



**HAL**  
open science

## The universal Sua5/TsaC family evolved different mechanisms for the synthesis of a key tRNA modification

Adeline Pichard-Kostuch, Violette da Cunha, Jacques Oberto, Ludovic Sauguet, Tamara Basta

### ► To cite this version:

Adeline Pichard-Kostuch, Violette da Cunha, Jacques Oberto, Ludovic Sauguet, Tamara Basta. The universal Sua5/TsaC family evolved different mechanisms for the synthesis of a key tRNA modification. *Frontiers in Microbiology*, 2023, 14, pp.1204045. 10.3389/fmicb.2023.1204045 . hal-04234441

**HAL Id: hal-04234441**

**<https://cnrs.hal.science/hal-04234441v1>**

Submitted on 23 Oct 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License



## OPEN ACCESS

## EDITED BY

Marleen van Wolferen,  
University of Freiburg, Germany

## REVIEWED BY

Lennart Randau,  
University of Marburg, Germany  
Jose Luis Llacer,  
Spanish National Research Council (CSIC),  
Spain

## \*CORRESPONDENCE

Tamara Basta  
✉ tamara.basta@i2bc.paris-saclay.fr

## PRESENT ADDRESS

Violette Da Cunha,  
Génomique Métabolique, Genoscope,  
Institut François Jacob,  
CEA, CNRS, Univ Evry,  
Université Paris-Saclay,  
Evry, France

RECEIVED 11 April 2023

ACCEPTED 02 June 2023

PUBLISHED 21 June 2023

## CITATION

Pichard-Kostuch A, Da Cunha V, Oberto J,  
Sauguet L and Basta T (2023) The universal  
Sua5/TsaC family evolved different  
mechanisms for the synthesis of a key tRNA  
modification.

*Front. Microbiol.* 14:1204045.

doi: 10.3389/fmicb.2023.1204045

## COPYRIGHT

© 2023 Pichard-Kostuch, Da Cunha, Oberto,  
Sauguet and Basta. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# The universal Sua5/TsaC family evolved different mechanisms for the synthesis of a key tRNA modification

Adeline Pichard-Kostuch<sup>1</sup>, Violette Da Cunha<sup>1†</sup>, Jacques Oberto<sup>1</sup>,  
Ludovic Sauguet<sup>2</sup> and Tamara Basta<sup>1\*</sup>

<sup>1</sup>CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, Gif-sur-Yvette, France, <sup>2</sup>Architecture and Dynamics of Biological Macromolecules, Institut Pasteur, Université Paris Cité, CNRS, UMR 3528, Paris, France

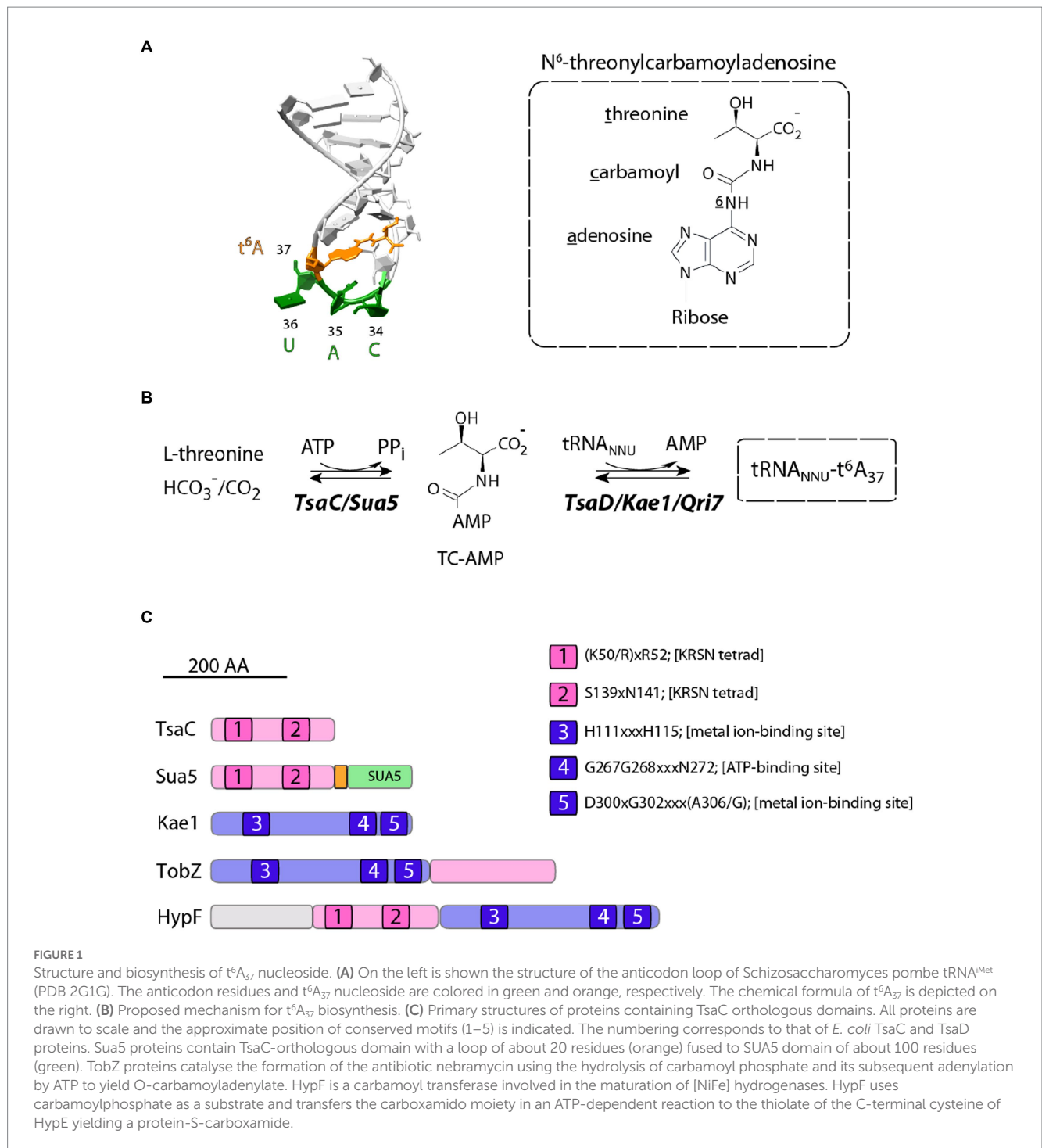
TsaC/Sua5 family of enzymes catalyzes the first step in the synthesis of N<sup>6</sup>-threonyl-carbamoyl adenosine (t<sup>6</sup>A) one of few truly ubiquitous tRNA modifications important for translation accuracy. TsaC is a single domain protein while Sua5 proteins contains a TsaC-like domain and an additional SUA5 domain of unknown function. The emergence of these two proteins and their respective mechanisms for t<sup>6</sup>A synthesis remain poorly understood. Here, we performed phylogenetic and comparative sequence and structure analysis of TsaC and Sua5 proteins. We confirm that this family is ubiquitous but the co-occurrence of both variants in the same organism is rare and unstable. We further find that obligate symbionts are the only organisms lacking *sua5* or *tsaC* genes. The data suggest that Sua5 was the ancestral version of the enzyme while TsaC arose via loss of the SUA5 domain that occurred multiple times in course of evolution. Multiple losses of one of the two variants in combination with horizontal gene transfers along a large range of phylogenetic distances explains the present day patchy distribution of Sua5 and TsaC. The loss of the SUA5 domain triggered adaptive mutations affecting the substrate binding in TsaC proteins. Finally, we identified atypical Sua5 proteins in Archaeoglobi archaea that seem to be in the process of losing the SUA5 domain through progressive gene erosion. Together, our study uncovers the evolutionary path for emergence of these homologous isofunctional enzymes and lays the groundwork for future experimental studies on the function of TsaC/Sua5 proteins in maintaining faithful translation.

## KEYWORDS

universal proteins, enzyme, Sua5, TsaC, t<sup>6</sup>A, tRNA, evolution

## Introduction

tRNA requires post-transcriptional maturation to be functional during the translation process. This involves *inter alia* the enzymatic modification of the canonical A, U, C and G bases to form modified nucleosides by addition of a variety of chemical groups (Väre et al., 2017). About hundred different modified nucleosides have been identified so far in tRNA, 20 of which are universal and were likely inherited from the Last Universal Common Ancestor (LUCA; Cantara et al., 2011; Machnicka et al., 2014). Among those, N<sup>6</sup>-threonyl-carbamoyl-adenosine (t<sup>6</sup>A) is found on the adenosine in position 37, next to the anticodon of almost all tRNAs that recognize ANN codons (where N = G, A, C or U; Figure 1A; Ishikura et al., 1969; Powers and Peterkofsky, 1972).



t<sup>6</sup>A is required for translation accuracy (Högenauer et al., 1972; Weissenbach and Grosjean, 1981; Stuart et al., 2000; Murphy et al., 2004; El Yacoubi et al., 2011), aminoacylation of some tRNA species (Nureki et al., 1994), translation initiation (Thiaville et al., 2016; Llácer et al., 2018) and translocation of the ribosome on mRNA (Phelps et al., 2004). The mutations in t<sup>6</sup>A synthetic genes were linked with a variety of phenotypes in eukaryotes and prokaryotes ranging from cell death or slow growth to short telomeres and defects in transcription probably reflecting the far-reaching indirect consequences of the unfaithful translation (El Yacoubi et al., 2009; Naor et al., 2012;

Beenstock and Sicheri, 2021). In humans, mutations in all t<sup>6</sup>A-synthetic genes were linked with Galloway-Mowat syndrome a severe genetic disease causing brain malformation and renal dysfunction (Braun et al., 2017; Edvardson et al., 2017; Arrondel et al., 2019).

t<sup>6</sup>A-tRNA biosynthesis proceeds in two steps and requires multiple proteins. The universal enzyme family Sua5/TsaC uses L-threonine, ATP and bicarbonate/carbon dioxide (HCO<sub>3</sub><sup>-</sup>/CO<sub>2</sub>) to synthesize the activated intermediate threonyl-carbamoyl-adenylate (TC-AMP) via a complex and poorly understood mechanism (Figure 1B; El Yacoubi et al., 2009; Lauhon, 2012; Perrochia et al.,

2013a,b; Harris et al., 2015). In the second step, the threonyl-carbamoyl moiety is transferred from the TC-AMP onto the substrate tRNA. This reaction is catalyzed by Kae1/TsaD/Qri7 universal protein family (Figure 1B) which associates with accessory proteins to form the DEZ complex in bacteria and the KEOPS complex in archaea and eukaryotes [reviewed in Thiaville et al. (2014b)].

