

# Identification of a Novel Epoxyqueuosine Reductase Family by Comparative Genomics

Rémi Zallot,<sup>\*,†</sup> Robert Ross,<sup>‡</sup> Wei-Hung Chen,<sup>§</sup> Steven D. Bruner,<sup>§</sup> Patrick A. Limbach,<sup>‡</sup> and Valérie de Crécy-Lagard<sup>\*,†</sup>

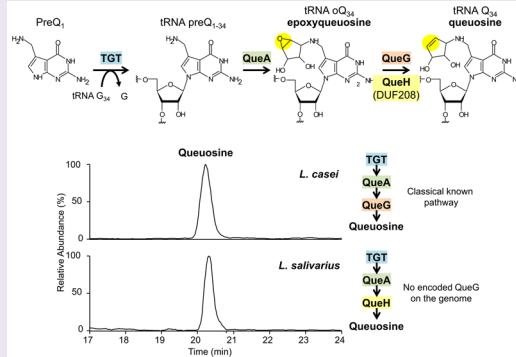
<sup>†</sup>Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611, United States

<sup>‡</sup>Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221, United States

<sup>§</sup>Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States

## S Supporting Information

**ABSTRACT:** The reduction of epoxyqueuosine (oQ) is the last step in the synthesis of the tRNA modification queuosine (Q). While the epoxyqueuosine reductase (EC 1.17.99.6) enzymatic activity was first described 30 years ago, the encoding gene *queG* was only identified in *Escherichia coli* in 2011. Interestingly, *queG* is absent from a large number of sequenced genomes that harbor Q synthesis or salvage genes, suggesting the existence of an alternative epoxyqueuosine reductase in these organisms. By analyzing phylogenetic distributions, physical gene clustering, and fusions, members of the Domain of Unknown Function 208 (DUF208) family were predicted to encode for an alternative epoxyqueuosine reductase. This prediction was validated with genetic methods. The Q modification is present in *Lactobacillus salivarius*, an organism missing *queG* but harboring the *duf208* gene. *Acinetobacter baylyi* ADP1 is one of the few organisms that harbor both *QueG* and DUF208, and deletion of both corresponding genes was required to observe the absence of Q and the accumulation of oQ in tRNA. Finally, the conversion oQ to Q was restored in an *E. coli queG* mutant by complementation with plasmids harboring *duf208* genes from different bacteria. Members of the DUF208 family are not homologous to *QueG* enzymes, and thus, *duf208* is a non-orthologous replacement of *queG*. We propose to name DUF208 encoding genes as *queH*. While *QueH* contains conserved cysteines that could be involved in the coordination of a Fe/S center in a similar fashion to what has been identified in *QueG*, no cobalamin was identified associated with recombinant *QueH* protein.



Queuosine (Q) is a complex modification of the base guanosine at the first position of tRNAs with the GUN anticodon (His, Tyr, Asn, and Asp). Q was discovered in the late 1960s,<sup>1–3</sup> and its structure was established eight years later.<sup>4</sup> The role of Q remained elusive for years, and there has been a renewed interest in this molecule, particularly as it appears to be a key micronutrient in humans.<sup>5–8</sup>

Q is found in eubacteria and eukaryotes, but only eubacteria are able to synthesize it.<sup>9</sup> However, not all bacteria are capable of Q *de novo* synthesis, and several rely on salvage of Q precursor(s) for the presence of Q in their tRNA.<sup>10</sup> Whether salvage or biosynthesis occurs, the tRNA-guanine(34) transglycosylase (EC 2.4.2.29), TGT, is the key enzyme in the pathway, as it is responsible for the exchange of guanine at position 34 of target tRNAs, with the precursor base preQ.<sup>11</sup> Two additional steps are required to finalize the synthesis (Figure 1A). In the penultimate step, epoxyqueuosine (oQ) is synthesized by the S-adenosylmethionine ribosyltransferase-isomerase enzyme *QueA* (EC 2.4.99.17).<sup>12</sup> The last step of the pathway is the conversion of oQ to Q. oQ was identified as an intermediate in the Q pathway in 1987.<sup>13</sup> It was shown a year

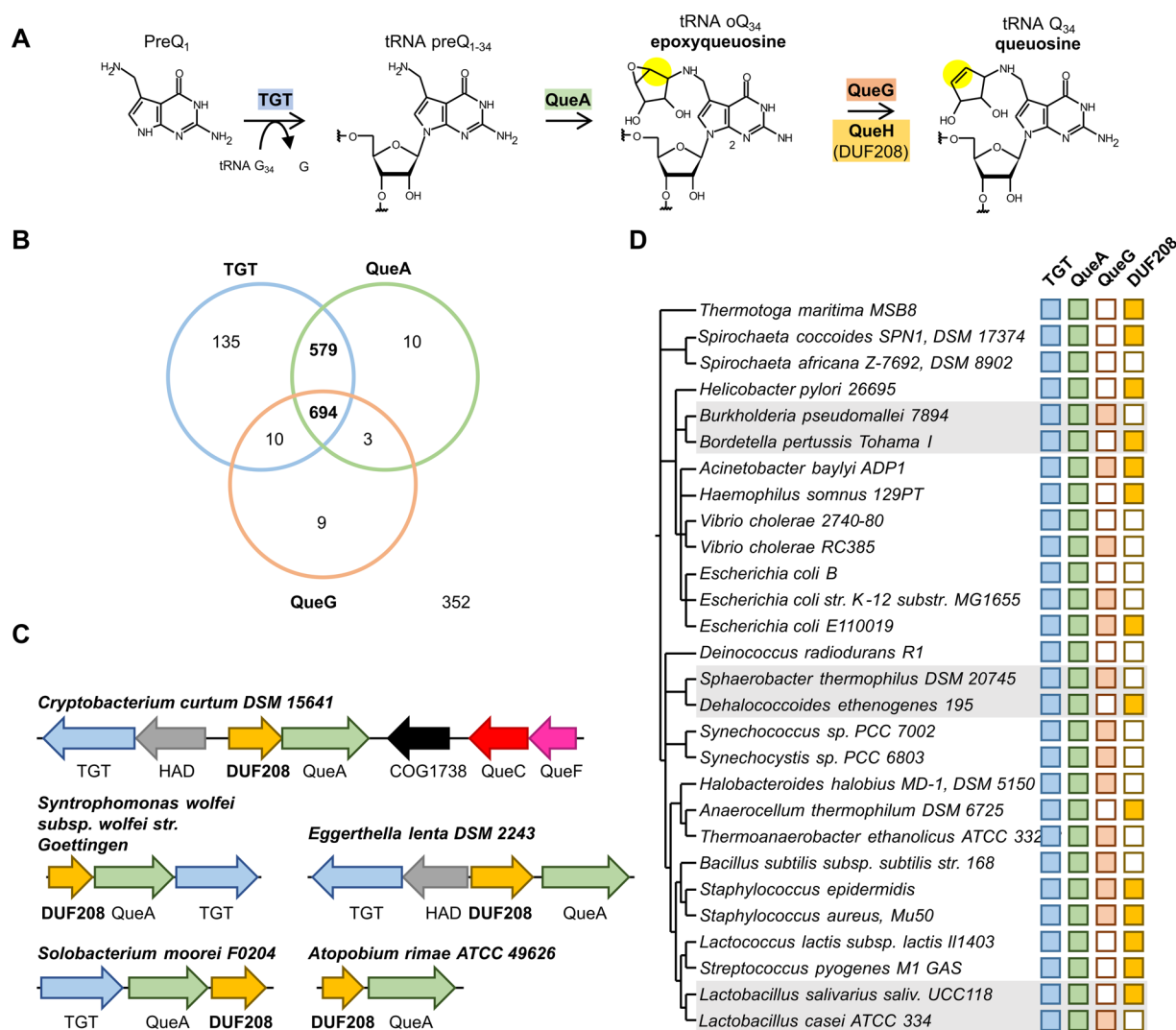
later, that the oQ reduction was performed by a cobalamin-dependent epoxyqueuosine reductase (EC 1.17.99.6) in *Escherichia coli*, at the time named *QueB*.<sup>9</sup> It then took 25 years for the epoxyqueuosine reductase activity to be assigned to the *yjeS* gene, as a result of a systematic screen of the *E. coli* Keio collection, and the enzyme was renamed *QueG*.<sup>14</sup> The *QueG* protein is homologous to B12-dependent iron–sulfur proteins involved in halorespiration.<sup>14</sup> *In vitro* activity was demonstrated with the *Bacillus subtilis* orthologous enzyme,<sup>14</sup> and other studies confirmed the Fe/S cluster role and the cobalamin requirement including for the orthologous enzyme from *Streptococcus thermophilus*.<sup>15–17</sup>

