A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification

Basma El Yacoubi^{1,*}, Isabelle Hatin²,
Christopher Deutsch³, Tamer Kahveci⁴,
Jean-Pierre Rousset², Dirk Iwata-Reuyl³,
Alexey G Murzin⁵ and Valérie
de Crécy-Lagard^{1,*}

¹Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA, ²University Paris-Sud and IGM, CNRS, UMR 8621, Orsay, France, ³Department of Chemistry, Portland State University, Portland, OR, USA, ⁴Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA and ⁵MRC Laboratory of Molecular Biology, Cambridge, UK

The YgjD/Kae1 family (COG0533) has been on the top-10 list of universally conserved proteins of unknown function for over 5 years. It has been linked to DNA maintenance in bacteria and mitochondria and transcription regulation and telomere homeostasis in eukaryotes, but its actual function has never been found. Based on a comparative genomic and structural analysis, we predicted this family was involved in the biosynthesis of N⁶-threonylcarbamoyl adenosine, a universal modification found at position 37 of tRNAs decoding ANN codons. This was confirmed as a yeast mutant lacking Kae1 is devoid of t⁶A. t⁶A[−] strains were also used to reveal that t⁶A has a critical role in initiation codon restriction to AUG and in restricting frameshifting at tandem ANN codons. We also showed that YaeZ, a YgjD paralog, is required for YgjD function *in vivo* in bacteria. This work lays the foundation for understanding the pleiotropic role of this universal protein family.

The EMBO Journal (2011) 30, 882–893. doi:10.1038/emboj.2010.363; Published online 1 February 2011

Subject Categories: RNA; proteins

Keywords: COG0533; comparative genomics; essential genes; tRNA modification

Introduction

Deciphering both the cellular and molecular roles of universally conserved proteins of unknown functions presents a unique challenge. Of the top-10 universal protein families of unknown function awaiting characterization listed by Galperin and Koonin (2004), only half have been assigned a cellular function to date (Galperin and Koonin, 2010). We have previously shown that one of those 10, the Sua5/YrdC family (YrdC), is involved in the biosynthesis of N⁶-threonylcarbamoyl adenosine (t⁶A) (El Yacoubi *et al*, 2009), a

universal modification found at position 37 of tRNAs decoding ANN codons. However, the t⁶A biosynthetic pathway has yet to be fully characterized. Early studies suggested that ATP, threonine and carbonate are required (Elkins and Keller, 1974); however, YrdC does not appear to bind threonine (El Yacoubi *et al*, 2009). Biochemical experiments including unsuccessful attempts to reconstitute the pathway *in vitro* strongly suggested that YrdC is not the sole enzyme required for t⁶A biosynthesis (El Yacoubi *et al*, 2009).

The YgjD/Kae1/Qri7 protein family (COG0533) is another member of the top-10 list of universally conserved enzymes of unknown function (Galperin and Koonin, 2004). Uncovering the function of this protein family has been used to illustrate the difficult path of ‘converting data into knowledge and knowledge into understanding’ (quote of Sydney Brenner in Galperin and Koonin (2010)). Indeed, the first attempt to functionally characterize a member of this family occurred 20 years ago when the bacterial member of the COG0533 was first annotated as Gcp for *O*-sialoglycoprotein endopeptidase (Abdullah *et al*, 1990, 1991), but this activity was never confirmed (Hecker *et al*, 2007). Since then members of this family have been studied in yeast, archaea and bacteria resulting in sometimes conflicting data, several proposed functions and annotations but no definitive characterization (Downey *et al*, 2006; Gavin *et al*, 2006; Kisseleva-Romanova *et al*, 2006; Hecker *et al*, 2007; Handford *et al*, 2009; Oberto *et al*, 2009). In bacteria, the *ygiD* gene appears essential (Arigoni *et al*, 1998) with conditional depletion leading to pleiotropic phenotypes including increased or reduced cell size (depending on the study), unusual distribution of DNA, nucleoid loss, and/or membrane/cell envelope defects (Handford *et al*, 2009; Oberto *et al*, 2009). In *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, loss of mitochondrial DNA integrity and aberrant mitochondrial morphology were observed under conditions where the YgjD mitochondrial targeted homologue Qri7 was depleted (Oberto *et al*, 2009). The *Escherichia coli* YgjD was found to form a complex with YeaZ (Handford *et al*, 2009), another essential protein. YeaZ and YgjD share 29% identity within their first 100 amino acids but unlike YgjD, YeaZ is only found in eubacteria (Hecker *et al*, 2007). In yeast, the cytoplasmic YgjD homologue Kae1 is not essential but *kae1* mutants display shortened telomeres and severe growth impairment phenotypes (Downey *et al*, 2006; Kisseleva-Romanova *et al*, 2006; Hecker *et al*, 2008). Intriguingly, Kae1 belongs to the EKC/KEOPS complex involved in telomere maintenance (Downey *et al*, 2006) and transcription of galactose- and pheromone-responsive genes (Kisseleva-Romanova *et al*, 2006). In the Archaea *Pyrococcus abyssi*, the COG0533 homologue (PAB1159) was suggested to be an atypical DNA-binding protein with apurinic-endonuclease activity *in vitro* (Hecker *et al*, 2007). Interestingly, in several Archaea, the Kae1 homologue is fused to an atypical small RIO-type kinase (Lopreiato *et al*, 2004) homologous to the yeast protein Bud32p (PRPK for p53-related protein kinase in

*Corresponding authors. B El Yacoubi or V de Crécy-Lagard, Department of Microbiology and Cell Science, University of Florida, PO Box 110700, Gainesville, FL, USA. Tel.: +1 352 262 1672; Fax: +1 352 392 5922; E-mail: basma@ufl.edu or Tel.: +1 352 392 9416; Fax: +1 352 392 5922; E-mail: vcrecy@ufl.edu

Received: 1 September 2010; accepted: 21 December 2010; published online: 1 February 2011

humans), neither of whose functions are known and for which there do not appear to be any bacterial counterpart. Despite these reports and findings, the actual function of the YgjD/Kae1/Qri7 family of proteins has yet to be described.

Using a combination of comparative genomic and experimental methods, we show here that the Kae1/YgjD/Qri7 family is (like the Sua5/YrdC family), directly involved in t⁶A biosynthesis. As previously predicted (El Yacoubi *et al*, 2009), our data indicate that, in yeast, the absence of t⁶A modification has profound effects on translation accuracy, affecting both initiation codon selection and frame maintenance. Our findings raise the possibility that at least some of the reported phenotypes of Sua5 and Kae1 depleted cells, as well as their pleiotropic nature, are simply indirect effects of severe translation defects and lay the foundations for finally understanding the cellular function(s) of this gene family.

Results

Identification of the YgjD/Kae1 family as a candidate for a missing t⁶A biosynthesis gene

t⁶A biosynthesis is predicted to require multiple enzymatic steps (Garcia and Goodenough-Lashua, 1998). We have already identified one partner, the YrdC/Sua5 family (El Yacoubi *et al*, 2009), and we reasoned that the missing enzymes should follow the same universal phylogenetic distribution. The YgjD/Kae1 (COG0533) family met this universality criterion and emerged as a candidate for a family involved in the t⁶A biosynthetic pathway. Indeed, sequence similarity searches showed that YrdC-like and YgjD-like domains were fused in two protein families, HypF and NodU/CmcH (Figure 1). The order of these domains differs in the HypF and NodU/CmcH (represented here by NovN) protein chains (Figure 1), but these independent fusion

events support a functional link between the YrdC and YgjD domains.

The hydrogenase maturation factor HypF is a multi-domain protein that consists of an N-terminal acylphosphatase domain, two repeats of a zinc finger-like motif, the YrdC-like domain and a C-terminal YgjD-like domain (Figure 1). HypF catalyses the transfer of carbamoyl group from carbamoylphosphate to the C-terminal cysteine residue of the hydrogenase accessory protein HypE (Paschos *et al*, 2002; Reissmann *et al*, 2003). This activity is dependent on ATP, which is hydrolyzed to AMP and pyrophosphate. The NodU/CmcH family members consist of a YgjD-like N-terminal domain and a YrdC-like C-terminal domain (Figure 1). These enzymes also catalyse the transfer of the carbamoyl group from carbamoylphosphate to hydroxyl groups of sugars and antibiotics (Brewer *et al*, 1980; Coque *et al*, 1995; Freil Meyers *et al*, 2004). As is the case for HypF, this activity is dependent on ATP, and is inhibited by pyrophosphate, but not phosphate anions, suggesting that both carbamoyltransferases may share similar enzymatic mechanisms.

