protein concentration. All buffers and solutions were accurately prepared, samples were maintained on ice to prevent protein degradation.

Polysome profiling

Polysome fractionation analysis was performed as described in (Charollais et al., 2003). *E. coli* strains were grown in LB medium to an absorbance at 600 nm of between 0.5 and 0.8. To avoid polysome run off, chloramphenicol was added to a final concentration of 100 µg/ml 5 min before harvesting. Cells were rapidly cooled, collected by centrifugation and washed with 1/50th volume of buffer A (10 mM Tris-HCl, pH 7.5, 60 mM KCl, 10 mM MgCl2). Then cells were lysed by adding lysozyme (0.4 mg/ml) freshly prepared in buffer A. After a freeze–thaw cycle, 1/200th volume of a solution containing detergents and DNase in buffer A was added (0.5% polyoxymethylene 20 cetyl ether, 0.5% deoxycholic acid and 0.1 unit/µl of RNase-free DNase). The mixture was incubated on ice for 20 min, and the lysate was clarified by centrifugation in a microfuge at 18 000xg for 10 min at 4°C. The extract concentration was estimated by measuring the absorbance at 260 nm. To analyze polysome profiles, about seven A260 units of lysate were layered onto a 10–40% (w/v) linear sucrose gradient in buffer B [10 mM Tris-HCl, pH 7.5, 50 mM NH4Cl, 10 mM MgCl2, 1 mM dithiothreitol (DTT)] and centrifuged at 38,000 r.p.m. for 2.5h at 4°C in a Beckman SW41 rotor. Gradients were analyzed with an ISCO UV detector with continuous monitoring absorbance at 254 nm.

Plasmid construction to test translation errors

pBAD33LacLuc vector was used to test the propensity of in silico-selected sequence to induce translation errors. This plasmid contains the lacZ-luc gene fusion inserted into the unique MscI site located in between an arabinose inducible promoter and terminator.

To biologically test the propensity of in silico-selected sequences and natural sequence Dnax ribosomal frameshifting element to induce translation errors complementary synthetic oligonucleotides were annealed together and cloned into the unique MscI site of the plasmid located in between the ß-galactosidase and luciferase open reading frames using the SLIC technic (Jeong et al., 2012). For all the tested sequences tested (Table 7), a control sequence was also constructed by adding TGG between the unique MscI site.

All plasmid constructs were cloned in E. coli DH5α strain. Transformants were grown in LB supplemented with chloramphenicol (40µg/mL). All constructs were verified by sequencing.
The *E. coli* strain wt BW2513, *ΔyhiN::kan, ΔrlhA::kan, ΔrlmL:kan* were used in this study. At least four independent transformants were grown overnight in 3 mL of LB medium supplemented with 0,2% arabinose. Culture aliquots were then diluted 1:100 in 5 mL of the same medium and grown for 3-5h to and centrifuged. Crude cellular extracts were obtained by adding 1ml of the lysis mix (1V Lysozyme 5mg/ml: 2V CCLR luciferase assay buffer X2: BSA 5mg/ml: 0,1M K_2HPO_4 pH7.8: 20mM EDTA) and vortexing cell pellet

Translation errors measurement

Luciferase and ß-galactosidase activities were measured from the same crude extracts. For luciferase activity measurement, 20 µL of crude extracts were supplemented with 50-100 µL of luciferase assay reagent. Light emission detection was performed with a Berthold Lumat LB9501 luminometer. Luciferase activity was calculated with the formula Luc = (Relative Light Units/Volume of crude extract). ß-galactosidase was assayed by adding 50 µL of crude extracts into 400 µL of Z-buffer (Miller 1972). The reaction was started by the addition 100 µL of a 4 mg/mL ONPG solution and incubated at 28°C. When the reaction mixture turned to a yellow coloration, the reaction was stopped by the addition of 250 µL of 1M Na2CO3 and the OD420nm of the reaction mixture was measured. The ß-galactosidase was established with the following formula: ß-Gal = 1000X [OD420 / (reaction time X volume of crude extract X OD600nm of the cultre)]. The error rate was established with the following formula: [(Luc/ß-Gal) target sequence / (Luc/ß-Gal) control sequence].

General conclusions

During my PhD, I focused on elucidating the biochemical and biological roles of ribosomal dihydrouridylation in *E. coli.* My research encompassed several key objectives, systematically addressed in Chapters II and III of this thesis.

-Chapter II: Exploring a Novel Class of Flavoenzymes: Identification and Biochemical Characterization of Ribosomal RNA Dihydrouridine Synthase. In this chapter, I investigated a previously unidentified ribosomal RNA modification enzyme. The primary accomplishments include:

I developed a novel method, RhoRT-PCR, for detecting dihydrouridine (D) in rRNA. This technique involves selective blocking of reverse transcriptase at rhodamine-labeled D2449, followed by PCR amplification.

Through genetic studies using *E. coli* deletion strains and advanced detection methods like MALDI-Mass spectrometry and AlkAniline-Seq, I identified the *yhiN* gene, now designated as RdsA, responsible for ribosomal dihydrouridine synthase.

Phylogenetic analysis demonstrated the widespread distribution of the *yhiN* gene across the bacterial kingdom, underscoring the evolutionary significance of rRNA dihydrouridylation.

I characterized RdsA as a novel flavoenzyme that catalyzes the reduction of the C5-C6 double bond of uridine in the 23S rRNA. This enzyme uses FAD and NADH for its redox-dependent catalytic activity and was found to be essential for the incorporation of dihydrouridine at position 2449.

-Chapter III: The Biological Significance of Ribosomal Dihydrouridylation and Dihydrouridine Synthase. This chapter explores the functional implications of ribosomal dihydrouridylation:

Post-transcriptional RNA modifications are crucial for the regulation of RNA structure, function, and stability. My research highlighted that ribosomal dihydrouridylation plays a fine-tuning role in translational activity and ribosome biogenesis.

Using polysome profiling and ribosome Cryo-EM, I investigated how dihydrouridine modifications influence ribosomal function and cellular processes. The presence of dihydrouridine near the peptidyl transferase center suggests a role in the structural dynamics and stability of rRNA, which may affect protein synthesis efficiency and fidelity.

Overall, my colleagues' contribution and mine brought insight into the understanding of ribosomal RNA modifications in *Escherichia coli.* The identification and characterization of the novel flavoenzyme RdsA and its role in dihydrouridylation have revealed critical insights into the molecular mechanisms underlying ribosome function and the evolutionary importance of RNA modifications. Remarkable progress has been made in the field, as many studies mentioned above contributed to the understanding of the function of this modification. However, more questions piled up, as to the means of transcriptional alterations we observed. The puzzling functional involvement of D2449 on translation process and its contribution in the response to different environmental stresses, especially oxidative stress, and finally elucidating the chemical mechanism that RdsA employs to reduce the U2449 residue, to understand how RdsA binds its substrate, and investigating the specificity of its enzymatic activity are still under investigation. Many leads remain to be pursued based on previous findings that I will gladly leave for the next doctoral students in my team to explore.

References

Abedeera, S. M., Jayalath, K. S., Xie, J., Rauff, R. M., & Abeysirigunawardena, S. C. (2023). Pseudouridine Synthase RsuA Confers a Survival Advantage to Bacteria under Streptomycin Stress. *Antibiotics*. https://doi.org/10.3390/antibiotics12091447

Agarwalla, S., Kealey, J. T., Santi, D. V., & Stroud, R. M. (2002). Characterization of the 23 S Ribosomal RNA m5U1939 Methyltransferase from Escherichia coli. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111825200

Aleksandrova, E. V., Wu, K. J. Y., Tresco, B. I. C., Syroegin, E. A., Killeavy, E. E., Balasanyants, S. M., Svetlov, M. S., Gregory, S. T., Atkinson, G. C., Myers, A. G., & Polikanov, Y. S. (2024). Structural basis of Cfr-mediated antimicrobial resistance and mechanisms to evade it. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-023-01525-w

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S. L., Hughes, T. R., Grayhack, E. J., & Phizicky, E. M. (2006). Rapid tRNA Decay Can Result from Lack of Nonessential Modifications. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2005.10.036

Alian, A., DeGiovanni, A., Griner, S. L., Finer-Moore, J. S., & Stroud, R. M. (2009). Crystal Structure of an RluF–RNA Complex : A Base-Pair Rearrangement Is the Key to Selectivity of RluF for U2604 of the Ribosome. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2009.03.029

Amitai, S., Kolodkin-Gal, I., Hananya-Meltabashi, M., Sacher, A., & Engelberg-Kulka, H. (2009). Escherichia coli MazF Leads to the Simultaneous Selective Synthesis of Both "Death Proteins" and "Survival Proteins". *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1000390

Anantharaman, V., Koonin, E. V., & Aravind, L. (2002). SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *Journal of Molecular Microbiology and Biotechnology*.