In this study, we show that not all organisms that harbor *tgt* and *queA* genes also harbor *queG* homologues and identify, using comparative genomics approaches combined with genetic

Received: December 12, 2016

Accepted: January 27, 2017

Published: January 27, 2017



**Figure 1.** DUF208 is identified as a possible replacement for QueG by comparative genomics. (A) The three last steps of the Queuosine biosynthesis pathway. (B) Venn diagram showing the distribution of the genes encoding TGT, QueA, and QueG. Only 55% of the genomes that have TGT and QueA have QueG. We hypothesized the existence of an unknown enzyme that could replace QueG. (C) DUF208 is often found clustered with genes involved in Q synthesis or salvage, including QueC, QueF, and COG1738 in genomes that do not have QueG. It is not known whether the haloacid dehydrogenase (HAD) is related to Q. (D) Phylogenetic distribution of TGT, QueA, QueG, and DUF208 show an inverted distribution pattern between QueG and DUF208, among various bacterial clades. Gray boxes highlight inversions among related organisms within a clade.

experiments, a non-orthologous family, DUF208, that replaces QueG in a large number of sequenced bacteria.

## RESULTS AND DISCUSSION

**Comparative Genomics Identify DUF208 As a Possible oQ Reductase.** Phylogenetic distribution of the three genes that encode the final steps of the Q biosynthesis pathway in eubacteria (*tgt*, *queA*, and *queG*) was analyzed in 1792 eubacterial genomes using the Microbes Online “Phylogenetic Profile for Cart” tool.<sup>18</sup> A total of 71% (1273) of the genomes analyzed harbor *tgt* and *queA* homologues. However, a *queG* ortholog can be identified in only 55% (694) of these genomes (Figure 1B and Supporting Information File 1). Because of the limited experimental data on Q presence in bacterial tRNAs from different sources,<sup>19</sup> it is not known if the bacteria that lack *queG* but harbor *tgt* and *queA* harbor Q or oQ in tRNA. The only published case of an organism that lacks *queG* and has lost the capacity to convert oQ to Q is *E. coli* B,<sup>20</sup> but this looks like a localized gene loss event. Therefore, we

postulated the existence of a gene that could encode an alternative epoxyqueuosine reductase among organisms lacking *queG*.

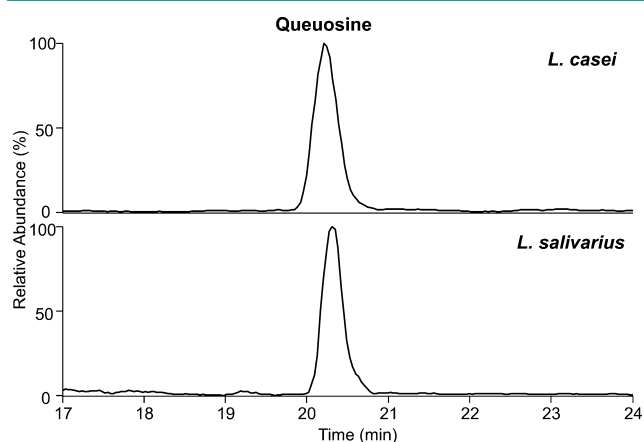
Because genes in a given pathway tend to physically cluster in bacterial genomes,<sup>21</sup> we investigated the gene neighborhoods of *tgt* and *queA* in organisms lacking *queG*. Using the SEED database and its tools,<sup>22</sup> we found that genes belonging to the family of Domain of Unknown Function 208 (DUF208) (also identified as members of COG1636 or Pfam02677 families) consistently clustered with genes involved in Q synthesis and salvage (Figure 1C). In addition, using the Pfam domain architecture tool (<http://pfam.xfam.org/family/PF02677#tabview=tab1>),<sup>23</sup> fusions between the DUF208 and QueA encoding genes were identified in an *Eggerthella* species (see Uniprot ID R7BMR5 as an example), reinforcing the link between DUF208 and the Q biosynthesis pathway.

To analyze the phylogenetic distribution of TGT, QueA, DUF208, and QueG families in eubacterial genomes, the “Phylogenetic Profile for Cart” from the Microbes online

database<sup>18</sup> was used (with DUF208 from *Streptococcus mitis* B6 as an entry point). Interestingly, DUF208 and QueG showed an inverted distribution pattern (Figure 1D and Supporting Information File S1). Several cases of closely related organisms within a given clade where one had QueG and the DUF208 could be identified (for example, among Burkholderiales, Spirocheteataceae, and Lactobacillus). A few organisms were also found to harbor both gene families (Acinetobacter, Staphylococcus, and Xanthomonas, for example, as seen in Figure 1D and Supporting Information File S1).