TsaC/Sua5 family of proteins is a part of a restricted set of about 60 proteins shared by all cellular organisms (Koonin, 2003; Galperin, 2004). Remarkably, this family is composed of two distinct variants, TsaC and Sua5, which share a homologous catalytic domain but differ in the presence of a second domain, named SUA5 (Figure 1C), found only in Sua5 proteins. Initial bioinformatics analysis reported no clear phyletic pattern underpinning the distribution of TsaC and Sua5 proteins and additionally showed that the simultaneous occurrence of both genes in a genome seems to be rare (Thiaville et al., 2014b). TsaC-like domain is also found fused to Kae1-like domain in HypF proteins involved in the maturation of [NiFe] hydrogenases and in TobZ proteins that synthesize the antibiotic nebramycin (Figure 1C). In these proteins, the TsaC-like domains use carbamoyl phosphate to catalyze a carbamoylation reaction similar to a putative step in TC-AMP synthesis (Petkun et al., 2011; Parthier et al., 2012).

The structures of TsaC from *E. coli* (*Ec*-TsaC; Teplova et al., 2000; Harris et al., 2015) and Sua5 from the *Sulfolobus tokodaii* (*St*-Sua5; Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012) and *Pyrococcus abyssi* (*Pa*-Sua5; Pichard-Kostuch et al., 2018) show that the catalytic domain adopts a globular twisted fold made of 7 parallel and anti-parallel  $\beta$ -strands bordered by 7  $\alpha$ -helices. The active site is in the cavity formed by this domain and contains highly conserved residues K<sup>56</sup>xR<sup>58</sup>/S<sup>143</sup>xN<sup>145</sup> and T<sup>33</sup>/S<sup>181</sup>/R<sup>195</sup> (*Pa*-Sua5 numbering) involved in ATP and threonine binding, respectively. The mutational analyses showed that these residues were all required for TC-AMP synthesis by Sua5 from *S. cerevisiae* (Wan et al., 2013).

The TsaC-like catalytic domain of Sua5 proteins carries a C-terminal extension of about 20 residues forming a flexible loop that contains a set of highly conserved residues, Pro<sup>228</sup>-Gly<sup>229</sup>-Met<sup>230</sup> and His<sup>234</sup>-Tyr<sup>235</sup> (*Pa*-Sua5 numbering; Pichard-Kostuch et al., 2018). The loop folds into the active site gorge and was suggested to act as a gate regulating the binding and release of ligands (Pichard-Kostuch et al., 2018).

The C-terminal extremity of the loop is anchored in the SUA5 domain which is composed of about 100 residues (Agari et al., 2008; Kuratani et al., 2011; Pichard-Kostuch et al., 2018). The domain exhibits an atypical Rossmann fold composed of five  $\beta$ -strands ( $\beta$ 12- $\beta$ 16) and three  $\alpha$ -helices ( $\alpha$ 9- $\alpha$ 11). SUA5 domain and the catalytic TsaC-like domain form a tight interface that was shown to be important for the activity (Pichard-Kostuch et al., 2018).

How TsaC and Sua5 emerged in course of evolution and what are the functional differences between these two different protein scaffolds remains elusive. To address these questions, we performed a comprehensive phylogenetic, structural, and sequence analysis of Sua5/TsaC proteins. The data suggest pre-LUCA origin of this family and a complex evolution including gene erosion, multiple gene losses and horizontal gene transfers. We identified conserved residues specific for TsaC or Sua5 proteins and tentatively assigned them a role in substrate binding. Finally, we identified atypical Sua5 proteins which seem to be in the process of losing the SUA5 domain thus potentially ongoing the transition toward TsaC homologs. Together,

this work provides testable hypothesis for uncovering the functional differences between TsaC and Sua5 proteins and, more generally, provides insights into the evolutionary mechanisms driving the emergence of isofunctional enzymes.

## Materials and methods

### Extraction of Sua5 and TsaC sequences and distribution analysis

Sua5 and TsaC sequences were extracted from UniProt (version 06/2021 or version 11/2022) and UniRef (version 11/2022). TsaC/Sua5 sequences in DPANN archaea were retrieved using BLASTp searches with *Pa*-Sua5 protein sequence as query against all genome assemblies available (nr database, version 11/2022). The hits were filtered for the presence of KxR/SxN tetrad and one sequence per species was chosen to reduce redundancy. Following this pipeline, 5,889 bacterial, 761 archaeal and 1,318 eukaryotic sequences were recovered. The identified Sua5 or TsaC orthologs were mapped on the RNA polymerase-based universal phylogenetic tree (Da Cunha et al., 2017) or on the previously established phylogeny of DPANN archaea (Moody et al., 2022). The organisms were ranked according to the NCBI taxonomy database (as of 11/2022) or, for DPANN organisms, based on (Moody et al., 2022). The accession numbers for all sequences are available at figshare repository (doi: 10.6084/m9.figshare.22283929).

### Phylogenetic analysis

For computing the universal phylogenetic tree we chose a set of species that has been previously vetted for its ability to produce a robust phylogenetic signal using universal proteins (Da Cunha et al., 2017). The corresponding Sua5/TsaC sequences (Supplementary Tables 2, 3) were extracted from Uniprot database, aligned using MAFFT v7 with auto settings (Yamada et al., 2016). The alignment was trimmed using BMGE (Criscuolo and Gribaldo, 2010) with BLOSUM30 matrix leaving 153 positions for tree construction, corresponding only to the catalytic domain residues (the SUA5 domain was removed by the trimming software). In addition, we generated an alignment of full length Sua5 sequences that contained 470 positions. The maximum likelihood (ML) trees were inferred using IQ-TREE v1.4.3 (Trifinopoulos et al., 2016) with the TESTNEW option for model selection. The branch support was obtained using nonparametric bootstrap (100 replicates), SH-aLRT test and ultrafast bootstrap approximation (1,000 replicates; Minh et al., 2013; Kalyanamoorthy et al., 2017) or booster (Lemoine et al., 2018).

Archaeoglobi Sua5 sequences were extracted from Uniprot (doi: 10.6084/m9.figshare.22283929) and the Maximum Likelihood tree was inferred using Phylogeny.fr web page (Dereeper et al., 2008).

### Horizontal gene transfer analysis

To screen for putative horizontally transferred genes in Archaea we extracted all sequences annotated as TsaC/Sua5 from

Uniprot database (9,553 sequences, database release 2019\_06). We filtered this dataset to remove sequences containing less than 100 amino acids or more than 400 amino acids as well as those without the KRSN tetrad. The remaining set of 1,275 archaeal TsaC/Sua5 sequences were classified according to their taxonomic affiliation to 13 different archaeal taxa (doi: 10.6084/m9.figshare.22283929). This allowed to identify candidate proteins in “mixed” groups containing both Sua5 and TsaC users. To reduce the impact of convergent evolution on small proteins, we focused our analyses on transfer of *sua5* genes. We used BLASTp searches against the nr protein database to identify the most similar sequences to the candidate query protein. If the best-hit species was only distantly related to the query species, the query protein was selected as candidate for HGT. The HGT candidate were subjected to phylogenetic analysis to determine their origin. The phylogenetic inferences were done as indicated above (phylogenetic analysis section).

## Protein sequence alignment

To identify Sua5- or TsaC-specific residues we aligned four representative sequences for TsaC and for TsaC-like domain from each domain of life (doi: 10.6084/m9.figshare.22283929) using MAFFT v7 with auto settings (Yamada et al., 2016).

To examine the distribution of the signature residue Pro<sup>143</sup>/Thr<sup>138</sup> we retrieved 26,036 TsaC and Sua5 sequences from the Complete Genome Data Bank of the NCBI (database release 2015\_05) using BLASTN, with Sua5 from *Pyrococcus abyssi* or TsaC from *Escherichia coli* as query. Using hmmsearch (Finn et al., 2011) we further retrieved 1,088 eukaryotic TsaC and Sua5 sequences from the Uniprot-Proteomes database (database release 2018\_11). The hit sequences missing the KRSN catalytic tetrad or partial sequences were eliminated from the analysis. One UniRef90 sequence per genus was kept for sequence alignment, 654 for prokaryotes and 564 for eukaryotes in total (accession numbers were deposited to doi: 10.6084/m9.figshare.22283929). Sequence alignment depiction was done using ESPript3 with default parameters (Robert and Gouet, 2014).

## Structure prediction and analysis

AlphaFold2 (Jumper et al., 2021) batch colab<sup>1</sup> was used to predict *de novo* structures of atypical Sua5 proteins from *Archaeoglobus* archaea: *Archaeoglobus profundus* (Ap-Sua5, UniProt D2RFV3), *A. veneficus* (Av-Sua5, UniProt F2KMZ1) and *A. fulgidus* (Af-Sua5, UniProt O29477). Crystal structures of several TsaC and Sua5 proteins were used for structural comparison. The PDB accession codes of those structures are 1HRU and 2MX1 for *Ec*-TsaC, 2EQA, 3AJE and 4E1B for *St*-Sua5, 6F87 for *Pa*-Sua5, 3VEZ for *St*-TobZ and 3TTC for *Ec*-HypF. Structures were visualized and edited with UCSF ChimeraX v. 1.4 (Pettersen et al., 2021).