Andersen, N. M., & Douthwaite, S. (2006). YebU is a m5C Methyltransferase Specific for 16 S rRNA Nucleotide 1407. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2006.04.007

Andries, O., Mc Cafferty, S., De Smedt, S. C., Weiss, R., Sanders, N. N., & Kitada, T. (2015). N1-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *Journal of Controlled Release*. https://doi.org/10.1016/j.jconrel.2015.08.051

Arai, T., Ishiguro, K., Kimura, S., Sakaguchi, Y., Suzuki, T., & Suzuki, T. (2015). Single methylation of 23S rRNA triggers late steps of 50S ribosomal subunit assembly. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1506749112

Arzumanian, V. A., Dolgalev, G. V., Kurbatov, I. Y., Kiseleva, O. I., & Poverennaya, E. V. (2022). Epitranscriptome : Review of Top 25 Most-Studied RNA Modifications. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms232213851

Åström, S. U., & Byström, A. S. (1994). Rit1, a tRNA backbone-modifying enzyme that mediates initiator and elongator tRNA discrimination. *Cell*. https://doi.org/10.1016/0092-8674(94)90262-3

Bacher, A., Eberhardt, S., Fischer, M., Kis, K., & Richter, G. (2000). Biosynthesis of Vitamin B 2 (Riboflavin). *Annual Review of Nutrition*. https://doi.org/10.1146/annurev.nutr.20.1.153

Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*. https://doi.org/10.1126/science.289.5481.905

Basturea, G. N., Dague, D. R., Deutscher, M. P., & Rudd, K. E. (2012). YhiQ Is RsmJ, the Methyltransferase Responsible for Methylation of G1516 in 16S rRNA of E. coli. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2011.10.044

Basturea, G. N., Rudd, K. E., & Deutscher, M. P. (2006). Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *RNA*. https://doi.org/10.1261/rna.2283106

Baudin-Baillieu, A., Fabret, C., Liang, X., Piekna-Przybylska, D., Fournier, M. J., & Rousset, J.-P. (2009). Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp816

Baxter-Roshek, J. L., Petrov, A. N., & Dinman, J. D. (2007). Optimization of Ribosome Structure and Function by rRNA Base Modification. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0000174

Benítez-Páez, A., Villarroya, M., & Armengod, M.-E. (2012a). Regulation of expression and catalytic activity of *Escherichia coli* RsmG methyltransferase. *RNA*. https://doi.org/10.1261/rna.029868.111

Benítez-Páez, A., Villarroya, M., & Armengod, M.-E. (2012b). The *Escherichia coli* RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy. *RNA*. https://doi.org/10.1261/rna.033266.112

Betteridge, T., Liu, H., Gamper, H., Kirillov, S., Cooperman, B. S., & Hou, Y.-M. (2007). Fluorescent labeling of tRNAs for dynamics experiments. *RNA*. https://doi.org/10.1261/rna.475407

Bishop, A. C., Xu, J., Johnson, R. C., Schimmel, P., & De Crécy-Lagard, V. (2002). Identification of the tRNA-Dihydrouridine Synthase Family. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M203208200

Blaby, I. K., Majumder, M., Chatterjee, K., Jana, S., Grosjean, H., De Crécy-Lagard, V., & Gupta, R. (2011). Pseudouridine formation in archaeal RNAs : The case of *Haloferax volcanii*. *RNA*. https://doi.org/10.1261/rna.2712811

Boal, A. K., Grove, T. L., McLaughlin, M. I., Yennawar, N. H., Booker, S. J., & Rosenzweig, A. C. (2011). Structural Basis for Methyl Transfer by a Radical SAM Enzyme. *Science*. https://doi.org/10.1126/science.1205358

Boccaletto, P., Stefaniak, F., Ray, A., Cappannini, A., Mukherjee, S., Purta, E., Kurkowska, M., Shirvanizadeh, N., Destefanis, E., Groza, P., Avşar, G., Romitelli, A., Pir, P., Dassi, E., Conticello, S. G., Aguilo, F., & Bujnicki, J. M. (2022). MODOMICS: a database of

RNA modification pathways. 2021 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab1083

Bou-Nader, C., Barraud, P., Pecqueur, L., Pérez, J., Velours, C., Shepard, W., Fontecave, M., Tisné, C., & Hamdane, D. (2019). Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1302

Bou-Nader, C., Montémont, H., Guérineau, V., Jean-Jean, O., Brégeon, D., & Hamdane, D. (2018). Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases : Perspectives on the evolution scenario. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1294

Bou-Nader, C., Pecqueur, L., Bregeon, D., Kamah, A., Guérineau, V., Golinelli-Pimpaneau, B., Guimarães, B. G., Fontecave, M., & Hamdane, D. (2015). An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv989

Brégeon, D., Pecqueur, L., Toubdji, S., Sudol, C., Lombard, M., Fontecave, M., De Crécy-Lagard, V., Motorin, Y., Helm, M., & Hamdane, D. (2022). Dihydrouridine in the Transcriptome : New Life for This Ancient RNA Chemical Modification. *ACS Chemical Biology*. https://doi.org/10.1021/acschembio.2c00307

Brunelle, J. L., Youngman, E. M., Sharma, D., & Green, R. (2006). The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity. *RNA*. https://doi.org/10.1261/rna.2256706

Bujnicki, J. M. (2000). Phylogenomic analysis of 16S rRNA:(guanine-N2) methyltransferases suggests new family members and reveals highly conserved motifs and a domain structure similar to other nucleic acid amino-methyltransferases. *The FASEB Journal*. https://doi.org/10.1096/fj.00-0076com

Byrne, R. T., Jenkins, H. T., Peters, D. T., Whelan, F., Stowell, J., Aziz, N., Kasatsky, P., Rodnina, M. V., Koonin, E. V., Konevega, A. L., & Antson, A. A. (2015). Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1500161112

Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F., & Rodnina, M. V. (2014). Programmed –1 Frameshifting by Kinetic Partitioning during Impeded Translocation. *Cell*. https://doi.org/10.1016/j.cell.2014.04.041

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A. P., Fabris, D., & Agris, P. F. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkq1028

Carell, T., Brandmayr, C., Hienzsch, A., Müller, M., Pearson, D., Reiter, V., Thoma, I., Thumbs, P., & Wagner, M. (2012). Structure and Function of Noncanonical Nucleobases. *Angewandte Chemie International Edition*. https://doi.org/10.1002/anie.201201193

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., & Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*. https://doi.org/10.1038/nature13802

Carr, D. O., & Grisolia, S. (1964). INCORPORATION OF DIHYDROURIDINE MONOPHOSPHATE AND URIDINE MONOPHOSPHATE INTO LIVER AND BRAIN RIBONUCLEIC ACID. *The Journal of Biological Chemistry*.

Cerutti, P., & Miller, N. (1967). Selective reduction of yeast transfer ribonucleic acid with sodium borohydride. *Journal of Molecular Biology*. https://doi.org/10.1016/0022-2836(67)90260-4

Chan, C. T. Y., Dyavaiah, M., DeMott, M. S., Taghizadeh, K., Dedon, P. C., & Begley, T. J. (2010). A Quantitative Systems Approach Reveals Dynamic Control of tRNA Modifications during Cellular Stress. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1001247

Charette, M., & Gray, M. W. (2000). Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life*. https://doi.org/10.1080/152165400410182

Charollais, J. (2004). CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkh603

Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M., & Iost, I. (2003). The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in *Escherichia coli*. *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.2003.03513.x

Chavatte, L., Frolova, L., Laugâa, P., Kisselev, L., & Favre, A. (2003). Stop Codons and UGG Promote Efficient Binding of the Polypeptide Release Factor eRF1 to the Ribosomal A Site. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(03)00813-1

Chen, X., Li, N., & Ellington, A. D. (2007). Ribozyme Catalysis of Metabolism in the RNA World. *Chemistry & Biodiversity*. https://doi.org/10.1002/cbdv.200790055

Chen, X., Sim, S., Wurtmann, E. J., Feke, A., & Wolin, S. L. (2014). Bacterial noncoding Y RNAs are widespread and mimic tRNAs. *RNA*. https://doi.org/10.1261/rna.047241.114

Chen, Y.-C., Brooks, A. F., Goodenough-Lashua, D. M., Kittendorf, J. D., Showalter, H. D., & Garcia, G. A. (2011). Evolution of eukaryal tRNA-guanine transglycosylase : Insight gained from the heterocyclic substrate recognition by the wild-type and mutant human and Escherichia coli tRNA-guanine transglycosylases. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkq1188

