**Step 1. Define the target protein family**

MnmC proteins are involved in the posttranscriptional modification of tRNAs. They modify the wobble position of the anticodon to produce variants of the xm5U nucleoside, required to recognize A/G-ending codons of mixed boxes. MnmC proteins are annotated as bifunctional enzymes with two very clear domains called MnmC1 (C-terminal) and MnmC2 (N-terminal), which carry out sequential and independent reactions. Evolutionary analyses indicate that MnmC proteins are not highly conserved, and that their phylogenetic distribution is confined to bacteria where they are present in a few groups. Simultaneously, these analyses reveal that MnmC proteins result from an early gene fusion. Consequently, the MnmC1 domain evolution relates to the Glycine D-Amino Acid Oxidase family of FAD-binding proteins (DAAD), whereas MnmC2 relates to the Polyamine Aminopropyl Transferase (PAPT) family of proteins [47,48]. Recent experimental data suggest that the MnmC1 and MnmC2 domains interact in vitro when they were cloned, overexpressed and purified separately [51]. According to the gene fusion hypothesis supporting the origin of MnmC proteins, a domain-domain interaction had to be evolutionarily gained to stabilise its global structure. Our aim was to detect the residues responsible for this interaction following a functional divergence (Type-I) approach [25].

**Step 2. Sequence selection**

The Escherichia coli MnmC sequence (Uniprot P77182) was used as a bait in a Blast search by employing default parameters. We collected 25 additional and non-redundant sequences. They were selected by gaining maximum variability by selection of sequences from most of representative bacterial groups, then, this avoided bias by phylogenetic clustering.

**Step 3. Multiple Sequence Alignment**

We locally employed the MUSCLE algorithm using default parameters to construct our MSA file. Visualisation of the MSA was assisted by the JalView viewer. Several conserved blocks along the alignment were observed and the fusion region was evidenced by its high variability in both sequence and length terms (see below).

**Step 4. Tree construction**

We submitted our MSA file to Protest Webserver [32] in order to achieve the best evolutionary model to explain the evolution of the MnmC family. Subsequently, we obtained the tree representation of the phylogenetic relationships in our target family according to the evolutionary model showing the lowest likelihood value based on Akaike Information Criterion (AIC).

**Step 5. Define clusters to compare**

Once we had both the MSA and tree for the MnmC family, we loaded those files into the DIVERGE tool [28]. In order to cluster the selection to be compared, we previously redrew our tree with the “NJ Tree-Making” option using Kimura’s distance method. If we use paralogues for this analysis, then we select the clusters according to molecular functions. Instead, we must select clusters according to the distant phylogeny if we wish to perform an intra-cluster analysis, which is the case of MnmC.

**Step 6. Perform the analysis**

As we are interested in determining a probable gain-of-function in some bacteria group, we have to analyse the functional divergence Type-I in our family of proteins. Consequently, DIVERGE retrieves several statistical values of interest, among which we find the coefficient of functional divergence (θ). This value between Group 1 and Group 2 was 0.202 ± 0.038, which suggests that a significantly altered functional constraint occurred between them (rejecting the null hypothesis θ = 0). Then, we plotted the posterior probability per site with a threshold of 0.70 or 0.85 (user defined) for residue selection.

**Step 7. Residue selection**

Normally, residues with the highest posterior probabilities (P(S1|X)) will be preferable candidates to study. However, we have to select residues according to the predicted function to be accomplished. In this particular case, we examined posterior probabilities only at the N-terminal MnmC2 domain because there are structural data available about this domain at the PDB (id 2QY6; released by Bonanno and coworkers, 2007). Moreover, we have to select high-score residues which are located on the protein surface, and which would probably indicate their involvement in the domain-domain interaction.
supplementary material S1

Step 8. Further evaluation
After filtering the list of selectable residues according to their localisation on the MnmC2 three-dimensional structure, we can perform additional analyses by integrating other useful computational methodologies to study the specific function predicted for our residue(s). Consequently, we predicted that W153 has a gain-of-function in Group 1 with a \( P(S1|X) \) of 0.73. Thus, it is probably a mediating interaction of MnmC2 with MnmC1 given its localisation on the surface. Then, we submitted the MnmC2 structure to SHARP2 webserver (http://www.bioinformatics.sussex.ac.uk/SHARP2/sharp2.html), which is specialised in predicting the location of protein interaction sites on the surface of 3D structures [52]. Among the top-score patches predicted by the SHARP2 tool, the W153 residue was always present. Furthermore, five other very close residues were predicted together.

Step 9. Experimental design
Once we have selected W153 as a candidate to mediate the interaction between the MnmC2 and MnmC1 domains, we must consider that one amino acid replacement is probably not enough to impair a molecular interaction between proteins. In this manner, we can design two strategies: i) a radical change of polarity in W153 which may not only avoid a hydrophobic interaction with the MnmC1 counterpart, but cause further steric hindrance; or ii) multiple amino acid replacements involving some other residues adjacent to W153 and predicted by SHARP2 by even considering radical polarity changes and steric hindrance.

Step 10. Experimental data
We are limited to presenting an experimental validation of our predictions (Moukadiri et al., 2011. in preparation). However, we found published experimental evidence supporting our predictions. Thusly, the full Escherichia coli MnmC structure was recently released at PDB (Kim & Almo 2010, id 3PS9). Then, we were able to map our candidate residues onto this structure to observe their localisation in relation to the MnmC1 domain (light grey domain at the structure showed below). We observed that all the candidate residues (orange sphere represented at the dark grey MnmC2 domain), including W153 (green sphere represented) were localised on the interface of both domains. Moreover, we realised that some interface residues from the MnmC1 domain were also high-scored in the Type-I analysis.