1 [https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/batch/AlphaFold2\\_batch.ipynb](https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/batch/AlphaFold2_batch.ipynb)

## Results

### The distribution of Sua5 and TsaC proteins does not follow a clear phyletic pattern

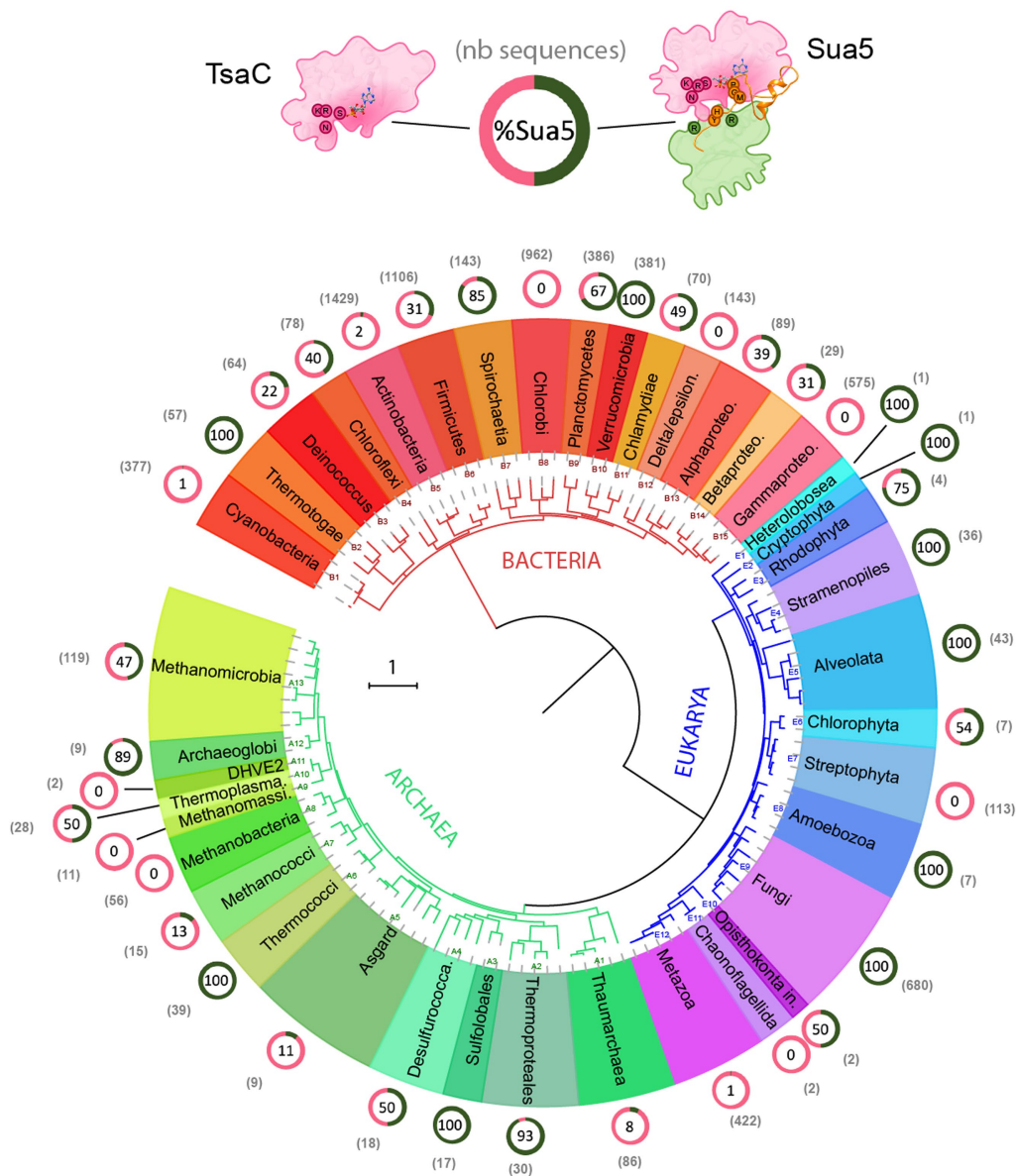
Previous work established that TsaC/Sua5 family of proteins (COG0009) is ubiquitous however the specific distribution of TsaC or Sua5 orthologs was only examined for a small number of representative isolated species.

Here, we exploited the recent massive increase of (meta)genome sequencing data to perform an up to date comprehensive analysis of the distribution of Sua5 and TsaC proteins among living organisms. To this end, we retrieved 5,889 bacterial, 761 archaeal and 1,318 eukaryotic sequences from public databases and mapped their presence on reference phylogenetic trees (Figure 2; Supplementary Table 1; Supplementary Figure 1).

The data confirm the ubiquitous distribution of the Sua5/TsaC family among organisms. We found that the obligate ectosymbiotic archaea *Nanoarchaeum equitans*, and *Nanobdella aerobiophila*, were the only cultured organisms that lack the genes encoding TsaC/Sua5 proteins. In addition, we failed to detect Sua5/TsaC proteins in Undinarchaeota and Huberarchaea lineages from DPANN archaeal superphylum (Supplementary Figure 1) but it is unclear whether they are genuinely missing or were missed during metagenomic assemblies. Notably, in the entire tree we detected few species encoding both TsaC and Sua5: in the *Leotiomyces* fungi clade (e.g., *Aspergillus bombycis* – Uniprot ID A0A1F7ZM40 and A0A1F8A8Q8) and in the *Vibrio* bacteria (e.g., *Vibrio aquaticus* – Uniprot A0A3S0QCR7 and A0A3S0V185). The data further show that out of 41 examined taxonomic groups depicted in Figure 2; Supplementary Figure 1, 10 groups contain species that are exclusively Sua5 users, 8 groups contain species that encode the TsaC orthologs only, 3 groups contain at least 98% of species that encode only the TsaC orthologs and the remaining 20 taxons contain species encoding Sua5 or TsaC. Intriguingly, however, the distribution of TsaC and Sua5 users across the tree shows no clear phyletic pattern suggesting a complex evolutionary history for the Sua5/TsaC family of proteins.

### The phylogeny suggests a pre-LUCA origin of TsaC/Sua5 family

To gain insight into the evolutionary history of Sua5/TsaC family we constructed a phylogenetic tree using balanced taxonomic sampling across the three domains of life (Supplementary Tables 2, 3). The tree was inferred from the alignment of TsaC proteins and the catalytic TsaC-like domain of Sua5 proteins. It exhibits bipartite topology whereby TsaC proteins and TsaC-like domains robustly (bootstrap values 100%) segregated into two monophyletic clades (Figure 3), each clade containing sequences from the three domains of life. The branches in the TsaC part of the tree are in general longer suggesting that these proteins evolved at a higher rate. While in the Sua5 part of the tree no clear phyletic pattern can be observed, the TsaC tree shows that eukaryotic sequences cluster together and seem to be more related to archaeal TsaC sequences thus recapitulating the established phylogeny of organisms. Similarly, the tree built from the alignment of full-length Sua5 sequences (Supplementary Figure 2) revealed a clear phyletic pattern whereby the bacterial and archaeal sequences formed two distinct groups while the eukaryotic sequences were split into two



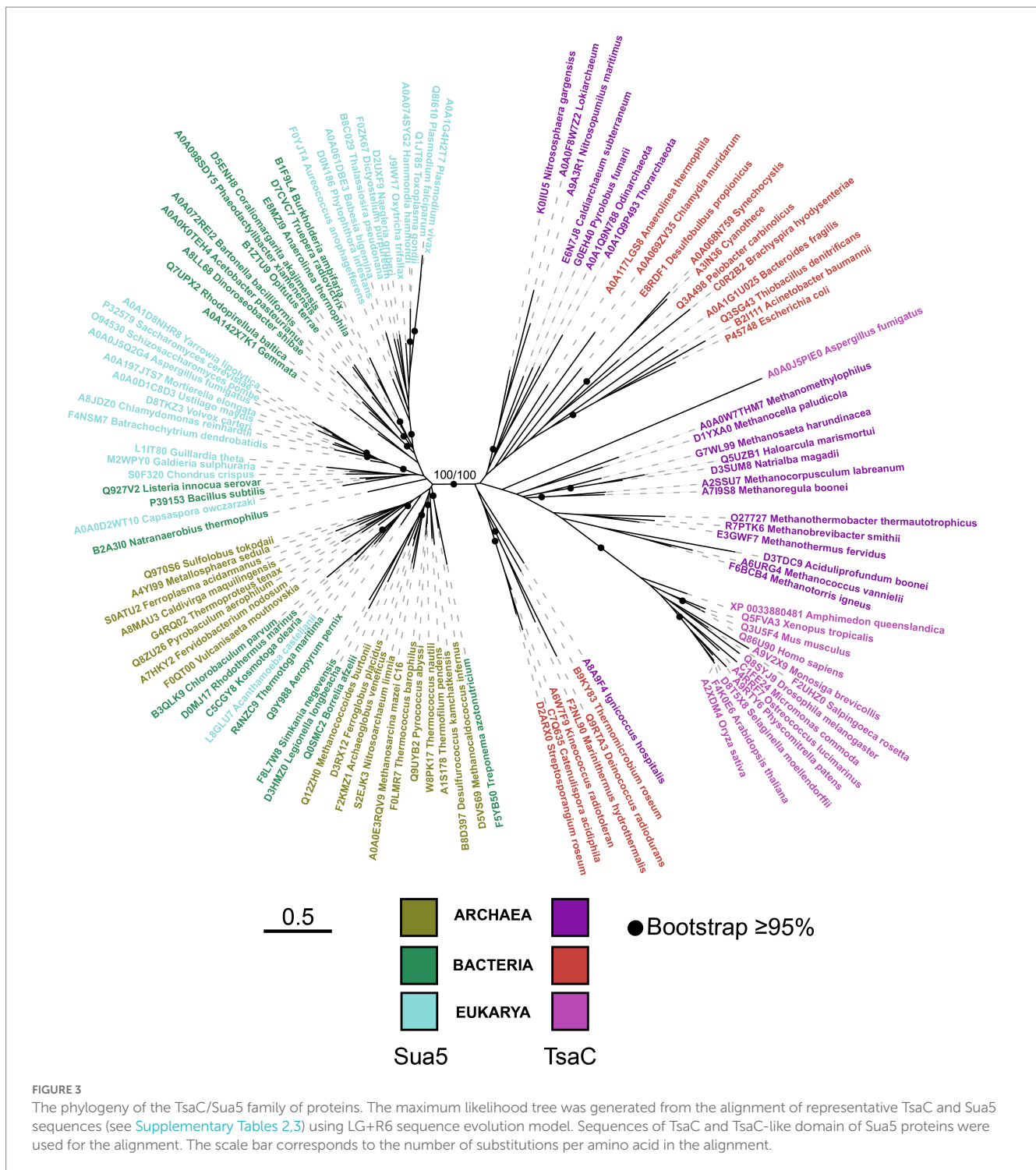
**FIGURE 2** Distribution of TsaC and Sua5 proteins across the tree of life. The universal phylogeny based on RNA polymerase sequences is depicted. The number of TsaC and Sua5 sequences is indicated in the brackets for each taxon. The ring graph indicates the ratio of TsaC (pink) and Sua5 (green) orthologs for a given taxon. The numbers in the ring correspond to the percentage of Sua5 sequences. The scale bar corresponds to the number of substitutions per position in the alignment.

clades. The bipartite tree topology suggests that Sua5 and TsaC emerged from an ancient duplication event that occurred before LUCA, followed by either SUA5 domain acquisition or loss to yield the two variants.

