Received: 11 May 2022 | Revised: 5 July 2022 | Accepted: 11 July 2022

DOI: 10.1002/pro.4393

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Early divergence of translation initiation and elongation factors

Evrim Fer[1,2,3] | **Kaitlyn M. McGrath**[1,3,4] | **Lionel Guy**[5] | **Adam J. Hockenberry**[6] | **Betül Kaçar**[1,3]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, Wisconsin, USA

[3]NASA Center for Early Life and Evolution, University of Wisconsin-Madison, Madison, Wisconsin, USA

[4]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, USA

[5]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[6]Department of Integrative Biology, The University of Texas at Austin, Austin, Texas, USA

Correspondence
Betül Kaçar, Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA.
Email: bkacar@wisc.edu

Funding information
John Templeton Foundation, Grant/Award Number: 61926; National Aeronautics and Space Administration; Human Frontiers in Science Program, Grant/Award Number: RGY0072; University of Wisconsin Foundation

Abstract

Protein translation is a foundational attribute of all living cells. The translation function carried out by the ribosome critically depends on an assortment of protein interaction partners, collectively referred to as the translation machinery. Various studies suggest that the diversification of the translation machinery occurred prior to the last universal common ancestor, yet it is unclear whether the predecessors of the extant translation machinery factors were functionally distinct from their modern counterparts. Here we reconstructed the shared ancestral trajectory and subsequent evolution of essential translation factor GTPases, elongation factor EF-Tu (aEF-1A/eEF-1A), and initiation factor IF2 (aIF5B/eIF5B). Based upon their similar functions and structural homologies, it has been proposed that EF-Tu and IF2 emerged from an ancient common ancestor. We generated the phylogenetic tree of IF2 and EF-Tu proteins and reconstructed ancestral sequences corresponding to the deepest nodes in their shared evolutionary history, including the last common IF2 and EF-Tu ancestor. By identifying the residue and domain substitutions, as well as structural changes along the phylogenetic history, we developed an evolutionary scenario for the origins, divergence and functional refinement of EF-Tu and IF2 proteins. Our analyses suggest that the common ancestor of IF2 and EF-Tu was an IF2-like GTPase protein. Given the central importance of the translation machinery to all cellular life, its earliest evolutionary constraints and trajectories are key to characterizing the universal constraints and capabilities of cellular evolution.

KEYWORDS

ancestral sequence reconstruction, elongation factor, initiation factor, LUCA, translation

1 | INTRODUCTION

The translation machinery is a network composed of proteins and RNA polymers that interact to read and decode nucleotide triplet codons into polypeptide outputs.[1] Thought to have evolved over ~3.8 billion years ago, the origin of the translation machinery represents an "evolutionary shockwave",[2] a key innovation and transition in the origin and early evolution of life.[3–5] Deciphering the key steps in the emergence of translation is considered to be one of the hardest problems in biology.[5,6]

Extant protein synthesis proceeds through four primary steps: initiation, elongation, termination, and ribosome recycling.[1,7,8] Each primary step of translation is carried out by largely universal protein factors: initiation factors (IF), elongation factors (EF), and termination or release factors (RF).[1] Among the steps in protein translation, the process of initiation, with the IF proteins as facilitators, is regarded as the most divergent across bacteria, eukaryotes, and archaea.[9] For instance, while bacteria require three IF proteins (IF1, IF2, IF3) to mediate translation initiation, archaea—and especially eukaryotic—translation initiation require more.[9,10] Of the IF proteins, initiation factor IF2 (IF2 in bacteria, aIF5B in archaea and eIF5B in eukaryotes; bacterial terminology used from here on)[11] is an essential and universal GTPase protein with a unique and crucial function in translation.[12] IF2 promotes the entire translation process by catalyzing the formation of a pre-initiation complex through its role in aligning the initiator tRNA—charged with a formylated methionine (fMet-tRNA)—to the start codon located at the P-site of the complex.[12,13] Furthermore, IF2 promotes the formation of a mature initiation complex by mediating the assembly of the small and large ribosomal subunits.[12,14] Following ribosomal subunit assembly, IF2 hydrolyzes GTP to GDP, thereby catalyzing IF2 disassociation. This permits the beginning of the elongation step in translation,[15] setting the stage for another multi-domain GTPase protein, elongation factor EF-Tu (EF-Tu in bacteria, aEF-1A in archaea and eEF-1A in eukaryotes; bacterial terminology used from here on).[8,16] EF-Tu is another essential and universal GTPase protein which shuttles the amino-acylated tRNAs to the A-site of the ribosome by binding to the aminoacylated elongator tRNAs and forming a tertiary complex.[17] EF-Tu is crucial for controlling elongation rate and accuracy.[18]

The IF2 and EF-Tu proteins from GTPase family harbor distinct functional sequence domains and motifs. IF2 is composed of three main domains: N-terminus, G-domain, and C-terminus, which are further divided into subdomains (Figure 1).[19] The N-terminal domain is not essential for IF2 protein function.[20] While its exact function is not clearly known, it is thought that the N-terminus assists in ribosomal subunit association.[14,21] The G2 and C2 subdomains of the G- and C-domains are crucial for GTP-binding and tRNA-binding, respectively.[19,22] EF-Tu is divided into three domains as well: G, D-II and D-III (Figure 1),[17,23] all of which interact with the elongator tRNAs.[17,23] The G-domain of EF-Tu is essential for GTP hydrolysis and protein function.[24,25] The D-II domain is crucial for tRNA-binding via interaction with the acceptor stem of the tRNA as well as the amino acid.[17,23] EF-Tu D-II domain shares significant structural similarity with the IF2 C2-subdomain.[26–28]

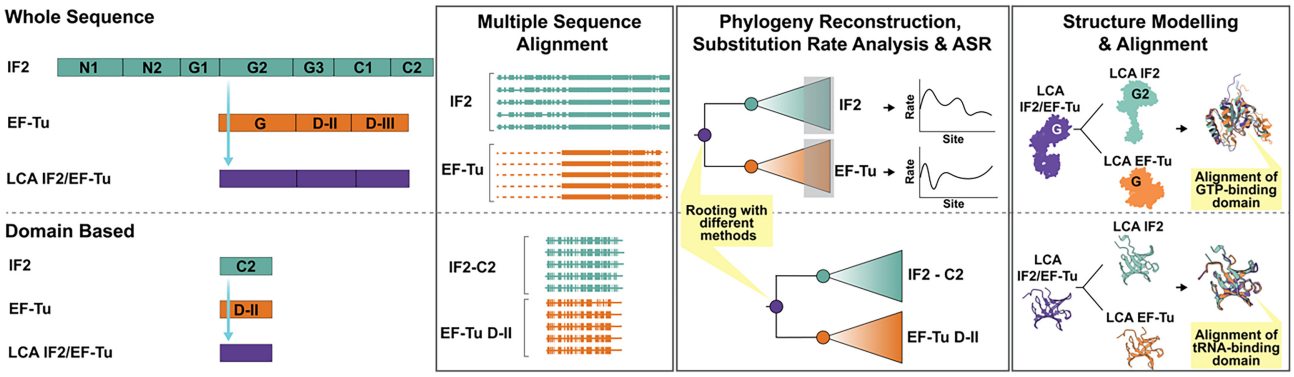

FIGURE 1 Methodology to investigate IF2/EF-Tu common ancestral traits. The domain representation of *Escherichia coli* (*E. coli*) IF2 (890 amino acids) and *E. coli* EF-Tu (394 amino acids) (left). Shown are the last common ancestor of IF2 (LCA IF2); the last common ancestor of EF-Tu (LCA EF-Tu); the last common ancestor of IF2 and EF-Tu (LCA IF2/EF-Tu). The GTP-binding domain of IF2/EF-Tu common ancestor was inferred using whole sequences of the extant IF2 and EF-Tu proteins. For the ancestral tRNA-binding domain, C2 and D-II fragments from extant IF2 and EF-Tu proteins were used. The analysis was performed as follows: multiple sequence alignment, phylogenetic reconstruction, ancestral sequence reconstruction (ASR), structural modeling, and structural alignment. IF2 and EF-Tu whole sequences were also used for relative substitution rate analysis. G and G2 indicate GTP-binding domains of EF-Tu and IF2.