Chimnaronk, S., Forouhar, F., Sakai, J., Yao, M., Tron, C. M., Atta, M., Fontecave, M., Hunt, J. F., & Tanaka, I. (2009). Snapshots of Dynamics in Synthesizing N^6 -Isopentenyladenosine at the tRNA Anticodon. *Biochemistry*. https://doi.org/10.1021/bi900337d

Chow, C. S., Lamichhane, T. N., & Mahto, S. K. (2007). Expanding the Nucleotide Repertoire of the Ribosome with Post-Transcriptional Modifications. *ACS Chemical Biology*. https://doi.org/10.1021/cb7001494

Cohn, W. E., & Volkin, E. (1957). The Nucleic Acids. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.bi.26.070157.002423

Connolly, K., Rife, J. P., & Culver, G. (2008). Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Molecular Microbiology*, *5*. https://doi.org/10.1111/j.1365-2958.2008.06485.x

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*. https://doi.org/10.1016/0022-2836(68)90392-6

Czerwoniec, A., & Bujnicki, J. M. (2011). Identification and modeling of a phosphatase-like domain in a tRNA 2′- *O* -ribosyl phosphate transferase Rit1p. *Cell Cycle*, *20*. https://doi.org/10.4161/cc.10.20.17857

Czudnochowski, N., Ashley, G. W., Santi, D. V., Alian, A., Finer-Moore, J., & Stroud, R. M. (2014). The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt1050

Dai, W., Li, A., Yu, N. J., Nguyen, T., Leach, R. W., Wühr, M., & Kleiner, R. E. (2021). Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-021-00874-8

Dalluge, J. (1996a). Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic Acids Research*. https://doi.org/10.1093/nar/24.6.1073

Dalluge, J. (1996b). Quantitative measurement of dihydrouridine in RNA using isotope dilution liquid chromatography-mass spectrometry (LC/MS). *Nucleic Acids Research*, *16*. https://doi.org/10.1093/nar/24.16.3242

Dalluge, J. J., Hamamoto, T., Horikoshi, K., Morita, R. Y., Stetter, K. O., & McCloskey, J. A. (1997). Posttranscriptional modification of tRNA in psychrophilic bacteria. *Journal of Bacteriology*. https://doi.org/10.1128/jb.179.6.1918-1923.1997

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Research*. https://doi.org/10.1093/nar/23.24.5020

Davis DR. (1998). *Biophysical and conformational properties of modified nucleotides in RNA. modification and editing of RNA: Vol. p. 85–102* (American Society for Microbiology Press).

Davis, F. F., & Allen, F. W. (1957). RIBONUCLEIC ACIDS FROM YEAST WHICH CONTAIN A FIFTH NUCLEOTIDE. *Journal of Biological Chemistry*. https://doi.org/10.1016/S0021-9258(18)70770-9

De Crécy-Lagard, V., Brochier-Armanet, C., Urbonavičius, J., Fernandez, B., Phillips, G., Lyons, B., Noma, A., Alvarez, S., Droogmans, L., Armengaud, J., & Grosjean, H. (2010). Biosynthesis of Wyosine Derivatives in tRNA: An Ancient and Highly Diverse Pathway in Archaea. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msq096

Decatur, W. A., & Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends in Biochemical Sciences*. https://doi.org/10.1016/S0968-0004(02)02109-6

Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S. C., Guymon, R., Crain, P. F., Mccloskey, J. A., & Ofengand, J. (2005). Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA*. https://doi.org/10.1261/rna.7209905

Deng, H.-Y., Odom, O. W., & Hardesty, B. (1986). Localization of L11 on the Escherichia coli ribosome by singlet-singlet energy transfer. *European Journal of Biochemistry*, 497-503. https://doi.org/10.1111/j.1432-1033.1986.tb09608.x

Denman, R., Weitzmann, C., Cunningham, P. R., Negre, D., Nurse, K., Colgan, J., Pan, Y. C., Miedel, M., & Ofengand, J. (1989). In vitro assembly of 30S and 70S bacterial

ribosomes from 16S RNA containing single base substitutions, insertions, and deletions around the decoding site (C1400). *Biochemistry*. https://doi.org/10.1021/bi00429a013

Desai, P. M., & Rife, J. P. (2006). The adenosine dimethyltransferase KsgA recognizes a specific conformational state of the 30S ribosomal subunit. *Archives of Biochemistry and Biophysics*. https://doi.org/10.1016/j.abb.2006.02.028

Dixit, S., & Jaffrey, S. R. (2022). Expanding the epitranscriptome : Dihydrouridine in mRNA. *PLOS Biology*. https://doi.org/10.1371/journal.pbio.3001720

Dobritzsch, D. (2001). Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *The EMBO Journal*. https://doi.org/10.1093/emboj/20.4.650

Dorner, S. (2003). Mononucleotide derivatives as ribosomal P-site substrates reveal an important contribution of the 2'-OH to activity. *Nucleic Acids Research*, *22*. https://doi.org/10.1093/nar/gkg842

Draycott, A. S., Schaening-Burgos, C., Rojas-Duran, M. F., Wilson, L., Schärfen, L., Neugebauer, K. M., Nachtergaele, S., & Gilbert, W. V. (2022). *Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA*. https://doi.org/10.1371/journal.pbio.3001622

Druzina, Z., & Cooperman, B. S. (2004). Photolabile anticodon stem–loop analogs of tRNA ^{Phe} as probes of ribosomal structure and structural fluctuation at the decoding center. *RNA*. https://doi.org/10.1261/rna.7930804

Dudzińska-Bajorek, B., Bąkowska, K., & Twardowski, T. (2006). Conformational changes of L-rRNA during elongation of polypeptide. *Journal of Plant Physiology*. https://doi.org/10.1016/j.jplph.2005.03.016

Dyubankova, N., Sochacka, E., Kraszewska, K., Nawrot, B., Herdewijn, P., & Lescrinier, E. (2015). Contribution of dihydrouridine in folding of the D-arm in tRNA. *Organic & Biomolecular Chemistry*. https://doi.org/10.1039/C5OB00164A

El Yacoubi, B., Bailly, M., & De Crécy-Lagard, V. (2012). Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-110711-155641

Emmerechts, G., Barbe, S., Herdewijn, P., Anne, J., & Rozenski, J. (s. d.). Post-transcriptional modification mapping in the Clostridium acetobutylicum 16S rRNA by mass spectrometry and reverse transcriptase assays. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm248

Ero, R., Leppik, M., Liiv, A., & Remme, J. (2010). Specificity and kinetics of 23S rRNA modification enzymes RlmH and RluD. *RNA*. https://doi.org/10.1261/rna.2234310

Ero, R., Leppik, M., Reier, K., Liiv, A., & Remme, J. (2024). Ribosomal RNA modification enzymes stimulate large ribosome subunit assembly in *E. coli*. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkae222

Faivre, B., Lombard, M., Fakroun, S., Vo, C.-D.-T., Goyenvalle, C., Guérineau, V., Pecqueur, L., Fontecave, M., De Crécy-Lagard, V., Brégeon, D., & Hamdane, D. (2021). Dihydrouridine synthesis in tRNAs is under reductive evolution in Mollicutes. *RNA Biology*. https://doi.org/10.1080/15476286.2021.1899653

Fasnacht, M., Gallo, S., Sharma, P., Himmelstoß, M., Limbach, P. A., Willi, J., & Polacek, N. (2022). Dynamic 23S rRNA modification ho5C2501 benefits *Escherichia coli* under oxidative stress. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab1224

Feder, M., Pas, J., Wyrwicz, L. S., & Bujnicki, J. M. (2003). Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2′-O-methyltransferases. *Gene*. https://doi.org/10.1016/S0378-1119(02)01097-1

Finet, O., Yague-Sanz, C., & Hermand, D. (2022). Epitranscriptomic mapping of RNA modifications at single-nucleotide resolution using rhodamine sequencing (Rho-seq). *STAR Protocols*. https://doi.org/10.1016/j.xpro.2022.101369

Finet, O., Yague-Sanz, C., Krüger, L. K., Tran, P., Migeot, V., Louski, M., Nevers, A., Rougemaille, M., Sun, J., Ernst, F. G. M., Wacheul, L., Wery, M., Morillon, A., Dedon, P., Lafontaine, D. L. J., & Hermand, D. (2022). Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2021.11.003

Fischer, M., & Bacher, A. (2011). Biosynthesis of Vitamin B $_2$: A Unique Way to Assemble a Xylene Ring. *ChemBioChem*. https://doi.org/10.1002/cbic.201000681