**Queuosine Is Detected in *L. salivarius*, an Organism Lacking QueG.** As mentioned above, the genus Lactobacillus is interesting as some species harbor QueG homologues and some do not. For example, while a QueG ortholog is encoded by the *Lactobacillus casei* ATCC 334 genome (YP\_807405), it is absent in *Lactobacillus salivarius* sp. *salivarius* ATCC 11741 that harbors a DUF208 family protein (WP\_035149309; Figure 1D). If our hypothesis is correct, both strains should harbor Q in their tRNAs. Both these organisms rely on salvage as they lack the Q precursor genes but have *tgt* and *queA*.<sup>10</sup> The cells were hence grown in a rich media, MRS, typically used for Lactobacillus growth.

As expected, Q was detected in a tRNA extracted from the positive control, *L. casei*, confirming the MRS medium contains a source of Q precursor (Figure 2). Q was also detected in



**Figure 2.** Queuosine (Q) detection by LC-MS in hydrolyzed bulk tRNA from *L. salivarius* and *L. casei*. The typical extracted ion chromatogram (XIC) at  $m/z$  410 for  $[M+H]^+$  for Q is detected in both organisms when only *L. casei* harbors a *queG* gene. The results of a typical experiment are presented.

tRNA extracted from *L. salivarius* that lacks a QueG homologue. This is the first time to our knowledge that Q is experimentally detected in an organism lacking QueG, and thus, another protein family must be performing the same reaction in this organism. The comparative genomic evidence presented above, combined with the presence of a *duf208* gene in *L. salivarius*, led us to propose that the DUF208 family fills that role, and we set out to test this hypothesis experimentally in two different genetic systems.

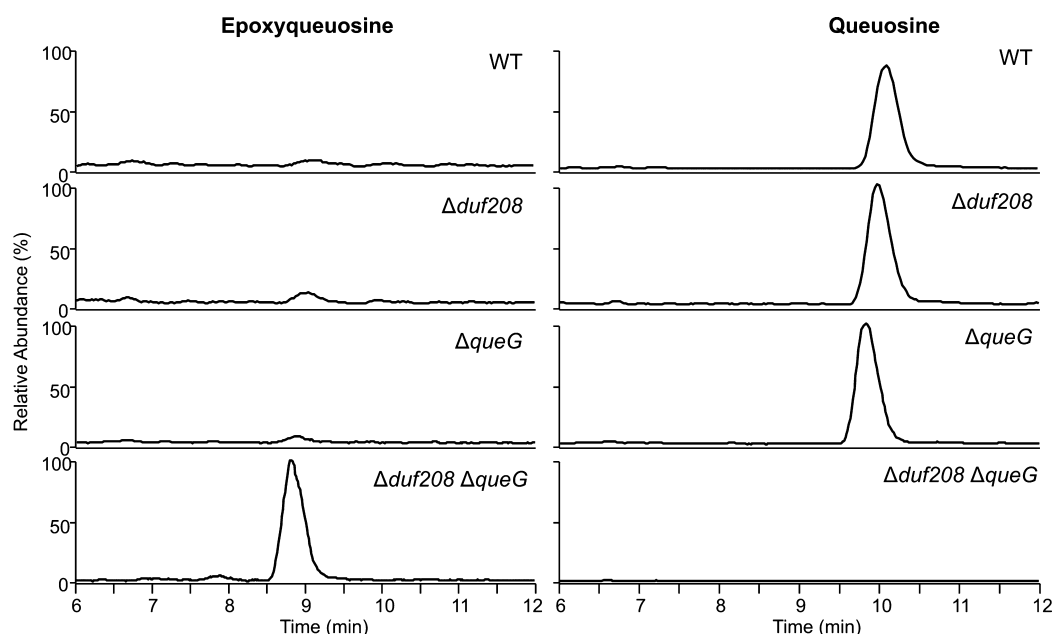
**DUF208 Is Needed for Epoxyqueuosine Reduction in *Acinetobacter baylyi* ADP1  $\Delta queG$ .** Because genetic manipulation of *L. salivarius* is far from straightforward, we decided to use another more genetically tractable model to test the hypothesis that DUF208 is the locally missing epoxy-Q reductase in many bacteria. *A. baylyi* is among the organisms that harbor both QueG and DUF208 encoding genes. It is very

easy to manipulate,<sup>24</sup> and a mutant collection is available.<sup>25</sup> As shown in Figure 3, Q was detected in tRNA extracted from the WT strain as well as from the single deletion strains ( $\Delta duf208$  or  $\Delta queG$ ). Accumulation of oQ was observed only in the  $\Delta duf208 \Delta queG$  strain, with the concomitant disappearance of Q. These results demonstrate that at least in *A. baylyi*, *duf208* encodes a protein involved in the formation of Q from oQ even in the absence of QueG. This role, however, could be indirect.

**In Vivo Functional Complementation of  $\Delta QueG$  *Escherichia coli* by DUF208.** To show that the DUF208 proteins are directly involved in the oQ to Q conversion, we turned to a heterologous complementation test. *Escherichia coli* BW25113 (WT strain of the Keio collection) encodes a QueG protein<sup>14</sup> but does not encode a DUF208 protein (Figure 1C). A functional complementation assay for the oQ reductase activity was hence developed in an *E. coli* strain carrying a deletion of the *queG* gene. As expected, tRNA extracted from the *E. coli*  $\Delta queG$  only harbors oQ (Figure 4). Expression of the native *E. coli queG* gene, in trans, restored the Q<sup>+</sup> phenotype (Figure 4). It was then tested whether expressing *duf208* genes from different species would restore the oQ to Q conversion, absent in the  $\Delta queG$  strain. As seen in Figure 4, complementation was observed with plasmids expressing *duf208* genes from very phylogenetically distant organisms including *Bordetella pertussis*, *Dehalococcoides ethenogenes* 195, *Haemophilus somnus* 129PT, *Staphylococcus aureus* subsp. *aureus* Mu50, *Staphylococcus epidermidis* ATCC 12228, *Streptococcus pyogenes*, *Thermotoga maritima*, and *Helicobacter pylori* 26695. The complementation efficiency did vary with the clone tested, as various levels of remaining oQ were seen in a few cases. These experiments, in combination with the *A. baylyi* ADP1 mutant analyses described above, clearly establish that *duf208* genes can functionally replace *queG* in vivo.