It has been hypothesized that IF2 and EF-Tu evolved from an EF-Tu like common ancestor based on sequence[29] and structural homology of their extant counterparts.[30] The shared key features include GTPase activity and tRNA binding in both IF2 and EF-Tu proteins, leading to a parsimonious explanation that these two proteins descended from an ancient, common ancestral protein. Other studies have shed light on the ancestry of EF-Tu[31–34] as well as its diverse role in extant cells.[16] Intriguingly, EF-Tu was suggested to be the first GTPase in early translation machinery.[17] However, the origins, early divergence and evolution of IF2 and EF-Tu proteins are not known with certainty.

In the present study, we utilized phylogenetic modeling, ancestral sequence reconstruction, structural and functional predictions to better understand the evolutionary origins and history of IF2 and EF-Tu, building on the hypothesis that the two evolved from a common ancestor in the last universal common ancestor (LUCA).[29,30] We reconstructed a shared IF2/EF-Tu phylogeny using bacterial and archaeal sequences and inferred sequences of IF2 and EF-Tu ancestors using different tree rooting methods. We further predicted ancient protein structures at target nodes using homology modeling, determined the site-specific evolutionary rates of IF2 and EF-Tu sequences, and comparatively interpreted the GTP and tRNA-binding functions of IF2 and EF-Tu proteins along their shared evolutionary timeline (Figure 1). Finally, we assessed the evolution of protein domain fragments for each of these factors by following the same analysis pipeline. We present an evolutionary scenario for the early origins and divergence of the initiation and elongation factors and discuss the implications for the characteristics of life's early translation machinery.

2 | RESULTS

2.1 | Reconstruction of the IF2/EF-Tu phylogeny

We reconstructed an IF2/EF-Tu phylogenetic tree by aligning the extant IF2 and EF-Tu sequence data (Figure 2a). Our dataset represents homolog sequences

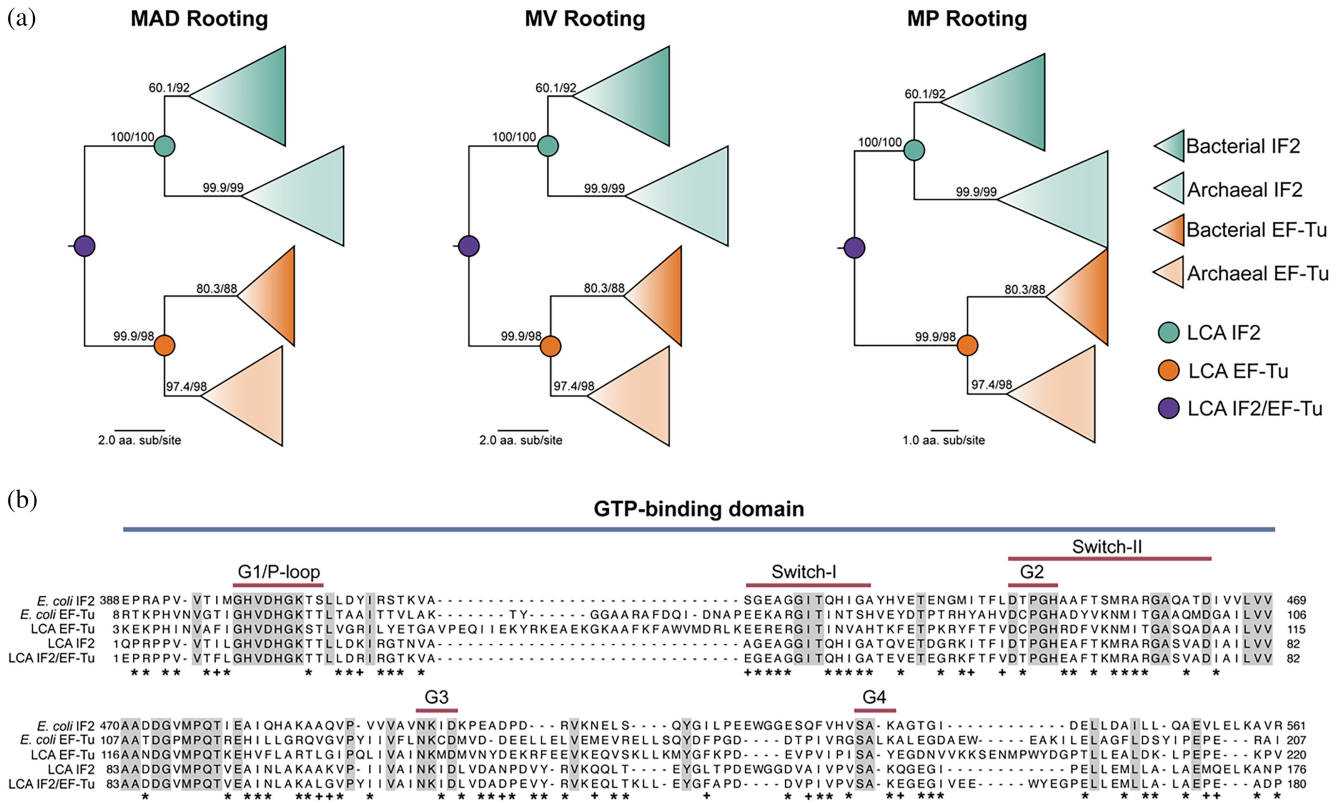

FIGURE 2 Phylogenetic and sequence analyses of the ancestral GTP-binding domain. (a) IF2 and EF-Tu joint tree (IF2/EF-Tu Tree-1, Table S2) constructed using different rooting methods. Branch supports are shown for the ancestral nodes. (b) Sequence alignment of ancestral and *Escherichia coli* IF2 and EF-Tu proteins GTP-binding regions. Gray columns highlight 100% conserved residues across all sequences. The G1/P-loop, Switch-I, Switch-II, G2, G3, and G4 motifs in the GTPase domain are shown with red bars. In the last common ancestor of IF2 and EF-Tu (LCA IF2/EF-Tu), sites identical to the last common ancestor of IF2 (LCA IF2) are shown with "*"; sites that are identical to the last common ancestor of EF-Tu (LCA EF-Tu) are shown with "+".

from bacterial and archaeal phyla (Table S1). We tested the IF2/EF-Tu joint tree robustness to alignment uncertainty and evolutionary model misspecification, resulting in four alternate trees (Table S2). Each resulting tree has a similar overall topology and IF2 and EF-Tu proteins are consistently separated into two major clades (Figure S1). In all trees, the support values for the IF2 and EF-Tu clades range between 98% and 100%. Each IF2 and EF-Tu major clade is further subdivided into bacterial and archaeal clades, except the tree constructed with MAFFT L-INS-i[35] alignment combined with the LG + F + G4 evolutionary model (IF2/EF-Tu Tree-2), in which archaeal IF2 sequences branch within bacterial IF2 with a support value of ∼50% (Figure S1D-F). Ultimately, we selected the IF2/EF-Tu phylogeny (IF2/EF-Tu Tree-1) represented by the highest likelihood score calculated by IQTREE[36] for subsequent analyses (Figure 2a, Figure S1A-C, Table S2). The highest likelihood IF2/EF-Tu tree constructed using MAFFT is in accordance with Vialle et al., who previously demonstrated that MAFFT alignments outperformed many other alignment methods in accuracy of ancestral sequences when using large, difficult-to-align datasets.[37] We rooted the IF2/EF-Tu tree using minimal ancestor deviation (MAD),[38] midpoint (MP),[39] and minimum variance (MV)[40] to test to what degree the rooting method affects the sequence prediction of the IF2/EF-Tu common ancestor (Figure 2a). Ancestral sequences for the last common ancestor of IF2/EF-Tu, as well as the last common ancestors for both IF2 (LCA IF2) and EF-Tu (LCA EF-Tu) major clades were inferred from each rooted IF2/EF-Tu tree (Figure 2a). The mean posterior probabilities of the IF2 and EF-Tu ancestors are ∼64.9% and ∼66.9% across each tree, respectively, and the posterior probability of the common ancestor sequence composition is robust to the rooting method. The mean posterior probabilities of the IF2/EF-Tu last common ancestor vary by ∼10% across differently rooted IF2/EF-Tu trees (Table S2). The sequence of the common ancestor of IF2/EF-Tu has ∼42.6% mean identity to the *Escherichia coli* (*E. coli*) IF2 and shares only ∼28.4% mean identity with *E. coli* EF-Tu across all trees (Figure S2A).