Freistroffer, D. V., Pavlov, M. Yu., MacDougall, J., Buckingham, R. H., & Ehrenberg, M. (1997). Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *The EMBO Journal*, 4126-4133. https://doi.org/10.1093/emboj/16.13.4126

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*. https://doi.org/10.1038/nrg.2016.47

Garbom, S., Forsberg, Å., Wolf-Watz, H., & Kihlberg, B.-M. (2004). Identification of Novel Virulence-Associated Genes via Genome Analysis of Hypothetical Genes. *Infection and Immunity*. https://doi.org/10.1128/IAI.72.3.1333-1340.2004

Ge, J., & Yu, Y.-T. (2013). RNA pseudouridylation : New insights into an old modification. *Trends in Biochemical Sciences*, 210-218. https://doi.org/10.1016/j.tibs.2013.01.002

Gerber, A. P., & Keller, W. (2001). RNA editing by base deamination : More enzymes, more targets, new mysteries. *Trends in Biochemical Sciences*. https://doi.org/10.1016/S0968-0004(01)01827-8

Golovina, A. Y., Dzama, M. M., Osterman, I. A., Sergiev, P. V., Serebryakova, M. V., Bogdanov, A. A., & Dontsova, O. A. (2012). The last rRNA methyltransferase of *E. coli* revealed : The *yhiR* gene encodes adenine-N6 methyltransferase specific for modification of A2030 of 23S ribosomal RNA. *RNA*. https://doi.org/10.1261/rna.034207.112

Grosjean, H., Auxilien, S., Constantinesco, F., Simon, C., Corda, Y., Becker, H. F., Foiret, D., Morin, A., Jin, Y. X., Fournier, M., & Fourrey, J. L. (1996). Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs : A review. *Biochimie*. https://doi.org/10.1016/0300-9084(96)84755-9

Grosjean, H, & Benne, R. (1998). *Modification and editing of RNA* (ASM Press).

Grove, T. L., Benner, J. S., Radle, M. I., Ahlum, J. H., Landgraf, B. J., Krebs, C., & Booker, S. J. (2011). A Radically Different Mechanism for *S* -Adenosylmethionine–Dependent Methyltransferases. *Science*. https://doi.org/10.1126/science.1200877

Grove, T. L., Livada, J., Schwalm, E. L., Green, M. T., Booker, S. J., & Silakov, A. (2013). A substrate radical intermediate in catalysis by the antibiotic resistance protein Cfr. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1251

Gu, C., Begley, T. J., & Dedon, P. C. (2014). tRNA modifications regulate translation during cellular stress. *FEBS Letters*, 4287-4296. https://doi.org/10.1016/j.febslet.2014.09.038

Gualerzi, C. O., Brandi, L., Caserta, E., Garofalo, C., Lammi, M., La Teana, A., Petrelli, D., Spurio, R., Tomsic, J., & Pon, C. L. (2001). Initiation Factors in the Early Events of mRNA Translation in Bacteria. *Cold Spring Harbor Symposia on Quantitative Biology*. https://doi.org/10.1101/sqb.2001.66.363

Gustafsson, C. (1996). Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic Acids Research*. https://doi.org/10.1093/nar/24.19.3756

Gustafsson, C., & Persson, B. C. (1998). Identification of the *rrmA* Gene Encoding the 23S rRNA m ^{1} G745 Methyltransferase in *Escherichia coli* and Characterization of an m ^{1} G745-Deficient Mutant. *Journal of Bacteriology*. https://doi.org/10.1128/JB.180.2.359-365.1998

Gutgsell, N. S., Deutscher, M. P., & Ofengand, J. (2005). The pseudouridine synthase RluD is required for normal ribosome assembly and function in *Escherichia coli*. *RNA*. https://doi.org/10.1261/rna.2550105

Guy, M. P., & Phizicky, E. M. (2014). Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. *RNA Biology*. https://doi.org/10.1080/15476286.2015.1008360

Hager, J., Staker, B. L., Bügl, H., & Jakob, U. (2002). Active Site in RrmJ, a Heat Shock-induced Methyltransferase. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M205423200

Halliday, N. M., Hardie, K. R., Williams, P., Winzer, K., & Barrett, D. A. (2010). Quantitative liquid chromatography–tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in Escherichia coli. *Analytical Biochemistry*. https://doi.org/10.1016/j.ab.2010.04.021

Hamdane, D., Grosjean, H., & Fontecave, M. (2016). Flavin-Dependent Methylation of RNAs : Complex Chemistry for a Simple Modification. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2016.10.031

Hamma, T., & Ferré-D'Amaré, A. R. (2006). Pseudouridine Synthases. *Chemistry & Biology*. https://doi.org/10.1016/j.chembiol.2006.09.009

Hansen, J. L., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2002). Structural insights into peptide bond formation. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.172404099

Hanukoglu, I. (2015). Proteopedia : Rossmann fold : A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education*. https://doi.org/10.1002/bmb.20849

Helm, M., & Alfonzo, J. D. (2014). Posttranscriptional RNA Modifications : Playing Metabolic Games in a Cell's Chemical Legoland. *Chemistry & Biology*. https://doi.org/10.1016/j.chembiol.2013.10.015

Helm, M., Giegé, R., & Florentz, C. (1999). A Watson−Crick Base-Pair-Disrupting Methyl Group (m^1 A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial $tRNA^{Lys}$. *Biochemistry*. https://doi.org/10.1021/bi991061g

Hill, C. H., & Brierley, I. (2023). Structural and Functional Insights into Viral Programmed Ribosomal Frameshifting. *Annual Review of Virology*. https://doi.org/10.1146/annurev-virology-111821-120646

Hoang, C., Chen, J., Vizthum, C. A., Kandel, J. M., Hamilton, C. S., Mueller, E. G., & Ferré-D'Amaré, A. R. (2006). Crystal Structure of Pseudouridine Synthase RluA: Indirect Sequence Readout through Protein-Induced RNA Structure. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2006.09.017

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., & Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*. https://doi.org/10.1126/science.147.3664.1462

House, C. H., & Miller, S. L. (1996). Hydrolysis of Dihydrouridine and Related Compounds. *Biochemistry*, 315-320. https://doi.org/10.1021/bi951577+

Ishiguro, K., Arai, T., & Suzuki, T. (2019). Depletion of S-adenosylmethionine impacts on ribosome biogenesis through hypomodification of a single rRNA methylation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz111

Ito, S., Akamatsu, Y., Noma, A., Kimura, S., Miyauchi, K., Ikeuchi, Y., Suzuki, T., & Suzuki, T. (2014). A Single Acetylation of 18 S rRNA Is Essential for Biogenesis of the Small Ribosomal Subunit in Saccharomyces cerevisiae. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M114.593996

Ito, S., Horikawa, S., Suzuki, T., Kawauchi, H., Tanaka, Y., Suzuki, T., & Suzuki, T. (2014). Human NAT10 Is an ATP-dependent RNA Acetyltransferase Responsible for N4-Acetylcytidine Formation in 18 S Ribosomal RNA (rRNA). *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.C114.602698

Jack, K., Bellodi, C., Landry, D. M., Niederer, R. O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A. M., Thompson, S. R., Ruggero, D., & Dinman, J. D. (2011). rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2011.09.017

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(61)80072-7

Jayalath, K., Frisbie, S., To, M., & Abeysirigunawardena, S. (2020). Pseudouridine Synthase RsuA Captures an Assembly Intermediate That Is Stabilized by Ribosomal Protein S17. *Biomolecules*. https://doi.org/10.3390/biom10060841

Jeong, J.-Y., Yim, H.-S., Ryu, J.-Y., Lee, H. S., Lee, J.-H., Seen, D.-S., & Kang, S. G. (2012). One-Step Sequence- and Ligation-Independent Cloning as a Rapid and Versatile Cloning Method for Functional Genomics Studies. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.00844-12

Jia, G., Fu, Y., & He, C. (2013). Reversible RNA adenosine methylation in biological regulation. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.11.003

Jiang, J., Aduri, R., Chow, C. S., & SantaLucia, J. (2014). Structure modulation of helix 69 from *Escherichia coli* 23S ribosomal RNA by pseudouridylations. *Nucleic Acids Research*, *6*. https://doi.org/10.1093/nar/gkt1329

Joosten, V., & Van Berkel, W. J. (2007). Flavoenzymes. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2007.01.010

Joyce, G. F., & Szostak, J. W. (2018). Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a034801

Jung, Y., & Goldman, D. (2018). Role of RNA modifications in brain and behavior. *Genes, Brain and Behavior*. https://doi.org/10.1111/gbb.12444