### Horizontal transfer of sua5 and tsaC genes occurs across a wide range of phylogenetic distances

The patchy distribution of Sua5 and TsaC proteins could, at least in part, be explained by horizontal gene transfers (HGT) of sua5 and tsaC genes. To identify such events, we performed an initial screen using BLASTp searches whereby we focused on “mixed” taxonomic groups containing Sua5 users and TsaC users. For example, within

Chlorophyta phylum (eukaryotes, green algae, E6 in Figure 2), the Chlorophyceae family (e.g., *Volvox carteri f. nagariensis* – Uniprot D8TKZ3) is the only one carrying Sua5 orthologs (Supplementary Table 1). This protein is most similar to the Sua5 from *Apophysomyces ossiformis* (Uniprot A0A162WET7, 57% sequence identity) which is a Mucoromycota fungus. Similarly, Rhodophyta (E3), are Sua5 users except for one species (*Rhodospirillum rubrum* – Uniprot A0A7S0G1B8) that encodes TsaC. This protein is most similar (49% identity) to the TsaC from Firmicutes bacteria (*Brochothrix campestris*). Leotiomyceta fungi such as *Aspergillus* are a peculiar case because several encode both Sua5 and TsaC proteins with a complete KxR/SxN motif suggesting that both orthologs are functional. However, the latter protein shows highest sequence similarity with Actinobacteria (*Amycolatopsis bartoniae*, 47.5% identity) and



Proteobacteria (*Verticella sediminum*, 41.9% identity) suggesting that this may be a case of inter-domain HGT. This initial screen identified several other candidate proteins in Bacteria and Archaea. Among those, we observed that all *Methanothermococcus* species encoded TsaC ortholog except *M. thermolithotrophicus* which carries a Sua5 ortholog (WP\_018153339.1). The sequence of this ortholog is identical to Sua5 from *Methanococcus maripaludis* KA1 and the gene is encoded on a ~14 kbp fragment that is 99% identical to *M. maripaludis* KA1 genome ([Supplementary Figure 3](#)). This suggests a recent acquisition of the whole fragment by horizontal gene transfer.

We next tested the robustness of the BLASTp results using phylogenetic analysis. As an example, we selected outlier Sua5 proteins from groups which are predominantly TsaC-users (the Methanococci archaea (A7), Thaumarchaea (A1) and Thermotogae bacteria (B2); [Supplementary Table 1](#)) and aligned those with representative Sua5 sequences from Bacteria and Archaea. The tree was resolved into two distinct clades containing bacterial or archaeal sequences allowing to identify the closest homologs for the outlier Sua5 proteins ([Figure 4A](#)). The tree topology showed that (i) the Sua5 proteins from *Methanococcus* species and Thaumarchaeota branched within Euryarchaeota; (ii) the

Sua5 protein from *Methanocaldococcus* archaeon branched within bacterial Sua5 orthologs; (iii) Sua5 proteins from Thermotoga bacteria branched within Crenarchaeal Sua5 orthologs. This suggested that the outlier Sua5 proteins were acquired by HGT from closely related (*Methanococcus*) or distantly related (Thaumarchaeota, *Methanocaldococcus*, and Thermotoga) organisms (Figure 4B).

Together, these analyses show that TsaC and Sua5 encoding genes have been submitted to horizontal transfers across different phylogenetic distances including at the highest taxonomic level.

## The identification of variant-specific and conserved residues indicates distinct modes of substrate recognition in TsaC and Sua5 proteins

The segregation of TsaC and TsaC-like sequences into two distinct phylogenetic clades indicated that each variant must contain a specific set of residues in addition to the residues conserved across the whole TsaC/Sua5 family. In search for such residues, we aligned TsaC or Sua5 sequences representative for each domain of life (Supplementary Figure 4). This showed that almost half out of approximately 200 positions in the alignment are similar, in line with a strong conservation of structure and function of the TsaC domain across the tree of life. In addition, we identified 15 conserved residues specific for Sua5 proteins (Supplementary Figure 4), 14 of which are variable among TsaC proteins. Notably, three of them, Pro<sup>59</sup>, Asn<sup>62</sup> and His<sup>67</sup> (*Pa*-Sua5 numbering) are found in the loop that forms the deep part of the catalytic cavity surface, with their side chains pointing towards the cavity (Figures 5A,C). His<sup>67</sup> side chain forms a H-bond (2.7 Å) with the hydroxyl function of L-threonine in the structures of *St*-Sua5 (Supplementary Figure 5A). In the structure of *Pa*-Sua5 the side chain nitrogen of Asn<sup>62</sup> acts as a donor to form a H-bond (2.4 Å) with PPI (Figure 5C) while its side chain oxygen acts as an acceptor to form a H-bond with the 3'-OH function of the ribosyl moiety of TC-AMP in the structure of *St*-Sua5 (2.6 Å; Supplementary Figure 5A). In several TsaC variants, Asn<sup>62</sup> is replaced by a lysine (Supplementary Figure 4). However, unlike asparagine, lysine can only act as a H-bond donor suggesting their respective role in TC-AMP synthesis may not be identical.

Most of the other Sua5 specific conserved residues form two motifs at the inter-domain interface: the KPSPT motif is composed of (Lys/Arg)<sup>149</sup>-Pro<sup>150</sup>-Ser<sup>151</sup>-Pro<sup>152</sup>-Thr<sup>153</sup> and the GVE motif is composed of Gly<sup>178</sup>-Hyd<sup>179</sup>-Glu<sup>180</sup> (where Hyd corresponds to the hydrophobic residue Val, Ile or Leu; Figure 5D). The motifs interact with one another via a H-bond (2.7 Å) formed by the side chains of Ser<sup>151</sup> and Glu<sup>180</sup>. The backbone chain of the KPSPT motif forms several H-bonds with *inter alia* the side chain of His<sup>157</sup>, another Sua5-specific residue, and the conserved Arg<sup>301</sup> present in the SUA5 domain. In addition, the Sua5-specific amino acids Asp<sup>161</sup> and Glu<sup>180</sup> are forming salt bridges with Arg<sup>301</sup> and Arg<sup>328</sup>, respectively (Figure 5D), thus likely playing a key role in the stabilization of the inter-domain interaction.

Remarkably, only one TsaC-specific conserved residue, Thr<sup>138</sup> (*Ec*-TsaC numbering), could be clearly identified (Supplementary Figure 4). This residue is found in the catalytic cavity of TsaC proteins and could potentially interact with substrate molecules (Figure 5B). To investigate this further we analyzed the structures of TsaC-like domains of TobZ and HypF proteins that co-crystallized with substrate molecules. The corresponding residue Thr<sup>529</sup> in TsaC-like domain of

TobZ protein forms a H-bond (2.8 Å) with the phosphor-moiety of the carbamoyl-phosphate a precursor for the formation of carbamoyl-adenylate (Figure 5F). This phosphor-moiety occupies the same position as the phosphor-moiety of TC-AMP and as the β phosphate of AMPPNP in the structures of *St*-Sua5 (Supplementary Figure 5B). In the TsaC-like domain of HypF from *E. coli*, the equivalent residue Thr<sup>321</sup> interacts via a H-bond (2.6 Å) with the β phosphates of the bound ADP (Supplementary Figure 5C).

Overall, the comparative sequence and structural analyses identified variant-specific conserved residues with putative key roles in substrate binding thus highlighting previously unrecognized functional differences between TsaC and Sua5 proteins.

## A single conserved residue constitutes a distinguishing feature between Sua5 and TsaC proteins

Structural and sequence analyses identified Thr<sup>138</sup> (*Ec*-TsaC numbering) as the only conserved residues specific for TsaC proteins. Intriguingly, the corresponding residue in Sua5 proteins is a conserved Proline (Pro<sup>143</sup> in *Pa*-Sua5; Supplementary Figure 4; Figure 5C) suggesting that this single and evolutionary conserved residue is a key determinant of specific mechanisms by which Sua5 and TsaC proteins catalyze the synthesis of TC-AMP. To test the relevance of this observation, we extracted 653 prokaryotic and 566 eukaryotic sequences from complete bacterial and archaeal genomes (one sequence per genus) and plotted the nature of the signature residue as a function of the size of the protein (Figure 6; Supplementary Figure 6). This showed that Sua5 proteins have exclusively Pro<sup>143</sup> at this position. In TsaC sequences, Thr<sup>138</sup> is predominant with up to 99% of occurrences in eukaryotes but is replaced by Ser<sup>138</sup> in 16% of bacterial sequences. The only exception to this rule are TsaC sequences from the bacterial genera *Brachyspira*, *Lactococcus*, *Erysipelothrix* as well as the entire *Aquificae* phylum where Thr<sup>138</sup> is replaced, intriguingly, by a Sua5-typical proline. These TsaC proteins lack the Sua5 specific motifs important for the interdomain interaction (Supplementary Figure 7) suggesting that these are genuine exceptions.

Together, these data confirm that the highly conserved Thr<sup>138</sup>/Ser<sup>138</sup> and the corresponding Pro<sup>143</sup> accurately identify TsaC and Sua5 proteins, respectively, in the vast majority of cases.