**Bioinformatics Analysis of the DUF208 Family.** The combination of comparative genomic and genetic evidence presented above strongly suggest that members of the DUF208 family functionally replace QueG in many bacteria. Previous literature on this family is very limited. DUF208 encoding genes had been reported to be found often physically clustered with either *yebC* or *ruwABC* genes involved in Holliday junction resolution, and a possible role of this family in Holliday junction resolution was proposed.<sup>26</sup> The *T. maritima* DUF208 protein (TM0731) was found to be insoluble when expressed in *E. coli* in studies aimed at understanding misfolded proteins.<sup>27,28</sup>

DUF208 proteins are not homologous to QueG proteins. *A priori*, no specific determinant for tRNA binding can be identified in DUF208 sequences. In the COG database, DUF208 (identified as COG1636) is annotated as “Predicted ATPase, Adenine nucleotide alpha hydrolases (AANH) superfamily [General function prediction only]” (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=COG1636>). Other members of the AANH superfamily are involved in tRNA modification. It is the case for MnmA,<sup>29</sup> involved in the formation of 5-methylaminomethyl-2-thiouridine (mnm<sup>5</sup>s<sup>2</sup>U), a modification at the wobble position of tRNAs; QueC, an enzyme of the Q biosynthesis pathway;<sup>30</sup> and ThiI, an enzyme involved in the formation of 4-thiouridine (s4U).<sup>31</sup> See <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=c100292> for the complete list. Hence, even if DUF208 is not similar to QueG, and does not harbor any known tRNA binding domain, a role in tRNA modification is clearly possible.



**Figure 3.** Epoxyqueuosine (oQ) and Queuosine (Q) detection by LC-MS in hydrolyzed bulk tRNA from *Acinetobacter baylyi* and derivatives. The typical XICs at  $m/z$  426 for oQ and  $m/z$  410 for Q were monitored in the WT,  $\Delta duf208$ ,  $\Delta queG$ , and  $\Delta duf208 \Delta queG$  strains. The results of a typical experiment are presented.

The sequence alignment also shows the conservation of six cysteine residues among various representative DUF208 proteins (Supporting Information Figure S2). This arrangement of conservation of cysteine residues is commonly associated with chelation of Fe/S clusters and redox capacity, as with QueG, and could provide for the two electron reduction of oQ to Q.<sup>14,15</sup> Cysteines could also be involved in other metal coordination,<sup>32</sup> as seen with the zinc containing QueC,<sup>33</sup> where the metal is proposed to play a structural (rather than catalytic) role.<sup>34</sup> Vitamin K epoxide reductases (VKORs) are integral membrane proteins that catalyze a similar reaction with the involvement of several cysteine residues and reducing equivalent from a thioredoxin-like domain, which is involved in the reduction of epoxides.<sup>35</sup>

No structural data are available for any member of the DUF208 family, but to try and predict further the role of the conserved cysteines, models were generated using Robetta<sup>36</sup> and I-TASSER.<sup>37</sup> Both programs converged to a model where the six conserved cysteines were clustered in groups of four and two (Supporting Information Figure S3). This is consistent with a 4( $\gamma$ -Cys)4Fe4S cluster and a pair of cysteines functioning in a mechanism similar to VKORs (Figure 5). Whether DUF208 proteins require a redox partner like VKORs remains uncertain.

**Cobalamin Seems Not to Be Associated with DUF208 Proteins.** In order to validate the oQ reductase activity of DUF208 proteins *in vitro*, we screened for overexpression and purification of candidate DUF208 proteins in *E. coli*, in diverse conditions and strains. Initial trials were unsuccessful, with insolubility issues arising, in accordance with previous reports on TmDUF208 (identified TM0731) cloned into pMH1.<sup>27,28</sup> However, recloning DUF208 encoding genes in pET28b to obtain C-terminal hexahistidine tag protein fusions reduced insolubility and allowed SpDUF208 and TmDUF208 to be partially in the soluble fraction.

Recombinant histidine-tagged proteins were partially purified using immobilized Ni<sup>2+</sup> affinity chromatography followed by an

anion exchange and size exclusion chromatography. Only small protein quantities could be obtained from large initial cultures, and the partially purified proteins (Supporting Information Figure S4) were shown to precipitate shortly after being concentrated. In all cases, eluted proteins showed a yellow-brown color, typical of iron–sulfur cluster containing proteins.

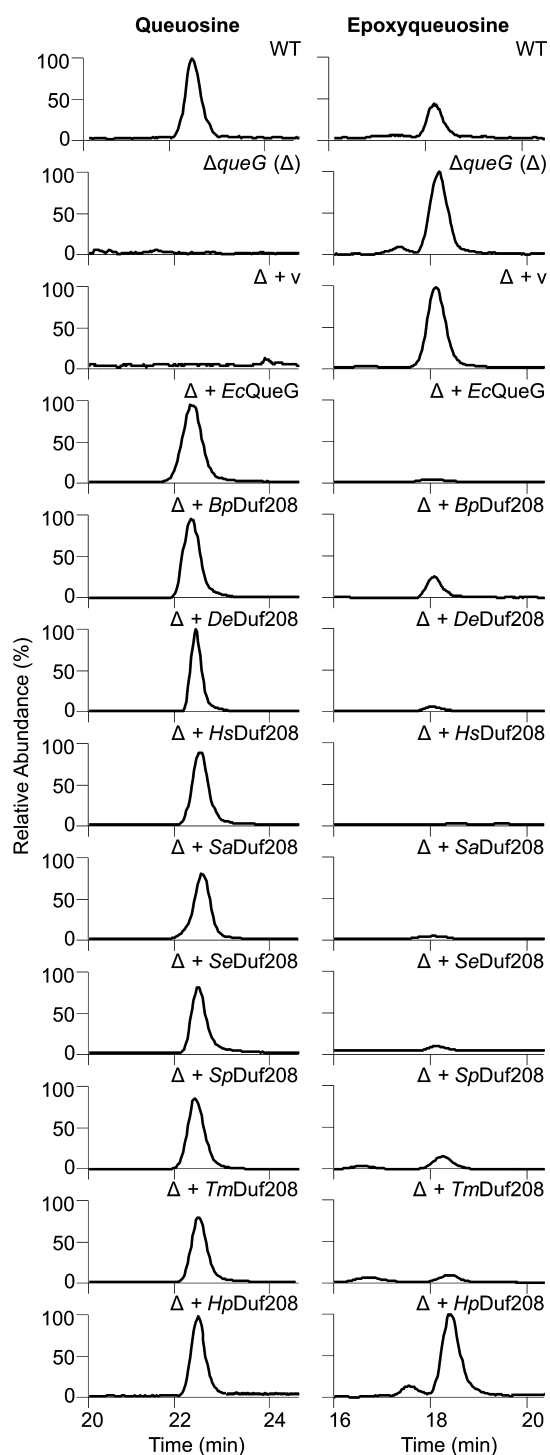
Because QueG is known to require cobalamin for its activity,<sup>15,16</sup> we wanted to test whether DUF208 proteins are also found associated with cobalamin. An approach previously used for QueG,<sup>16</sup> that required a small amount of protein, was used: purified proteins DUF208 from *T. maritima* (TmDUF208) or *S. pyogenes* (SpDUF208) were precipitated in the presence of potassium cyanide. No compound harboring spectral properties similar to hydroxocobalamin, used as a positive control, was identified in the supernatant (Supporting Information Figure S5). We were not able to detect cobalamin associated with these DUF208 proteins.