Only the structure of the HypF N-terminal domain has been determined so far (Rosano *et al*, 2002). Meanwhile, the available structures of YrdC and YgjD homologs can guide sequence–structure analysis of the related domains of the HypF and NodU/CmcH families. The crystal structure of YrdC homologue Sua5 from *Sulfolobus tokodaii* (PDB id: 2eqa (Yoshihiro *et al*, 2008)) contains an AMP molecule, fortuitously bound in the putative active site. This protein was reported to hydrolyze ATP to AMP (Yoshihiro *et al*, 2008). The four conserved residues, Lys50, Arg52, Ser139, and Asn141 (using the *E. coli* YrdC numbering), designated the KRSN tetrad, cluster in the structure in the vicinity the α -phosphate group and are essential for YrdC function (El Yacoubi *et al*, 2009). The YrdC-like domains of HypF and NodU/CmcH families share the KRSN tetrad, suggesting that it has a similar function in all three families (Figure 1). The remaining part of putative active site is family specific; it is formed by residues that are conserved within each family, but not between families. In the Sua5 structure, the additional C-terminal domain is situated next to this part of the active site, expanding the YrdC-specific site.

The crystal structures of Kae1 proteins (YgjD homologs) from three different archaeal species have been determined, confirming that the YgjD/Kae1 family belongs to the superfamily of ATPases with an actin-like fold (Hecker *et al*, 2007, 2008; Mao *et al*, 2008). The *P. abyssi* Kae1 structure has been determined in complex with ATP and metal ion (Fe II) (Hecker *et al*, 2007). It shares the ATP-binding site with several other ATPases of the same superfamily. However, the metal ion-binding site, located near the γ -phosphate group, is unique to the YgjD/Kae1 family. The metal ion-binding residues (His111, His115, and Asp300, *E. coli* numbering), contribute to the two conserved motifs, which are shared by the YgjD-like domains of the HypF and NodU/CmcH families (Figure 1). The bound metal ion may have similar role in all three families, for example, it may position the acceptor group of a substrate molecule for γ -phosphate group transfer. In the Kae1 structure, the ion sits at the bottom of a groove, which is formed by the residues conserved in the YgjD/Kae1 family, but not the other two families (HypF and NodU/CmcH families). The bacterial/mitochondrial YgjD subfamily contains a conserved His

Figure 1 Domain organization of the YrdC, YgjD, and Sua5 domain containing proteins. (A) Each protein is shown as a bar, the length of which is approximately proportional to the number of residues; the YrdC-like domains are shaded dark grey, and the YgjD-like domains are shaded light grey. Positions of the conserved motifs in the two domain families are marked by numbered boxes. In YeaZ, only the N-terminal region has homology to the YgjD domains, and there are no conserved motifs identifiable in the rest of its chain. (B) Sequence signatures of the conserved motifs (marked on panel A). All numbering is based on the *E. coli* YgjD protein (NP_417536.1).

residue (His 139, *E. coli* numbering), whereas in the archaeal/eukaryotic Kae1 subfamily, the equivalent position is occupied by a conserved Asn residue.

In combination with the universal codistribution of members of the YrdC and YgjD family, domain organization and structural examinations of the proteins above described led us to propose that the YgjD/Kae1 family is also involved in t⁶A biosynthesis. This hypothesis was tested experimentally and the results are reported here.

Kae1/Qri7 are involved in the formation of t⁶A in tRNA of yeast

Two members of the COG0533 family are encoded in the *S. cerevisiae* genome: Kae1 (YKR038C) and Qri7 (YDL104C). Qri7 has been shown to localize to the mitochondria and Kae1 to the nucleus and cytoplasm (Huh *et al*, 2003; Hecker *et al*, 2007). In order to address the respective involvement of these two homologs in the biosynthesis of t⁶A, the *kae1Δ::KanMX4* strain was constructed (as described in Supplementary data) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis of bulk tRNA extracted from WT and *kae1Δ* strains performed as described

previously (El Yacoubi *et al*, 2009). This analysis revealed that the 25.5-min peak detected on the UV trace in the WT strain, corresponding to the protonated molecular weight of t⁶A (MH⁺ = 413 m/z), is absent from the tRNA isolated from the strain carrying the *kae1Δ* allele (Figure 2A). The t⁶A peak was restored when this strain was transformed with a plasmid carrying the wild-type KAE1 gene (plasmid pYES-KAE1_{Sc}) (Figure 2A). To test if Qri7 can functionally replace Kae1, QRI7 lacking its mitochondrial targeting peptide was expressed *in trans* (pYES-QRI7_{Sc}) in *kae1Δ::KanMX4* (see Supplementary data for plasmid and strain construction) and bulk tRNA t⁶A levels analysed. As shown in Figure 2B, when expressed *in trans* from pYES2, QRI7 was able to restore t⁶A levels to wild type, establishing Qri7 as a functional homologue of Kae1 in t⁶A biosynthesis. We also tested the ability of QRI7 to complement the growth defect of the *kae1Δ* strain. As depicted in Figure 1C, the growth defects of *kae1Δ* was restored in strains expressing KAE1 or QRI7 *in trans* (pYES-QRI7_{Sc}, pYES-KAE1_{Sc}). These functional complementations together with the localization of both Kae1 (cytoplasmic and nuclear (Huh *et al*, 2003)) and Qri7 (mitochondrial (Huh *et al*, 2003; Hecker *et al*, 2007)) indicate that Kae1 is

Figure 2 LC–MS/MS analysis of yeast tRNA extracted from different strains carrying homologs of the Kae1/YgjD/Qri7 family. Extracted ion chromatograms for 413 m/z are shown for each strain under the UV traces. (A) t⁶A profiles linking the disappearance of the t⁶A peak to the deletion of KAE1. WT (BY4741) (left panels), *kae1Δ* (middle panels), and *kae1Δ* carrying KAE1 *in trans* (right panels). (B) t⁶A profiles of *kae1Δ* expressing the homologue from *E. coli* YgjD (pYESYgjD) (left panels) or the yeast mitochondria located QRI7 (pYESQRI7) (right panels). Positive and negative controls were run as presented in (A). (C) Growth phenotypes of yeast strains lacking a functional KAE1 and transformed with pYES, pYESKAE1_{Sc}, or pYESQRI7_{Sc}. The parent BY4741 was transformed with pYES and used as reference. Each growth curve presented here is an average of 10 independent growth curves. Error bars represent s.d. The growth conditions are described in Supplementary data.

Figure 3 Role of t⁶A in the frequency +1 and -1 frameshifts. (A) Sequence of the frameshift-inducing targets. (B) Frameshifting frequencies for target sequence BLV, Prsv, and EST3. Measurements are average of five independent measurements using a dual reporter construct with the target sequence inserted between the *lacZ* and firefly luciferase ORFs and a control constructs containing the corresponding in-frame sequence inserted in between. Efficiency of frameshifting, expressed as percentage, was calculated by dividing the firefly luciferase/ β -galactosidase ratio obtained for each test construct by the same ratio obtained for the corresponding in-frame control construct. Errors bars represent s.d.

involved in the modification of cytoplasmic tRNA while Qri7 is involved in the modification of the mitochondrial tRNAs.

To test the universality of this gene family, the homologs from *E. coli* (*ygjD_{Ec}*) and *Bacillus subtilis* (hereafter *ydiE*=*ygjD_{Bs}*) were expressed *in trans* (pYES-*ygjD_{Ec}* and pYES-*ygjD_{Bs}*, see Supplementary data for plasmid constructions) in *kae1Δ::KanMX4*. However, none of the tested bacterial *ygjD* genes complemented the t⁶A⁻ phenotype of the *kae1Δ* yeast strain (Figure 2B; Supplementary Figure S1).