## Atypical Sua5 proteins of Archaeoglobi archaea carry functionally relevant deletions in the SUA5 domain

The identification of TsaC and Sua5 specific residues allowed us to spot atypical sequences such as those of Pro<sup>143</sup>-containing bacterial TsaC proteins. In addition to these outliers, our attention was drawn to the Sua5 sequence from the archaeon *Archaeoglobus profundus* (*Ap*-Sua5) that carried Thr<sup>138</sup> instead of Pro<sup>143</sup> signature residue and was significantly shorter (289 AA) than a classical Sua5 protein (average size 331 AA). Extending the analysis to six other isolated Archaeoglobi species showed that, in addition to *Ap*-Sua5, the Sua5 proteins from *A. veneficus* (*Av*-Sua5) and *A. fulgidus* (*Af*-Sua5) carry deletions in the interdomain loop and in the SUA5 domain (Supplementary Figure 8). Around 50% of the interdomain loop residues are missing in these



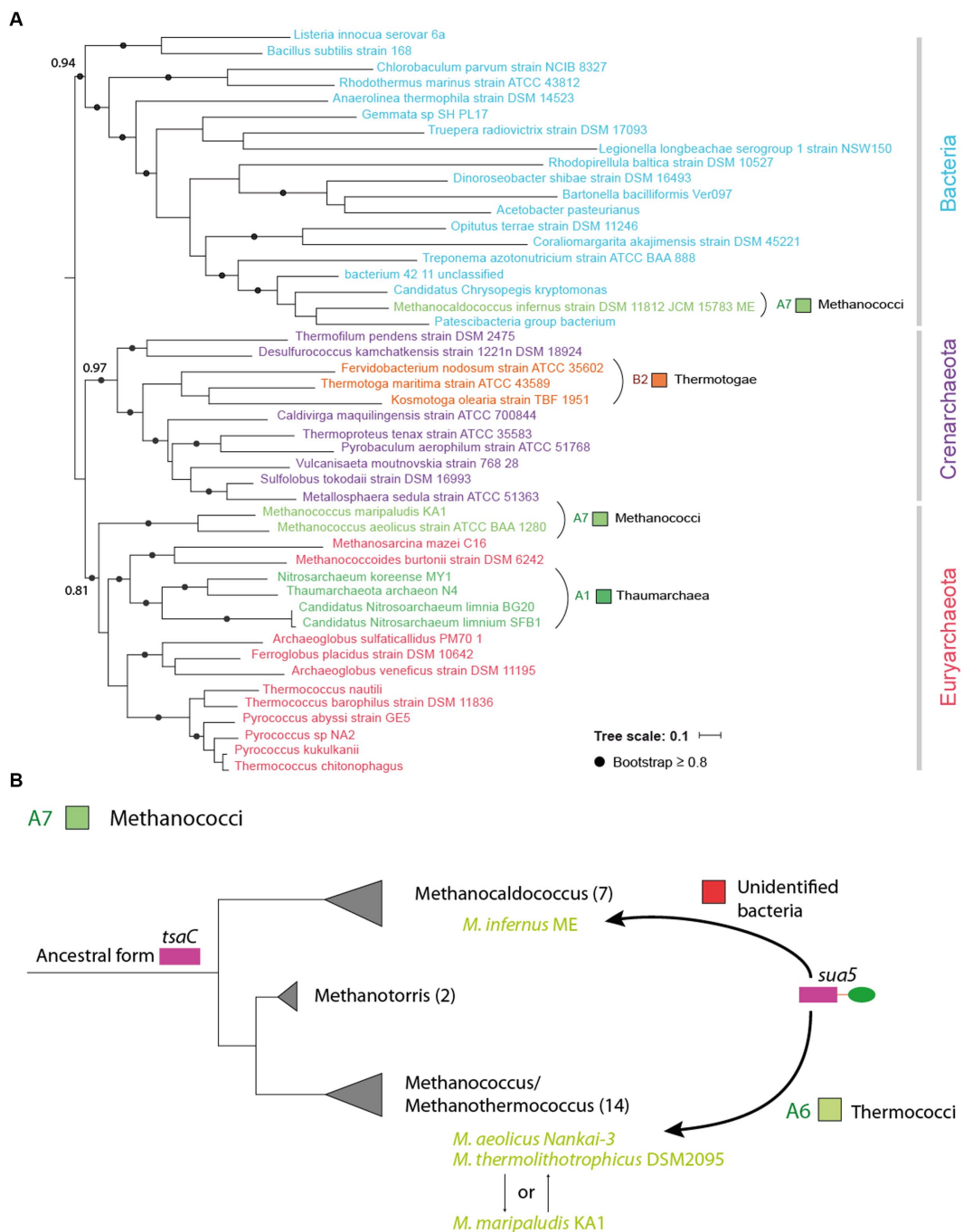


FIGURE 4

Horizontal transfer of *sua5* genes occurs across different phylogenetic distances. (A) Maximum likelihood phylogenetic tree of *Sua5* sequences from a wide range of bacterial taxa and from representative sequences of Crenarchaeota and Euryarchaeota. The tree was arbitrarily rooted between Bacteria and Archaea. The *Sua5* sequences from Methanococci, Thaumarchaea and Thermotogae are indicated in colors corresponding to their taxonomic group (see Supplementary Table 1). The scale bar corresponds to the number of substitutions per amino acid. (B) Evolutionary scenario for the acquisition of *sua5* genes in Methanococci archaea. The established phylogeny of the four main genera of Methanococci is shown. The number of species in each genus is indicated in brackets. All species carry *TsaC* orthologs except for the four species (indicated in green color) that carry *Sua5* orthologs. This suggests that the common ancestor of Methanococci was a *TsaC*-user and that the four species acquired the *sua5* gene by HGT. The putative donor of these *sua5* genes is indicated above the arrows.

proteins and among those the highly conserved and functionally important motif Pro<sup>228</sup>-Gly<sup>229</sup>-Met<sup>230</sup>. Moreover, when comparing with the archetypal *Sua5* from *A. sulfaticallidus* these three proteins lack 12%, (*Av-Sua5*), 13% (*Af-Sua5*), and 36% (*Ap-Sua5*) of residues in the SUA5 domain.

To get insight into the history of the acquisition of the detected mutations and their potential consequences for the function of these atypical *Sua5* proteins, we performed phylogenetic analyses and AlphaFold2 modelling. The sequences that accumulated the highest number of mutations clustered within a single clade (Figure 7A). Of note,