## CONCLUSIONS

A combination of comparative genomics and genetic analyses shows that members of the DUF208 family replace QueG as the missing epoxyqueuosine reductase. Because DUF208s are not similar in sequence, this is a clear case of non-orthologous replacement of QueG, and we propose to rename this family QueH. This finding greatly reduces the number of genomes that are missing the enzyme catalyzing the last step from 45% to 19% in the set of genomes analyzed (Supporting Information Figure S1). This work also reinforces the powers of using comparative genomic based associations to predict the function of unknowns.<sup>38,39</sup>

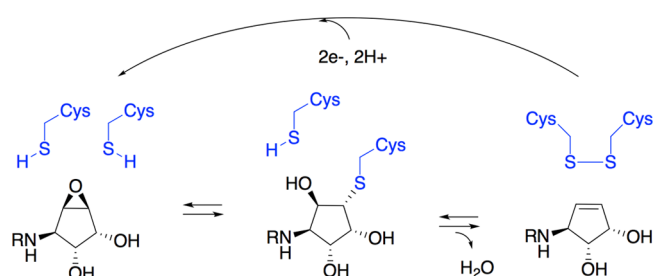
Unfortunately, our attempts to confirm the epoxyqueuosine reductase activity of QueH proteins *in vitro* have failed, and strategies that use anaerobic conditions that might help with solubility and activity are ongoing.

The absence of cobalamin associated with DUF208 suggests a reaction mechanism unique from QueG. We propose a mechanism where two cysteines acts in a similar way to



**Figure 4.** Queuosine (Q) and Epoxyqueuosine (oQ) detection by LC-MS in digested tRNA from *Escherichia coli* WT,  $\Delta queG$ , or  $\Delta queG$  transformed with pBAD24 or pBAD24 carrying *duf208* genes from various bacteria. The typical XICs at  $m/z$  426 for oQ and  $m/z$  410 for Q were monitored. The WT strain accumulates Q, whereas oQ only can be detected in the  $\Delta queG$  (abbreviated  $\Delta$ ) and  $\Delta queG$  carrying the empty vector pBAD24 (abbreviated  $\Delta + (v)$ ). All the genes *duf208* from *B. pertussis*, *D. ethenogenes*, *H. somnus*, *S. aureus*, *S. epidermidis*, *S. pyogenes*, *T. maritima*, and *H. pylori* allows the restoration of the presence of Q into the  $\Delta queG$  strain when present on a pBAD24 plasmid.

VKORs, and the reducing equivalents are transferred from an iron/sulfur cluster (Figure 5). The fact that QueG is cobalamin-



**Figure 5.** Proposed reaction mechanism for the reduction of epoxyqueuosine to queuosine by QueH, in a cobalamin independent manner, based on the VKORs proposed mechanism. For clarity, only the cyclopentane moiety is shown, and the rest of the epoxyqueuosine, queuosine, and the intermediate molecules are abbreviated “R.”

dependent, whereas QueH apparently is not, may explain the inverted distribution pattern and the fact that some organisms have both. It is not the first time that non-orthologous enzymes with different requirements for cobalamin have been identified in bacteria. The best known examples are the two methionine synthase families: the cobalamin-dependent MetH and the cobalamin-independent MetE. Their presence/absence is correlated with the B12 synthesis or salvage capabilities of a given organism.<sup>40</sup>

## MATERIALS AND METHODS

**Comparative Genomics.** The BLAST tools<sup>41</sup> and resources at NCBI (<http://www.ncbi.nlm.nih.gov>) were routinely used. Multiple sequence alignments were built using Clustal Omega<sup>42</sup> or Multalin.<sup>43</sup> Protein domain analysis was performed using the Pfam database tools (<http://pfam.xfam.org>).<sup>44</sup> Analysis of the phylogenetic distribution and physical clustering was performed in the SEED<sup>22</sup> and Microbes online (<http://www.microbesonline.org/>)<sup>18</sup> databases. Physical clustering was analyzed with the SEED subsystem coloring tool or the Seedviewer Compare region tool<sup>22</sup> as well as on the Microbesonline tree based genome browser.<sup>18</sup> Mapping of gene distribution profile to taxonomic trees was generated using the iTOL suite (<http://itol.embl.de>).<sup>45</sup>

**Strains and Growth Conditions.** *Lactobacillus salivarius* sp. *salivarius* ATCC 11741 was obtained from ATCC. *Lactobacillus casei* ATCC 334 was kindly provided by Dr. Graciela Lorca (University of Florida). *Lactobacillus* strains were grown in de Man, Rogosa, and Sharpe (MRS) broth (Thermo Scientific Remel) at 37 °C in still-standing closed 15 mL conical tubes.

*E. coli* strains were grown in Luria–Bertani medium (LB) at 37 °C. Media were solidified with 15 g/L of agar. Kanamycin (Kan, 50  $\mu$ g/mL), ampicillin (Amp, 100  $\mu$ g/mL), spectinomycin (Sp, 50  $\mu$ g/mL), and chloramphenicol (Cm, 25  $\mu$ g/mL) were added as required. *E. coli* strain  $\Delta queG::Kan^R$  (JW4124, abbreviated  $\Delta queG$ ) and the corresponding wild type strain BW25113 were from the KEIO collection.<sup>46</sup> These strains were verified by PCR (primers listed in Table S1). A  $\Delta queG$  strain that had lost Kan resistance was used for *in vivo* complementation. Briefly, the  $\Delta queG$  strain was first transformed with pCP20 to eliminate the  $Kan^R$  cassette by F1p recombinase.<sup>47</sup> Transformants were selected on LB plus Amp at 30 °C, restreaked twice on LB, grown at 42 °C, and then checked for the absence of Kan and Amp resistance.

*Acinetobacter baylyi* ADP1 strains WT and  $\Delta duf208::Kan^R/tdk$  (abbreviated  $\Delta duf208$ ; gene ACIAD2098) were kindly provided by Dr. Véronique de Berardinis (GENOSCOPE, Institut de Genomique, CEA).<sup>25</sup> The single mutant  $\Delta queG::Tet^R/tdk$  (abbreviated  $\Delta queG$ ; gene ACIAD2043) and the double mutant  $\Delta duf208::Kan^R/tdk \Delta queG::Tet^R/tdk$  (abbreviated  $\Delta duf208 \Delta queG$ ) were constructed in the WT strain and in the  $\Delta duf208::Kan^R/tdk$  strain following established protocols, by PCR amplification of the  $Tet^R/tdk$  cassette from  $\Delta ilvC::Tet^R/tdk$ <sup>24</sup> and overlapping fragment of genomic DNA at

the 5' and 3' of the *queG* (ACIAD2043) gene, followed by sawing, transformation, and selection on LB containing tetracycline (Tet-20  $\mu\text{g}/\text{mL}$ ).<sup>24</sup> All strains were verified by PCR and growth phenotype on LB with appropriate antibiotics. For tRNA preparation, strains were grown overnight at 30 °C. Primers used for amplifications are listed in Supporting Information Table S1.