Karikó, K., Buckstein, M., Ni, H., & Weissman, D. (2005). Suppression of RNA Recognition by Toll-like Receptors : The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity*. https://doi.org/10.1016/j.immuni.2005.06.008

Karimi, R., Pavlov, M. Y., Buckingham, R. H., & Ehrenberg, M. (1999). Novel Roles for Classical Factors at the Interface between Translation Termination and Initiation. *Molecular Cell*, 601-609. https://doi.org/10.1016/S1097-2765(00)80353-6

Kasai, H., Nakanishi, K., Macfarlane, R. D., Torgerson, D. F., Ohashi, Z., McCloskey, J. A., Gross, H. J., & Nishimura, S. (1976). The structure of Q* nucleoside isolated from rabbit liver transfer ribonucleic acid. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja00432a071

Kasprzak, J. M., Czerwoniec, A., & Bujnicki, J. M. (2012). Molecular evolution of dihydrouridine synthases. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-13-153

Kaur, J., Raj, M., & Cooperman, B. S. (2011). Fluorescent labeling of tRNA dihydrouridine residues : Mechanism and distribution. *RNA*. https://doi.org/10.1261/rna.2670811

Keffer-Wilkes, L. C., Soon, E. F., & Kothe, U. (2020). The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa548

Keffer-Wilkes, L. C., Veerareddygari, G. R., & Kothe, U. (2016). RNA modification enzyme TruB is a tRNA chaperone. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1607512113

Kiel, M. C., Raj, V. S., Kaji, H., & Kaji, A. (2003). Release of Ribosome-bound Ribosome Recycling Factor by Elongation Factor G. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M304834200

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974). Three-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA. *Science*. https://doi.org/10.1126/science.185.4149.435

Kimura, S., Ikeuchi, Y., Kitahara, K., Sakaguchi, Y., Suzuki, T., & Suzuki, T. (2012). Base methylations in the double-stranded RNA by a fused methyltransferase bearing unwinding activity. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkr1287

Kimura, S., Sakai, Y., Ishiguro, K., & Suzuki, T. (2017). Biogenesis and iron-dependency of ribosomal RNA hydroxylation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx969

Kimura, S., & Suzuki, T. (2010). Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp1073

Kirpekar, F., Hansen, L. H., Mundus, J., Tryggedsson, S., Teixeira Dos Santos, P., Ntokou, E., & Vester, B. (2018). Mapping of ribosomal 23S ribosomal RNA modifications in *Clostridium sporogenes*. *RNA Biology*. https://doi.org/10.1080/15476286.2018.1486662

Klaholz, B. P., Myasnikov, A. G., & Van Heel, M. (2004). Visualization of release factor 3 on the ribosome during termination of protein synthesis. *Nature*. https://doi.org/10.1038/nature02332

Kowalak, J. A., Bruenger, E., & McCloskey, J. A. (1995). Posttranscriptional Modification of the Central Loop of Domain V in Escherichia coli 23 S Ribosomal RNA. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.270.30.17758

Kressler, D. (1999). Spb1p is a putative methyltransferase required for 60S ribosomal subunit biogenesis in Saccharomyces cerevisiae. *Nucleic Acids Research*. https://doi.org/10.1093/nar/27.23.4598

Kusuba, H., Yoshida, T., Iwasaki, E., Awai, T., Kazayama, A., Hirata, A., Tomikawa, C., Yamagami, R., & Hori, H. (2015). *In vitro* dihydrouridine formation by tRNA dihydrouridine synthase from *Thermus thermophilus* , an extreme-thermophilic eubacterium. *Journal of Biochemistry*. https://doi.org/10.1093/jb/mvv066

Kyuma, T., Kimura, S., Hanada, Y., Suzuki, T., Sekimizu, K., & Kaito, C. (2015). Ribosomal RNA methyltransferases contribute to *Staphylococcus aureus* virulence. *The FEBS Journal*. https://doi.org/10.1111/febs.13302

Lafontaine, D. L. J., & Tollervey, D. (2001). The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology*, 514-520. https://doi.org/10.1038/35080045

Lafontaine, D., Vandenhaute, J., & Tollervey, D. (1995). The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes & Development*. https://doi.org/10.1101/gad.9.20.2470

Lamond, A. I. (1990). The trimethylguanosine cap is a nuclear targeting signal for snRNPs. *Trends in Biochemical Sciences*. https://doi.org/10.1016/0968-0004(90)90292-J

Lapeyre, B., & Purushothaman, S. K. (2004). Spb1p-Directed Formation of Gm2922 in the Ribosome Catalytic Center Occurs at a Late Processing Stage. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2004.10.022

Lee, C.-H., & Tinoco, I. (1977). Studies of the conformation of modified dinucleoside phosphates containing 1,N ^{6} -ethenoadenosine and 2'-O-methylcytidine by 360-MHz proton nuclear magnetic resonance spectroscopy. Investigation of the solution conformation of dinucleoside phosphates. *Biochemistry*. https://doi.org/10.1021/bi00644a001

Lee, T. T., Agarwalla, S., & Stroud, R. M. (2004). Crystal Structure of RumA, an Iron-Sulfur Cluster Containing E. coli Ribosomal RNA 5-Methyluridine Methyltransferase. *Structure*. https://doi.org/10.1016/j.str.2004.02.009

Lee, T. T., Agarwalla, S., & Stroud, R. M. (2005). A Unique RNA Fold in the RumA-RNA-Cofactor Ternary Complex Contributes to Substrate Selectivity and Enzymatic Function. *Cell*. https://doi.org/10.1016/j.cell.2004.12.037

Leppik, M., Liiv, A., & Remme, J. (2017). Random pseuoduridylation in vivo reveals critical region of Escherichia coli 23S rRNA for ribosome assembly. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx160

Lesnyak, D. V., Sergiev, P. V., Bogdanov, A. A., & Dontsova, O. A. (2006). Identification of Escherichia coli m2G methyltransferases : I. The ycbY Gene Encodes a Methyltransferase Specific for G2445 of the 23 S rRNA. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2006.09.009

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., & Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1836

Liljeruhm, J., Leppik, M., Bao, L., Truu, T., Calvo-Noriega, M., Freyer, N. S., Liiv, A., Wang, J., Blanco, R. C., Ero, R., Remme, J., & Forster, A. C. (2022). Plasticity and conditional essentiality of modification enzymes for domain V of *Escherichia coli* 23S ribosomal RNA. *RNA*. https://doi.org/10.1261/rna.079096.121

Liu, M., & Douthwaite, S. (2002). Methylation at nucleotide G745 or G748 in 23S rRNA distinguishes Gram-negative from Gram-positive bacteria. *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.2002.02866.x

Liu, M., Novotny, G. W., & Douthwaite, S. (2004). Methylation of 23S rRNA nucleotide G745 is a secondary function of the RlmA ^{I} methyltransferase. *RNA*. https://doi.org/10.1261/rna.7820104

Lombard, M., & Hamdane, D. (2017). Flavin-dependent epitranscriptomic world. *Archives of Biochemistry and Biophysics*. https://doi.org/10.1016/j.abb.2017.06.011

Lombard, M., Reed, C. J., Pecqueur, L., Faivre, B., Toubdji, S., Sudol, C., Brégeon, D., De Crécy-Lagard, V., & Hamdane, D. (2022). Evolutionary Diversity of Dus2 Enzymes Reveals Novel Structural and Functional Features among Members of the RNA Dihydrouridine Synthases Family. *Biomolecules*. https://doi.org/10.3390/biom12121760

Lorenz, C., Lünse, C., & Mörl, M. (2017). tRNA Modifications : Impact on Structure and Thermal Adaptation. *Biomolecules*. https://doi.org/10.3390/biom7020035

Lövgren, J. M., & Wikström, P. M. (2001). The *rlmB* Gene Is Essential for Formation of Gm2251 in 23S rRNA but Not for Ribosome Maturation in *Escherichia coli*. *Journal of Bacteriology*. https://doi.org/10.1128/JB.183.23.6957-6960.2001

Macheroux, P., Kappes, B., & Ealick, S. E. (2011). Flavogenomics – a genomic and structural view of flavin-dependent proteins. *The FEBS Journal*. https://doi.org/10.1111/j.1742-4658.2011.08202.x

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K. M., Helm, M., Bujnicki, J. M., & Grosjean, H. (2012). MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks1007

Madsen, C. T. (2003). Identifying the methyltransferases for m5U747 and m5U1939 in 23S rRNA using MALDI mass spectrometry. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkg657

Marchand, V., Ayadi, L., Ernst, F. G. M., Hertler, J., Bourguignon-Igel, V., Galvanin, A., Kotter, A., Helm, M., Lafontaine, D. L. J., & Motorin, Y. (2018). AlkAniline-Seq :

Profiling of m ^{7} G and m ^{3} C RNA Modifications at Single Nucleotide Resolution. *Angewandte Chemie International Edition*. https://doi.org/10.1002/anie.201810946

Martínez Giménez, J. A., Sáez, G. T., & Seisdedos, R. T. (1998). On the Function of Modified Nucleosides in the RNA World. *Journal of Theoretical Biology*. https://doi.org/10.1006/jtbi.1998.0770

Massey, V. (2000). The chemical and biological versatility of riboflavin. *Biochemical Society Transactions*.