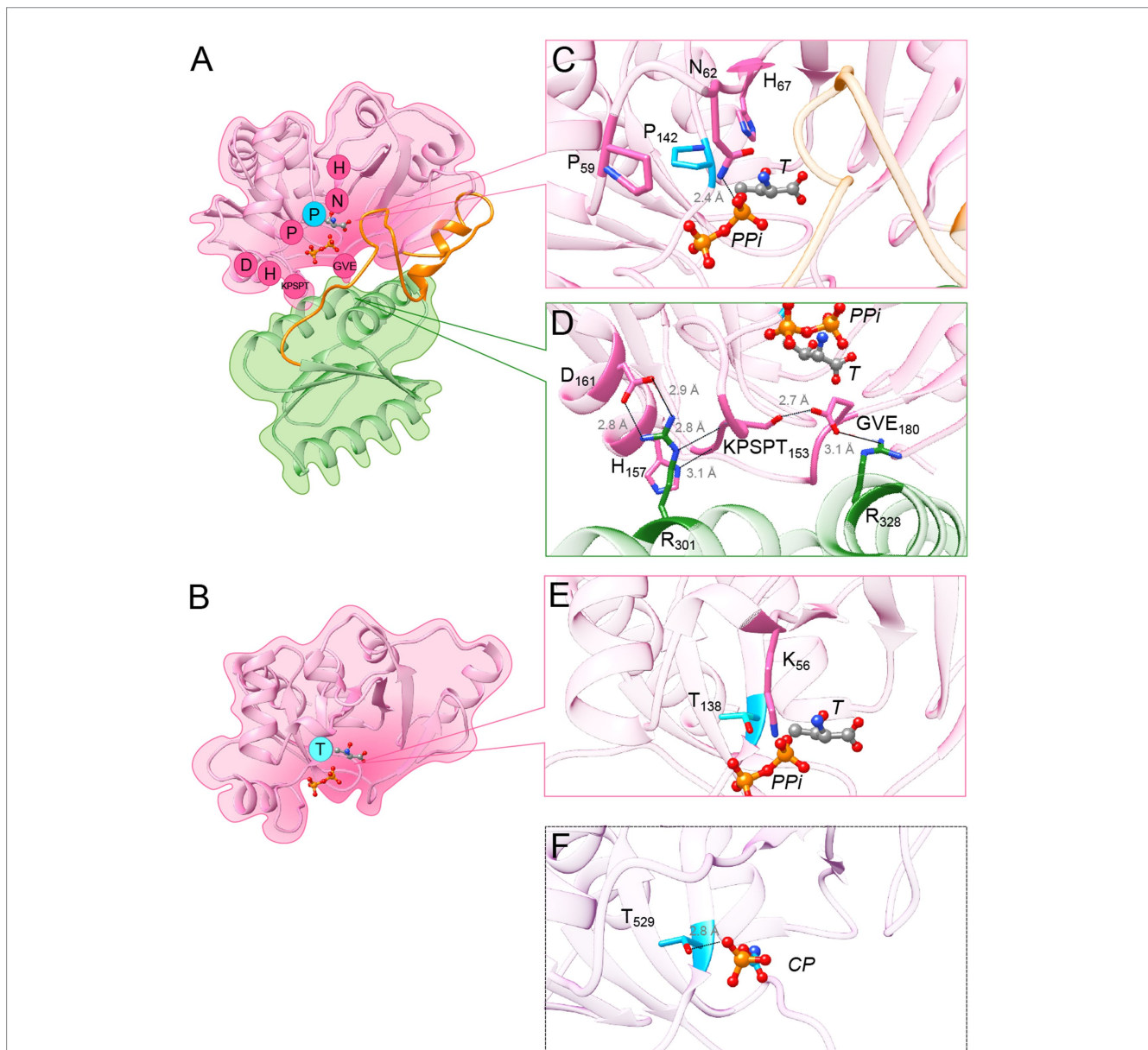


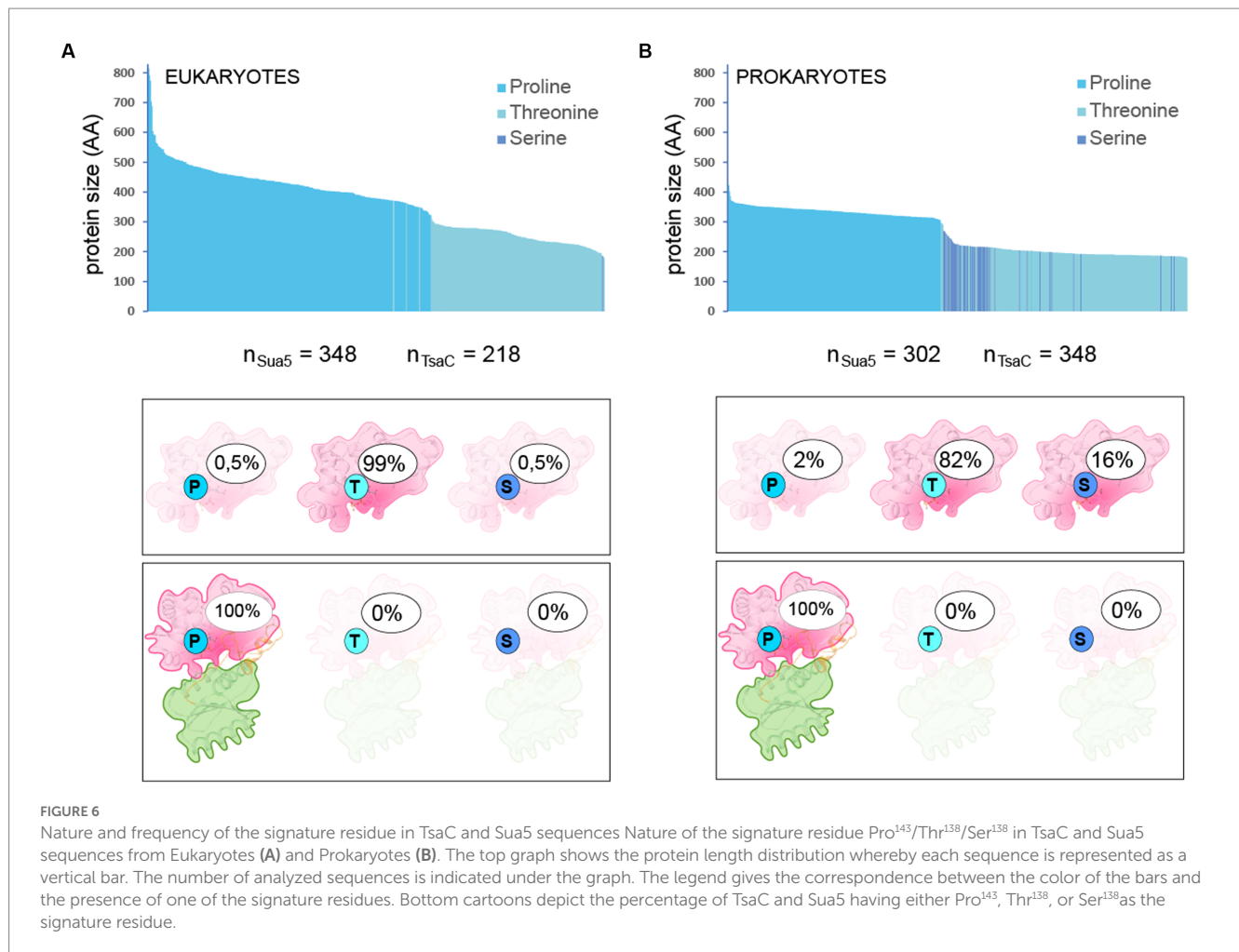
FIGURE 5

Variant-specific conserved residues interact with substrates and stabilize interdomain interface. **(A)** Cartoon showing the structure of Pa-Sua5 in complex with threonine and PPI shown as stick presentation. The TsaC-like domain is in pink, SUA5 domain is in green and the flexible loop is in orange. The Sua5-specific conserved residues are indicated with one letter code in pink circles or in cyan circle for the signature residue. **(B)** Cartoon showing the crystal structure of Ec-TsaC. Threonine and PPI molecules were modelled in the active site by superposing the Ec-TsaC with Pa-Sua5. The signature residue Thr is indicated with one code letter in a cyan-colored circle. **(C)** Zoom in the active site of Pa-Sua5. Sua5-specific conserved residues Pro<sup>59</sup>, Asn<sup>62</sup>, His<sup>67</sup>, and Pro<sup>142</sup> are highlighted. Asn<sup>62</sup> forms a hydrogen bond (indicated as dotted line) with the PPI molecule. **(D)** Zoom at the interface between the TsaC-like and SUA5 domains. The KPSPT motif and the GVE motif are shown as well as the network of H-bonds formed between these residues. Two salt bridges formed by Asp<sup>161</sup> and Glu<sup>180</sup> with Arg<sup>301</sup> and Arg<sup>328</sup>, respectively, likely stabilize the two domains. **(E)** Zoom in the active site cavity of Ec-TsaC. Threonine and PPI molecules were modelled in the active site by superposing the Ec-TsaC and Pa-Sua5 that co-crystallized with these ligands. Side chain of the signature residue Thr<sup>138</sup> is shown as stick model. **(F)** Zoom in the active site of TsaC-like domain of TobZ protein from *Streptoalloteichus tenebrarius* bound to its substrate carbamoyl-phosphate (CP). The side chain of Thr<sup>529</sup> (the equivalent of Thr<sup>138</sup> in Ec-TsaC) is indicated as stick model. The H-bond between Thr<sup>529</sup> and the phosphate moiety is indicated by a dotted line.

they all carry a deletion in the interdomain loop leading to the loss of the PGM motif and, concomitantly, the substitution of Asn<sup>62</sup> to Lys and His<sup>67</sup> to Gly/Val, the residues that we have tentatively identified as being involved in interaction with substrates (Figures 5C, 7A). We next modelled the structure of *Ap*-Sua5 and compared it to the crystal structure of *Pa*-Sua5. This showed that the TsaC-like domain adopts a typical fold whereby Lys and Val residues superpose well with Asn<sup>62</sup> and His<sup>67</sup>, respectively, suggesting that these mutations affect the binding of

substrates (Figure 7B). Despite a significant shortening, the SUA5 domain of *Ap*-Sua5 was modelled as a SUA5-like globular fold with the two arginines (Arg<sup>257</sup> and Arg<sup>279</sup> in *Ap*-Sua5) being correctly positioned to build salt bridges with the catalytic domain (Figure 7B). However, the interdomain loop seem to adopt a much more linear conformation leaving the entrance to the catalytic cavity open to the solvent (Figure 7C).

Together, the data suggest a progressive accumulation of mutations in interdomain loop and SUA5 domain of Archaeoglobi



Sua5 with *Ap*-Sua5 being the extreme case. Consequently, these atypical Sua5 proteins seem to have acquired compensatory mutations in the catalytic domain that probably affect the binding to substrates and/or TC-AMP molecules.

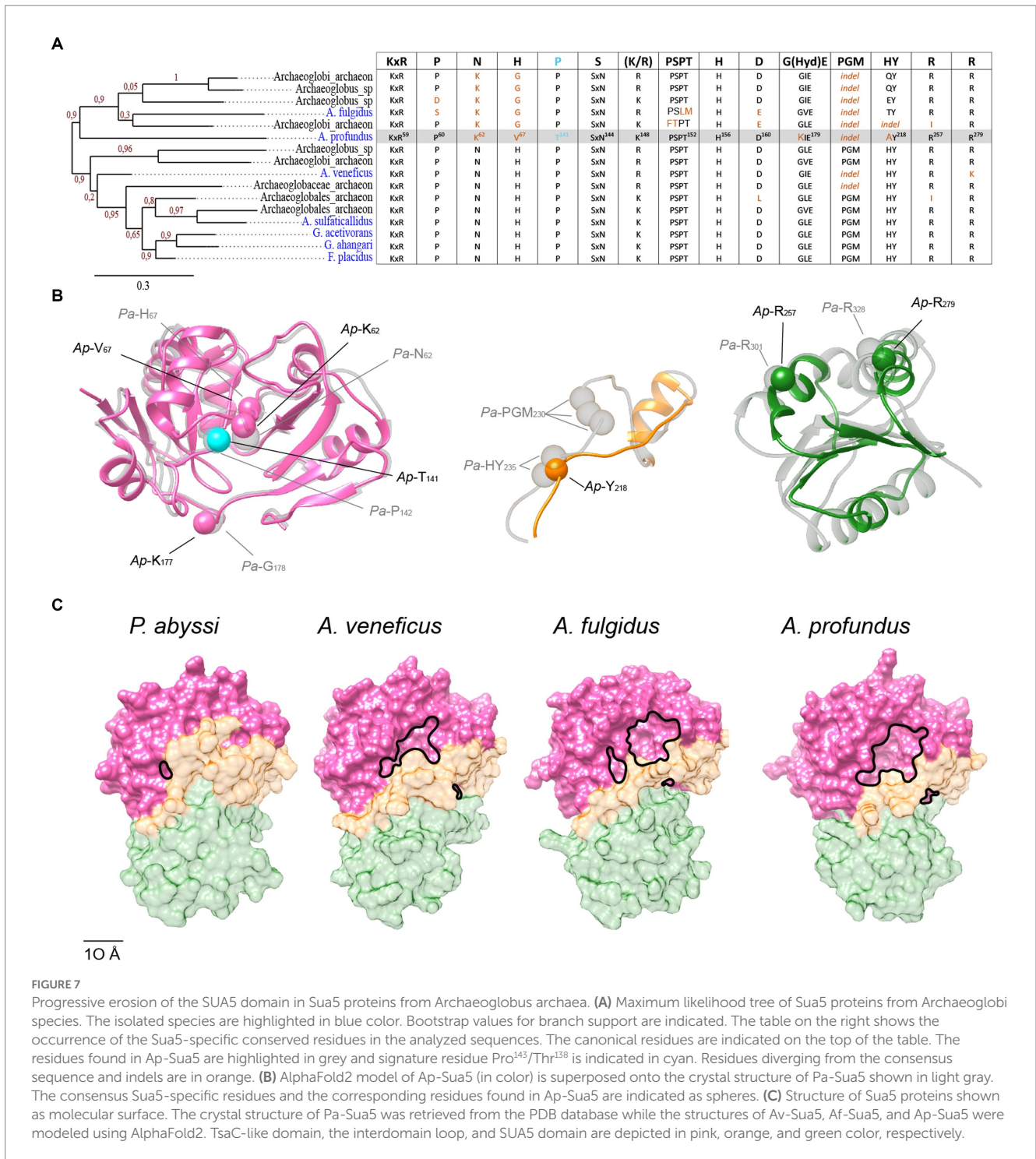
## Discussion

In the present work, we delineated the evolutionary pathway that led to the existence of two partially homologous proteins performing the same essential function in all extant organisms. We show that the evolution resulted in two different substrate binding modalities with potential consequences for the catalytic mechanism and/or efficiency of Sua5 and TsaC enzymes.