**Complementation Vectors.** *E. coli queG* (EcQueG) was amplified from genomic DNA. Primers used for amplifications are listed in Table S1. EcQueG was cloned into the *NcoI* and *HindIII* restriction sites of pBAD24<sup>48</sup> and used as a positive control for functional complementation assays. All the other complementation plasmids were obtained from DNASU Plasmid Repository at the The Biodesign Institute from Arizona State University (<https://dnasu.org/DNASU/Home.do>). They consisted of BpCD00282666 containing NP\_880031 from *Bordetella pertussis* and abbreviated BpDUF208, HsCD00091938 containing NC\_008309 from *Haemophilus somnus* 129PT and abbreviated HsDUF208, TmCD00090076 containing NP\_228540 from *Thermotoga maritima* and abbreviated TmDUF208, DeCD00093123 containing YP\_181349 from *Dehalococcoides ethenogenes* 195 and abbreviated DeDUF208, HpCD00396352 containing NP\_206900 from *Helicobacter pylori* 26695 and abbreviated HpDUF208, SpCD00283267 containing NP\_268593 from *Streptococcus pyogenes* and abbreviated SpDUF208, SeCD00085816 containing NP\_765690 from *Staphylococcus epidermidis* ATCC 12228 and abbreviated SeDUF208, and SaCD00282781 containing NP\_647328 from *Staphylococcus aureus* subsp. *aureus* Mu50 and abbreviated SaDUF208. The DUF208 genes encoded on these plasmids were all verified by sequencing.

**Functional Complementation Experiments.** The *E. coli*  $\Delta queG$  deletant strain was transformed with pBAD24 alone (negative control) or with pBAD24 containing EcQueG (positive control), BpDUF208, HsDUF208, TmDUF208, DeDUF208, HpDUF208, SpDUF208, SeDUF208, or SaDUF208. Complementation tests were made by growing selected transformants in LB broth containing the appropriate antibiotic for plasmid selection and 0.2% arabinose at 37 °C until OD<sub>600 nm</sub> reached 0.6.

**Preparation of tRNA Enriched Fractions.** Cells from a 10 mL culture of *E. coli*, *A. baylyi*, and *Lactobacillus* species were harvested by centrifugation for 10 min at 3700g at 4 °C. Pellets were resuspended in 1 mL of Trizol (Life Technologies) and kept on a rocking shaker at moderate speed for 2 to 4 h before further processing. Small RNA were extracted using a Purelink miRNA Isolation kit (Invitrogen) according to the manufacturer's protocol. The purified RNA were eluted in 50  $\mu\text{L}$  of RNase free water. This extraction method was proven efficient for the purification of tRNA enriched fractions,<sup>49</sup> elsewhere referenced in this manuscript as tRNAs.

**Enzymatic Digestion and LC-MS Analysis of RNA Hydrolysates.** Purified tRNAs were denatured by heating for 3 min at 100 °C and then rapidly chilled. The purified tRNAs were buffered by adding a 1/10 volume of 0.1 M ammonium acetate. Next, 0.02U (units) of nuclease P1 was added for every microgram of tRNAs, and the mixture was incubated at 45 °C for 2 h. After incubation, a 1/10 volume of 1 M ammonium carbonate was added. Then,  $1.2 \times 10^{-4}$  U of snake venom phosphodiesterase (Phosphodiesterase I) and 0.03 U of bacterial alkaline phosphatase were added per microgram of tRNA. The solution was incubated at 37 °C for 2 h. Enzymes were separated from nucleosides with a 3 kDa filter (micropore) and lyophilized for storage at -80 °C.<sup>50</sup>

Lyophilized nucleosides samples were reconstituted in 5 mM ammonium acetate at pH 5.3. Two liquid chromatography systems were used for their separation. Retention times of oQ and Q vary between these systems. Nucleosides were separated with a Thermo Finnigan Surveyor HPLC system equipped with a Phenomenex Hydro-RP (100  $\times$  250) 3  $\mu\text{m}$  column and a gradient system composed with 5 mM ammonium acetate at pH 5.3 (mobile phase A) and 40% aqueous ACN (mobile phase B) at a flow rate of 100  $\mu\text{L}$  per min. The gradient consisted of 1% B at t0 and increasing at the following points: 5% at 5 min, 2% at 9 min, 3% at 11 min, 5% at 13 min, 25% at 32 min, 50% at 38 min, 75% at 43 min, 75% at 45 min, 99% at 50 min, 99% at 55 min, and 1% at 60 min. Nucleosides were also separated with a

Waters nanoAcquity UPLC system equipped with a Waters HSS T3 (1  $\times$  100 mm) 1.8  $\mu\text{m}$  column and a gradient system composed of 5 mM ammonium acetate at pH 5.3 (mobile phase A) and 30% aqueous ACN (mobile phase B) at a flow rate of 30  $\mu\text{L}$  per minute. The gradient consisted of 1% B at t0 and increasing at the following points, 1% B at 5 min, 7% B at 2 min, 10% B at 5 min, 25% B at 25 min, 50% B at 30 min, and 75% B at 35 min, with a hold for 2 min, and then 99% B at 40 min, with a hold for 5 min and then 1% B at 46 min.

For the detection of nucleosides, a Thermo LTQ-XL ion trap mass spectrometer equipped with an ion max electrospray source was used for the LC-MS identification of nucleosides. Mass spectra were recorded in positive polarity over an *m/z* range of 105–1200. The capillary temperature was set to 275 °C, spray voltage to 3.5 kV, tube lens to 150 V, and 30, 15, and 10 arbitrary flow units of sheath, auxiliary, and sweep gas, respectively. Extracted ion chromatograms for *m/z* 410 and *m/z* 426 were generated using Thermo Xcalibur 3.0 software.