2.2 | Evolution of the GTP- and tRNA-binding domains of IF2 and EF-Tu

GTP-binding domains of *E. coli* IF2 and EF-Tu share 26% sequence identity (Figure S2B). The reconstructed GTP-binding domain of the IF2/EF-Tu common ancestor shares high sequence identity with the IF2 ancestor (∼85% mean identity across all trees, ∼43% mean identity

TABLE 1 The list of residues and domains in IF2 and EF-Tu proteins that interact with GTP and tRNA—and are assumed to be homologous—and the likely state of that position (including posterior probabilities, post. prob.) for the IF2/EF-Tu common ancestor (LCA IF2/EF-Tu)

Residue			Domain		Associated function	
IF2	EF-Tu	LCA IF2/EF-Tu	IF2	EF-Tu	IF2	EF-Tu
V400	V20	V13 (94.9%)	G2	G	In G1/P-loop. Essential for α- and β-phosphate binding[42–49]	
H448	H84	H61 (99.7%)			In G2/switch II. Essential for GTP-hydrolysis-induced conformational shift[42–49]	
D501	D138	D115 (98.8%)			In G3 motif. Essential for guanine nucleotide recognition[42–49]	
F804	F219	F15 (99.8%)	C2	D-II	Within lining of fMet binding-pocket[27,28]	Within lining of amino acid pocket[17,27]
A813	V228	V24 (82.1%)			Interaction with acceptor arm of fMet-tRNA[55]	Unknown
G814	T229	G25 (74.5%)			Within lining of fMet binding-pocket[54]	Within lining of amino acid pocket[17,27]
R847	M261	H57 (32.2%)			Within lining of fMet binding-pocket[28,54,55]	Unknown
K849	R263	K59 (48.0%)			Interaction with formyl group and lining of fMet binding-pocket[27,28,54,55]	Interaction with acceptor arm of tRNA[39,56]
E860	N274	E70 (68.3%)			Within lining of fMet binding-pocket[27,28,54,55]	Within lining of amino acid pocket[39,56]
C861	V275	V71 (53.4%)			Stabilization of amino acid pocket by forming di-sulfide bond[27,28,54,55]	Unknown
G862	G276	G72 (93.2%)			Within lining of fMet binding-pocket[28,54,55]	Unknown

Note: Sequence site numbers are based on *E. coli*: accession AIS23657.1 (IF2) and WP_000031784.1 (EF-Tu). Where relevant, associated function lists references to publications describing the role of individual residues in the associated protein family.

with EF-Tu ancestor) (Figure S2B). The IF2 and EF-Tu G-domain, like all GTPases, contains consensus motifs important for GTPase activity, such as the "GHVDHGK" (G1 motif/P-loop), "DXPGH" (G2 motif) in Switch-II region, "NKXD" (G3 motif),[41–43] and others[41–44] (Figure 2b). Specifically, Val-400 in G1/P-loop, His-448 in G2 motif, and Asp-501 in G3 motif in *E. coli* IF2 are essential for α- and *β*-phosphate binding, GTP-hydrolysis-induced conformational shift, and guanine nucleotide recognition, respectively[43,45–49] (Table 1). All the consensus GTPase motifs are conserved in the ancestral sequences (Figure 2b). Especially, residues of functional importance—Val-400, His-448 and Asp-501 (Table 1)—exhibit high posterior probabilities in the IF2/EF-Tu common ancestor (94–99%, Table 1, Figure 2b). The GTP-binding domain Switch-II region in the IF2/EF-Tu common ancestor is completely identical to Switch-II region in the ancestral IF2, whereas 5 out of 20 residues that compose Switch-II differ from modern IF2 (Figure 2b). On the other hand, almost half of the Switch-II region in IF2/EF-Tu common ancestor is different from ancestral and *E. coli* EF-Tu (Figure 2b).

Consistently, the GTP-binding domain of extant IF2 and EF-Tu proteins exhibits the slowest substitution rates across each respective protein—indicating higher sequence conservation (Figure 3). Comparing across the protein families, the mean relative substitution rates of the EF-Tu GTP-binding domain is ∼3.7-fold higher than the IF2 GTP-binding domain (Figure 3, Table S3). Also, the relative substitution rate in the tRNA-binding domain is ∼2-fold higher for EF-Tu compared to IF2 (Figure 3, Table S3).

Shared function and structure are often thought to be an indicator of shared history, despite low sequence similarity.[50,51] IF2 C2 and EF-Tu D-II domains are crucial for tRNA binding,[17,19,23] are structural homologs, and have the same domain folding (OB-fold) despite shared sequence identity of only 13%.[27,28] Tracking the origin of IF2 C2 and EF-Tu D-II domains may provide clues about the establishment of tRNA specificity and the evolution of the binding envelope. This is particularly important because what distinguishes IF2 and EF-Tu is the type of tRNA they bind to, a fact that has also garnered interest from early life studies focusing on possible links between tRNA emergence and amino acid availability implicated in RNA world scenarios.[52]

We thus explicitly focused on the C2 and D-II domains to study the likely origin and divergence of the tRNA binding domain and generated a C2/D-II phylogeny. We tested various alignment methods and evolutionary models (Table S4), resulting in four trees (Figure S3). In all trees, except the tree constructed with MUSCLE alignment and the LG + R8 model, EF-Tu bacterial and archaea sequences form a monophyletic group (Figure S3). We selected the C2/D-II phylogeny (CD/D-II Tree-1) represented by the highest likelihood score calculated by IQTREE[36] (Table S4). The selected C2/D-II tree

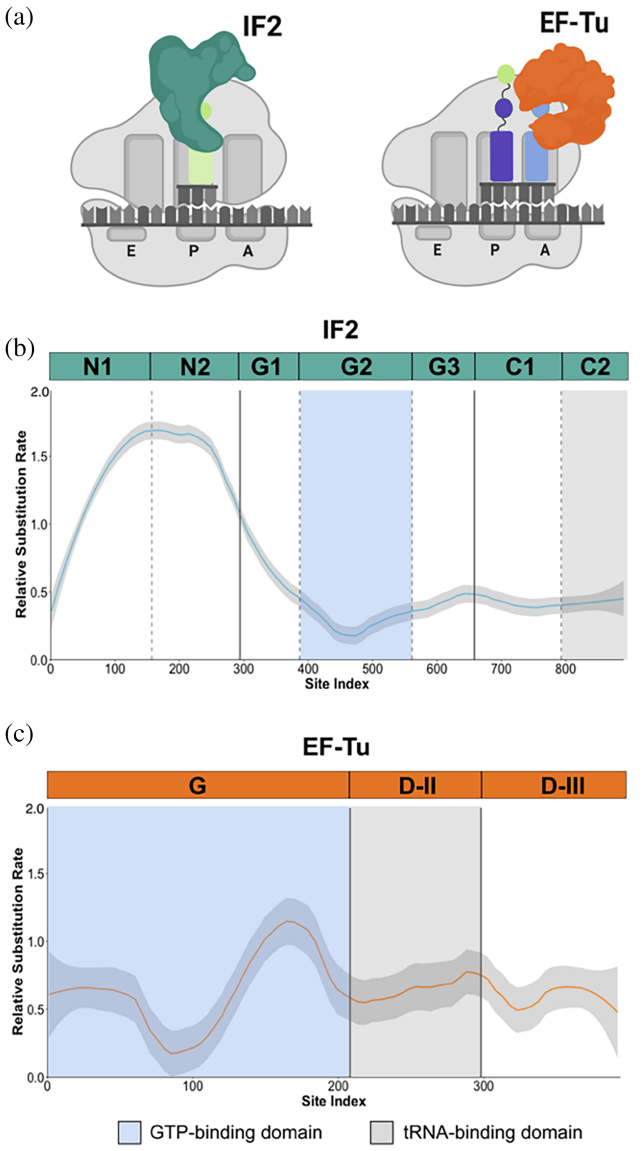

FIGURE 3 Relative sitewise substitution rates across IF2 and EF-Tu proteins. (a) IF2 recruits initiator-tRNA to the P-site of the ribosome. EF-Tu brings elongator-tRNAs to the A-site of the ribosome. (b) Relative substitution rates in IF2 protein. A value less than 1 indicates more slowly evolving than the average site, and a value more than 1 indicates more rapidly evolving than the average site. The domain names are given at the top (N1: 1–157, N2: 158–294, G1: 295–387, G2: 388–561, G3: 562–658, C1: 659–793, C2: 794–890; the residue numbers are based on *Escherichia coli* IF2). (c) Relative substitution rates in EF-Tu. The domain names are given at the top (G: 1–208, D-II: 209–298, D-III: 299–394; the residue numbers are based on *E. coli* EF-Tu). Blue and gray shades represent the GTP-binding and tRNA-binding regions in IF2 and EF-Tu proteins, respectively.

uses LG + R8 model and shows support values greater than 97% for separation of C2 and D-II (Figure 4a). We tested different rooting methods (i.e., MAD, MP, and MV) to assess the degree to which rooting may impact our inference of ancestral sequences from the C2/D-II phylogeny as well (Figure 4a). Different rooting methods do not substantially affect mean ancestral sequence posterior probabilities, which range between ~46 and 49% for the C2/D-II common ancestor (Table S4). *E. coli* IF2 C2 and *E. coli* EF-Tu D-II domains have 18% sequence identity. The sequence of the common ancestor of C2/D-II has ~32% identity to both *E. coli* IF2 C2 and *E. coli* EF-Tu D-II (Figure 4b). We further calculated the posterior probability of functionally distinct sites—defined as the sites impacting tRNA interaction (Table 1). Specifically, five distinct residues in the common ancestor, Phe-15, Val-24, Gly-25, Glu-70, and Gly-72 (corresponding to Phe-804, Ala-813, Gly-814, Glu-860, and Gly-862 in IF2 sites) are constructed with low ambiguity (Figure 4b)—the likelihood of the second best residue at these sites does not exceed the 20% posterior probability threshold.[53] While only Gly-814 and Gly-862 are conserved in bacterial IF2 (Table S5), all five corresponding residues help form the binding pocket for fMet.[27,28,46,54,55] Gly-814 and Gly-862 correspond to Thr-229 and Asn-274 in EF-Tu, respectively. EF-Tu's Thr-229 and Asn-274 (conserved in bacteria; Table S6) help form the binding pocket for amino acids on tRNA.[28,55]

2.3 | Predicted structure of the IF2/EF-Tu common ancestor

We modeled and predicted the structure of the IF2/EF-Tu common ancestor as well as IF2 and EF-Tu ancestors from the selected IF2/EF-Tu and C2/D-II trees (see Section 4; Table S2, S4). For each ancestor, five structures were predicted using the five best templates found from the PDB database by trRosetta[57] (see Section 4). The quality of each predicted structure was assessed via QMeanDisCo,[58] where a quality score greater than 0.6 indicates a high-confidence structure. The mean quality scores for the predicted structures of each IF2 ancestor, EF-Tu ancestor, and the IF2/EF-Tu common ancestor from IF2/EF-Tu tree are ~0.69, ~0.79, and ~0.57, respectively. The predicted structures from C2/D-II tree have mean quality scores of ~0.68 for C2 ancestor, ~0.85 for D-II ancestor, and ~0.75 for C2/D-II common ancestor.

The predicted ancestral EF-Tu structure shares ~68% sequence similarity with the previously crystallized ancient EF-Tu (PDB: 5W76)[32] and ~64% sequence similarity with the extant EF-Tu structure (PDB: 1EFC).[59]

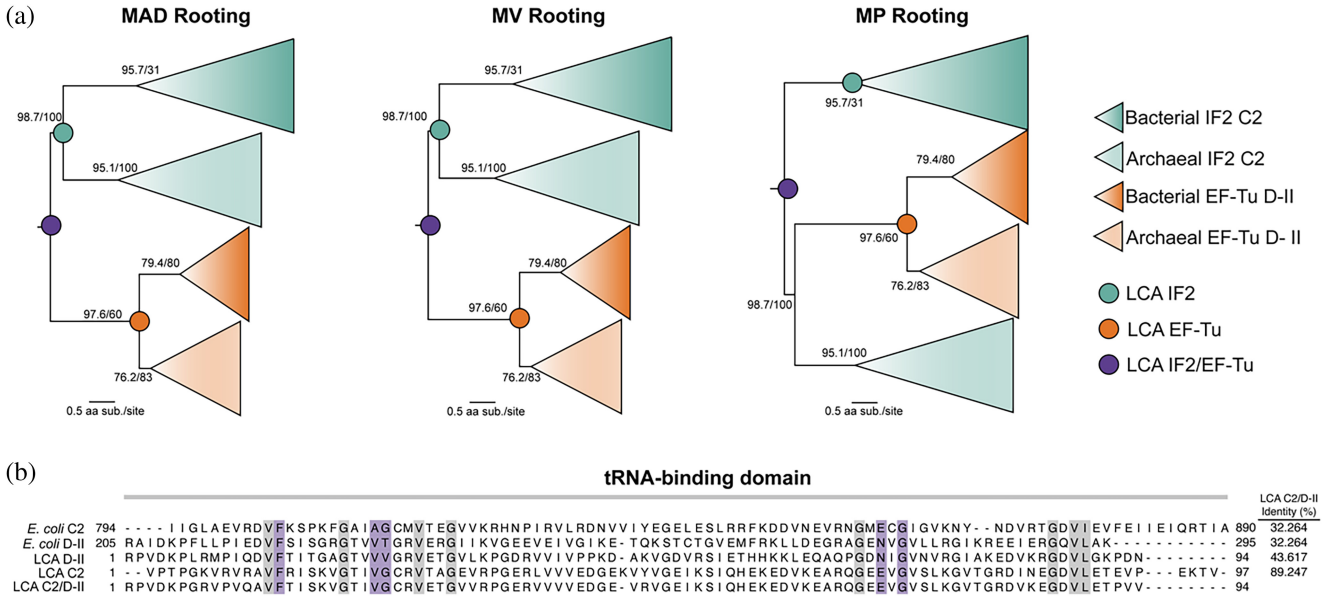

FIGURE 4 Phylogenetic and sequence analyses of the ancestral tRNA-binding domain. (a) IF2-C2 and EF-Tu D-II joint trees (C2-D-II Tree-1, Table S4), constructed using different rooting methods. (b) Sequence alignment of *Escherichia coli* IF2 C2 and EF-Tu D-II domains and ancestral tRNA-binding regions. Gray columns highlight 100% conserved residues across all sequences. Purple columns highlight functionally distinct residues having high posterior probabilities (Table 1). Minimal ancestor deviation (MAD), minimum variance (MV), and midpoint (MP) rooting. The circles show the ancestral nodes for LCA IF2 as green, LCA EF-Tu as orange LCA IF2/EF-Tu as purple. Branch supports are shown for the ancestral nodes

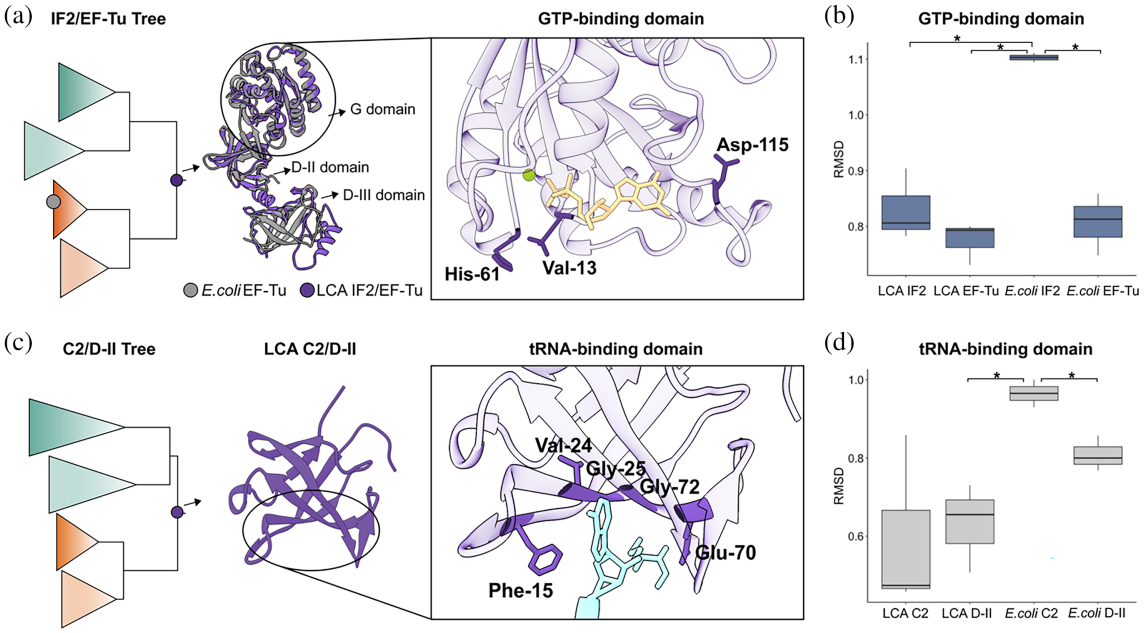

FIGURE 5 Ancestral protein structure homology modeling. (a) The superimposition of G, D-II, and D-III domains of *Escherichia coli* EF-Tu (dark gray) to the predicted structure of IF2/EF-Tu common ancestor (LCA IF2/EF-Tu; purple). The GTP-binding domain is enlarged in the top inset and conserved functional residues Val-13, His-61, and Asp-115 (site numbers in the ancestral sequence) are shown. The GTP molecule is represented in yellow. Mg^{2+} ion is shown in green. (b) RMSD values calculated from the structural alignment of GTP-binding domains between IF2/EF-Tu common ancestor and IF2 last common ancestor (LCA IF2), EF-Tu last common ancestor (LCA EF-Tu), *E. coli* IF2 (PDB ID: 3JCJ) and *E. coli* EF-Tu (PDB ID: 1EFC). (c) The tRNA-binding sequence of the common ancestor from C2/D-II joint tree is structurally modeled. The tRNA-binding domain is focused and conserved functional residues Phe-15, Val-24, Gly-25, Glu-70, and Gly-72 (site numbers in the ancestral sequence) are shown. A representative fMet-tRNA is shown with blue. (d) RMSD values from structural alignment of tRNA-binding domains between C2/D-II common ancestor and C2 last common ancestor (LCA C2), D-II last common ancestor (LCA D-II), *E. coli* C2 (PDB: 3JCJ)[28] and *E. coli* D-II (PDB: 1EFC).[59] "*" signs indicate statistical significance tested by Student's *t*-test with Bonferroni adjustment.

The IF2/EF-Tu common ancestor has the same domain organization as the predicted and modern EF-Tu structure and contains all G, D-II, and D-III domains (Figure 5a; Figure S4A-C). To assess structural similarity, we calculated root mean square deviation (RMSD), where lower scores indicate more similar structures.[60] Specifically, RMSD between the GTP-binding domain of IF2/EF-Tu common ancestor from each rooted IF2/EF-Tu tree and GTP-binding domains of *E. coli* IF2 (PDB: 3JCJ)[28] and EF-Tu (PDB: 1EFC)[59] were calculated. The mean RMSD between the GTP-binding domain of the common ancestor and the *E. coli* IF2 GTP-binding domain (PDB: 3JCJ)[28] is ~1.1 Å, while the mean RMSD between the IF2/EF-Tu common ancestor and the *E. coli* EF-Tu GTP-binding (PDB: 1EFC)[59] is ~0.8 Å (Figure 5b). The mean RMSD between the tRNA-binding domain of the common ancestor and the *E. coli* IF2 tRNA-binding domain (C2) (PDB: 3JCJ)[28] is ~0.965 Å, while mean RMSD between the tRNA binding-domain of the common ancestor and the *E. coli* EF-Tu tRNA-binding domain (D-II) (PDB: 1EFC)[59] is ~0.8 Å (Figure 5d). According to statistical tests (Student's *t*-test with Bonferroni *p*-value adjustment), the RMSD between GTP and tRNA-domains of the common ancestor and modern EF-Tu are significantly lower (p-value = .01) than the RMSD between GTP and tRNA-binding domains of the common ancestor and modern IF2 (Figure 5b,d), indicating that GTP and tRNA-binding domain structure in the last common ancestor are more similar to modern EF-Tu than modern IF2.

3 | DISCUSSION

Using phylogenetics, ancestral sequence reconstruction, as well as structural and evolutionary models, we inferred the common ancestor of translation IF2 and EF-Tu protein functionality. Phylogenies derived from whole sequences from both protein families (IF2/EF-Tu tree) (Figure 2a) as well as single protein domains (C2/D-II) (Figure 4a) provide a robust backbone to understand the GTP- and tRNA-binding domain ancestry of IF2 and EF-Tu proteins. To remain as agnostic as possible about the order of emergence of IF2 and EF-Tu, we employed

various rooting methods in place of an imposed outgroup. These methods consistently root the tree between IF2 and EF-Tu (Figures 2a and 4a). Our results tracking the GTP- and tRNA-binding domain ancestry suggest that the common ancestor exhibited IF2-like properties.

The GTP-binding domain is thought to be one of the oldest protein relics[61,62] and GTP activation in translational GTPases promotes changes in overall domain conformation and interactions in the functional domains by the help of Switch-I and Switch-II loops.[26,63] The Switch-II region—including the "DXPGH" motif—largely contributes to the conformational change that is triggered by the exchange of GTP and GDP both in IF2 and EF-Tu.[22] EF-Tu Switch-II undergoes a large conformational change which enables EF-Tu to maintain stable interaction with amino acylated tRNAs,[23,26,56] critically impacting EF-Tu activation and function.[26,56] Unlike EF-Tu, the GTP-triggered conformational changes are limited in IF2[22]; it was shown that GTP-conformation is not as "critical for productive interactions between eIF5B and its effector molecules as it is for EF-Tu".[26] Our results show that the GTP-binding domain sequence of the reconstructed common ancestor is more identical to modern IF2 than EF-Tu as opposed to the previous suggestions for EF-Tu-like ancestry of translational GTPases.[17] To be able to make a functional inference, we focused on the residues known to impact protein activity with high posterior probabilities in the common ancestor sequence (Table 1). The GTP-binding domain in the IF2/EF-Tu common ancestor still retains the conserved sequence motifs and functional residues (i.e., Val-400, His-448, and Asp-501 in *E. coli* IF2) for GTPase activity with high posterior probabilities (≥94%) (Figure 2b, Table 1). In the common ancestor's GTP-binding domain sequence, the Switch-II region—which affects conformational changes between GTP/GDP bound states[22,63]—is similar to IF2 (Figure 2b). This indicates that analogous to modern IF2, the common ancestor may have a limited conformational change between GTP bound/unbound states. Further, crucial tRNA-binding domain fMet binding pocket residues of IF2 are identical in IF2/ EF-Tu ancestor (Figure 4b, Table 1).

While our inferences strongly indicate that the IF2/EF-Tu ancestor primarily exhibited IF2-like properties, we cannot overlook that one feature, namely the structure of the binding pocket of the common ancestor, is more comparable to modern EF-Tu. Structural prediction for reconstructed ancient proteins is still in its infancy.[64,65] Nevertheless, such a possibility may have interesting implications for the IF2/EF-Tu common ancestor. For instance, the IF2/EF-Tu ancestor may have capitalized on IF2-like functionality, while accumulating mutations that would later prime its descendants for another function. We thus submit that following gene duplication from an IF2/EF-Tu ancestor, one of the duplicated copies, the IF2 ancestor, acquired substitutions in the tRNA-binding domain, fine-tuning IF2 recognition of the initiator fMet-tRNA. These changes further optimized the translation initiation accuracy and efficiency by specializing for this role while maintaining the original limited conformation at the GTPase domain.[66] The other copy, the EF-Tu ancestor, coevolved with tRNAs, gaining flexible conformational function in the GTP-binding domain, and acquired mutations to bind a wider variety of tRNA molecules, specializing for EF-Tu functionality. Intriguingly, it was recently shown that ancestral EF-Tu proteins were generalists in terms of their compatibility with ribosomes, compared to the modern EF-Tus.[31] The translation system consists of numerous essential molecules and the complete picture of evolutionary trajectory of translation over geologic time will require an integrated view into other components such as the other translation GTPases (e.g., EF-G, RF3) and tRNAs[67] with the underlying chemical systems. Further, ancestral tRNA-IF2 interactions may differ from those of extant translation interaction partners. In any case, a multitasking bifunctional factor operating within ancestral translation machinery could suggest intriguing possibilities for the earliest translation. Whether the shared ancestor was able to conduct both initiation and elongation functions efficiently as a single proto-translation factor, whether mutations drove a tradeoff between structural conformation versus function, and the degree to which these traits were interrelated, are to be studied in greater detail.

It is generally accepted that the ribosome and its protein factors were already established in LUCA.[4,6] Expansion of the translation machinery, on the other hand, is thought to have occurred prior to LUCA[68] potentially following the rRNA core center of the ribosome formation,[31,69] strengthening the hypothesis that, unlike the ribosome,[70] the supporting translational factors were not yet fully established in the early cellular era. While we do not know if LUCA had adapted a system that is largely similar to the extant translation machinery, or whether translation-as-we-know-it subsequently emerged in a stepwise fashion or through simultaneous maturation of supporting components,[71,72] tracing the evolutionary history of early translation factors will unravel the fascinating story of life's core information processing machinery. Further studies exploiting protein evolution should be implemented to empirically recreate distinct, yet extinct sequences of proteins that once operated at the heart of an ancestral translation machinery.

4 | METHODS

4.1 | Sequence curation

IF2 protein sequences (NCBI: AIS23657.1) were curated from the National Center for Biotechnology Information (NCBI) non-redundant (nr) database.[73] IF2 protein sequences from 17 different bacterial phyla (Actinobacteria, Aquificae, Armatimonadetes, Bacteroidetes, Chlamydiae, Chlorobi, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Fusobacteria, Planctomycetes, Proteobacteria, Spirochaetes, Synergistetes, Tenericutes, Thermotogae) were retrieved using the Basic Local Alignment Search Tool, BLASTp[74] with an e-value threshold of <1e−5 (Last accessed November 2021). The dataset with all homologous sequences and taxon-specific insertions was filtered manually to remove partial, duplicate, or non-IF2 sequences. The remaining sequences were aligned and used to construct a maximum-likelihood tree with FastTree v.2.1.12[75] to inspect sequence alignment and phylogeny. Sequences having distant sequence identity (<65%) to their corresponding phyla were manually removed. The alignment, tree reconstruction, and sequence removing steps were iterated until all the sequences were grouped in their corresponding phyla. The final IF2 dataset included 268 bacterial taxa (Table S1). EF-Tu sequences from the same taxa list were additionally retrieved from NCBI BLASTp search using bacterial EF-Tu sequence as query (BLASTp query: WP_000031784.1). From the BLAST outputs, EF-Tu sequences for the interested taxa were kept and non-EF-Tu sequences were filtered out. To collect archaeal IF2 and EF-Tu sequences to be used as outgroup, separate PSSMs using standalone PSI-BLAST were performed.[76] The gap-only columns in each alignment were removed using Seaview v4.[77] BLAST searches using an IF2 query returned EF-Tu sequence hits. Sequences were clustered with CD-HIT.[78] The duplicated sequences were removed from the dataset. The final dataset included archaeal IF5B and EF-1A sequences from Euryarchaeota, TACK, Candidatus thermoplasmata, DPANN and Asgard phyla groups (Table S1).

4.2 | Phylogenetic analysis and ancestral sequence reconstruction

To minimize bias of the method, different alignment algorithms were used to generate alternative trees following Garcia et al.[79] The combined phylogeny of IF2 and EF-Tu was constructed using the same 408 IF2 and 408 EF-Tu sequences. In total, 816 sequences were aligned together using MAFFT L-INS-i[35] and MUSCLE v3.8 alignment.[80] The alternate four phylogenies were reconstructed using IQTREE v1.6.11.[36] Ultrafast bootstrap (UFBoot)[81] and Shimodaira-Hasegawa approximate likelihood-rate test (SH-aLRT)[82] were implemented. The reconstructed trees were rooted using (i) Minimal Ancestor Deviation (MAD),[38] (ii) midpoint rooting (MP),[39] (iii) Minimum Variance (MV) rooting methods.[40] The rooted trees were visualized using FigTree v1.4.4.[83]

The combined phylogeny of IF2-C2 and EF-Tu D-II was constructed using the same C2 fragments from 405 IF2 and D-II fragments from 405 EF-Tu sequences. From the previous dataset, IF2 and EF-Tu sequences were removed from three archaeal taxa (*Candidate M. washburnensis*, *Candidate P. syntrophicum* and *C. tenax*) since they branched outside of archaeal clade. In total, 810 sequences were aligned using MAFFT L-INS-i[35] and MUSCLE v3.8.[80] Model fitting, rooting, and visualization were performed as described above. Each rooted IF/EF-Tu and C2/D-II maximum likelihood (Tables S2, S4) tree was used for ancestral sequence inference by PAML.[84]

4.3 | Structure prediction of ancestral proteins

The structures of ancestral proteins were modelled using trRosetta.[57] The qualities of the predicted structures were calculated using QMeanDisCo.[58] For structural alignment, *E. coli* IF2 (PDB: 3JCJ)[28] and *E. coli* EF-Tu (PDB: 1EFC)[59] crystallized structures were used. The structural alignment scores were calculated as RMSD scores in UCSF Chimera v1.14.[85] Statistical tests were performed using the Student's *t*-test with "Bonferroni" *p*-adjustment method.

4.4 | Substitution rate analysis

The site-specific evolutionary rates per protein sites were calculated using all extant sequences present in the alignments described above (Figure S5). Substitution rate per site was calculated by IQTREE v1.6.11 using an empirical Bayesian approach. The rates were indexed based on the *E. coli* IF2 and EF-Tu protein sites. The site-specific substitution rates were smoothed and represented with a curve using non-parametric local polynomial regression fitting (i.e., loess) method. The difference between the mean of the site-specific rates of the domains was tested using the pairwise *t*-test with "Bonferroni" p-adjustment method.

AUTHOR CONTRIBUTIONS

Evrim Fer: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); validation (lead); visualization (lead); writing – original draft

(equal); writing – review and editing (equal). **Kaitlyn McGrath:** Data curation (supporting); formal analysis (supporting); investigation (supporting); writing – review and editing (supporting). **Lionel Guy:** Investigation (supporting); methodology (supporting); validation (supporting); writing – review and editing (supporting). **Adam J. Hockenberry:** Investigation (supporting); methodology (supporting); writing – review and editing (supporting). **Betül Kaçar:** Conceptualization (equal); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (lead); supervision (lead); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

Funding is provided by the UW Foundation Hiroshi and Sugiyama Fund for Graduate Studies (EF), a NASA Early Career Collaboration Award (KMM), an NIH Training Grant (#T32 GM136536) (KMM), the NASA Exobiology and Evolutionary Biology Program (#16-IDEAS16-0003) and the John Templeton Foundation (#61926) (BK). We thank Jaime Fraser, Lynn Kamerlin, Colin Jackson, and Nobuhiko Tokuri for the invitation to contribute to the Dan Tawfik Special Issue.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Evrim Fer https://orcid.org/0000-0003-2731-1207
Betül Kaçar https://orcid.org/0000-0002-0482-2357

REFERENCES

1. Ramakrishnan V. Ribosome structure and the mechanism of translation. Cell. 2002;108(4):557–572. https://doi.org/10.1016/s0092-8674(02)00619-0.
2. Williamson JR. The ribosome at atomic resolution. Cell. 2009;139(6):1041–1043. https://doi.org/10.1016/j.cell.2009.11.028.
3. Goldman AD, Kacar B. Cofactors are remnants of Life's origin and early evolution. J Mol Evol. 2021;89(3):127–133. https://doi.org/10.1007/s00239-020-09988-4.
4. Woese CR. Translation: In retrospect and prospect. RNA. 2001;7(8):1055–1067. https://doi.org/10.1017/s1355838201010615.
5. Crick FH, Brenner S, Klug A, Pieczenik G. A speculation on the origin of protein synthesis. Orig Life. 1976;7(4):389–397. https://doi.org/10.1007/bf00927934.
6. Wolf YI, Koonin EV. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. Biol Direct. 2007;2:14. https://doi.org/10.1186/1745-6150-2-14.
7. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. Microbiol Mol Biol Rev. 2005;69(1):101–123. https://doi.org/10.1128/mmbr.69.1.101-123.2005.
8. Rodnina MV. Translation in prokaryotes. Cold Spring Harb Perspect Biol. 2018;10(9):a032664. https://doi.org/10.1101/cshperspect.a032664.
9. Marintchev A, Wagner G. Translation initiation: Structures, mechanisms and evolution. Q Rev Biophys. 2004;37(3–4):197–284. https://doi.org/10.1017/s0033583505004026.
10. Londei P. Evolution of translational initiation: New insights from the archaea. FEMS Microbiol Rev. 2005;29(2):185–200. https://doi.org/10.1016/j.femsre.2004.10.002.
11. Benelli D, Londei P. Translation initiation in archaea: Conserved and domain-specific features. Biochem Soc Trans. 2011;39(1):89–93. https://doi.org/10.1042/bst0390089.
12. Roll-Mecak A, Shin BS, Dever TE, Burley SK. Engaging the ribosome: Universal IFs of translation. Trends Biochem Sci. 2001;26(12):705–709. https://doi.org/10.1016/s0968-0004(01)02024-2.
13. Boelens R, Gualerzi CO. Structure and function of bacterial initiation factors. Curr Protein Pept Sci. 2002;3(1):107–119. https://doi.org/10.2174/1389203023380765.
14. Simonetti A, Marzi S, Billas IM, et al. Involvement of protein IF2 N domain in ribosomal subunit joining revealed from architecture and function of the full-length initiation factor. Proc Natl Acad Sci U S A. 2013;110(39):15656–15661. https://doi.org/10.1073/pnas.1309578110.
15. Marshall RA, Aitken CE, Puglisi JD. GTP hydrolysis by IF2 guides progression of the ribosome into elongation. Mol Cell. 2009;35(1):37–47. https://doi.org/10.1016/j.molcel.2009.06.008.
16. Harvey KL, Jarocki VM, Charles IG, Djordjevic SP. The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. Front Microbiol. 2019;10:2351. https://doi.org/10.3389/fmicb.2019.02351.
17. Nissen P, Kjeldgaard M, Thirup S, et al. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. Science. 1995;270(5241):1464–1472. https://doi.org/10.1126/science.270.5241.1464.
18. Wohlgemuth I, Pohl C, Mittelstaet J, Konevega AL, Rodnina MV. Evolutionary optimization of speed and accuracy of decoding on the ribosome. Philos Trans R Soc Lond B Biol Sci. 2011;366(1580):2979–2986. https://doi.org/10.1098/rstb.2011.0138.
19. Caserta E, Tomsic J, Spurio R, La Teana A, Pon CL, Gualerzi CO. Translation initiation factor IF2 interacts with the 30 S ribosomal subunit via two separate binding sites. J Mol Biol. 2006;362(4):787–799. https://doi.org/10.1016/j.jmb.2006.07.043.
20. Laalami S, Sacerdot C, Vachon G, et al. Structural and functional domains of *E. coli* initiation factor IF2. Biochimie. 1991;73(12):1557–1566. https://doi.org/10.1016/0300-9084(91)90191-3.
21. Laursen BS, Kjaergaard AC, Mortensen KK, Hoffman DW, Sperling-Petersen HU. The N-terminal domain (IF2N) of bacterial translation initiation factor IF2 is connected to the conserved C-terminal domains by a flexible linker. Protein Sci. 2004;13(1):230–239. https://doi.org/10.1110/ps.03337604.
22. Roll-Mecak A, Cao C, Dever TE, Burley SK. X-ray structures of the universal translation initiation factor IF2/eIF5B: Conformational changes on GDP and GTP binding. Cell. 2000;103(5):781–792. https://doi.org/10.1016/s0092-8674(00)00181-1.

23. Berchtold H, Reshetnikova L, Reiser CO, Schirmer NK, Sprinzl M, Hilgenfeld R. Crystal structure of active elongation factor Tu reveals major domain rearrangements. Nature. 1993; 365(6442):126–132. https://doi.org/10.1038/365126a0.
24. Parmeggiani A, Swart GW, Mortensen KK, et al. Properties of a genetically engineered G domain of elongation factor Tu. Proc Natl Acad Sci U S A. 1987;84(10):3141–3145. https://doi.org/10.1073/pnas.84.10.3141.
25. Schuette JC, Murphy FVT, Kelley AC, et al. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. EMBO J. 2009;28(6):755–765. https://doi.org/10.1038/emboj.2009.26.
26. Kuhle B, Ficner R. eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. EMBO J. 2014; 33(10):1177–1191. https://doi.org/10.1002/embj.201387344.
27. Meunier S, Spurio R, Czisch M, et al. Structure of the fMet-tRNA(fMet)-binding domain of B. stearothermophilus initiation factor IF2. EMBO J. 2000;19(8):1918–1926. https://doi.org/10.1093/emboj/19.8.1918.
28. Sprink T, Ramrath DJ, Yamamoto H, et al. Structures of ribosome-bound initiation factor 2 reveal the mechanism of subunit association. Sci Adv. 2016;2(3):e1501502. https://doi.org/10.1126/sciadv.1501502.
29. Sacerdot C, Dessen P, Hershey JW, Plumbridge JA, Grunberg-Manago M. Sequence of the initiation factor IF2 gene: Unusual protein features and homologies with elongation factors. Proc Natl Acad Sci U S A. 1984;81(24):7787–7791. https://doi.org/10.1073/pnas.81.24.7787.
30. Hartman H, Smith TF. GTPases and the origin of the ribosome. Biol Direct. 2010;5:36. https://doi.org/10.1186/1745-6150-5-36.
31. De Tarafder A, Parajuli NP, Majumdar S, Kaçar B, Sanyal S. Kinetic analysis suggests evolution of ribosome specificity in modern elongation factor-Tus from "generalist" ancestors. Mol Biol Evol. 2021; 38(8):3436–3444. https://doi.org/10.1093/molbev/msab114.
32. Okafor CD, Pathak MC, Fagan CE, et al. Structural and dynamics comparison of Thermostability in ancient, modern, and consensus elongation factor Tus. Structure. 2018;26(1): 118–129.e113. https://doi.org/10.1016/j.str.2017.11.018.
33. Zhou Y, Asahara H, Gaucher EA, Chong S. Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. Nucleic Acids Res. 2012;40(16):7932–7945. https://doi.org/10.1093/nar/gks568.
34. Kacar, B, Ge, X, Sanyal, S, Gaucher, E. Experimental Evolution of Escherichia coli Harboring an Ancient Translation Protein. J Mol Evol. 2017;84:69–84. https://doi.org/10.1007/s00239-017-9781-0.
35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780. https://doi.org/10.1093/molbev/mst010.
36. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1): 268–274. https://doi.org/10.1093/molbev/msu300.
37. Vialle RA, Tamuri AU, Goldman N. Alignment modulates ancestral sequence reconstruction accuracy. Mol Biol Evol. 2018;35(7):1783–1797. https://doi.org/10.1093/molbev/msy055.
38. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol. 2017;1:193. https://doi.org/10.1038/s41559-017-0193.
39. Farris JS. Estimating phylogenetic trees from distance matrices. Am Nat. 1972;106(951):645–668. https://doi.org/10.1086/282802.
40. Mai U, Sayyari E, Mirarab S. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. PLoS One. 2017;12(8):e0182238. https://doi.org/10.1371/journal.pone.0182238.
41. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: Conserved structure and molecular mechanism. Nature. 1991;349(6305):117–127. https://doi.org/10.1038/349117a0.
42. Dever TE, Glynias MJ, Merrick WC. GTP-binding domain: Three consensus sequence elements with distinct spacing. Proc Natl Acad Sci U S A. 1987;84(7):1814–1818. https://doi.org/10.1073/pnas.84.7.1814.
43. Liljas A, Ehrenberg M. Structural aspects of protein synthesis. 2nd ed. Singapore: World Scientific Publishing Company, 2013.
44. Åqvist J, Kamerlin SCL. The conformation of a catalytic loop is central to GTPase activity on the ribosome. Biochemistry. 2015; 54(2):546–556. https://doi.org/10.1021/bi501373g.
45. Ge X, Mandava CS, Lind C, Åqvist J, Sanyal S. Complementary charge-based interaction between the ribosomal-stalk protein L7/12 and IF2 is the key to rapid subunit association. Proc Natl Acad Sci U S A. 2018;115(18):4649–4654. https://doi.org/10.1073/pnas.1802001115.
46. Gualerzi CO, Brandi L, Caserta E, et al. Translation initiation in bacteria. The ribosome. Washington: American Society for Microbiology 2000; p. 475–494.
47. Lee JH, Choi SK, Roll-Mecak A, Burley SK, Dever TE. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. Proc Natl Acad Sci U S A. 1999;96(8):4342–4347. https://doi.org/10.1073/pnas.96.8.4342.
48. Luchin S, Putzer H, Hershey JW, Cenatiempo Y, Grunberg-Manago M, Laalami S. In vitro study of two dominant inhibitory GTPase mutants of *Escherichia coli* translation initiation factor IF2. Direct evidence that GTP hydrolysis is necessary for factor recycling. J Biol Chem. 1999;274(10):6074–6079. https://doi.org/10.1074/jbc.274.10.6074.
49. Tomsic J, Smorlesi A, Caserta E, Giuliodori AM, Pon CL, Gualerzi CO. Disparate phenotypes resulting from mutations of a single histidine in switch II of *Geobacillus stearothermophilus* translation initiation factor IF2. Int J Mol Sci. 2020;21(3): 735. https://doi.org/10.3390/ijms21030735.
50. Brenner SE, Chothia C, Hubbard TJP, Murzin AG. [37] Understanding protein structure: Using scop for fold interpretation. Methods Enzymol. 1996;266:635–643.
51. Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol. 2007;8(12):995–1005. https://doi.org/10.1038/nrm2281.
52. Eargle J, Black AA, Sethi A, Trabuco LG, Luthey-Schulten Z. Dynamics of recognition between tRNA and elongation factor Tu. J Mol Biol. 2008;377(5):1382–1405. https://doi.org/10.1016/j.jmb.2008.01.073.
53. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW. Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol Biol Evol. 2017;34(2): 247–261. https://doi.org/10.1093/molbev/msw223.

54. Gualerzi CO, Pon CL. Initiation of mRNA translation in bacteria: Structural and dynamic aspects. Cell Mol Life Sci. 2015; 72(22):4341–4367. https://doi.org/10.1007/s00018-015-2010-3.
55. Guenneugues M, Caserta E, Brandi L, et al. Mapping the fMet-tRNA(f)(met) binding site of initiation factor IF2. EMBO J. 2000; 19(19):5233–5240. https://doi.org/10.1093/emboj/19.19.5233.
56. Schmeing TM, Voorhees RM, Kelley AC, et al. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. Science. 2009;326(5953):688–694. https://doi.org/10.1126/science.1179700.
57. Du Z, Su H, Wang W, et al. The trRosetta server for fast and accurate protein structure prediction. Nat Protoc. 2021;16(12): 5634–5651. https://doi.org/10.1038/s41596-021-00628-9.
58. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—Distance constraints applied on model quality estimation. Bioinformatics. 2020;36(6):1765–1771. https://doi.org/10.1093/bioinformatics/btz828.
59. Song H, Parsons MR, Rowsell S, Leonard G, Phillips SE. Crystal structure of intact elongation factor EF-Tu from *Escherichia coli* in GDP conformation at 2.05A resolution. J Mol Biol. 1999; 285(3):1245–1256. https://doi.org/10.1006/jmbi.1998.2387.
60. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. J Mol Biol. 1994;235(2):625–634. https://doi.org/10.1006/jmbi.1994.1017.
61. Atkinson GC. The evolutionary and functional diversity of classical and lesser-known cytoplasmic and organellar translational GTPases across the tree of life. BMC Genom. 2015;16(1): 78. https://doi.org/10.1186/s12864-015-1289-7.
62. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol. 2002;317(1):41–72. https://doi.org/10.1006/jmbi.2001.5378.
63. Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. Science. 2001;294(5545):1299–1304. https://doi.org/10.1126/science.1062023.
64. Wilke CO. Bringing molecules back into molecular evolution. PLoS Comput Biol. 2012;8(6):e1002572. https://doi.org/10.1371/journal.pcbi.1002572.
65. Kacar B, Guy L, Smith E, Baross J. Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. Phil Trans R Soci A: Math Phys Eng Sci. 2017;375(2109):20160352. https://doi.org/10.1098/rsta.2016.0352.
66. Antoun A, Pavlov MY, Lovmar M, Ehrenberg M. How initiation factors maximize the accuracy of tRNA selection in initiation of bacterial protein synthesis. Mol Cell. 2006;23(2):183–193. https://doi.org/10.1016/j.molcel.2006.05.030.
67. Fournier GP, Andam CP, Alm EJ, Gogarten JP. Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. Origins Life Evol Biospheres. 2011;41(6):621–632. https://doi.org/10.1007/s11084-011-9261-2.
68. Goldman AD, Samudrala R, Baross JA. The evolution and functional repertoire of translation proteins following the origin of life. Biol Direct. 2010;5:15. https://doi.org/10.1186/1745-6150-5-15.
69. Fox GE, Naik AK. The evolutionary history of the translation machinery. In: Ribas de Pouplana L, editor. The genetic code and the origin of life. Springer US: Boston, MA, 2004; p. 92–105.
70. Hsiao C, Mohan S, Kalahar BK, Williams LD. Peeling the onion: Ribosomes are ancient molecular fossils. Mol Biol Evol. 2009; 26(11):2415–2425. https://doi.org/10.1093/molbev/msp163.
71. Venkataram S, Monasky R, Sikaroodi SH, Kryazhimskiy S, Kacar B. Evolutionary stalling and a limit on the power of natural selection to improve a cellular module. Proc Natl Acad Sci U S A. 2020;117(31):18582–18590. https://doi.org/10.1073/pnas.1921881117.
72. Fried SD, Fujishima K, Makarov M, Cherepashuk I, Hlouchova K. Peptides before and during the nucleotide world: An origins story emphasizing cooperation between proteins and nucleic acids. J R Soci Interface. 2022;19(187):20210641. https://doi.org/10.1098/rsif.2021.0641.
73. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44(D1):D733–D745. https://doi.org/10.1093/nar/gkv1189.
74. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: Architecture and applications. BMC Bioinform. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
75. Price MN, Dehal PS, Arkin AP. FastTree 2—Approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. https://doi.org/10.1371/journal.pone.0009490.
76. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–3402. https://doi.org/10.1093/nar/25.17.3389.
77. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221–224. https://doi.org/10.1093/molbev/msp259.
78. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
79. Garcia AK, Kolaczkowski B, Kaçar B. Reconstruction of Nitrogenase predecessors suggests origin from maturase-like proteins. Genome Biol Evol. 2022;14(3):evac031. https://doi.org/10.1093/gbe/evac031.
80. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5): 1792–1797. https://doi.org/10.1093/nar/gkh340.
81. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5): 1188–1195. https://doi.org/10.1093/molbev/mst024.
82. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–321. https://doi.org/10.1093/sysbio/syq010.
83. Rambaut A. FigTree v1. 3.1. Institute of Evolutionary Biology. Edinburgh: University of Edinburgh, 2009 http://tree.bio.ed.ac.uk/software/figtree/.
84. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.
85. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF

Chimera. BMC Bioinform. 2006;7:339. https://doi.org/10.1186/1471-2105-7-339.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fer E, McGrath KM, Guy L, Hockenberry AJ, Kaçar B. Early divergence of translation initiation and elongation factors. Protein Science. 2022;31(9):e4393. https://doi.org/10.1002/pro.4393