Early comparative genomics analyses identified TsaC/Sua5 family of proteins as part of a small set of about 60 truly ubiquitous proteins that were probably present in the Last Universal Common Ancestor (LUCA) of all extant organisms (Galperin, 2004). Our up to date distribution analysis shows that this is still true. The only fully sequenced organisms lacking *tsaC* or *sua5* genes are obligate ectosymbionts such as the DPANN archaea *N. equitans* and *N. aerobiophila*. These organisms carry greatly reduced genomes (0.49 Mbp and 0.67 Mbp, respectively) and depend on their hosts for the uptake of essential metabolites such as nucleotides, lipids and ATP (Huber et al., 2002; Waters et al., 2003; Kato et al., 2022). This suggests

that the absence of *tsaC* and *sua5* genes in a genome could be a marker of symbiotic or parasitic lifestyle. It remains to be investigated whether these organisms adapted to translate their genetic information faithfully in the absence of t<sup>6</sup>A-modified tRNA or if this function is supplied by the host. The latter may be the case for *N. equitans* which encodes the KEOPS complex (Waters et al., 2003) suggesting that it synthesizes t<sup>6</sup>A-tRNA but needs the host to supply TC-AMP or the TsaC/Sua5 enzyme by some yet unknown mechanism. The transport of TC-AMP from the host *Acidianus hospitalis* to *N. equitans* seems unlikely since this molecule is highly unstable at 90°C, the optimal growth temperatures of these organisms (Lauhon, 2012). The export of Sua5 enzyme from the cytoplasm to mitochondria via a signal sequence has, however, been reported, in *S. cerevisiae* (Thiaville et al., 2014a) suggesting that uptake of Sua5/TsaC by *N. equitans* could be a possible mechanism.

Our data further show that the concomitant presence of the *tsaC* and *sua5* genes in a genome is rare in accordance with previous work reporting the co-occurrence of these genes only in *Acetobacterium woodii* and *Vibrio cholerae* strains among 9,200 analyzed genomes (Thiaville et al., 2014b). In the additional cases we detected, we found that the second gene was always of exogenous origin suggesting that most of the co-occurrence cases are explained by recent horizontal gene transfers (HGT). These transfers can occur even over highest phylogenetic distances, between two different domains, suggesting that TsaC/Sua5 proteins are not dependent upon a particular cellular context and/or partners to function. This is further corroborated by



**FIGURE 7**  
 Progressive erosion of the SUA5 domain in Sua5 proteins from Archaeoglobli archaea. **(A)** Maximum likelihood tree of Sua5 proteins from Archaeoglobli species. The isolated species are highlighted in blue color. Bootstrap values for branch support are indicated. The table on the right shows the occurrence of the Sua5-specific conserved residues in the analyzed sequences. The canonical residues are indicated on the top of the table. The residues found in Ap-Sua5 are highlighted in grey and signature residue Pro<sup>143</sup>/Thr<sup>139</sup> is indicated in cyan. Residues diverging from the consensus sequence and indels are in orange. **(B)** AlphaFold2 model of Ap-Sua5 (in color) is superposed onto the crystal structure of Pa-Sua5 shown in light gray. The consensus Sua5-specific residues and the corresponding residues found in Ap-Sua5 are indicated as spheres. **(C)** Structure of Sua5 proteins shown as molecular surface. The crystal structure of Pa-Sua5 was retrieved from the PDB database while the structures of Av-Sua5, Af-Sua5, and Ap-Sua5 were modeled using AlphaFold2. Tsac-like domain, the interdomain loop, and SUA5 domain are depicted in pink, orange, and green color, respectively.

functional complementation experiments showing that *tsaC* and *sua5* genes originating from very distantly related organisms (yeast, bacteria and archaea) are all interchangeable *in vitro* and *in vivo* (El Yacoubi et al., 2009; Perrochia et al., 2013a). Such “independency” would probably facilitate horizontal gene transfers of *sua5* and *tsaC* genes and this may explain, at least in part, the patchy distribution of these genes across the universal tree.

Present day distribution of the Tsac/Sua5 family is difficult to reconcile with a simple evolutionary scenario. The bipartite tree topology of this family resembles that of translation elongation factors

EF-Tu and EF-G which are paralogs that duplicated before the divergence of all extant organismal lineages (Iwabe et al., 1989; Baldauf et al., 1996). Therefore, the phylogeny indicates that Tsac/Sua5 family emerged in a pre-LUCA ancestor and that at the time of LUCA both *tsaC* and *sua5* existed. However, several observations we made contradict this scenario. First, Tsac-like domain of Sua5 proteins contains 15 conserved residues (10% of the alignment positions used for tree construction) that are variable in the Tsac proteins. We therefore cannot exclude the possibility that the bipartite tree topology we observe is artificial and simply reflects the fact that

TsaC and TsaC-like sequences are “mechanically” segregated. Second, we identified several clear cases of HGT of *sua5* or *tsaC* genes whereby the endogenous gene was lost suggesting that the concomitant presence of the two genes carrying out the same function is not advantageous for an organism. The presence in many bacterial lineages of an inactive paralog of TsaC called YciO (Thiaville et al., 2014b) further supports this idea. Finally, the concomitant presence of both enzymes in LUCA would imply numerous independent gene loss events throughout the tree of life in order to explain the presence of only one of the two genes in extant organisms. In the light of these observations we therefore favour a more parsimonious scenario whereby LUCA encoded one of the two enzymes.

Which of the two variants is then the version that was present in LUCA? One possibility is the emergence of *tsaC* first and *sua5* post-LUCA by acquisition of an additional SUA5 domain. Domain fusion is a frequent method for new proteins to arise (Bornberg-Bauer et al., 2010), with evolution famously being described as a tinkerer rather than an inventor (Jacob, 1977). The SUA5 domain contains an atypical Rossmann fold which is an ancient and widely distributed protein fold (Ma et al., 2008) and therefore we cannot exclude a reassignment of this domain into a SUA5 domain. Of note, the presence of highly conserved motifs at the interface between the two domains supports the idea that all current Sua5 proteins originated from one common ancestral protein suggesting that the SUA5 domain acquisition, if it occurred, was a unique event.

The emergence of *tsaC* from *sua5* could have happened through the introduction of a premature stop codon. However, such drastic event may disrupt stability and/or activity of a protein (Weiner et al., 2006) and we observed this for truncated version of *Pa*-Sua5 whereby the whole SUA5 domain was removed (Pichard-Kostuch et al., 2018). Rather, a progressive accumulation of deletions in SUA5 domain such as observed in atypical Archaeoglobi Sua5 could be a mechanism that initially led to emergence of TsaC. This process would need to be accompanied by mutations compensating for the loss of the loop and the SUA5 domain. Archaeoglobi Sua5 indeed presents non-neutral mutations of highly conserved Sua5-specific residues (Asn<sup>62</sup>/Lys<sup>62</sup> and His<sup>67</sup>/Gly or Val<sup>67</sup>). Given their position in the active site, we speculate that these mutations affect the substrates and/or product binding.

The gene erosion to yield shorter TsaC variant may have occurred several times in the course of evolution and each time a different compensatory set of mutations could have arisen thus explaining the absence of highly conserved TsaC-specific residues. This makes the TsaC-specific and conserved Thr/Ser<sup>138</sup>, the sole exception to this rule, particularly interesting as it suggests that this residue is a key prerequisite for becoming the shorter version of the enzyme. The comparative structural analysis performed here suggests that Thr/Ser<sup>138</sup> contacts the alpha phosphate moiety of ATP/TC-AMP. Intriguingly, the corresponding Sua5 residue Pro143 would not be able to establish such a contact. However, the highly conserved histidine in the HY motif found in the interdomain loop was shown to be important for the activity of *Pa*-Sua5, likely by being involved in the binding of PPI/ATP (Pichard-Kostuch et al., 2018). Notably, none among the thousands of Sua5/TsaC sequences we screened in our analysis displayed both of these residues. It is therefore tempting to suggest that simultaneous presence of Thr/Ser<sup>138</sup> and HY motifs, both of which may interact with PPI/ATP, could have a negative impact on catalysis. Consequently, this combination was counterselected, resulting in the two variant-specific signature residues. Some evidence that this may be a plausible scenario comes from the atypical Sua5 protein of

*Archaeoglobus profundus*, the most advanced case of SUA5 domain erosion, where the Pro<sup>143</sup> signature residue was replaced by a threonine.

The Sua5 protein of *A. profundus* would be a good candidate to test whether a Sua5 protein could be evolved to become TsaC. Using directed mutagenesis it could be possible to generate even shorter active variants and ultimately remove the SUA5 domain completely. If our hypothesis is true than the complete loss of the SUA5 domain would only be possible in combination with further mutations of substrate binding residues to their TsaC-like counterparts. As a further line of experimental studies, it would be interesting both from the evolutionary and mechanistic standpoint, to compare the specific activities of TsaC and Sua5 proteins from closely related organisms.

In conclusion, we suggest that Sua5 was the ancestral version of the extant TC-AMP synthetic enzyme family. By articulating our previous experimental observations (Pichard-Kostuch et al., 2018) with our new in-depth sequence-structure analysis of this family, we suggest that the SUA5 domain is still essential for the activity of Sua5 variant by *inter alia* ensuring the right positioning of the catalytically important linker. However, key mutations in the TsaC-like domain can reduce this dependency and lead to the emergence of TsaC from Sua5 through SUA5 domain loss. Although it remains to be tested how these mutations affect the activity, this scenario combined with HGT events would account for the present day broad but inconsistent distribution of *tsaC* and *sua5* genes among different lineages. We suggest that SUA5 domain loss occurred at several independent occasions and that the resulting TsaC proteins adapted by evolving different interfaces for binding to substrates and/or products. Thus, TsaC proteins could be a more “advanced” version of the TC-AMP producing enzyme suggesting that, to quote Leonardo da Vinci, “simplicity is the ultimate sophistication” when it comes to the evolution of Sua5/TsaC family.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://figshare.com/>, doi: 10.6084/m9.figshare.22283929.

## Author contributions

APK, VDC and TBL conceived the study and all authors analysed the data. APK, VDC and TBL prepared the figures, wrote the draft and finalized the manuscript. All authors proofread the manuscript and approved the submitted version.

## Funding

This work was funded by Agence Nationale de la Recherche, grant number ANR-18-CE11-0018 to TB.

## Acknowledgments

The authors wish to thank Ryan Catchpole and Patrick Forterre for stimulating discussions during the early stages of this work.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1204045/full#supplementary-material>

## References

- Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., et al. (2008). X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins Struct. Funct. Bioinform.* 70, 1108–1111. doi: 10.1002/prot.21794
- Arrondel, C., Missoury, S., Snoek, R., Patat, J., Menara, G., Collinet, B., et al. (2019). Defects in t(6)a tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome. *Nat. Commun.* 10:3967. doi: 10.1038/s41467-019-11951-x
- Baldauf, S. L., Palmer, J. D., and Doolittle, W. F. (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7749–7754. doi: 10.1073/pnas.93.15.7749
- Beenstock, J., and Sicheri, F. (2021). The structural and functional workings of KEOPS. *Nucleic Acids Res.* 49, 10818–10834. doi: 10.1093/nar/gkab865
- Bornberg-Bauer, E., Huylmans, A.-K., and Sikosek, T. (2010). How do new proteins arise? *Curr. Opin. Struct. Biol.* 20, 390–396. doi: 10.1016/j.sbi.2010.02.005
- Braun, D. A., Rao, J., Mollet, G., Schapiro, D., Daugeron, M.-C., Tan, W., et al. (2017). Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat. Genet.* 49, 1529–1538. doi: 10.1038/ng.3933
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., et al. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028
- Criscuolo, A., and Gribaldo, S. (2010). BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210
- Da Cunha, V., Gaia, M., Gabelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13:e1006810. doi: 10.1371/journal.pgen.1006810
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). [Phylogeny.fr](http://Phylogeny.fr): robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180
- Edvardson, S., Prunetti, L., Arraf, A., Haas, D., Bacusmo, J. M., Hu, J. F., et al. (2017). tRNA N6-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3 (OSGEP) mutation manifest by neurodegeneration and renal tubulopathy. *Eur. J. Hum. Genet.* 25, 545–551. doi: 10.1038/ejhg.2017.30
- EL Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., et al. (2009). The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Res.* 37, 2894–2909. doi: 10.1093/nar/gkp152
- El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., et al. (2011). A role for the universal Kae1/Qri7/YgdJ (COG0533) family in tRNA modification: t<sup>6</sup> a biosynthesis. *EMBO J.* 30, 882–893. doi: 10.1038/emboj.2010.363
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Galperin, M. Y. (2004). “Conserved hypothetical” proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–5463. doi: 10.1093/nar/gkh885
- Harris, K. A., Bobay, B. G., Sarachan, K. L., Sims, A. F., Billible, Y., Deutsch, C., et al. (2015). NMR-based structural analysis of Threonylcarbamoyl-AMP synthase and its substrate interactions. *J. Biol. Chem.* 290, 20032–20043. doi: 10.1074/jbc.M114.631242
- Högenauer, G., Turnowsky, F., and Unger, F. M. (1972). Codon-anticodon interaction of methionine specific tRNAs. *Biochem. Biophys. Res. Commun.* 46, 2100–2106. doi: 10.1016/0006-291X(72)90765-6
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67. doi: 10.1038/417063a
- Ishikura, H., Yamada, Y., Murao, K., Saneyoshi, M., and Nishimura, S. (1969). The presence of N-[9-(β-D-ribofuranosyl)purin-6-ylcarbamoyl]threonine in serine, methionine and lysine transfer RNAs from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 37, 990–995. doi: 10.1016/0006-291X(69)90229-0
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9355–9359. doi: 10.1073/pnas.86.23.9355
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kato, S., Ogasawara, A., Itoh, T., Sakai, H. D., Shimizu, M., Yuki, M., et al. (2022). Nanobdella aerobiophila gen. Nov., sp. nov., a thermoacidophilic, obligate ectosymbiotic archaeon, and proposal of Nanobdellaceae fam. Nov., Nanobdellales Ord. nov. and Nanobdellia class. Nov. *Int. J. Syst. Evol. Microbiol.* 72, 005489. doi: 10.1099/ijsem.0.005489
- Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136. doi: 10.1038/nrmicro751
- Kuratani, M., Kasai, T., Akasaka, R., Higashijima, K., Terada, T., Kigawa, T., et al. (2011). Crystal structure of *Sulfolobus tokodaii* Sua5 complexed with L-threonine and AMPNP. *Proteins* 79, 2065–2075. doi: 10.1002/prot.23026
- Laumon, C. T. (2012). Mechanism of N6-Threonylcarbamoyladenine (t<sup>6</sup> a) biosynthesis: isolation and characterization of the intermediate Threonylcarbamoyl-AMP. *Biochemistry* 51, 8950–8963. doi: 10.1021/bi301233d
- Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-0
- Llácer, J. L., Hussain, T., Saini, A. K., Nanda, J. S., Kaur, S., Gordiyenko, Y., et al. (2018). Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *elife* 7:e39273. doi: 10.7554/eLife.39273
- Ma, B.-G., Chen, L., Ji, H.-F., Chen, Z.-H., Yang, F.-R., Wang, L., et al. (2008). Characters of very ancient proteins. *Biochem. Biophys. Res. Commun.* 366, 607–611. doi: 10.1016/j.bbrc.2007.12.014
- Machnicka, M. A., Olchowik, A., Grosjean, H., and Bujnicki, J. M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* 11, 1619–1629. doi: 10.4161/15476286.2014.992273
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Moody, E. R., Mahendrarajah, T. A., Dombrowski, N., Clark, J. W., Petitjean, C., Offre, P., et al. (2022). An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *elife* 11:e66695. doi: 10.7554/eLife.66695
- Murphy, F. V., Ramakrishnan, V., Malkiewicz, A., and Agris, P. F. (2004). The role of modifications in codon discrimination by tRNA<sup>Lys</sup>UUU. *Nat. Struct. Mol. Biol.* 11, 1186–1191. doi: 10.1038/nsmb861
- Naor, A., Thiaville, P. C., Altman-Price, N., Cohen-Or, I., Allers, T., de Crécy-Lagard, V., et al. (2012). A genetic investigation of the KEOPS complex in Halophilic Archaea. *PLoS One* 7:e43013. doi: 10.1371/journal.pone.0043013
- Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., et al. (1994). Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. *J. Mol. Biol.* 236, 710–724. doi: 10.1006/jmbi.1994.1184
- Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Bräuer, U., Fandrich, U., et al. (2012). The O-Carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem. Int. Ed.* 51, 4046–4052. doi: 10.1002/anie.201108896
- Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., et al. (2013a). In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.* 41, 1953–1964. doi: 10.1093/nar/gks1287

- Perrochia, L., Guetta, D., Hecker, A., Forterre, P., and Basta, T. (2013b). Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t(6)a tRNA modification. *Nucleic Acids Res.* 41, 9484–9499. doi: 10.1093/nar/gkt720
- Petkun, S., Shi, R., Li, Y., Asinas, A., Munger, C., Zhang, L., et al. (2011). Structure of Hydrogenase maturation protein HypF with reaction intermediates shows two active sites. *Structure* 19, 1773–1783. doi: 10.1016/j.str.2011.09.023
- Petersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943
- Phelps, S. S., Malkiewicz, A., Agris, P. F., and Joseph, S. (2004). Modified nucleotides in tRNA<sup>Lys</sup> and tRNA<sup>Val</sup> are important for translocation. *J. Mol. Biol.* 338, 439–444. doi: 10.1016/j.jmb.2004.02.070
- Pichard-Kostuch, A., Zhang, W., Liger, D., Dageron, M.-C., Létoquart, J., Li de la Sierra-Gallay, I., et al. (2018). Structure–function analysis of Sua5 protein reveals novel functional motifs required for the biosynthesis of the universal t6A tRNA modification. *RNA* 24, 926–938. doi: 10.1261/rna.0666092.118
- Powers, D. M., and Peterkofsky, A. (1972). The presence of N-(Purin-6-ylcarbamoyl) threonine in transfer ribonucleic acid species whose codons begin with adenine. *J. Biol. Chem.* 247, 6394–6401. doi: 10.1016/S0021-9258(19)44706-6
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Stuart, J. W., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., et al. (2000). Functional anticodon architecture of human tRNA<sup>Lys3</sup> includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t<sup>6</sup>a<sup>1</sup>. *Biochemistry* 39, 13396–13404. doi: 10.1021/bi0013039
- Teplova, M., Tereshko, V., Egli, M., Sanishvili, R., Joachimiak, A., Bushueva, T., et al. (2000). The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. *Protein Sci.* 9, 2557–2566. doi: 10.1110/ps.9.12.2557
- Thiaville, P. C., el Yacoubi, B., Perrochia, L., Hecker, A., Prigent, M., Thiaville, J. J., et al. (2014a). Cross kingdom functional conservation of the Core universally conserved Threonylcarbamoyladenine tRNA synthesis enzymes. *Eukaryot. Cell* 13, 1222–1231. doi: 10.1128/EC.00147-14
- Thiaville, P. C., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2014b). Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t(6)a), a universal modification of tRNA. *RNA Biol.* 11, 1529–1539. doi: 10.4161/15476286.2014.992277
- Thiaville, P. C., Legendre, R., Rojas-Benítez, D., Baudin-Baillieu, A., Hatin, I., Chalancon, G., et al. (2016). Global translational impacts of the loss of the tRNA modification t(6)a in yeast. *Microb Cell* 3, 29–45. doi: 10.15698/mic2016.01.473
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256
- Väre, V. Y. P., Erusyal, E. R., Narendran, A., Sarachan, K. L., and Agris, P. F. (2017). Chemical and conformational diversity of modified nucleosides affects tRNA structure and function. *Biomol. Ther.* 7:29. doi: 10.3390/biom7010029
- Wan, L. C. K., Mao, D. Y. L., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., et al. (2013). Reconstitution and characterization of eukaryotic N6-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Res.* 41, 6332–6346. doi: 10.1093/nar/gkt322
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12984–12988. doi: 10.1073/pnas.1735403100
- Weiner, J., Beaussart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. *FEBS J.* 273, 2037–2047. doi: 10.1111/j.1742-4658.2006.05220.x
- Weissenbach, J., and Grosjean, H. (1981). Effect of Threonylcarbamoyl modification (t6A) in yeast tRNA<sup>ArgIII</sup> on codon-anticodon and anticodon-anticodon interactions. A thermodynamic and kinetic evaluation. *Eur. J. Biochem.* 116, 207–213. doi: 10.1111/j.1432-1033.1981.tb05320.x
- Yamada, K. D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251. doi: 10.1093/bioinformatics/btw412