**Cloning, Expression, and Purification of DUF208 Enzyme.** All the vectors obtained from the Biodesign Institute from Arizona State University (<https://dnasu.org/DNASU/Home.do>) were transformed into BL21-CodonPlus (DE3)-RIPL (Agilent), Rosetta 2(DE3) pLysS (EMD Millipore), and OverExpress C41(DE3) (Sigma) *E. coli* expression strains, and various conditions for protein expression were tested. Bacteria were lysed using a French press apparatus. Lysates were cleared by centrifugation for 20 min at 4 °C at 16 000g, and the soluble fractions were loaded and run on SDS-PAGE gels. Whereas candidate proteins can be visualized as overexpressed, no combination of strain, plasmid, and conditions was identified giving satisfactory soluble expression. It was decided to reclone the genes encoding the diverse DUF208 obtained (listed in the above section Complementation Vectors) in pET28b in fusion with a C-terminal hexahistidine tag. Primers used for amplifications are listed in Table S1. All plasmids were sequence validated, transformed into *E. coli* expression strains, and screened for protein expression and solubility.

DUF208 from *S. pyogenes*, *T. maritima*, *S. epidermidis*, and *S. aureus* cloned into pET28b were identified as partly soluble when grown at 37 °C until optical density reached 0.4–0.5 and then cooled down to 18 °C, induced with 250  $\mu\text{M}$  IPTG at optical density 0.7–0.8 for overnight induction. The overnight culture (1 L) was spun down and resuspended in 35 mL of lysis buffer (20 mM Tris at pH 7.5, 500 mM NaCl, 1 mM  $\beta\text{ME}$ ) and lysed with a microfluidizer at 14 000 psi through a nitrogen-pressure microfluidizer cell (M-110L Pneumatic). The lysate was cleared by centrifugation with 14 000 rpm at 4 °C for 30 min. The supernatant was incubated with 0.3 mL of Ni-NTA resin for 45 min at 4 °C. The resin was washed with 15 mL of 20 mM Tris at pH 7.5 and 500 mM NaCl and then twice with 15 mL of 20 mM Tris at pH 7.5, 500 mM NaCl, and 25 mM imidazole. The protein was eluted from the resin by three 5 mL elution buffer washes consisting of 20 mM Tris at pH 7.5, 500 mM NaCl, and 250 mM imidazole. The elution fractions were pooled and dialyzed against 1 L of 20 mM Tris at pH 7.5 and 100 mM NaCl. Further purification was performed with a HiTrap-Q (GE Biosciences) ion exchange column and HiLoad 16/60 SuperDex 200 (GE Bioscience) gel filtration column. For ion exchange, the gradient was 0–80% buffer B over 30 min, at a flow rate of 2 mL per minute (buffer A: 50 mM Tris-HCl at pH 7.5, 1 mM  $\beta\text{ME}$ ; buffer B: 1 M NaCl, 20 mM Tris-HCl at pH 7.5, 1 mM  $\beta\text{ME}$ ). The buffer used for gel filtration was 20 mM Tris-HCl, 100 mM NaCl at pH 7.5, and 1 mM  $\beta\text{ME}$ .

The presence of cobalamin associated with purified proteins was tested following the method previously described<sup>16</sup> with an Agilent 8453 UV-vis spectrophotometer.

**Modeling of *S. aureus* DUF208.** The protein sequence of *S. aureus* DUF208 (NP\_647328) was used to generate *de novo* models using Robetta<sup>36</sup> and I-TASSER,<sup>37</sup> using default parameters.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.6b01100.

Supporting Figures S1–S5 and Supporting Information Table S1 (PDF)

Excel document corresponding to the original file phyloprofile.14544.txt obtained from microbesonline.com (07/31/2015) and following analysis showing the inverted distribution pattern between *QueG* and *Duf208* (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: remizallot@ufl.edu.

\*Phone: +1 352 392 9416. E-mail: vcrecy@ufl.edu.

### ORCID

Valérie de Crécy-Lagard: 0000-0002-9955-3785

### Author Contributions

All authors have given approval to the final version of the manuscript.

### Funding

This work was supported by the National Institutes of Health (grant R01 GM70641 to V.d.C.-L.), the National Science Foundation (grant CHE1507357 to P.A.L.), and the University of Florida (to S.D.B.).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We would like to thank G. Marcial and G. Lorca for providing *L. casei* ATCC 334 and helping growing and processing the *Lactobacillus* species. We would like to thank V. de Berardinis for providing the *A. baylyi* ADP1 mutant strains.

## ■ REFERENCES

- (1) RajBhandary, U. L., and Ghosh, H. P. (1969) Studies on polynucleotides. XCI. Yeast methionine transfer ribonucleic acid: purification, properties, and terminal nucleotide sequences. *J. Biol. Chem.* 244, 1104–1113.
- (2) Goodman, H. M., Abelson, J., Landy, A., Brenner, S., and Smith, J. D. (1968) Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature* 217, 1019–24.
- (3) Doctor, B. P., Loebel, J. E., Sodd, M. A., and Winter, D. B. (1969) Nucleotide sequence of *Escherichia coli* tyrosine transfer ribonucleic acid. *Science* 163, 693–5.
- (4) Kasai, H., Ohashi, Z., Harada, F., Nishimura, S., Oppenheimer, N. J., Crain, P. F., Liehr, J. G., Von Minden, D. L., and McCloskey, J. A. (1975) Structure of the modified nucleoside Q isolated from *Escherichia coli* transfer ribonucleic acid. 7-(4,5-cis-Dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deazaguanosine. *Biochemistry* 14, 4198–4208.
- (5) Nishimura, S. (1983) Structure, biosynthesis, and function of queuosine in transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 28, 49–73.
- (6) Vinayak, M., and Pathak, C. (2010) Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci. Rep.* 30, 135–148.
- (7) Fergus, C., Barnes, D., Alqasem, M. A., and Kelly, V. P. (2015) The queuine micronutrient: Charting a course from microbe to man. *Nutrients* 7, 2897–2929.
- (8) Hutinet, G., Swarjo, M. A., and de Crécy-Lagard, V. (2016) Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. *RNA Biol.* 1–10.
- (9) Frey, N., McCloskey, J., Kersten, W., and Kersten, H. (1988) New function of vitamin B12: Cobamide-dependent reduction of epoxyqueuosine to queuosine in tRNAs of *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* 170, 2078–2082.
- (10) Iwata-Reuyl, D., and de Crécy-Lagard, V. (2009) Enzymatic formation of the 7-deazaguanosine hypermodified nucleosides of tRNA. *DNA RNA Modif. Enzym. Struct. Mech. Funct. Evol.*; Landes Bioscience: Austin, TX.
- (11) Noguchi, S., Nishimura, Y., Hirota, Y., and Nishimura, S. (1982) Isolation and characterization of an *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J. Biol. Chem.* 257, 6544–6550.
- (12) Slany, R. K., Bösl, M., and Kersten, H. (1994) Transfer and isomerization of the ribose moiety of AdoMet during the biosynthesis of queuosine tRNAs, a new unique reaction catalyzed by the QueA protein from *Escherichia coli*. *Biochimie* 76, 389–393.
- (13) Phillipson, D. W., Edmonds, C. G., Crain, P. F., Smith, D. L., Davis, D. R., and McCloskey, J. A. (1987) Isolation and structure elucidation of an epoxide derivative of the hypermodified nucleoside queuosine from *Escherichia coli* transfer RNA. *J. Biol. Chem.* 262, 3462–3471.
- (14) Miles, Z. D., McCarty, R. M., Molnar, G., and Bandarian, V. (2011) Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7368–72.
- (15) Miles, Z. D., Myers, W. K., Kincannon, W. M., Britt, R. D., and Bandarian, V. (2015) Biochemical and Spectroscopic Studies of Epoxyqueuosine Reductase: A Novel Iron-Sulfur Cluster- and Cobalamin-Containing Protein Involved in the Biosynthesis of Queuosine. *Biochemistry* 54, 4927–4935.
- (16) Payne, K. A. P., Fisher, K., Sjuts, H., Dunstan, M. S., Bellina, B., Johannissen, L., Barran, P., Hay, S., Rigby, S. E. J., and Leys, D. (2015) Epoxyqueuosine reductase structure suggests a mechanism for cobalamin-dependent tRNA modification. *J. Biol. Chem.* 290, 27572–27581.
- (17) Dowling, D. P., Miles, Z. D., Köhrer, C., Maiocco, S. J., Elliott, S. J., Bandarian, V., and Drennan, C. L. (2016) Molecular basis of cobalamin-dependent RNA modification. *Nucleic Acids Res.* gkw806.
- (18) Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J., and Arkin, A. P. (2010) MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–400.
- (19) Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., and Pütz, J. (2009) tRNADB 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* 37, D159–62.
- (20) Dineshkumar, T. K., Thanedar, S., Subbulakshmi, C., and Varshney, U. (2002) An unexpected absence of queuosine modification in the tRNAs of an *Escherichia coli* B strain. *Microbiology* 148, 3779–3787.
- (21) Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23, 324–328.
- (22) Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goessmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rülckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, S691–S702.
- (23) Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein

families database: Towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285.

(24) Metzgar, D., Bacher, J. M., Pezo, V., Reader, J., Döring, V., Schimmel, P., Marlière, P., and de Crécy-Lagard, V. (2004) *Acinetobacter* sp. ADP1: An ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res.* 32, 5780–5790.

(25) de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., Durot, M., Kreimeyer, A., Le Fèvre, F., Schächter, V., Pezo, V., Döring, V., Scarpelli, C., Médigue, C., Cohen, G. N., Marlière, P., Salanoubat, M., and Weissenbach, J. (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* 4, 174.

(26) Zhang, Y., Lin, J., and Gao, Y. (2012) In silico identification of a multi-functional regulatory protein involved in Holliday junction resolution in bacteria. *BMC Syst. Biol.* 6 (Suppl 1), S20.

(27) Peti, W., Etezady-Esfarjani, T., Herrmann, T., Klock, H. E., Lesley, S. A., and Wüthrich, K. (2004) NMR for structural proteomics of *Thermotoga maritima*: Screening and structure determination. *J. Struct. Funct. Genomics* 5, 205–215.

(28) Lesley, S. A., Graziano, J., Cho, C. Y., Knuth, M. W., and Klock, H. E. (2002) Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Eng., Des. Sel.* 15, 153–160.

(29) Hagervall, T. G., Edmonds, C. G., McCloskey, J. A., and Björk, G. R. (1987) Transfer RNA(5-methylaminomethyl-2-thiouridine)-methyltransferase from *Escherichia coli* K-12 has two enzymatic activities. *J. Biol. Chem.* 262, 8488–8495.

(30) Gaur, R., and Varshney, U. (2005) Genetic analysis identifies a function for the *queC* (*ybaX*) gene product at an initial step in the queosine biosynthetic pathway in *Escherichia coli*. *J. Bacteriol.* 187, 6893–6901.

(31) You, D., Xu, T., Yao, F., Zhou, X., and Deng, Z. (2008) Direct evidence that *thil* is an ATP pyrophosphatase for the adenylation of uridine in 4-thiouridine biosynthesis. *ChemBioChem* 9, 1879–1882.

(32) Auld, D. S. (2001) Zinc coordination sphere in biochemical zinc sites. *BioMetals* 14, 271–313.

(33) Cicmil, N., and Huang, R. H. (2008) Crystal structure of QueC from *Bacillus subtilis*: An enzyme involved in preQ1 biosynthesis. *Proteins: Struct., Funct., Genet.* 72, 1084–1088.

(34) Winkler, M., Dokulil, K., Weber, H., Pavkov-Keller, T., and Wilding, B. (2015) The Nitrile-Forming Enzyme 7-Cyano-7-Deazaguanine Synthase from *Geobacillus kaustophilus*: A Reverse Nitrilase? *ChemBioChem* 16, 2373–2378.

(35) Goodstadt, L., and Ponting, C. P. (2004) Vitamin K epoxide reductase: Homology, active site and catalytic mechanism. *Trends Biochem. Sci.* 29, 289–292.

(36) Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 32, W526–31.

(37) Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* 12, 7–8.

(38) Niehaus, T. D., Thamm, A. M. K., de Crécy-Lagard, V., and Hanson, A. D. (2015) Proteins of Unknown Biochemical Function: A Persistent Problem and a Roadmap to Help Overcome It. *Plant Physiol.* 169, 1436–1442.

(39) Hanson, A. D., Pribat, A., Waller, J. C., and de Crécy-Lagard, V. (2010) Unknown proteins and “orphan” enzymes: the missing half of the engineering parts list—and how to find it. *Biochem. J.* 425, 1–11.

(40) Foster, M. a, Tejerina, G., Guest, J. R., and Woods, D. D. (1964) Two enzymic mechanisms for the methylation of homocysteine by extracts of *Escherichia coli*. *Biochem. J.* 92, 476–488.

(41) Altschul, S. F., Gish, W., Miller, W. T., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

(42) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality

protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.

(43) Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890.

(44) Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Heeger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014) Pfam: The protein families database. *Nucleic Acids Res.* 42, D222–30.

(45) Letunic, I., and Bork, P. (2011) Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–8.

(46) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.

(47) Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158, 9–14.

(48) Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *J. Bacteriol.* 177, 4121–4130.

(49) Endres, L., Begley, U., Clark, R., Gu, C., Dziergowska, A., Malkiewicz, A., Melendez, J. A., Dedon, P. C., Begley, T. J., and Santos, J. (2015) Alkbh8 regulates selenocysteine-protein expression to protect against reactive oxygen species damage. *PLoS One* 10, e0131335.

(50) Ross, R., Cao, X., Yu, N., and Limbach, P. A. (2016) Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* 107, 73–78.