The absence of Sua5 or Kae1 leads to similar defects in frame maintenance

To evaluate the effects of the lack of t⁶A on translation in yeast, we used the previously characterized t⁶A⁻ deletion mutant, *sua5Δ*, as well as the newly constructed *kae1Δ* mutant. We reasoned that hypomodified ANN decoding tRNAs would stimulate frameshifting at shift prone sites comprised of ANN codons. To address this, sequences known to promote either -1 or +1 frameshifting (Figure 3A) inserted between the β -galactosidase and firefly luciferase open reading frames were used to quantify frameshifting efficiency of t⁶A⁻ and wild-type BY4741 strains as described in the Materials and methods section.

The -1 frameshifting efficiency using the BLV recoding sequence (pGKU-BLV), containing tandem t⁶A-dependent

codons, was significantly higher in both t⁶A deficient strains (60% in *sua5Δ* and 42% in *kae1Δ* compared with 8% observed in wild-type strain BY4741, see Figure 3B). Frameshift efficiency was restored to wild-type levels in the *sua5Δ* pRS313SUA5 and *kae1Δ* pYESLEU2-KAE1 backgrounds (Figure 3B), confirming that the observed differences were due to the absence of KAE1 or SUA5 in each respective deletion strain. For the recoding sequence Prsv (pGKUPrrsv), which contains a single t⁶A-dependent codon, no differences were observed between WT and t⁶A deficient strains, (4% frameshift efficiency in BY4741, 7% in *sua5Δ*, and 5% in *kae1Δ*; Figure 3B). The +1 frameshifting levels of frameshift promoting sequence EST3 in the wild-type strain BY4741 and in both mutants were also measured (Figure 3B). The +1 frameshifting efficiency was higher for the t⁶A⁻ strains (13% in *sua5Δ* and 22% in *kae1Δ*) when compared with wild type (6%) for the EST3 target (Figure 3B). These effects were complemented when the WT genes were expressed *in trans*. In all, these results suggest that the absence of t⁶A₃₇ leads to higher frameshift levels, particularly at sequences containing tandem ANN codons.

To begin addressing the question of how the lack of t⁶A may affect the yeast proteome in *sua5Δ* and *kae1Δ* mutant strains, we estimated the frequency and the distribution of ANN codons (number and length of stretches of ANN

Figure 4 Lack of t⁶A in yeast affects initiation codon selection *in vivo*. The initiation efficiency is defined as the ratio of the firefly luciferase activity initiated on AUG or GUG codon to the one of the renilla luciferase activity initiated on AUG codon. The relative activities presented correspond to the ratios for the t⁶A[−] strains (deletion mutants *sua5*Δ and *kae1*Δ) normalized by the ratio obtained for wild-type strain BY4741 with the firefly luciferase and the renilla luciferase both initiated on AUG. All the measurements were the average of five independent extractions. Errors bars represent standard deviations.

stretches) in yeast coding sequences, using an in-house program (the code is available at <http://bioinformatics.cise.ufl.edu/codon.html>). Our hypothesis was that frameshift defects, and low translation efficiency of translation due to lack of t⁶A have a greater impact on proteins whose coding sequences are heavily depending on ANN decoding tRNAs. We ranked all yeast proteins according to the number of ANN codons in their cDNAs. In general, the coding sequences that contained the highest number of ANN codons were also those with the highest number of tandem ANN codons (Supplementary Table S1). Highly represented proteins include those involved in RNA processing, cytoskeletal/intracellular trafficking, ribosomal biogenesis and telomere homeostasis. Additionally, the retrotransposon genes TYA Gag and TYB Pol (similar to retroviral genes) were also identified. These genes are transcribed/translated as one unit and the polypeptide is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (Kim *et al*, 1998).

The absence of Sua5 or Kae1 leads to high levels of misinitiation at GUG codons

The cytoplasmic initiator tRNA (met-tRNA_i) of all eukaryotes examined to date contains the t⁶A modification, whereas it is absent in the formylmethionine-tRNA_i (fmet-tRNA_i) of prokaryotes, Archaea, and organelles. It has been proposed that one of the roles of t⁶A₃₇ is to prevent mispairing between the first base of the codon and the third base of the anticodon (Dube *et al*, 1968). This was suggested based on the observation that *E. coli* fmet-tRNA_i recognizes not only the AUG, GUG, and UUG as initiator codons, whereas yeast met-tRNA_i that contains the t⁶A₃₇ modification recognizes only AUG (Baralle and Brownlee, 1978; Schneider *et al*, 1986). To test if the presence of t⁶A is involved in restricting initiator codon choice *in vivo*, the efficiency of translation initiation at GUG using a dual luciferase reporter system (see Supplementary

Figure 5 Genetic complementation of the *ygiD* essentiality phenotype in *E. coli* carrying the *ygiD* gene under the control of the P_{TET} promoter. (A) Phenotype of the P_{TET}:*ygiD* strain transformed with control plasmid pBAD24 or plasmid pBAD*ygiDEc* (pBAD24 carrying *ygiDEc*) grown under inducing conditions (aTc and ara) or repressing condition (glu). (B) Lack of complementation of the essentiality phenotype of *ygiDEc* by the yeast homologs *KAE1* and *QRI7* (lacking the sequence of the 30 AA targeting signal) and the archaeal homologue for *M. maripaludis* *kae1-prpk*. Cells grew in the presence of aTc allowing the expression of the chromosomal *ygiDEc* allele but failed to grow in the absence of aTc and the presence of arabinose allowing the expression of the plasmid borne alleles.

data for plasmid construction) was measured in the *sua5*Δ and *kae1*Δ t⁶A[−] yeast strains. We first measured the ratio of firefly luciferase activity to renilla luciferase activity, in which translation of both genes was initiated at AUG codons (see Figure 4). Similar activity ratios were obtained in the WT, *sua5*Δ, and *kae1*Δ strains (91 and 101%, respectively). Changing the firefly luciferase initiation codon from AUG to GUG led to a 50-fold decrease in the luciferase activity ratio in the WT strain (Figure 4). In the mutant strains lacking *SUA5* or *KAE1*, the ratio increased only two- to three-fold (Figure 4). Thus, in the absence of t⁶A, initiation at GUG codons is substantially increased. These data indicate that one of the roles of t⁶A₃₇ is indeed to prevent mispairing between the first base of the codon and the third base of the anticodon as previously proposed (Dube *et al*, 1968).

Complementation of the *E. coli* *ygiD* essentiality phenotype with bacterial orthologs requires coexpression of heterologous *ygiD*/*yeaZ* pairs

Because the *ygiD* orthologue is essential in *E. coli*, a strain was constructed in which the expression of the chromosomal *ygiD* gene was placed under the P_{TET} promoter (see Materials and methods section for strain construction). This strain required the presence of anhydrotetracycline (aTc) for growth (Figure 5). Transformation of this strain with *pygiDEc*, a

Figure 6 The *B. subtilis* *ygjD* homologue functionally replaces the *E. coli* homologue only when coexpressed with the *yeaZ* gene from *B. subtilis*. (A) The physical clustering deduced from analysing the genetic context of both *ygjD* and *yeaZ* suggest that these two genes function in the same pathway. (B) Model of YgjD–YeaZ heterodimer. The N-terminal domain of the YgjD homologue Kae1 (orange, PDB 2IVP) is superimposed on the related domain of one subunit of the YeaZ homodimer (silver; PDB 1OKJ), resulting in the conserved interface with the other YeaZ subunit (blue). (C) Genetic complementation of the essentiality of *E. coli* *ygjD* by coexpression of the *B. subtilis* *ygjD* and *YeaZ* in an operonic structure. The P_{TET} :*ygjD* strain was transformed with plasmid pBAD*ygjD_{Bs}* carrying the *yeaZ_{Bs}* in an operonic structure (pBAD*ygjD_{Bs}yeaZ_{Bs}*) or in the reverse orientation (pBAD*ygjD_{Bs}oppyeaZ_{Bs}*) and tested for growth under inducing conditions (aTc and ara) or repressing condition (glu). pBAD*ygjD_{Ec}*, pBAD*ygjD_{Ec}yeaZ_{Ec}*, and pBAD24 were used as positive and negative controls, respectively.

pBAD24 derivative expressing the *E. coli* *ygjD* gene under the control of the P_{BAD} promoter, allowed for growth in the absence of aTc and in the presence of arabinose after 16 h (Figure 5). pBAD24 derivatives carrying *ygjD* homologs from yeast (*QRI7* lacking its secretion signal and KAE1), *B. subtilis*, or *Methanococcus maripaludis* (*kae1-prpk*, encoding the fused Kae1-PRPK protein) were similarly transformed into the P_{TET} :*ygjD* strain but none complemented the essentiality phenotype and all behaved similar to the negative control transformed with the empty vector pBAD24 (Figure 5; Supplementary Figure S2).