McCown, P. J., Ruszkowska, A., Kunkler, C. N., Breger, K., Hulewicz, J. P., Wang, M. C., Springer, N. A., & Brown, J. A. (2020). Naturally occurring modified ribonucleosides. *WIREs RNA*. https://doi.org/10.1002/wrna.1595

Metodiev, M. D., Lesko, N., Park, C. B., Cámara, Y., Shi, Y., Wibom, R., Hultenby, K., Gustafsson, C. M., & Larsson, N.-G. (2009). Methylation of 12S rRNA Is Necessary for In Vivo Stability of the Small Subunit of the Mammalian Mitochondrial Ribosome. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2009.03.001

Meyer, B., Wurm, J. P., Sharma, S., Immer, C., Pogoryelov, D., Kötter, P., Lafontaine, D. L. J., Wöhnert, J., & Entian, K.-D. (2016). Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw244

Michel, G., Sauvé, V., Larocque, R., Li, Y., Matte, A., & Cygler, M. (2002). The Structure of the RlmB 23S rRNA Methyltransferase Reveals a New Methyltransferase Fold with a Unique Knot. *Structure*. https://doi.org/10.1016/S0969-2126(02)00852-3

Moazed, D., & Noller, H. F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature*. https://doi.org/10.1038/342142a0

Motorin, Y., & Helm, M. (2011). RNA nucleotide methylation. *WIREs RNA*. https://doi.org/10.1002/wrna.79

Moukadiri, I., Prado, S., Piera, J., Velázquez-Campoy, A., Björk, G. R., & Armengod, M.-E. (2009). Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp762

Mueller EG, & Ferré-D'Amaré AR. (2000). *Pseudouridine Formation, the Most Common Transglycosylation in RNA*. Landes Bioscience. https://www.ncbi.nlm.nih.gov/books/NBK6205/

Müller, F. (1987). Flavin radicals : Chemistry and biochemistry. *Free Radical Biology and Medicine*. https://doi.org/10.1016/0891-5849(87)90009-8

Müller, U. F. (2006). Re-creating an RNA world. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-006-6047-1

Muth, G. W., Ortoleva-Donnelly, L., & Strobel, S. A. (2000). A Single Adenosine with a Neutral p K_a in the Ribosomal Peptidyl Transferase Center. *Science*. https://doi.org/10.1126/science.289.5481.947

Narayan, G., Gracia Mazuca, L. A., Cho, S. S., Mohl, J. E., & Koculi, E. (2023). RNA Post-transcriptional Modifications of an Early-Stage Large-Subunit Ribosomal Intermediate. *Biochemistry*. https://doi.org/10.1021/acs.biochem.3c00291

Nawrot, B., Sochacka, E., & Düchler, M. (2011). tRNA structural and functional changes induced by oxidative stress. *Cellular and Molecular Life Sciences*, 4023-4032. https://doi.org/10.1007/s00018-011-0773-8

Nelson, J. W., & Breaker, R. R. (2017). The lost language of the RNA World. *Science Signaling*. https://doi.org/10.1126/scisignal.aam8812

Nguyen, T.-Q., & Nicolet, Y. (2022). Structure and Catalytic Mechanism of Radical SAM Methylases. *Life*. https://doi.org/10.3390/life12111732

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*. https://doi.org/10.1126/science.289.5481.920

Noeske, J., Wasserman, M. R., Terry, D. S., Altman, R. B., Blanchard, S. C., & Cate, J. H. D. (2015). High-resolution structure of the Escherichia coli ribosome. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/nsmb.2994

Noller, H. F., Hoffarth, V., & Zimniak, L. (1992). Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science*. https://doi.org/10.1126/science.1604315

Noller, H. F, & M. Nomura. (1996). *Ribosomes. In F. C* (Neidhardt).

Noller, H. F., Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Lancaster, L., Dallas, A., Fredrick, K., Earnest, T. N., & Cate, J. H. D. (2001). Structure of the Ribosome at 5.5 A Resolution and Its Interactions with Functional Ligands. *Cold Spring Harbor Symposia on Quantitative Biology*. https://doi.org/10.1101/sqb.2001.66.57

Noma, A., Kirino, Y., Ikeuchi, Y., & Suzuki, T. (2006). Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA. *The EMBO Journal*. https://doi.org/10.1038/sj.emboj.7601105

Nomura, Y., Ohno, S., Nishikawa, K., & Yokogawa, T. (2016). Correlation between the stability of tRNA tertiary structure and the catalytic efficiency of a tRNA -modifying enzyme, archaeal tRNA -guanine transglycosylase. *Genes to Cells*. https://doi.org/10.1111/gtc.12317

Noon, K. R., Guymon, R., Crain, P. F., McCloskey, J. A., Thomm, M., Lim, J., & Cavicchioli, R. (2003). Influence of Temperature on tRNA Modification in Archaea : *Methanococcoides burtonii* (Optimum Growth Temperature [T_{opt}], 23°C) and *Stetteria hydrogenophila* (T_{opt} , 95°C). *Journal of Bacteriology*. https://doi.org/10.1128/JB.185.18.5483-5490.2003

O'Connor, M. (2001). Mutagenesis of the peptidyltransferase center of 23S rRNA: the invariant U2449 is dispensable. *Nucleic Acids Research*. https://doi.org/10.1093/nar/29.3.710

O'Connor, M., & Gregory, S. T. (2011). Inactivation of the RluD Pseudouridine Synthase Has Minimal Effects on Growth and Ribosome Function in Wild-Type *Escherichia coli* and *Salmonella enterica*. *Journal of Bacteriology*. https://doi.org/10.1128/JB.00970-10

O'Connor, M., Leppik, M., & Remme, J. (2018). Pseudouridine-Free Escherichia coli Ribosomes. *Journal of Bacteriology*. https://doi.org/10.1128/JB.00540-17

Ofengand, J., & Del Campo, M. (2004). Modified Nucleosides of *Escherichia coli* Ribosomal RNA. *EcoSal Plus*. https://doi.org/10.1128/ecosalplus.4.6.1

Ohira, T., & Suzuki, T. (2024). Transfer RNA modifications and cellular thermotolerance. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2023.11.041

Oldenburg, M., Krüger, A., Ferstl, R., Kaufmann, A., Nees, G., Sigmund, A., Bathke, B., Lauterbach, H., Suter, M., Dreher, S., Koedel, U., Akira, S., Kawai, T., Buer, J., Wagner, H., Bauer, S., Hochrein, H., & Kirschning, C. J. (2012). TLR13 Recognizes Bacterial 23 *S* rRNA Devoid of Erythromycin Resistance–Forming Modification. *Science*. https://doi.org/10.1126/science.1220363

Orgel, L. E. (1963). THE MAINTENANCE OF THE ACCURACY OF PROTEIN SYNTHESIS AND ITS RELEVANCE TO AGEING. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.49.4.517

Osterman, I. A., Dontsova, O. A., & Sergiev, P. V. (2020). rRNA Methylation and Antibiotic Resistance. *Biochemistry (Moscow)*. https://doi.org/10.1134/S000629792011005X

Osterman, I. A., Sergiev, P. V., Tsvetkov, P. O., Makarov, A. A., Bogdanov, A. A., & Dontsova, O. A. (2011). Methylated 23S rRNA nucleotide m2G1835 of Escherichia coli ribosome facilitates subunit association. *Biochimie*. https://doi.org/10.1016/j.biochi.2010.12.016

Pape, T. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E.coli ribosome. *The EMBO Journal*. https://doi.org/10.1093/emboj/17.24.7490

Perche-Letuvée, P., Molle, T., Forouhar, F., Mulliez, E., & Atta, M. (2014). Wybutosine biosynthesis : Structural and mechanistic overview. *RNA Biology*. https://doi.org/10.4161/15476286.2014.992271

Persson, B. C., Esberg, B., Ólafsson, Ó., & Björk, G. R. (1994). Synthesis and function of isopentenyl adenosine derivatives in tRNA. *Biochimie*. https://doi.org/10.1016/0300-9084(94)90044-2