Previously, *E. coli* *ygjD* was found to interact with *yeaZ* (Handford *et al*, 2009), and in ~20% of bacterial genomes available in the SEED database (closely related genomes were omitted from the analysis), *yeaZ* is physically clustered with *ygjD* (see Figure 6A for representative genomes). YeaZ adopts the same actin-like fold (Nichols *et al*, 2006; Xu *et al*, 2010) as YgjD, but lacks both its ATP-binding and metal ion-binding sites. The region of strong sequence and structural similarity between the two proteins is limited to their N-terminal domains (Figures 1 and 6). Interestingly, in all YeaZ crystal structures, there are similar dimeric contacts via the N-terminal domains (in the structure of *Thermotoga maritima* YeaZ (PDB id: 2a6a), only one of the two molecules in the asymmetric unit makes such a contact). At the two-fold axis, the two symmetry-related helices are closely packed through the interface of small residues (Gly96, Gly100, and Gly104 in *E. coli* YeaZ). The small sizes of residues occupying the equivalent sites are conserved in the YgjD subfamily (Thr93, Ser97, and Ala101, *E. coli* numbering) but not in the Kae1 subfamily. This suggests the likelihood of heterodimerization of the bacterial YgjD

and YeaZ through the same interface (Figure 6). This interface also includes the C-terminal tail of YgjD (Figure 6), suggesting that YeaZ binding may affect the relative orientation of the YgjD domains. In other words, based upon the bioinformatic and structural analyses described above, we postulated that the essential function of YgjD in *E. coli* requires YeaZ. Thus, our previous attempts to complement the YgjD essentiality in *E. coli* using homologs failed due to the absence of the corresponding YeaZ partner.

To test this hypothesis, *B. subtilis* *ygjD* and *yeaZ* homologs were cloned in an operonic structure in pBAD24 (see Supplementary data for details) for coexpression. The resulting plasmid, *pygjd_{Bs}yeaZ_{Bs}* was transformed in the *E. coli* P_{TET} :*ygjD* strain and was now able to complement the YgjD essentiality phenotype as described below. The resulting strain grew in the presence of aTc or arabinose but not in the presence of added glucose (Figure 6). In addition, growth in the presence of arabinose was slightly delayed compared with control strain (i.e. P_{TET} :*ygjD* transformed with *pygjd_{Ec}*) (Figure 6). When restreaked on media containing aTc, arabinose, or glucose, the P_{TET} :*ygjD*/*pygjd_{Bs}yeaZ_{Bs}* clones displayed growth patterns consistent with those of the original transformants that is growth in the presence of aTc or arabinose but no growth in the presence of glucose (see Supplementary Figure S3). The complementation was strictly dependent on the expression of both genes as transformation with derivatives carrying *yeaZ_{Bs}* alone or *ygjD_{Bs}* with *yeaZ_{Bs}* in the opposite orientation (*pygjd_{Bs}oppyeaZ_{Bs}*) did not lead to growth even in the presence of the inducer arabinose (Figure 6). These results show that the essential function of *ygjD* in bacteria is dependent on *yeaZ*.

Discussion

Universality of the Kae1/Qri7/YgjD family function

Here, we show that the Kae1/YgjD family is involved in t⁶A biosynthesis; deletion of the yeast *KAE1* gene leads to the absence of t⁶A in bulk tRNA. With the previously identified Sua5/YrdC family, this brings to two the number of universal protein families involved in the t⁶A biosynthetic pathway (El Yacoubi *et al*, 2009).

The question of compartmentalization of enzymes of the t⁶A biosynthesis pathway remains open. The LC-MS/MS analysis method used did not allow us to detect mitochondrial modifications (see Supplementary data). This could explain why no t⁶A was detected in *kae1Δ::KanMX4* even in the presence of the wild-type *QRI7* allele.

In order to evaluate the role of Qri7 in mitochondrial t⁶A biosynthesis, we did attempt to purify mitochondrial tRNA from the *qri7Δ* strain as described previously (Umeda *et al*, 2005). We were unable to purify sufficient quantities of tRNA for analysis, which could be expected given the mitochondrial defects of this strain (Oberto *et al*, 2009). Interestingly, there is only one Sua5 homologue in yeast (the other enzyme required for t⁶A formation), and therefore we expect this protein to display dual localization as has been observed with other essential translation enzymes (Natsoulis *et al*, 1986; Chatton *et al*, 1988; Souciet *et al*, 1999; Lee *et al*, 2007).

Archaeal, bacterial, and eukaryotic Sua5/YrdC homologs are functionally interchangeable (El Yacoubi *et al*, 2009). We did not observe within the YgjD/Kae1 family the universal complementation capacity we had previously observed for the Sua5/YrdC family. First, in yeast, despite the fact that the t⁶A⁻ phenotype of yeast *kae1Δ* is complemented by expressing *in trans* the yeast Qri7 homologue (member of the bacterial YgjD subfamily) lacking its mitochondrial targeting signal (Hecker *et al*, 2007), similar complementation attempts failed when using bacterial homologs from *E. coli* or *B. subtilis*. Second, in *E. coli* no complementation of the essentiality phenotype of the *ygjD* deletion was observed with the *QRI7/KAE1/ygjD_{Ec}/ygjD_{Bs}/kae1-prpk_{Mm}* genes. In our hands, *QRI7* did not complement the essentiality phenotype of *ygjD* in *E. coli* even though this complementation was previously reported (Oberto *et al*, 2009). This discrepancy may be due to differences in design between the two reports. In this study, *QRI7* was expressed *in trans* under P_{BAD} regulation while the chromosomal *ygjD* was under the tight P_{TET} regulation. Complementation was tested in the absence of aTc and in the presence or absence of the inducer arabinose. In the previous study (Oberto *et al*, 2009), both *QRI7* and *ygjD_{Ec}* were expressed *in trans* using two compatible plasmids while the chromosomal *ygjD* was deleted. *QRI7* was constitutively expressed and *ygjD_{Ec}* was under the control of the P_{BAD} promoter. Complementation by *QRI7* was observed in the absence of arabinose; however, it is possible that low levels of *ygjD* expression even under uninduced conditions allowed for *QRI7* complementation and therefore growth in the absence of inducer.

The lack of complementation by *QRI7* is in agreement with our results showing that the essentiality of *ygjD* depletion was only complemented with coexpression of the *B. subtilis* YeaZ/YgjD pair. This is also confirming the biological relevance of the recent data showing that only the YeaZ–YgjD pairs from closely related organisms could form complexes *in vitro*

(Rajagopala *et al*, 2010) and thus explaining the absence of *in vivo* cross-species complementation for the Kae1/Qri7/YjD protein family.

Our results combined with the study by Rajagopala *et al* (2010) strongly suggest that YeaZ and YgjD are subunits of a bacterial heterooligomeric enzyme that have coevolved together. There are several known heterooligomeric enzymes composed of homologous subunits, including the bacterial luciferase LuxAB (Fisher *et al*, 1996) and the sulphur transfer mediator TusBCD in the s²U modification (Numata *et al*, 2006). Their structures display pseudo-symmetrical arrangements, with the subunit interfaces being the most conserved parts, suggesting that these heterooligomers probably have evolved from homooligomers by gene duplication. In general, only one of the subunits retains the catalytic site, whereas the others provide new surfaces for additional interactions. Some of the isolated individual subunits may form similar homooligomers with the same symmetry; others may not form active oligomeric species without partners (Thoden *et al*, 1997). The relationship between bacterial YeaZ and YgjD fits well into the same scenario. Our model of the YgjD/YeaZ complex suggests that it forms a stable heterodimer similar to the common homodimer observed in the YeaZ structures. A coevolution of YeaZ and YgjD could explain why both corresponding genes from the same or closely related bacterial species are required to observe functional complementation. Despite the overall similarity, there are subtle differences in the dimerization interfaces of *Ec* and *Tm* YeaZ, so there could be small but significant differences between the YeaZ–YgjD interfaces from different organisms.

The biosynthesis of t⁶A requires ATP, bicarbonate, and free threonine with hydrolysis of at least two ATP molecules needed for the formation of two carbon–nitrogen bonds (Chheda *et al*, 1972; Powers and Peterkofsky, 1972; Elkins and Keller, 1974; Korner and Söll, 1974). As both YrdC and YgjD appear to bind adenosine nucleotides, each could synthesize one such bond. Based on our structural analysis, we predict the following: YgjD is likely to be a ligase for bicarbonate and threonine (Figure 7C). In the first step, YgjD catalyses transfer of a γ-phosphate group from ATP to bicarbonate, yielding carboxyphosphate and ADP. Carboxyphosphate would react with the threonine amine, producing N-carbamoylthreonine and phosphate. Both substrate molecules can be readily accommodated in the YgjD/Kae1 active site: bicarbonate is next to the γ-phosphate group and bound metal ion, and threonine is on the top of bicarbonate in the conserved part of the groove (Figure 7A). YrdC would then catalyse the activation of N-carbamoylthreonine via the formation of an acyladenylate with the release of pyrophosphate, followed by the transfer of carbamoylthreonine to the N⁶ group of A37 (Figure 7C). The conserved KRSN tetrad in YrdC is predicted to stabilize the leaving pyrophosphate group in the first step, whereas N-carbamoylthreonine can be readily accommodated in a YrdC-specific part of the putative active site (Figure 7B). In the Sua5 homologue of YrdC, the Sua5 C-terminal domain probably assists in the correct positioning of the tRNA adenine base (A37), which is to be modified. These biochemical hypotheses require experimental validation; however, the complementation results presented here combined with our inability to reconstitute t⁶A formation *in vitro* using purified *E. coli* YrdC and YgjD proteins (see Supplementary data) suggest that at least one

Figure 7 Fitting ligands into the putative active sites of the YrdC/Sua5 and YgjD/Kae1 families. **(A)** The product carbamoylthreonine (stick, yellow carbons) is in the active site of Kae1 (2IVP). The carbamoyl group contacts the bound metal ion (magenta sphere). Those conserved residues not in the ATP or metal ion-binding sites are shown with opaque spheres, cyan carbons). **(B)** ATP (stick, green carbons) and the substrate carbamoylthreonine (stick, yellow carbons) are in the active site of the YrdC domain of Sua5 (PDB id: 2EQA). The AMP group of the ATP molecule is taken from the Sua5 structure; the pyrophosphate group is docked with the KRSN tetrad (opaque spheres, magenta carbons). Carbamoylthreonine sits in the pocket of conserved residues with carbamoyl group contacting the α -phosphate group. **(C)** Proposed t⁶A biosynthesis pathway.

other protein component is required for the synthesis of this complex modification. Obvious candidates for the missing t⁶A enzymes are YeaZ (as our genetic results suggest) and YjeE as members of this essential bacterial-specific gene family have been shown to interact with YeaZ (Handford *et al*, 2009). Candidates for missing components in Eukaryotes and Archaea are the proteins of the KEOPS/EKC complex Pcc1, Bud32, and Cgi121 that interact with Kae1 (Kisseleva-Romanova *et al*, 2006).

The pleiotropic phenotypes of Kae1 depleted strains can be explained by the lack of t⁶A

Previous studies have suggested that t⁶A contributes to accurate codon–anticodon recognition mainly by preventing the formation of U33–A37 across-the-loop base pairing interaction, and allowing cross-strand stacking of A₃₈ and t⁶A₃₇ with the first position of the codon (Weissenbach and Grosjean, 1981; Stuart *et al*, 2000; Yarian *et al*, 2000; Murphy *et al*, 2004; Lescrinier *et al*, 2006). We had hypothesized (El Yacoubi *et al*, 2009) that the lack of t⁶A should severely affect translation as it is found in seven *E. coli* and 12

S. cerevisiae tRNAs (four of which are mitochondrial) (Jühling *et al*, 2009). Such an effect would be pleiotropic, since anticodon-surrounding modifications are context dependent (Weissenbach and Grosjean, 1981; Stuart *et al*, 2000; Yarian *et al*, 2000; Murphy *et al*, 2004; Lescrinier *et al*, 2006). In *S. cerevisiae*, loss of *SUA5* increased the level of leaky scanning through start codons (Lin *et al*, 2009). Our result shows that in both *S. cerevisiae* *sua5* Δ and *kae1* Δ strains, a robust increase in the frequency of –1 frameshifts at tandem AAA codons occurred. The +1 frameshift was less robust, in agreement with observations also made by others (Lin *et al*, 2009). Our data also show an increase in misinitiation at GUG codons in both *kae1* Δ and *sua5* Δ t⁶A[–] yeast strains. Because both mutants have similar translation defects, it is likely that these are due to the absence of t⁶A and not to other potential roles of the Sua5 and Kae1 proteins that have been implicated in rRNA maturation (Kaczanowska and Ryden-Aulin, 2004; Kaczanowska and Rydén-Aulin, 2005) or transcriptional regulation (Kisseleva-Romanova *et al*, 2006). Thus, loss of t⁶A could lead to expression of non-functional proteins and/or aberrant and truncated proteins some of which may display

dominant negative effects leading to the observed pleiotropic phenotypes. One interesting candidate is EST3. This protein is required for telomere maintenance *in vivo* and its translation is dependent upon a +1 programmed frameshift (Morris and Lundblad, 1997). Both *sua5Δ* and *kae1Δ* cells display progressively shortened telomeres during early passages (Downey *et al*, 2006; Meng *et al*, 2009). However, we have tested the frameshift promoting ability of the EST3 sequence and showed that in the absence of t⁶A, frameshifting is actually increased (Figure 3). Therefore, our data indicate that the absence of t⁶A should not lead to a loss of EST3 translation and therefore cannot account for the observed phenotypes of the t⁶A⁻ mutants.

Our analysis of the distribution of ANN codons in *S. cerevisiae* did however identify several coding sequences with high numbers of ANN codons. The efficiency/accuracy of translation of these proteins could be affected in t⁶A⁻ strains. ANN-rich candidates included YBL004W (Bernstein *et al*, 2004), YKL014C (Dez *et al*, 2004; Rosado and De La Cruz, 2004), YJL109C (Dragon *et al*, 2002), YMR229C (Venema and Tollervey, 1995), which are all involved in the biogenesis of ribosomes, mostly through their role in processing of rRNA (see Supplementary Table S1). YHR165C and YER172C are two other high ANN-containing genes and both belong to the U4/U6 spliceosome. This suggests that ribosome biogenesis and RNA processing might be affected because of mistranslation of key proteins involved in these processes, which in turn affects the generalized protein biosynthesis and/or integrity.

However, the association of Kae1 and Sua5 with other complexes in yeast also suggests that these proteins might together and/or individually contribute to essential cellular functions other than tRNA modification. Both *sua5Δ* and *kae1Δ* cells have a shortened-telomere phenotype and KAE1 and SUA5 have been directly associated with telomere homeostasis (Downey *et al*, 2006; Meng *et al*, 2009). Telomeres are essential nucleoprotein complexes found at the end of linear chromosomes that fulfil two critical functions. They recruit and activate telomerase, a RT that polymerizes short TG-rich repeats at the 3' end of chromosomes and protect chromosome ends (telomere capping) (Blackburn, 2001). t⁶A deficient mutants and telomerase deficient mutants have similar growth defects (Steinmetz *et al*, 2002; Deutschbauer *et al*, 2005; Matsui and Matsuura, 2010). As deletions of either *SUA5* or *KAE1* lead to very similar telomere defects, we propose two possible scenarios for a role of t⁶A in telomere maintenance: (1) the translation of a specific protein required for telomerase function may be affected by the absence of the modification on tRNAs (high level of ANN codons in its sequence in addition to general translation defects associated with lack of translation t⁶A) and/or (2) a component of the telomerase RNA or protein is directly modified with a t⁶A derivative. The latter scenario therefore postulates that Kae1 and Sua5 have other RNAs and/or proteins as substrates in addition to tRNAs. Precedent for a tRNA modification enzyme that can modify other RNA or DNA substrates has already been established (Wrzesinski *et al*, 1995; Nonekowsky *et al*, 2002). Regarding the former possibility, a number of genes implicated in telomere maintenance including TEL1 (Ritchie and Petes, 2000), RIF1 (Hardy *et al*, 1992), MLP1 and MLP2 (Hediger *et al*, 2002), MEC1 (Takata *et al*, 2004), and YRF1 (present in several

copies in the genome) (Yamada *et al*, 1998) are all among the top-100 genes with the highest levels of ANN tandem codons. This supports the idea that mistranslation of one or several of these genes is at the origin of the telomere maintenance phenotype observed for t⁶A⁻ strains.

For prokaryotes, it is still not possible to conclude definitively that the essentiality of the *ygjD* gene in bacteria is due to a role of YgjD in t⁶A synthesis. Further work examining translation efficiency, ribosome biogenesis, and possibly maintenance of DNA integrity is warranted.

The Kae1/Qri7/YgjD is universal and can be used as a marker for life

The consequences of the absence of t⁶A on translation is quite profound, explaining why t⁶A₃₇, like m¹G37 or Ψ55, is one of the tRNA modifications that has not been lost through evolution, even in organisms with small genomes such as *Mycoplasma* (de Crécy-Lagard *et al*, 2007). Of note is the lack of *yrdC* and *ygjD* homologs in symbionts such as *Carsonella rudii* and *Sulcia muelleris*. While it is possible that these organisms lack the t⁶A modification in their tRNA, given the fact that genes encoding other essential components of bacterial translation such as tRNA synthetases and ribosomal proteins are also absent from the genomes of these organisms (Nakabachi *et al*, 2006; McCutcheon and Moran, 2007; Tamames *et al*, 2007; McCutcheon *et al*, 2009; Woyke *et al*, 2010), and that homologs of YrdC and YgjD are found in all other sequenced organisms, we favour the hypothesis that YrdC and YgjD, like the other components of translation, are dependent on protein import from the host, supporting the status of these organisms as organelles in the making (Koonin and Wolf, 2008). We therefore conclude that YgjD/Kae1 and YrdC/Sua5 families are truly universal.

Materials and methods

Bioinformatics

The Blast tools and resources at NCBI were used (Altschul *et al*, 1997). Multiple protein alignments were performed with the ClustalW tool (Chenna *et al*, 2003) in the SEED database or the MultiAlign software (<http://omics.pnl.gov/>). The 3D models were generated using the protein-fold recognition protocols of Phyre (Kelley and Sternberg, 2009) based on one- and three-dimensional sequence profiles, coupled with secondary structure and solvation potential information (using the Phyre server available on the World Wide Web at <http://www.sbg.bio.ic.ac.uk/~phyre/>). Structure-based alignment of a subset was performed using the ESPript platform (Gouet *et al*, 1999) through the webinterface (<http://esript.ibcp.fr/ESPript/ESPript/>).

Strains, plasmids media, and growth conditions

The strains mentioned in this work are listed in Supplementary Table S2, the plasmids are listed in Supplementary Table S3, and the primers are listed in Supplementary Table S4. Growth conditions for *E. coli* and yeast derivatives are described in Supplementary data.

tRNA extraction and analysis

Bulk tRNA was prepared, hydrolyzed, and analysed by LC-MS/MS as described previously (El Yacoubi *et al*, 2009). To compare tRNA concentrations, we compared the ratio of the levels of the modified base (245 *m/z*) in each sample by integrating the peak area from the extracted ion chromatograms. The MS/MS fragmentation data were also used to confirm the presence of t⁶A. All tRNA extractions and analysis were performed at least twice independently.

Initiation and frameshifting efficiency

Plasmid pRaugFFgtg (Kolitz *et al*, 2009) expresses *Renilla* and firefly luciferase as separate messages, with *LUC_{renilla}* under the control of

the *ADH1* promoter and *HIS* terminator, *LUC_{firefly}* under the *GPD* promoter and *CYC1* terminator, and the start codon of the *LUC_{firefly}* ORF altered to GTG. The reporter plasmid for frameshifting measurement pGK007 was derived from the pAC vector (Bidou *et al*, 2000; Baudin-Baillieu *et al*, 2009). The promoter SV40 in the pAC vector was replaced by the yeast PGK promoter to reach a higher expression level and the selection marker *LEU2* was replaced by *URA3*. The target sequences were inserted at the *MscI* restriction site located at the junction between β -galactosidase and firefly luciferase coding sequences.

The β -galactosidase and firefly luciferase activities were quantified in the same crude extract as described previously (Stahl *et al*, 1995) for standard growth conditions. All the quantifications were the mean of five independent measurements. The efficiency is defined as the ratio of firefly luciferase activity to β -galactosidase activity. To establish the relative activities of β -galactosidase and firefly luciferase, the ratio of firefly luciferase activity to β -galactosidase activity from an in-frame control plasmid was taken as a reference for each strain tested. Efficiency of frameshifting, expressed as percentage, was calculated by dividing the firefly luciferase/ β -galactosidase ratio obtained from each test construct by the same ratio obtained with an in-frame control construct for each strain (Bidou *et al*, 2000).

For quantification of initiation efficiency at AUG or GUG codons, the renilla luciferase and firefly luciferase activities were measured in the same crude extract as already described (Bidou *et al*, 2004) for standard growth conditions using the Dual luciferase kit from Promega. All the measurements were the mean of five independent extractions. The initiation efficiency was defined as the ratio of firefly luciferase activity initiated at AUG or GUG codons to renilla luciferase activity initiated at an AUG codon (Cheung *et al*, 2007; Kolitz *et al*, 2009). To establish the relative activities, this ratio was normalized by the same ratio obtained in the wild-type strain BY4741 with the firefly luciferase and the renilla luciferase both initiated at an AUG codon.

Construction of *P_{TET}::ygiD_{Ec}* and complementation assays

In order to place chromosomal target gene *ygiD_{Ec}* under the control of the REXTETrbs (rbs for ribosomal-binding site) cassette, we inserted the cassette just upstream of the start codon of *ygiD* in BW25113 by homologous recombination. To do so, forward and reverse primers were designed to carry 50 bp homology 5' and 3' of the native start codon, respectively, followed by 15 bp homology to the REXTET cassette. DNA from strain MG1655*kmREXTETrbs-lacZ*

(Da Re *et al*, 2007) was used as template to amplify the REXTETrbs component of the construct. The PCR product was then electroporated in *E. coli* competent cells as previously described (Baba and Mori, 2008). Cells were plated on Kan and aTc. Positive clones were then tested for dependence on aTc for growth as well as by PCR and sequencing to verify the accurate insertion of the cassette.

For *ygiD* complementation assays, cells were electroporated with 100 ng of plasmid, then recovered with 1 ml of LB and placed at 37°C for 1 h with shaking. After that, 100 μ l of cells were plated on each of three different media: LB Km Amp supplemented with aTc (50 ng/ml), glu 0.2%, or ara 0.2%. Km Amp resistant colonies from each condition were reisolated and checked for growth phenotype on appropriate inducer/repressor media. A gene was considered a functional homologue when transformants grew on media supplemented with aTc (for induction of expression of chromosomal *ygiD_{Ec}*) and ara (for induction of expression of the gene tested for complementation *in trans*) but with glu (for repression of *trans*-gene expression and no induction of chromosomal *ygiD_{Ec}*) both after initial transformation and replating.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

This work was supported by the US Department of Energy (grant no. DE-FG02-07ER64498) and by the National Institutes of Health (grant no. R01 GM70641-01) to V de C-L. This work was also supported by funds from ANR (contract ANR-06-BLAN-0391 to JPR). We thank Kenneth Stadman for *S. sulfolobus* genomic DNA, Jon Lorsch and Julie Takacs for pRaugFFaug and pRaugFFgug plasmids, Marc Bailly and Gabriella Phillips for help with tRNA preparations, Sophie Alvarez for the LC-MS/MS analysis, and Henri Grosjean for constant discussion, support, and critical reading of the manuscript. A special acknowledgement is given to Helen McGuirk and Paula Ruiz for their valuable technical help and to Nemat Keyhani for critical reading of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Abdullah KM, Lo RY, Mellors A (1990) Distribution of glycoprotease activity and the glycoprotease gene among serotypes of *Pasteurella haemolytica*. *Biochem Soc Trans* **18**: 901–903
- Abdullah KM, Lo RY, Mellors A (1991) Cloning, nucleotide sequence, and expression of the *Pasteurella haemolytica* A1 glycoprotease gene. *J Bacteriol* **173**: 5597–5603
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Arigoni F, Talabot F, Peitsch M, Edgerton MD, Meldrum E, Allet E, Fish R, Jamotte T, Curchod ML, Loferer H (1998) A genome-based approach for the identification of essential bacterial genes. *Nat Biotechnol* **16**: 851–856
- Baba T, Mori H (2008) The construction of systematic in-frame, single-gene knockout mutant collection in *Escherichia coli* K-12. *Methods Mol Biol* **416**: 171–181
- Baralle FE, Brownlee GG (1978) AUG is the only recognisable signal sequence in the 5' non-coding regions of eukaryotic mRNA. *Nature* **274**: 84–87
- Baudin-Baillieu A, Fabret C, Liang X-H, Piekna-Przybylska D, Fournier MJ, Rousset J-P (2009) Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res* **37**: 7665–7677
- Bernstein KA, Gallagher JEG, Mitchell BM, Granneman S, Baserga SJ (2004) The small-subunit processome is a ribosome assembly intermediate. *Eukaryotic Cell* **3**: 1619–1626
- Bidou L, Hatin I, Perez N, Allamand V, Panthier JJ, Rousset JP (2004) Premature stop codons involved in muscular dystrophies show a broad spectrum of readthrough efficiencies in response to gentamicin treatment. *Gene Ther* **11**: 619–627
- Bidou L, Stahl G, Hatin I, Namy O, Rousset JP, Farabaugh PJ (2000) Nonsense-mediated decay mutants do not affect programmed -1 frameshifting. *RNA* **6**: 952–961
- Blackburn EH (2001) Switching and signaling at the telomere. *Cell* **106**: 661
- Brewer SJ, Taylor PM, Turner MK (1980) An adenosine triphosphate-dependent carbamoylphosphate-3-hydroxymethylcephem O-carbamoyltransferase from *Streptomyces clavuligerus*. *Biochem J* **185**: 555–564
- Chatton B, Walter P, Ebel JP, Lacroute F, Fasiolo F (1988) The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. *J Biol Chem* **263**: 52–57
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497–3500
- Cheung Y-N, Maag D, Mitchell SF, Fekete CA, Algire MA, Takacs JE, Shirokikh N, Pestova T, Lorsch JR, Hinnebusch AG (2007) Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection *in vivo*. *Genes Dev* **21**: 1217–1230
- Chheda GB, Hong CI, Piskorz CF, Harmon GA (1972) Biosynthesis of N-(purin-6-ylcarbamoyl)-L-threonine riboside. Incorporation of L-threonine *in vivo* into modified nucleoside of transfer ribonucleic acid. *Biochem J* **127**: 515–519

- Coque JJ, Pérez-Llarena FJ, Enguita FJ, Fuente JL, Martin JF, Liras P (1995) Characterization of the *cmcH* genes of *Nocardia lactam-durans* and *Streptomyces clavuligerus* encoding a functional 3'-hydroxymethylcephem O-carbamoyltransferase for cephamycin biosynthesis. *Gene* **162**: 21
- Da Re S, Le Quéré B, Ghigo JM, Beloin C (2007) Tight modulation of *Escherichia coli* bacterial biofilm formation through controlled expression of adhesion factors. *Appl Environ Microbiol* **73**: 3391–3403
- de Crécy-Lagard V, Marck C, Brochier-Armanet C, Grosjean H (2007) Comparative RNomics and Modomics in Mollicutes: prediction of gene function and evolutionary implications. *IUBMB Life* **1–25**
- Deuschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Gaeveer G (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**: 1915–1925
- Dez C, Froment C, Noaillac-Depeyre J, Monsarrat B, Caizergues-Ferrer M, Henry Y (2004) Npa1p, a component of very early pre-60s ribosomal particles, associates with a subset of small nucleolar RNPs required for peptidyl transferase center modification. *Mol Cell Biol* **24**: 6324–6337
- Downey M, Houlsworth R, Maringe L, Rollie A, Brehme M, Galicia S, Guillard S, Partington M, Zubko MK, Krogan NJ, Emili A, Greenblatt JF, Harrington L, Lydall D, Durocher D (2006) A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell* **124**: 1155
- Dragon F, Gallagher JEG, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlege RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF, Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**: 967–970
- Dube SK, Marcker KA, Clark BF, Cory S (1968) Nucleotide sequence of N-formyl-methionyl-transfer RNA. *Nature* **218**: 232–233
- El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crécy-Lagard V (2009) The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Res* **37**: 2894–2909
- Elkins BN, Keller EB (1974) The enzymatic synthesis of N-(purin-6-ylcarbamoyl)threonine, an anticodon-adjacent base in transfer ribonucleic acid. *Biochemistry* **13**: 4622–4628
- Fisher AJ, Thompson TB, Thoden JB, Baldwin TO, Rayment I (1996) The 1.5-Å resolution crystal structure of bacterial luciferase in low salt conditions. *J Biol Chem* **271**: 21956–21968
- Freel Meyers CL, Oberthur M, Xu H, Heide L, Kahne D, Walsh CT (2004) Characterization of NovP and NovN: completion of novobiocin biosynthesis by sequential tailoring of the noviosyl ring. *Angew Chem Int Ed Engl* **43**: 67–70
- Galperin MY, Koonin EV (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res* **32**: 5452–5463
- Galperin MY, Koonin EV (2010) From complete genome sequence to 'complete' understanding? *Trends Biotechnol* **28**: 398–406
- Garcia GA, Goodenough-Lashua DM (1998) Mechanisms of RNA-modifying and -editing enzymes. In *Modification and Editing of RNA*, Grosjean H, Benne R (eds) pp 135–168. Washington, DC: ASM Press
- Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M *et al* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 631–636
- Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305–308
- Handford JI, Ize B, Buchanan G, Butland GP, Greenblatt J, Emili A, Palmer T (2009) Conserved network of proteins essential for bacterial viability. *J Bacteriol* **191**: 4732–4749
- Hardy CF, Sussel L, Shore D (1992) A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev* **6**: 801–814
- Hecker A, Leulliot N, Gadelle D, Graille M, Justome A, Dorlet P, Brochier C, Quevillon-Cheruel S, Le Cam E, van Tilbeurgh H, Forterre P (2007) An archaeal orthologue of the universal protein Kae1 is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity *in vitro*. *Nucleic Acids Res* **35**: 6042–6051
- Hecker A, Lopreiato R, Graille M, Collinet B, Forterre P, Libri D, van Tilbeurgh H (2008) Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex. *EMBO J* **27**: 2340
- Hediger F, Dubrana K, Gasser SM (2002) Myosin-like proteins 1 and 2 are not required for silencing or telomere anchoring, but act in the Tel1 pathway of telomere length control. *J Struct Biol* **140**: 79–91
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* **425**: 686–691
- Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J (2009) tRNADB 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**(Suppl 1): D159–D162
- Kaczanowska M, Ryden-Aulin M (2004) Temperature sensitivity caused by mutant release factor 1 is suppressed by mutations that affect 16S rRNA maturation. *J Bacteriol* **186**: 3046–3055
- Kaczanowska M, Rydén-Aulin M (2005) The YrdC protein—a putative ribosome maturation factor. *Biochim Biophys Acta* **1727**: 87
- Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protocols* **4**: 363
- Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* Genome sequence. *Genome Res* **8**: 464–478
- Kisseleva-Romanova E, Lopreiato R, Baudin-Baillieu A, Rousselle JC, Ilan L, Hofmann K, Namane A, Mann C, Libri D (2006) Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J* **25**: 3576–3585
- Kolitz SE, Takacs JE, Lorsch JR (2009) Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* **15**: 138–152
- Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688–6719
- Korner A, Söll D (1974) N-(purin-6-ylcarbamoyl)threonine: biosynthesis *in vitro* in transfer RNA by an enzyme purified from *Escherichia coli*. *FEBS Lett* **39**: 301–306
- Lee C, Kramer G, Graham DE, Appling DR (2007) Yeast mitochondrial initiator tRNA is methylated at guanosine 37 by the Trm5-encoded tRNA (Guanine-N1-)-methyltransferase. *J Biol Chem* **282**: 27744–27753
- Lesclapier E, Nauwelaerts K, Zanier K, Poesen K, Sattler M, Herdewijn P (2006) The naturally occurring N6-threonyl adenine in anticodon loop of *Schizosaccharomyces pombe* tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res* **34**: 2878–2886
- Lin CA, Ellis SR, True HL (2009) The Sua5 protein is essential for normal translational regulation in yeast. *Mol Cell Biol* **30**: 354–363
- Lopreiato R, Facchin S, Sartori G, Arrigoni G, Casonato S, Ruzzene M, Pinna LA, Carignani G (2004) Analysis of the interaction between pD261/Bud32, an evolutionarily conserved protein kinase of *Saccharomyces cerevisiae*, and the Grx4 glutaredoxin. *Biochem J* **377** (Part 2): 395–405
- Mao DY, Neculai D, Downey M, Orlicky S, Haffani YZ, Ceccarelli DF, Ho JSL, Szilard RK, Zhang W, Ho CS, Wan L, Fares C, Rumpel S, Kurinov I, Arrowsmith CH, Durocher D, Sicheri F (2008) Atomic Structure of the KEOPS complex: an ancient protein kinase-containing molecular machine. *Mol Cell* **32**: 259
- Matsui A, Matsuura A (2010) Cell size regulation during telomere-directed senescence in *Saccharomyces cerevisiae*. *Biosci Biotechnol Biochem* **74**: 195–198
- McCutcheon JP, McDonald BR, Moran NA (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* **106**: 15394–15399
- McCutcheon JP, Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci U S A* **104**: 19392–19397
- Meng F-L, Hu Y, Shen N, Tong X-J, Wang J, Ding J, Zhou J-Q (2009) Sua5p a single-stranded telomeric DNA-binding protein facilitates telomere replication. *EMBO J* **28**: 1466
- Morris DK, Lundblad V (1997) Programmed translational frame-shifting in a gene required for yeast telomere replication. *Curr Biol* **7**: 969
- Murphy FVT, Ramakrishnan V, Malkiewicz A, Agris PF (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat Struct Mol Biol* **11**: 1186–1191

- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**: 267
- Natsoulis G, Hilger F, Fink GR (1986) The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of *Saccharomyces cerevisiae*. *Cell* **46**: 235–243
- Nichols CE, Johnson C, Lockyer M, Charles IG, Lamb HK, Hawkins AR, Stammers DK (2006) Structural characterization of *Salmonella typhimurium* YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog. *Proteins* **64**: 111–123
- Nonekowsky ST, Kung F-L, Garcia GA (2002) The *Escherichia coli* tRNA-guanine transglycosylase can recognize and modify DNA. *J Biol Chem* **277**: 7178–7182
- Numata T, Fukai S, Ikeuchi Y, Suzuki T, Nureki O (2006) structural basis for sulfur relay to RNA mediated by heterohexameric TusBCD complex. *Structure* **14**: 357–366
- Oberto J, Breuil N, Hecker A, Farina F, Brochier-Armanet C, Culetto E, Forterre P (2009) Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance. *Nucleic Acids Res* **37**: 5343–5352
- Paschos A, Bauer A, Zimmermann A, Zehelein E, Böck A (2002) HypF, a carbamoyl phosphate-converting enzyme involved in [NiFe] hydrogenase maturation. *J Biol Chem* **277**: 49945–49951
- Powers DM, 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
- Rajagopala S, Yamamoto N, Zweifel A, Nakamichi T, Huang H-K, Mendez-Rios J, Franca-Koh J, Boorgula M, Fujita K, Suzuki K-i, Hu J, Wanner B, Mori H, Uetz P (2010) The *Escherichia coli* K-12 ORFeome: a resource for comparative molecular microbiology. *BMC Genomics* **11**: 470
- Reissmann S, Hochleitner E, Wang H, Paschos A, Lottspeich F, Glass RS, Böck A (2003) Taming of a poison: biosynthesis of the NiFe-hydrogenase cyanide ligands. *Science* **299**: 1067–1070
- Ritchie KB, Petes TD (2000) The Mre11p/Rad50p/Xrs2p complex and the Tel1p function in a single pathway for telomere maintenance in yeast. *Genetics* **155**: 475–479
- Rosado INV, De La Cruz J (2004) Np1p is an essential trans-acting factor required for an early step in the assembly of 60S ribosomal subunits in *Saccharomyces cerevisiae*. *RNA* **10**: 1073–1083
- Rosano C, Zuccotti S, Bucciantini M, Stefani M, Ramponi G, Bolognesi M (2002) Crystal structure and anion binding in the prokaryotic hydrogenase maturation factor HypF acylphosphatase-like domain. *J Mol Biol* **321**: 785
- Schneider TD, Stormo GD, Gold L, Ehrenfeucht A (1986) Information content of binding sites on nucleotide sequences. *J Mol Biol* **188**: 415–431
- Souciet G, Menand B, Ovesna J, Cosset A, Dietrich A, Wintz H (1999) Characterization of two bifunctional *Arabidopsis thaliana* genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. *FEBS J* **266**: 848–854
- Stahl G, Bidou L, Rousset JP, Cassan M (1995) Versatile vectors to study recoding: conservation of rules between yeast and mammalian cells. *Nucleic Acids Res* **23**: 1557–1560
- Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002) Systematic screen for human disease genes in yeast. *Nat Genet* **31**: 400–404
- Stuart JW, Gdaniec Z, Guenther R, Marszalek M, Sochacka E, Malkiewicz A, Agris PF (2000) Functional anticodon architecture of human tRNA^{Lys3} includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, ^t6A. *Biochemistry* **39**: 13396–13404
- Takata H, Kanoh Y, Gunge N, Shirahige K, Matsuura A (2004) Reciprocal association of the budding yeast ATM-related proteins Tel1 and Mec1 with telomeres *in vivo*. *Mol Cell* **14**: 515–522
- Tamames J, Gil R, Latorre A, Pereto J, Silva FJ, Moya A (2007) The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol Biol* **7**: 181
- Thoden JB, Holden HM, Fisher AJ, Sinclair JF, Wesenberg G, Baldwin TO, Rayment I (1997) Structure of the β 2 homodimer of bacterial luciferase from *Vibrio harveyi*: X-ray analysis of a kinetic protein folding trap. *Protein Sci* **6**: 13–23
- Umeda N, Suzuki T, Yukawa M, Ohya Y, Shindo H, Watanabe K, Suzuki T (2005) Mitochondria-specific RNA-modifying enzymes responsible for the biosynthesis of the wobble base in mitochondrial tRNAs. *J Biol Chem* **280**: 1613–1624
- Venema J, Tollervey D (1995) RRP5 is required for formation of both 18S and 5.8S rRNA in yeast. *EMBO J* **15**: 5701–5714
- Weissenbach J, Grosjean H (1981) Effect of threonylcarbamoyl modification (^t6A) in yeast tRNA Arg III on codon-anticodon and anticodon-anticodon interactions. A thermodynamic and kinetic evaluation. *Eur J Biochem* **116**: 207–213
- Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng JF (2010) One bacterial cell, one complete genome. *PLoS One* **5**: e10314
- Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J (1995) A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). *RNA* **1**: 437–448
- Xu Q, McMullan D, Jaroszewski L, Krishna SS, Elsliger M-A, Yeh AP, Abdubek P, Astakhova T, Axelrod HL, Carlton D, Chiu H-J, Clayton T, Duan L, Feuerhelm J, Grant J, Han GW, Jin KK, Klock HE, Knuth MW, Miller MD *et al* (2010) Structure of an essential bacterial protein YeaZ (TM0874) from *Thermotoga maritima* at 2.5 Å resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**: 1230–1236
- Yamada M, Hayatsu N, Matsuura A, Ishikawa F (1998) Y'-Help1, a DNA helicase encoded by the yeast subtelomeric Y' element, is induced in survivors defective for telomerase. *J Biol Chem* **273**: 33360–33366
- Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF (2000) Modified nucleoside dependent Watson-Crick and Wobble codon binding by tRNA^{LysUUU} species. *Biochemistry* **39**: 13390
- Yoshihiro A, Shinya S, Taisuke W, Yoshitaka B, Akio E, Shigeyuki Y, Seiki K, Akeo S (2008) X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins* **70**: 1108–1111