Piano, V., Palfey, B. A., & Mattevi, A. (2017). Flavins as Covalent Catalysts : New Mechanisms Emerge. *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2017.02.005

Pintard, L. (2002). MRM2 encodes a novel yeast mitochondrial 21S rRNA methyltransferase. *The EMBO Journal*. https://doi.org/10.1093/emboj/21.5.1139

Pletnev, P., Guseva, E., Zanina, A., Evfratov, S., Dzama, M., Treshin, V., Pogorel'skaya, A., Osterman, I., Golovina, A., Rubtsova, M., Serebryakova, M., Pobeguts, O. V., Govorun, V. M., Bogdanov, A. A., Dontsova, O. A., & Sergiev, P. V. (2020). Comprehensive Functional Analysis of Escherichia coli Ribosomal RNA Methyltransferases. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2020.00097

Polacek, N., Gaynor, M., Yassin, A., & Mankin, A. S. (2001). Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature*. https://doi.org/10.1038/35078113

Polikanov, Y. S., Melnikov, S. V., Söll, D., & Steitz, T. A. (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/nsmb.2992

Popova, A. M., & Williamson, J. R. (2014). Quantitative Analysis of rRNA Modifications Using Stable Isotope Labeling and Mass Spectrometry. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja412084b

Proux, F., Dreyfus, M., & Iost, I. (2011). Identification of the sites of action of SrmB, a DEAD-box RNA helicase involved in Escherichia coli ribosome assembly : rRNA mutations

that bypass SrmB requirement. *Molecular Microbiology*. https://doi.org/10.1111/j.1365-2958.2011.07779.x

Purta, E., Kaminska, K. H., Kasprzak, J. M., Bujnicki, J. M., & Douthwaite, S. (2008). YbeA is the m ^{3} Ψ methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA*. https://doi.org/10.1261/rna.1198108

Purta, E., O'Connor, M., Bujnicki, J. M., & Douthwaite, S. (2008). YccW is the m5C Methyltransferase Specific for 23S rRNA Nucleotide 1962. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2008.08.061

Reid, R., Greene, P. J., & Santi, D. V. (1999). Exposition of a family of RNA m5C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Research*. https://doi.org/10.1093/nar/27.15.3138

Robertson, M. P., & Joyce, G. F. (2012). The Origins of the RNA World. *Cold Spring Harbor Perspectives in Biology*, a003608-a003608. https://doi.org/10.1101/cshperspect.a003608

Rodnina, M. V., Daviter, T., Gromadski, K., & Wintermeyer, W. (2002). Structural dynamics of ribosomal RNA during decoding on the ribosome. *Biochimie*. https://doi.org/10.1016/S0300-9084(02)01409-8

Rodnina, M. V., Savelsbergh, A., Katunin, V. I., & Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature*. https://doi.org/10.1038/385037a0

Roy-Burman, P., Roy-Burman, S., & Visser, D. W. (1965). Incorporation of 5,6-dihydrouridine triphosphate into ribonucleic acid by DNA-dependent RNA polymerase. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/0006-291X(65)90362-1

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., & Schwartz, S. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*. https://doi.org/10.1038/nature24456

Schaening-Burgos, C., Li, G.-W., & Gilbert, W. (2023). *RluA is the major mRNA pseudouridine synthase in* Escherichia coli. https://doi.org/10.1101/2023.12.08.570740

Schubert, H. L., Blumenthal, R. M., & Cheng, X. (2003). Many paths to methyltransfer : A chronicle of convergence. *Trends in Biochemical Sciences*, 329-335. https://doi.org/10.1016/S0968-0004(03)00090-2

Schultz, S. K., & Kothe, U. (2024). RNA modifying enzymes shape tRNA biogenesis and function. *Journal of Biological Chemistry*. https://doi.org/10.1016/j.jbc.2024.107488

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., & Cate, J. H. D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*. https://doi.org/10.1126/science.1117230

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., Engreitz, J. M., Guttman, M., Satija, R., Lander, E. S., Fink, G., & Regev, A. (2014). Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell*. https://doi.org/10.1016/j.cell.2014.08.028

Seffouh, A., Trahan, C., Wasi, T., Jain, N., Basu, K., Britton, R. A., Oeffinger, M., & Ortega, J. (2022). RbgA ensures the correct timing in the maturation of the 50S subunits functional sites. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkac059

Sergeeva, O. V., Bogdanov, A. A., & Sergiev, P. V. (2015). What do we know about ribosomal RNA methylation in Escherichia coli? *Biochimie*. https://doi.org/10.1016/j.biochi.2014.11.019

Sergiev, P. V., Aleksashin, N. A., Chugunova, A. A., Polikanov, Y. S., & Dontsova, O. A. (2018). Structural and evolutionary insights into ribosomal RNA methylation. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.2569

Sergiev, P. V., Lesnyak, D. V., Bogdanov, A. A., & Dontsova, O. A. (2006). Identification of Escherichia coli m2G methyltransferases : II. The ygjO Gene Encodes a Methyltransferase Specific for G1835 of the 23 S rRNA. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2006.09.008

Sergiev, P. V., Serebryakova, M. V., Bogdanov, A. A., & Dontsova, O. A. (2008). The ybiN Gene of Escherichia coli Encodes Adenine-N6 Methyltransferase Specific for Modification of A1618 of 23 S Ribosomal RNA, a Methylated Residue Located Close to the Ribosomal Exit Tunnel. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2007.10.051

Sheng, J., Larsen, A., Heuberger, B. D., Blain, J. C., & Szostak, J. W. (2014). Crystal Structure Studies of RNA Duplexes Containing s ^{2} U:A and s ^{2} U:U Base Pairs. *Journal of the American Chemical Society*, 13916-13924. https://doi.org/10.1021/ja508015a

Shi, H., Wei, J., & He, C. (2019). Where, When, and How : Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular Cell*, 640-650. https://doi.org/10.1016/j.molcel.2019.04.025

Shi, R., Villarroya, M., Ruiz-Partida, R., Li, Y., Proteau, A., Prado, S., Moukadiri, I., Benítez-Páez, A., Lomas, R., Wagner, J., Matte, A., Velázquez-Campoy, A., Armengod, M.-E., & Cygler, M. (2009). Structure-Function Analysis of *Escherichia coli* MnmG (GidA), a Highly Conserved tRNA-Modifying Enzyme. *Journal of Bacteriology*. https://doi.org/10.1128/JB.00650-09

Shigi, N. (2014). Biosynthesis and functions of sulfur modifications in tRNA. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2014.00067

Shine, J., & Dalgarno, L. (1974). The 3′-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences*, *71*. https://doi.org/10.1073/pnas.71.4.1342

Siibak, T., & Remme, J. (2010). Subribosomal particle analysis reveals the stages of bacterial ribosome assembly at which rRNA nucleotides are modified. *RNA*. https://doi.org/10.1261/rna.2160010

Sipa, K., Sochacka, E., Kazmierczak-Baranska, J., Maszewska, M., Janicka, M., Nowak, G., & Nawrot, B. (2007). *Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA*. https://doi.org/10.1261/rna.538907

Sirand-Pugnet, P., Brégeon, D., Béven, L., Goyenvalle, C., Blanchard, A., Rose, S., Grosjean, H., Douthwaite, S., Hamdane, D., & De Crécy-Lagard, V. (2020). Reductive Evolution and Diversification of C5-Uracil Methylation in the Nucleic Acids of Mollicutes. *Biomolecules*. https://doi.org/10.3390/biom10040587

Soderberg, T., & Poulter, C. D. (2000). *Escherichia coli* Dimethylallyl Diphosphate:tRNA Dimethylallyltransferase : Essential Elements for Recognition of tRNA Substrates Within the Anticodon Stem−Loop. *Biochemistry*. https://doi.org/10.1021/bi992775u

Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., Suter, C. M., & Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks144

Srinivasan, P. R., & Borek, E. (1964). Enzymatic Alteration of Nucleic Acid Structure : Enzymes put finishing touches characteristic of each species on RNA and DNA by insertion of methyl groups. *Science*. https://doi.org/10.1126/science.145.3632.548

Stojković, V., Myasnikov, A. G., Young, I. D., Frost, A., Fraser, J. S., & Fujimori, D. G. (2020). Assessment of the nucleotide modifications in the high-resolution cryo-electron microscopy structure of the Escherichia coli 50S subunit. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa037

Su, J., Yang, J., Zhao, D., Kawula, T. H., Banas, J. A., & Zhang, J.-R. (2007). Genome-Wide Identification of *Francisella tularensis* Virulence Determinants. *Infection and Immunity*. https://doi.org/10.1128/IAI.01865-06

Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, J. J., Kim, S. H., & Rich, A. (1974). Three-dimensional structure of yeast phenylalanine transfer RNA at 3. 0Å resolution. *Nature*. https://doi.org/10.1038/248020a0

Sudol, C., Kilz, L.-M., Marchand, V., Thullier, Q., Guérineau, V., Goyenvalle, C., Faivre, B., Toubdji, S., Lombard, M., Jean-Jean, O., de Crécy-Lagard, V., Helm, M., Motorin, Y., Brégeon, D., & Hamdane, D. (2024). Functional redundancy in tRNA dihydrouridylation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkae325

Sun, J., Kinman, L. F., Jahagirdar, D., Ortega, J., & Davis, J. H. (2023). KsgA facilitates ribosomal small subunit maturation by proofreading a key structural lesion. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/s41594-023-01078-5

Sunita, S., Purta, E., Durawa, M., Tkaczuk, K. L., Swaathi, J., Bujnicki, J. M., & Sivaraman, J. (2007). Functional specialization of domains tandemly duplicated within 16S rRNA methyltransferase RsmC. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm411

Suzuki, T. (2002). Taurine as a constituent of mitochondrial tRNAs : New insights into the functions of taurine and human mitochondrial diseases. *The EMBO Journal*. https://doi.org/10.1093/emboj/cdf656

Suzuki, T. (2021). The expanding world of tRNA modifications and their disease relevance. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-021-00342-0

Suzuki, T., Ueda, H., Okada, S., & Sakurai, M. (2015). Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nature Protocols*. https://doi.org/10.1038/nprot.2015.037

Teufel, R. (2024). Oxygen-transfer reactions by enzymatic flavin-N5 oxygen adducts—Oxidation is not a must. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2024.102464

Tillault, A.-S., Schultz, S. K., Wieden, H.-J., & Kothe, U. (2018). Molecular Determinants for 23S rRNA Recognition and Modification by the E. coli Pseudouridine Synthase RluE. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2018.03.011

Tissières, A., & Hopkins, J. W. (1961). FACTORS AFFECTING AMINO ACID INCORPORATION INTO PROTEINS BY ESCHERICHIA COLI RIBOSOMES. *Proceedings of the National Academy of Sciences*, *12*. https://doi.org/10.1073/pnas.47.12.2015

Tissières, A., & Watson, J. D. (1958). Ribonucleoprotein Particles from Escherichia Coli. *Nature*. https://doi.org/10.1038/182778b0

Toh, S.-M., Xiong, L., Bae, T., & Mankin, A. S. (2008). The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA. *RNA*. https://doi.org/10.1261/rna.814408

Torres, A. G., Piñeyro, D., Filonava, L., Stracker, T. H., Batlle, E., & Ribas De Pouplana, L. (2014). A-to-I editing on tRNAs : Biochemical, biological and evolutionary implications. *FEBS Letters*. https://doi.org/10.1016/j.febslet.2014.09.025

Trahan, C., & Oeffinger, M. (2023). The Importance of Being RNA-est : Considering RNA-mediated ribosome plasticity. *RNA Biology*. https://doi.org/10.1080/15476286.2023.2204581

Tsai, K., Stojković, V., Noda-Garcia, L., Young, I. D., Myasnikov, A. G., Kleinman, J., Palla, A., Floor, S. N., Frost, A., Fraser, J. S., Tawfik, D. S., & Fujimori, D. G. (2022). Directed evolution of the rRNA methylating enzyme Cfr reveals molecular basis of antibiotic resistance. *eLife*. https://doi.org/10.7554/eLife.70017

Vaidyanathan, P. P., Deutscher, M. P., & Malhotra, A. (2007). RluD, a highly conserved pseudouridine synthase, modifies 50S subunits more specifically and efficiently than free 23S rRNA. *RNA*. https://doi.org/10.1261/rna.711207

Van Buul, C. P. J. J., & Van Knippenberg, P. H. (1985). Nucleotide sequence of the ksgA gene of Escherichia coli : Comparison of methyltransferases effecting dimethylation of adenosine in ribosomal RNA. *Gene*. https://doi.org/10.1016/0378-1119(85)90204-5

Voorhees, R. M., & Ramakrishnan, V. (2013). Structural Basis of the Translational Elongation Cycle. *Annual Review of Biochemistry*, *1*. https://doi.org/10.1146/annurev-biochem-113009-092313

Wang, X., & He, C. (2014). Dynamic RNA Modifications in Posttranscriptional Regulation. *Molecular Cell*, 5-12. https://doi.org/10.1016/j.molcel.2014.09.001

Watanabe, K., Yokoyama, S., Hansske, F., Kasai, H., & Miyazawa, T. (1979). CD and NMR studies on the conformational thermostability of 2-thioribothymidine found in the TψC loop of thermophile tRNA. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/0006-291X(79)91574-2

Wei, Y., Zhang, H., Gao, Z.-Q., Wang, W.-J., Shtykova, E. V., Xu, J.-H., Liu, Q.-S., & Dong, Y.-H. (2012). Crystal and solution structures of methyltransferase RsmH provide basis for methylation of C1402 in 16S rRNA. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2012.04.011

Weinger, J. S., Parnell, K. M., Dorner, S., Green, R., & Strobel, S. A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/nsmb841

Weitzmann, C., Tumminia, S. J., Boublik, M., & Ofengand, J. (1991). A paradigm for local conformational control of fucntion in the ribosome : Binding of ribosomal protein S19 to Escherichia coli 16S rRNA in the presence of S7 is required for methylation of m ^{2}

G966 and blocks methylation of m ^{5} C967 by their respective methyltransferases. *Nucleic Acids Research*. https://doi.org/10.1093/nar/19.25.7089

Wierenga, R. K. (2001). The TIM-barrel fold : A versatile framework for efficient enzymes. *FEBS Letters*. https://doi.org/10.1016/S0014-5793(01)02236-0

Williamson, J. R. (2003). After the ribosome structures : How are the subunits assembled? *RNA*. https://doi.org/10.1261/rna.2164903

Wilson, D. N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro3155

Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, R., Gutell, R., Hogan, J. J., & Noller, H. F. (1980). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Research*. https://doi.org/10.1093/nar/8.10.2275

Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G., & Ofengand, J. (1995). Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from Escherichia coli. *Biochemistry*. https://doi.org/10.1021/bi00027a043

Wrzesinski, J., Bakin, A., Ofengand, J., & Lane, B. G. (2000). Isolation and Properties of Escherichia coli 23S-RNA Pseudouridine 1911, 1915, 1917 Synthase (RluD). *IUBMB Life*. https://doi.org/10.1080/15216540050176566

Xing, F., Hiley, S. L., Hughes, T. R., & Phizicky, E. M. (2004). The Specificities of Four Yeast Dihydrouridine Synthases for Cytoplasmic tRNAs. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M401221200

Yanas, A., & Liu, K. F. (2019). RNA modifications and the link to human disease. In *Methods in Enzymology*. Elsevier. https://doi.org/10.1016/bs.mie.2019.08.003

Yi, C., & Pan, T. (2011). Cellular Dynamics of RNA Modification. *Accounts of Chemical Research*. https://doi.org/10.1021/ar200057m

Yokoyama, S. (1987). Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. *Advances in Biophysics*. https://doi.org/10.1016/0065-227X(87)90006-2

Yu, F., Tanaka, Y., Yamashita, K., Suzuki, T., Nakamura, A., Hirano, N., Suzuki, T., Yao, M., & Tanaka, I. (2011). Molecular basis of dihydrouridine formation on tRNA. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1112352108

Yusupov, M. M., Yusupova, G. Zh., 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. *Science*. https://doi.org/10.1126/science.1060089

Zaher, H. S., Shaw, J. J., Strobel, S. A., & Green, R. (2011). The 2′-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC: Effects of dA76 modification of the peptidyl-tRNA. *The EMBO Journal*. https://doi.org/10.1038/emboj.2011.142

Zhao, M., Wang, L., Zhang, H., Dong, Y., Gong, Y., Zhang, L., & Wang, J. (2014). Purification, crystallization and preliminary crystallographic analysis of the 16S rRNA methyltransferase RsmI from *Escherichia coli*. *Acta Crystallographica Section F Structural Biology Communications*. https://doi.org/10.1107/S2053230X14016999

Zhou, C., & Huang, R. H. (2008). Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: Insight into tRNA recognition and reaction mechanism. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0805680105

Zhou, Y., Kierzek, E., Loo, Z. P., Antonio, M., Yau, Y. H., Chuah, Y. W., Geifman-Shochat, S., Kierzek, R., & Chen, G. (2013). Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt352